FINAL REPORT

Contract 05-303

Follow-on Development of CARBITS: A Response Model for the California Passenger Vehicle Market

Principal Investigator

Dr. David S. Bunch

(dsbunch@ucdavis.edu)

Amine Mahmassani

Prepared for:

State of California Air Resources Board
Research Division
P. O. Box 2815
Sacramento, CA 95812

Prepared by:

University of California, Davis
Institute of Transportation Studies
One Shields Avenue
2028 Academic Surge
Davis, CA 95616

April 30, 2009

Disclaimer

Acknowledgments

This Report was submitted in fulfillment of contract 05-303, “Follow-on Development of CARBITS: A Response Model for the California Passenger Vehicle Market” by the Institute of Transportation Studies, University of California, Davis, under sponsorship of the California Air Resources Board. Work was completed as of April 30, 2009.

Table of Contents

ABSTRACT

CARBITS is a market simulation model for the passenger vehicle market in California. Professor David S. Bunch developed CARBITS for the ARB during 2003-2004 under a contract with the University of California, Davis. Its primary purpose is as a scenario analysis tool to evaluate market response under alternative regulation scenarios. For purposes of this Final Report, the version of CARBITS developed during 2003-2004 will be referred to as "CARBITS 1.0." CARBITS 1.0 was requested by the ARB to meet specific needs for their work under AB 1493 regulating motor vehicle greenhouse gas emissions, and was developed within a short time frame to accommodate their schedule. The project was feasible because it was possible to base CARBITS development on pre-existing research results developed under an earlier University of California-Institute of Transportation Studies research program. Although time and monetary constraints prevented development of a full range of features, ARB staff successfully used CARBITS 1.0 in support of the climate change regulation adopted by the Board in September 2004.

This project has produced an updated version of CARBITS ("CARBITS 2.0") with a number of improvements and new features to address specific perceived "deficiencies" identified by ARB staff during the collaboration with Prof. Bunch. Some of these represented desired extensions based on experience in using the model. A related area of concern is the ever-present potential for criticism by the hired consultants of various stakeholders. The original project proposal identifies a list of specific goals:

1. Estimate a new set of vehicle choice models using more recent datasets.
2. Specifically address the issue of vehicle market exit/scrappage.
3. Develop re-calibration procedures to update certain model constants based on aggregate-level vehicle counts.
4. Include the capability to address hybrid electric vehicles.
5. Address issues of statistical noise and runtimes.

These specific goals have been addressed by this project. A new set of vehicle choice models has been estimated using data from the 2000-2001 Caltrans Statewide Travel Survey. This data source (although a few years old) is attractive due to its large sample size and high-quality sampling and weighting characteristics. In conjunction with using these data (which include information on vehicle holdings, but not transactions), CARBITS was converted from a transactions microsimulation model to a vehicle holdings model. This approach directly addresses the issue of statistical noise and run times: holdings models use analytical computations that yield deterministic (noise free) results requiring relatively short run times. Substantial effort was invested in data compilation and cleaning for this project. In particular, procedures for using DMV data routinely accessible to ARB were developed to address needs for periodic re-calibration using updated vehicle counts, patterns of vehicle market exit, and recent penetration of hybrid electric vehicles. Aside from meeting specific project goals, the substantial amount of work on data development, and the formulation of a generic vehicle market model framework, will provide additional benefits to ARB in succeeding projects.

EXECUTIVE SUMMARY

CARBITS is a market simulation model for the personal vehicle market in California. Professor David S. Bunch developed CARBITS for the ARB during 2003-2004 under a contract with the University of California, Davis. Its primary purpose is as a scenario analysis tool to evaluate market response under alternative regulation scenarios. For purposes of this Final Report, the version of CARBITS developed during 2003-2004 will be referred to as "CARBITS 1.0." CARBITS 1.0 was commissioned by the ARB to meet specific needs for their work under AB 1493 regulating motor vehicle greenhouse gas emissions, and was developed under a short time frame. For practical reasons, it was based on an existing model developed under an earlier University of California-Institute of Transportation Studies research program. Although time and monetary constraints prevented development of a full range of features, ARB staff successfully used CARBITS 1.0 in support of the climate change regulation adopted by the Board in September 2004.

Experience in working CARBITS as part of the 1493 rulemaking process led to some ideas for potential improvements. The overall stated objective of this project is to update and extend existing CARBITS model based on these experiences. Briefly, the stated goals of this project are:

1. Estimate new vehicle choice models using more recently collected datasets.
2. Address issues of statistical noise and runtimes.
3. Specifically address the issue of vehicle market exit/scrappage.
4. Develop re-calibration procedures to update certain model constants based on aggregate-level vehicle counts.
5. Include the capability to address hybrid electric vehicles

To illuminate these goals, we first review some details about CARBITS 1.0. As noted, CARBITS 1.0 was created using a pre-existing model. During the period 1992-1995, a team of Institute for Transportation (ITS) researchers at University of California (Davis and Irvine campuses) pursued a multi-year research program involving data collection and vehicle choice modeling. The California Energy Commission (CEC) provided much of the motivation for this work, which was targeted at exploring the future market for alternative fuel vehicles in California, including: battery-powered electric vehicles, compressed natural gas (dedicated and dual fuel versions), and alcohol/flex fuel. A major task was fielding a panel survey of California households that included stated choice questions on alternative fuel vehicles. One research goal was to explore household demand models based on *transaction* choices (e.g., vehicle replacement, addition, or disposal decisions) as an alternative to vehicle *holdings* models (the usual state of practice). The results of this project were used to develop CARBITS 1.0 to meet the needs of ARB.

The experiences and insights gained during the development and use of CARBITS 1.0 led to a number of ideas that were the motivation for this project. We briefly review these here. More details are included in the main report. First, from the very beginning of the earlier project, concerns were raised about the dataset being "old." This is a standard criticism for any model like CARBITS, given the expense and difficulty of collecting large-scale data sets on a regular basis. Regardless of whether there are technical merits to this narrow argument, it provides an opening to criticism by hired consultants. Second, the transactions models adapted from the earlier research required the use of pure microsimulation. This means that the model does not produce deterministic, analytical results, and it also requires special expertise (and long run times) to produce results in the proper manner. One example of why this can be an issue occurred during the 1493 rulemaking. Auto industry consultants (either accidentally or intentionally) produced results using CARBITS 1.0 that did not use enough replications to produce stable results, and then used these in an attempt to undermine CARBITS. A more practical concern is that using CARBITS 1.0 requires very long run times, making analysis more burdensome to the user.

A related issue is that the original modeling approach was primarily concerned with evaluating the entry of new types of vehicles (none of which, by the way, were hybrid electric—see below), with much less emphasis on vehicle exit and scrappage. CARBITS 1.0 takes an approach where vehicles exit the market "implicitly," based on the dynamics of vehicle replacement. In contrast, other approaches use aggregate data to estimate models that explicitly address vehicle exit. There are pros and cons to each method; however, because the latter method is easier to understand, it is typically used by outside consultants. Moreover, the AB 1493 experience suggests that a more complex model like CARBITS is vulnerable to criticism through both misapplication of the model and misrepresentation of results. Finally, there is the issue of hybrid electric vehicles. The recent penetration of hybrid electric vehicles makes it obvious that future policy analyses may need to address this new type of vehicle.

These specific goals listed above been addressed by this project. With regard to introducing new data, various options were considered. Maintaining and updating CARBITS 1.0 as a transactions-based model would require a new source of household panel data that includes details on vehicle transactions. This type of data is very expensive to collect and difficult to come by. Moreover, experience suggested that the transactions-based approach was the common source of a number of the issues this project was intended to address. Based on multiple factors, we decided to update CARBITS using the 2000-2001 Caltrans Statewide Travel Survey.

These data (although a few years old) are attractive for a number of reasons. For a household survey of this type it has a very large sample size (over 17,000 households, all from California), and uses high-quality sampling and weighting procedures. In conjunction with using these data (which include information on vehicle holdings, but not transactions), CARBITS was converted from a transactions microsimulation model to a vehicle holdings model. This approach directly addresses the issue of statistical noise and run times, since holdings models can be implemented using analytical computations that yield deterministic (noise free) results requiring relatively short run times.

Although it is less than obvious from the stated project goals, the decision to estimate a completely new model for CARBITS (regardless of which household dataset was chosen) created a whole host of additional data requirements. Substantial effort was invested in data compilation and cleaning for this project. One area requiring a large amount of work was the development of a Vehicle Technology Database. Vehicle choice models have a number of requirements for characterizing the vehicle choices faced by consumers in the marketplace. These include such things as market prices, vehicle body types and sizes, fuel economy, performance characteristics, and others. No one data source includes all of these information items. This requires creating a large database by merging together data from multiple data sources. Because each data source has its own way of defining vehicles (which includes character string data describing the make and model of vehicle), cleaning and merging these data is a herculean task.

In addition to vehicle technology data, there are multiple aspects of the project that require aggregate data on multiple aspects of the vehicle market. For example, models like CARBITS (which are estimated on the basis of household survey data) must periodically be re-calibrated so that the vehicle distributions for the model base year match the aggregated vehicle totals from an outside source (project goal 3). In addition, estimating a model of vehicle exit requires some type of data set that tracks the entry and exit of vehicles from the market (project goal 2). Finally, in recent years hybrid electric vehicles have been entering the market. Survey data cannot possibly have the sample size to obtain accurate measurements of this aggregate phenomenon (project goal 4). To address these data needs, procedures for processing Department of Motor Vehicles (DMV) registrations data were developed.

We emphasize the data collection and cleaning aspect of this project because (i) a substantial amount of the contract effort was devoted to it, and (ii) we consider the outcome of this effort to be a major side benefit of this project that goes beyond the narrow statement of the project goals. In a similar vein, our approach to creating the new version of CARBITS ("CARBITS 2.0") incorporated system design concepts such as object-oriented analysis and object-oriented programming. Specifically, rather than program CARBITS 2.0 as a stand-alone one-time effort, we decided to create a generic system framework for "CARBITS-like models," and then implement CARBITS 2.0 as a specific "instance" within this framework. The system framework and CARBITS 2.0 were implemented using the object-oriented features of MATLAB. (In contrast, CARBITS 1.0 was written in FORTRAN.) This approach will make any future efforts to modify or update CARBITS much easier.

To summarize, the project outcomes include the following:

1. CARBITS was updated using a more recent data set (2000-2001 Caltrans Travel Survey)
2. CARBITS was converted to a holdings-based model from the original transactions-based model.

3. Outcomes 1 and 2 directly address the issue of model runtimes and statistical noise by using an approach that produces results based on deterministic computations.
4. DMV data were developed as a source of data on aggregate vehicle counts, vehicle entry and exit statistics, and penetration of hybrid electric vehicles.
5. Outcome 4 supported the development of procedures to re-calibrate model constants to match aggregate vehicle totals, the estimation of a vehicle market exit model, and the capability to incorporate data on hybrid electric vehicles.
6. A substantial amount of effort on compiling and cleaning data (including many data sets on vehicle prices and technology) yielded an additional side benefit for future work by ARB.
7. CARBITS 2.0 was developed using object-oriented analysis and programming methods. A generic system framework for "CARBITS-like models" was established, and then CARBITS 2.0 was coded as a special case.

BACKGROUND

In late 2002, ARB staff approached the Institute of Transportation Studies (ITS) at University of California, Davis (UC Davis) to discuss a number of research needs related to its charge to perform rulemaking under AB 1493 (Pavley). One such need was for a scenario analysis tool to provide a quantitative assessment of the effects of alternative regulatory policies on the personal vehicle market in California over the medium and long term. For example, manufacturers would be expected to change their vehicle offerings in order to comply with a regulation. The operating characteristics, and new vehicle prices would be expected to change. This, in turn, would elicit a response from the vehicle market. Prof. David S. Bunch agreed to develop such a model under as part of a larger research project performed during 2003-2004. Both time and budget requirements precluded a major research effort, e.g., fielding a household survey, collecting data, and developing an entirely new model. The proposed solution was to adapt models developed under an earlier research program.

The earlier research involved data collection and vehicle choice modeling for the California market. It was performed during the mid 1990's by a team of ITS researchers (including Prof. Bunch) from two University of California campuses (Davis and Irvine). The program was a multi-year effort with funding from multiple sources. The California Energy Commission provided much of the motivation for this work. In addition to funding a pilot project, they coordinated efforts for a sequence of projects funded first by Southern California Edison, and then Pacific Gas & Electric. In addition, the research team received pass-through federal funding from the ISTEA program.

One component of the project was a panel survey of California households. The desire was to get observations from the same household at multiple points in time in order to trace the transaction dynamics of their vehicle purchases. In addition, the survey involved the application of stated preference methods to collect data on hypothetical choice of alternative fuel vehicles, including battery-powered electric vehicles, compressed natural gas (dedicated and dual fuel versions), and alcohol/flex fuel. The two main goals of the research were to (i) produce models of "transaction choice," based on the argument that such models could be superior to the more traditional vehicle holdings models that were in use at that time, and (ii) support the analysis of policies related to the introduction of alternative fuel vehicles into the California market. The products of this research program were used in developing the original version of CARBITS. For purposes of this Final Report, the original version of CARBITS developed in 2003-2004 will be referred to as "CARBITS 1.0."

ARB staff used CARBITS 1.0 when developing greenhouse gas regulations to meet AB 1493 requirements. Although staff's use of the model was considered successful, there was also a desire to upgrade the model to address some perceived "deficiencies." Some issues arose directly from the decision to rely on the earlier research results. For example, the behavioral models used in CARBITS were based on the panel survey of California households collected in the mid 1990's, so some critics considered the data to be "old." However, most of the motivation for this project was based on experience and insight gained while developing and using the model. In what follows, we give additional background on this motivation.

As noted above, CARBITS 1.0 was developed by adaption of pre-existing behavioral models. A key component was a transactions choice model estimated by a PhD student at UC Irvine as a major part of her thesis (Sheng). The original dataset used for estimating this model was no longer available, so re-estimation or other approaches were not possible. The most important feature of this model was that it was based on modeling household-level vehicle *transactions* using observations collected from the same sample of households at two points in time. (In addition, responses from a stated choice experiment were incorporated.) This model structure required that vehicle market forecasts be computed using pure microsimulation. Specifically, the model was populated by a large database of households. Results were obtained by repeated simulation of individual transaction *events*, and taking averages. This approach required very long computer run times. In particular, a very large number of replications are required to produce results with the required level of smoothness.

In addition to creating something of a burden for staff, the CARBITS 1.0 approach is vulnerable to criticism from outside consultants. The model is relatively complex and can be readily misrepresented. For example, auto industry consultants gained access to CARBITS 1.0 and (either accidentally or intentionally) generated model runs without using sufficient simulation replications. They then used the output to claim that CARBITS 1.0 performs poorly. A related issue is that CARBITS 1.0 follows a practice of modeling vehicle scrappage as an implicit outcome of choices made in the used vehicle market. Alternative approaches model vehicle scrappage explicitly, giving the modeler greater control over how model output is generated.

Other items are more practical, and support the ongoing use of CARBITS for other types of analysis. The original CARBITS model was put together to meet the immediate needs of ARB staff. It was calibrated "by hand" to match vehicle count data corresponding to the time period of the original survey data. A desirable enhancement would be to create procedures for automated re-calibration of model constants when updated vehicle count data become available. Finally, in looking ahead to future applications, it is clear that the recent and ongoing penetration of hybrid electric vehicles could be an important factor in formulating vehicle-related policies.

The goals for this project as based on the above discussion may be briefly summarized as:

1. Estimate new vehicle choice models using more recently collected datasets.
2. Address issues of statistical noise and runtimes.
3. Specifically address the issue of vehicle market exit/scrappage.
4. Develop re-calibration procedures to update certain model constants based on aggregate-level vehicle counts.
5. Include the capability to address hybrid electric vehicles

With this as background, we give an overview of key decisions and elements of the project, as an introduction to the remainder of the report.

1. As indicated, CARBITS 1.0
    a. Is a transactions model requiring pure microsimulation.
    b. Is based on a special-purpose panel survey collected in the mid 1990's.

2. Goals for CARBITS 2.0 include
    a. Estimating models using more recent data.
    b. Reducing statistical noise and run times.

3. The two previous goals can both be met by:
    a. Updating CARBITS using the 2000-2001 Caltrans Statewide Travel Survey.
    b. Converting CARBITS from a transactions model to a holdings model

In this project, various options for updating CARBITS using "new data" were considered. This project represented an option to directly address the issues described above. CARBITS 1.0 was, by necessity, a transactions model. A straightforward update of CARBITS without any changes to the modeling structure would require a panel data set with details on vehicle transactions. Although there were some possible data sources to support this (i.e., the Consumer Expenditure Survey), the most attractive data set in terms of sample size and quality is the Caltrans Travel Survey. However, this is a standard cross-sectional data set (not a panel data set) and can only support the estimation of a holdings model. At the same time, the transactions model in CARBITS 1.0 requires pure microsimulation, which is the source of the run time and statistical noise problems to be addressed. The decision to adopt the Caltrans Travel Survey and develop holdings models allows us to adopt the highest quality data, with the largest sample size (all of which comes from California), and also eliminates problems with run times and statistical noise.

The other high-level goals for this project (develop procedures for regular model recalibration, incorporate vehicle scrappage, expand the model to address hybrid electric vehicles) have a more general theme: Placing CARBITS on a footing whereby it can be regularly updated and improved by incorporating new data. In conducting this project, we strove to take a broader view to address this general theme, i.e., perform activities in this project to enhance the ongoing viability of CARBITS. In this regard, we approached the work to update CARBITS in two ways:

1. Designing a generic system/framework for "CARBITS-type" models.
2. Identifying and compiling data sources and procedures to support future updating of CARBITS.

Regardless of the details of our approach, we remark here that the implications for the data requirements in this project may not have been readily apparent from a discussion of the high-level goals. The wholesale estimation of new models creates requirements for vehicle data, not just household data. Specifically, choice models assume that households make vehicle choices based on vehicle attributes. These include both vehicle technology characteristics, and vehicle market *prices*. A substantial amount of effort in this project was expended on the collection, cleaning, and integration of vehicle data. Similarly, model calibration and estimation of vehicle market exit rates require data on the vehicle population at large, at multiple points in time. In this regard, this project also required the processing and analysis of large DMV data files.

The main body of this report provides more detailed discussion and documentation of Project Outcomes. Project Outcomes are presented in a series of separate sections. In accordance with the approach described here, Section 1 presents a generic framework for what we are calling "CARBITS-type models." The basic framework has been implemented in MATLAB, using principles of object-oriented analysis and programming. One benefit of this approach is the reusability of computer code, and the flexibility to easily alter models, update models, create multiple versions of models for comparison and testing purposes, etc.

Specific frameworks can be defined by adopting a particular set of definitions for model inputs and outputs. Within a given framework, many different models can be implemented as long as they use the same inputs and outputs. CARBITS-type models require input data related to household characteristics and vehicle classes/attributes. A critical requirement for this project was to adopt a specific set of Vehicle Class definitions (with an identified set of vehicle attributes) to provide a basis for vehicle demand modeling. Vehicle Class definitions are discussed in Section 2. Section 3 reviews information about the Caltrans Travel Survey data that form the basis for the new CARBITS 2.0 models. Section 4 discusses the household vehicle demand models developed for CARBITS 2.0. It provides a review of vehicle choice models, including a discussion of transaction versus holdings models, and then gives results for the vehicle holdings models estimated using the Caltrans data. Section 5 gives an overview of DMV data. Section 6 discusses a vehicle market exit model estimated using DMV data. Section 7 discusses calibration. Section 8 contains remarks on remaining project issues.

Section 9 is the bibliography. Appendix A provides background on database related issues related to vehicle technology, vehicle prices, and vehicle count data.

PROJECT OUTCOMES

1. A Generic Framework for CARBITS-type Models

The basic function of a CARBITS-type model is to simulate the behavior of the California personal vehicle market over a specified period of time, and to do so in a way that will support the analysis of alternative policy scenarios. There are many possible ways to do this, and a fully documented description of any specific model's implementation could be rather technical, and contain a high level of detail. However, it is also possible (and helpful) to formulate a *generic* framework for modeling a "vehicle market system" in terms of key components and their relationships. The basic structure would be applicable across a wide range of models, but at the same time, many of the technical details might be different, e.g., *within* a given component. In our work we have been approaching the development of CARBITS-type models using object-oriented modeling and programming techniques. Although a full discussion of such methods is beyond the scope of this report, the idea is that a logical system constructed of "entities" (e.g., households, vehicles) and "relationships" (vehicle ownership) can be implemented as modules where the internal detailed workings of the various components are "encapsulated." The model can be continually updated and improved in a variety of ways with minimal changes to the system. For example, a specific behavioral model related to household vehicle choice can be changed, improved, etc., by upgrading the internal workings of a single module. This framework also offers the possibility of creating multiple alternative models by substitution of modules, and comparing them on the results they produce. These are capabilities that could be used for future improvements or research activities. For this project, a single model ("CARBITS 2.0") has been created using this framework.

In this section, we review (informally) the generic features of what we are now calling "CARBITS-type" models. In addition to establishing a framework that can support ongoing technical development, this provides useful background for later discussion. The following is a list of basic assumptions underlying CARBITS-type models:

1. The entity that is the source of vehicle market demand is the *Household*.

2. Total demand in the vehicle market is the result of an aggregation of decisions made at the individual household level.

3. In each period of a "market simulation," households make decisions about their vehicle fleet. (The details of what decisions are made, and how, can vary depending on what type of behavioral model is used.)

4. In each period, both new and used vehicles are available in the market. Manufacturers introduce new vehicle offerings in each model year. New vehicles purchased in a model year become part of the used vehicle market in later years.

5. Households make decisions on the basis of "utility maximization," and have preference functions that capture their evaluation of vehicles that are available in the market.

6. Household preferences are formed on the basis of vehicle characteristics, including a vehicle's technical specifications and its market price. The fuel operating cost of a vehicle is based on its fuel economy, but also on the price of fuel during the period.

7. Household preferences are also a function of household demographics, such as income, household size, age, etc.

To implement a model based on the above assumptions, the following elements are required:

1. A Base Calendar Year (a.k.a., "**Base Year**").

2. A database of **Households** that represents California for the Base Year.

3. A system for defining Vehicles that represent the unique choice "options" in the market. Although vehicles could be defined at the Year-Make-Model level, the large number of such vehicles makes this impractical. The usual practice is to define a set of **Vehicle Classes** to represent the types of vehicles available in the market.

4. A **Vehicle Technology Database** that provides vehicle technical specifications ("attributes") and new vehicle prices for Vehicle Class offerings (typically by model year). This requires *historical* data for vehicles available in the Base Year. In addition, a *forecast* of available Vehicle Classes and vehicle attributes is required for future years.

5. A **Fuel Forecast** specifying fuel prices for the Base Year and all future years covered by the simulation.

6. A method for "**aging**" the Household database to reflect population growth and shifts in demographic distributions in the future.

7. **Behavioral models** for representing Household vehicle-related decisions.

8. A method of **setting vehicle prices** that "clears the market" that balances vehicle supply (new and used vehicles) with Household demand. (This also includes scrappage of old vehicles.)

Vehicle Market Behavior over a multiple-year time period is "simulated" by the following procedure (which assumes one-year time intervals):

1. For Base Year, initialize:
    a. Households
    b. Current Market Vehicles
    c. Current Vehicle Counts
    d. Current Year = Base Year
2. Begin Loop
    a. Previous Vehicle Count = Current Vehicle Count
    b. Current Year = Current Year + 1
    c. Lookup Current Fuel Costs
    d. Age Households
    e. Update ("age") Current Market Vehicles
        i. Introduce New Vehicles for Model Year = Current Year
        ii. Update Vehicle Characteristics (e.g., re-compute fuel operating costs using current fuel prices)
    f. Simulate Vehicle Market Behavior for Current Year
    g. Summarize Current Vehicle Counts, and report results.
3. Does Current Year = Final Year?
    a. If Yes, Stop
    b. If No, Go To Step 2

The above procedure is generic, in that it is consistent with a wide variety of specific model implementations. By adopting a specific set of data elements for key model *inputs* and *outputs*, it is possible to create a well-defined "platform" for model development and implementation of multiple CARBITS-type models. Data elements can be selected so that the same input and output formats can be re-used for a variety of models. For purposes of this project, we have established conventions for inputs and outputs, and have implemented a "CARBITS Vehicle Market Simulation Framework." The issue of Model Inputs is discussed in more detail in the next sub-section. The portion of the process denoted "Simulate Personal Vehicle Market Behavior for Current Year" represents a "module" that can be implemented using, e.g., different types of household behavioral models. This module can be further decomposed into additional sub-modules that address such questions as how household vehicle-related decisions will be modeled (e.g., as transaction choices, holdings choices, etc.), how the market is cleared, prices changed, etc.

Behavioral models in this CARBITS framework are based on household-level survey data. The availability of data and other considerations have an effect on the Base Year and options for specific behavioral models. This is briefly addressed in sub-section 1.2, as well as other parts of this report.

1.1 Model Inputs

In the current implementation of CARBITS, the main inputs that are typically used for policy analysis are the ***Vehicle Technology Database*** (VehTechDB) and the ***Fuel Forecast***. The VehTechDB includes *historical* data on vehicles corresponding to the Base Year. However, scenario analysis is based on simulating how the *future* vehicle market will behave in response to changes in regulations. This requires the user to provide a *forecast* of vehicle technology offerings for future years. In many cases, regulations might require vehicle manufacturers to change their offerings. If so, this must be reflected in the model inputs provided by the user. The model then simulates how the market would behave under this scenario.

A key design issue for a CARBITS-like model is the definition of ***Vehicle Classes***, and the identification of *vehicle attributes* to be included in the VehTechDB. Deciding on these elements is important, because they represent the only information that can be used as inputs to Vehicle Demand Models. Section 2 discusses the Vehicle Class definitions adopted for CARBITS 2.0. In addition to Vehicle Class, household vehicle choice is assumed to depend on three attributes: Market Price of the vehicle, Fuel Operating Cost (in cents per mile), and Acceleration (seconds for 0 to 60 miles per hour). Fuel Operating Cost in any given year is computed from Fuel Economy and the Fuel Cost for that year (provided in the Fuel Forecast).

Although this may seem to be a straightforward proposition, rigorously establishing vehicle attributes for each Vehicle Class requires a procedure for aggregating data from the large number of individual makes and models that are available in the market. Generally speaking, weighted averages of attributes are required, which in turn requires data on the *distribution* of vehicles in the market. This project required integration of data from the following sources:

Chrome VINMatch data
Chrome New Vehicle Data (NVD)
National Automobile Dealers Association VINPrefix Solution
California Department of Motor Vehicles (DMV) registration data
California Bureau of Automobile Regulation (BAR) Smog Check data
EPA Fuel Economy Guide
Wards Automotive Yearbook Vehicle Specifications

The commercially available data sets are sources of vehicle specification and market price data. The DMV and BAR data provide weighting information to allow attributes to be averaged over vehicle classes. In addition, the DMV data provide information on actual vehicle counts in the California fleet, and data on the rate at which vehicles exit the market. Appendix A provides more details about these data sources.

1.2 Base Year, Household Data, and Models

The nominal Base Year for CARBITS 2.0 is 2001, which corresponds to the household database used for this project: The 2001 Caltrans Travel Survey. These data are used as the Household database in the above simulation framework, and, in addition, were used for estimating the household vehicle demand behavior models used in Step 6. The database is discussed in Section 3, and details on the behavioral models are given in Section 4.

2. Vehicle Class Definitions and Attributes

To begin, we review the vehicle classification scheme from CARBITS 1.0:

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

Table 2.1 CARBITS 1.0 Body Type and Size Classes

CARBITS 1.0 uses this classification scheme because it was based on a model developed by an Irvine-Davis ITS team under a program sponsored by the California Energy Commission (CEC). This was the classification schemed used by the CEC at that time in their CalCars model. At the time, a substantial amount of effort had been expended in structuring vehicle technology data (i.e., attributes, prices, etc.) according to this framework, both in a historical context as well as in the form of technology forecasts. In addition, the CEC had a substantial investment in generating DMV vehicle counts using this framework.

One main concern with this approach is that it represents a market structure that, while appropriate in the 70's and 80's, might no longer be an adequate representation. Specifically, during that period in history the term "luxury car" was generally associated a type vehicle with a particular set of characteristics and a well-established image in the minds of consumers. These vehicles were generally larger than other vehicles, and much more expensive with certain types of interior features. Representative vehicles would be the offerings from nameplates such as Cadillac, Lincoln, and Mercedes. The market is now more differentiated so that each size class has both "high-end" and "low-end" vehicles. The high-end vehicles are typically represented by a more "prestigious" brand name, have higher performance characteristics (and lower fuel economy), and are more expensive. In our approach, we have adopted the term "Prestige" (rather than "Luxury") to characterize these high-end vehicles.

A similar, overlapping concern has to do with the use of the term "sports car." Finding an objective standard to classify vehicles into this category is problematic, and this term no longer means what it once did. There are also challenges associated with vehicles in the "Mini" (or, "Mini-subcompact") category. In the range of years for which data are currently available for updating CARBITS, there has been very low demand for these vehicles. (However, it seems likely that this class will be making a comeback in the near future.)

This project represented an opportunity to re-examine these issues related to vehicle classification, because a number of the project goals were already going to require the type of data collection that could support the development and testing of alternative vehicle classification schemes. Having said this, once the data had been collected and reviewed, a greater appreciation for the practical issues associated with vehicle classification became apparent. In what follows, we review some of the details related to vehicle classification that were explored for this project.

2.1 Issues to Consider when Classifying Vehicles

The notion of vehicle classification can be tricky, since the concept relates both to a consumer's *conception* of what a vehicle "is" and what it can be used for (which drives vehicle demand), and the physical and technological features that a vehicle may incorporate. The latter relate to a number of issues, including the basis for how regulations are formulated, and how vehicles can be characterized in terms of attributes in quantitative demand models. After a detailed review, we came away with a greater appreciation of the practical role that data availability can play in formulating vehicle classification schemes. Briefly stated, we have adopted a scheme whereby vehicles are characterized along three dimensions:

1. Body Type
2. Size
3. Prestige

We also consider the issue of hybrid electric vehicles, but this will be addressed elsewhere. Because CARBITS must address both used and new vehicle markets, there will also be a vintage/age dimension. In what follows, Body Type and Size will be discussed together.

2.1.1 *Body-Type-Size Classes*

For our purposes, "Body Type" refers to the physical configuration of a vehicle whereby it has a specific type of general functionality. For historical reasons, there is now a strong bifurcation between two basic configurations: Passenger Car, and Light-Duty Truck (LDT).

Passenger cars can be subdivided in a number of ways according to "Body Style" (e.g., sedan, hatchback, coupe) where the most important differences occur in the case of station wagons, two-seaters (roadsters), and perhaps convertibles. In our work we collected data at the level of body style, but decided that the following three categories represented the most fundamental distinction in terms of functionality: Car, Station Wagon, and Two-seater. We also considered "convertible," but with very few exceptions convertibles overlapped heavily with Two-seaters.

Light-Duty Trucks are now generally sub-divided into Pickups, Vans, and (Sports) Utility Vehicles (SUVs). In terms of functionality, there is a clear difference between Pickups, which have an open bed and limited seating, versus Vans and SUVs, which are enclosed and have more seating but can also be re-configured to one degree or another for carrying cargo. The SUV has other distinguishing features that might be more related to a type of product image that appeals to a particular type of consumer. In considering specific makes and models of vehicles over time, there can be some ambiguity in how to classify certain vehicles based on their physical configurations, since many could qualify either as a station wagon, a minivan, or an SUV. Most recently, Crossover vehicles have created additional confusion.

It turns out that the above discussion combined with other issues (including data availability) has led us to a vehicle classification scheme that is not dramatically different from CARBITS 1.0, or others used in the academic literature (and for similar reasons). In general, the basis for most of these is a vehicle classification scheme that has long been used by EPA, which interacts the Body Types discussed above with some particular definitions of Size. (In addition, LDTs are divided into 2-wheel drive and 4-wheel drive versions). The full EPA scheme has changed some over the years: Prior to 1998 the non-Pickup LDTs that would generally be classified today as SUVs or Vans were characterized as "Special Purpose Vehicles." The terms "Sport Utility Vehicle" and "Minivan" were introduced in 1998 as a substitute.

Another factor is that a major source of vehicle attributes for this project, the Chrome databases (see Appendix) uses a MarketClass variable that is a slight extension of the EPA Class (it adds in the number of passenger doors for cars, i.e., 2 or 4), and, most importantly, it appears to maintain complete consistency with the EPA data. In our work, we begin with Body-Type-Size Definition 1 ("BTS1") classes that are based on EPA and Chrome. See Table 2.2. Differences are: (1) doors and drive train information are removed, and (2) Special Purpose Vehicles prior to 1998 are re-classified as Minivans or SUVs.

EPA/Chrome = BTS1	BTS2	BTS3 (CARBITS 2.0)
Two-seater Passenger Car	Two-seater	1. Two-seater
Mini-Compact Passenger Car Sub-Compact Passenger Car	Mini-compact Car Subcompact Car	2. Small Car
Compact Passenger Car Small Station Wagon	Compact Car Small SW	3. Compact Car
Midsize Passenger Car Midsize Station Wagon	Midsize Car Midsize SW	4. Midsize Car
Large Passenger Car Large Station Wagon	Large Car Large SW	5. Large Car
Small Pickup Trucks	Small Pickup	6. Small Pickup
Standard Pickup Trucks	Standard Pickup	7. Standard Pickup
Minivans*	Minivans*	8. Minivan
Large Passenger Vans Cargo Vans	Large Passenger Vans Cargo Vans	9. Full-size Van
Sport Utility*	Small SUV Midsize SUV Large SUV	10. Small SUV 11. Midsize SUV 12. Large SUV

Table 2.2 Development of Body-Type-Size (BTS) Definitions for CARBITS 2.0

One major issue with EPA/Chrome/BTS1 classes is that SUVs are not assigned to size classes in these published databases. However, EPA frequently must address vehicle size issues in various publications. For example, the following definitions appear in EPA (2007, page 5):

	Small	Midsize	Large
Pickup	< 105”	105” to 115”	> 115”
Van	< 109”	109” to 124”	> 124”
SUV	< 100”	100” to 110”	> 110”

Table 2.3 Wheelbase-based Size Definitions for Light-Duty Trucks

Note that defining the size of Pickups based on wheelbase is a different approach from EPA’s classification system—see column 1 of Table 2.2. In BTS1, Pickups are classified as Small and Standard Pickups based on gross vehicle weight rating (GVWR). In the standard EPA classification system, Vans are classified into Minivans, Large Passenger Vans, and Cargo Vans (based on definitions that we have as yet been unable to locate). Also, we remark that the definitions in Table 2.3 were taken from a 2007 EPA publication, but that these values could be different in publications from other years.

Definition BTS2 in Table 2.2 is obtained by adding SUV size classes to BTS1 based on the definition in Table 2.3. Definition BTS3 is obtained by merging together some BTS2 classes to obtain fewer categories. BTS3 generally looks like other classifications found in the literature, and it is based on similar concerns and considerations:

1. We have elected to merge Large Passenger Vans and Cargo Vans into the more generic “Full-Size Van.” Two reasons for this are: (1) the total demand

for these vehicles by households is rather small, (2) based on only make and model information is very difficult to distinguish between these two when working with most data sets.

2. In the choice modeling literature there has almost always been question of what to do about station wagons. Although they have some functional differences with, e.g., sedans, the sales volumes for station wagons are relatively small. Including them increases the number of categories. We adopted the usual practice of merging station wagons with standard cars of similar size in order to reduce the number of categories.

3. Mini-Compact cars have been absorbed into Small cars. (As discussed previously, demand for minicompacts has been extremely small for many years. Essentially all published choice models typically eliminate these as a separate class.)

3. Two-seater has been preserved as a separate class. It is an easily identifiable *physical* characteristic (in contrast to an image-based concept) that generally couples small size with a significant configuration feature (limited seating and luggage space) that is easier to identify than the less-well-defined concept "sports car."

2.1.2 *Prestige*

For this project we elected to define Prestige on the basis of vehicle brand name, incorporating the notion of "brand equity" frequently used in the marketing literature. Certain brand names are clearly associated with an image that incorporates a combination of such things as quality, reputation, a consistently high level of amenities and features offered as standard equipment, etc. One advantage of this approach is that it represents an "attribute" that is easily identifiable and readily assigned to each vehicle. Moreover, vehicles grouped together using this dimension share a number of similarities, resulting in more homogeneous groups (see discussion below). Finally, it generalizes the concept of "luxury" that previously was assigned to a very specific type of vehicle. One unfortunate, but unavoidable complication of this dimension is a higher degree of correlation between purchase price and other attributes (e.g., fuel economy and performance), which can complicate model estimation (see section 4).

Another dimension under consideration was "Country/Region of Manufacturer" (e.g., "Domestic versus Foreign," or, "Domestic-Asia-Europe"). There is little doubt that this dimension can have some explanatory power. Many years ago, studies seemed to support the idea that domestic consumers would prefer to "buy American" all else equal. Unfortunately, in more recent years this dimension as become convoluted with "reputation for quality" (see Train and Winston 2007), with many foreign manufacturers having a reputation for higher quality than their domestic competitors. Moreover, the foreign-domestic distinction has become less clear, with the advent of foreign manufacturers locating manufacturing plants in the U. S., and domestic manufacturers importing some of its product lines.

Prestige Brands	Region			
	Domestic	Europe	Asia	Total
Acura			13.50%	13.50%
Audi		1.50%		1.50%
BMW		11.60%		11.60%
Cadillac	14.00%			14.00%
Infiniti			6.60%	6.60%
Land Rover		1.90%		1.90%
Lexus			15.10%	15.10%
Lincoln	9.80%			9.80%
Mercedes Benz		18.60%		18.60%
Saab		1.20%		1.20%
Volvo		6.20%		6.20%
	23.90%	40.90%	35.20%	100.00%
Non-Prestige Brands	Region			
	Domestic	Europe	Asia	Total
Buick	2.80%			2.80%
Chevrolet	11.70%			11.70%
Chrysler	2.00%			2.00%
Dodge	5.90%			5.90%
Eagle	0.20%			0.20%
Ford	21.30%			21.30%
Geo		1.00%		1.00%
GMC	0.70%			0.70%
Honda			11.70%	11.70%
Hyundai			0.80%	0.80%
Isuzu			0.60%	0.60%
Jeep	2.10%			2.10%
Mazda			2.80%	2.80%
Mercury	2.40%			2.40%
Mitsubishi			1.80%	1.80%
Nissan			6.20%	6.20%
Oldsmobile	2.30%			2.30%
Plymouth	1.80%			1.80%
Pontiac	2.50%			2.50%
Saturn	2.40%			2.40%
Subaru			0.20%	0.20%
Suzuki			0.10%	0.10%
Toyota			14.80%	14.80%
Volkswagen		2.00%		2.00%
	57.90%	3.00%	39.10%	100.00%

Table 2.4 Distribution of Vehicles by Manufacturer (Classified by Prestige versus Region) in the California Personal Vehicle Fleet (October 2001)

Table 2.4 explores the two dimensions "Prestige" and "Region" on the basis of vehicle count distributions in California in Fall 2001 (October). These figures are based on October 2001 DMV data that were assembled to match the timeframe of the most recent Caltrans Travel Survey, and are intended to reflect the personal vehicle market—see Section 3. In this table, we have included breakdowns by region of origin, and report the percentage of the California vehicle fleet within each category (Prestige versus Non-Prestige) for model years 1989-2002. Prestige vehicles made up about 15% of the California fleet.

The percentages of Prestige versus Region are highly correlated. Domestic vehicles made up 58% of the non-Prestige fleet, but only 24% of the Prestige fleet. European vehicles had the largest share of the Prestige fleet (41%), and essentially none of the non-Prestige fleet (3%). It is important to note that, since these figures pool together model years 1989-2002, they do not illustrate more recent trends in Domestic versus non-Domestic new vehicle sales. However, even in 2001, the percentage of Lexus vehicles on the road had reached 15%, second only to Mercedes.

2.2. *CARBITS 2.0 Vehicle Classes (Historical)*

Taken together, Tables 2.2 and 2.4 illustrate some of the challenges in developing vehicle choice models for practical use in policy analysis. BTS1 includes 17 body-type-size classes. If one were to include ten vehicle manufacturers and 20 model years, the total number of make/vehicle-class/vintage combinations would be 17 x 10 x 20 = 3,400. (This is for gasoline vehicles only, i.e., it ignores the "dimension" of fuel/fuel technology type. Moreover, using the model to evaluate the impact of policies 20 years into the future requires *forecasts* of vehicle classes and attributes over this range of years. Determining the level of detail required for policy analysis is always a difficult judgment call.

The Vehicle Classes adopted for CARBITS 2.0 (for the case of historical data) are represented in Table 2.5. The table is based on scenario requirements for estimating choice models using Caltrans Travel Survey data, where the vehicle model year window begins in 1982 and goes through 2001. Certain Vehicle Classes do not exist over the full range of years (1982-2001). See Table 2.5. All Car types have both Non-Prestige and Prestige versions over the entire range of years; however, there are no Prestige Pickup Trucks or Minivans. There are no Midsize or Large SUVs included prior to 1985. Prestige SUVs begin in 1996. There are 350 combinations in all. Note: In reality, there are very small numbers of some vehicle types in some years that are not included in this table. However, they have been eliminated for modeling purposes.

The main purpose of defining vehicle classes is to provide a structure for modeling vehicle choice. Consumer choice of a vehicle class as defined in Table 2.5 is based on *preference* for vehicle configuration, size, prestige level, and also vintage. However, vehicle classes will also vary on other important attributes. Chief among these are market price, fuel operating cost, and performance. These would be expected to vary across vehicle class. This is illustrated next.

BTS3	Non-Prestige	Prestige
1. Two-seater	All Years*	All
2. Small Car	All	All
3. Compact Car	All	All
4. Midsize Car	All	All
5. Large Car	All	All
6. Small Pickup	All	[None]
7. Standard Pickup	All	[None]
8. Minivan	All	[None]
9. Full-size Van	All	[None]
10. Small SUV	All	1996-2001
11. Midsize SUV	1985-2001	1996-2001
12. Large SUV	1985-2001	1998-2001

Table 2.5 CARBITS 2.0 Vehicle Classes (*1982-2001)

2.2.1. Vehicle Attributes for CARBITS 2.0 Vehicle Classes (Historical)

This subsection reviews historical patterns of vehicle attributes for the Vehicle Classes defined previously. As has been noted, the key attributes used for consumer choice modeling in this project are market price, fuel operating cost, and performance. When consumers decide to make a vehicle purchase, they take possession of a specific year-make-model vehicle with well-defined physical characteristics. However, estimating choice models at Vehicle Class level does not support this level of detail, and requires representative attribute values that are typically obtained by taking *averages* over the individual vehicle offerings in a class. (Usually these are sales-weighted averages.)

There are many issues and details associated with the construction of Vehicle Technology databases that are too numerous to discuss here. This information is included in Appendix A. However, we provide some very brief remarks here:

1. Market price data for this study come from the National Automobile Dealer Association (NADA) VIN Prefix solution. These data include estimates of market prices for both new and used vehicles for a particular month and year, at the level of an individual VIN Prefix (which captures information on make, model, style, engine, and other characteristics). See Appendix A.

2. Because fuel operating cost (measured in cents per mile) is a function of both fuel efficiency (mpg) *and* fuel price ($ per gallon), the relevant vehicle *technology* variable is fuel efficiency. The original source of mpg ratings is the EPA fuel economy guide data, which are also replicated in other vehicle specification databases. EPA provides three ratings: city, highway, and combined. When representative values are called for, we used the combined mpg estimate.

3. There are many possible choices for measuring vehicle performance, including: horsepower, horsepower-to-weight ratio, top speed, etc. In this project, we use a

measure called “EPA_0_60,” i.e., time (in seconds) to accelerate from 0 to 60 miles-per-hour. However, this is not a direct measure. This measure is computed using a formula from an EPA publication that converts horsepower-to-weight ratio into an estimated acceleration time. The measure is computed at a high level of detail, requiring knowledge of the transmission type. These figures are then averaged, as discussed in Appendix.

Average market prices as a function of Model Year in December 2001 for various combinations of Vehicle Classes are shown in Figures 2.1 and 2.2. Figure 2.1 gives average market prices by major body type (Car, Pickup, Van, SUV). Curves for Car and SUV are similar to one another from 2001 to 1996, as are Pickups-Vans. For earlier model years SUV prices drop to a point intermediate between Cars and Pickups/Vans. As model years get older, prices for all body types converge.

Figure 2.2 gives more detail on market prices to illustrate a point. In this figure, vehicles are further divided into Prestige versus Non-Prestige. There are no Prestige Pickups or Vans. The only Prestige SUVs begin in model year 1996, which explains the pattern in Figure 2.1. With the additional level of detail in Figure 2.2, it can be seen that prices for Non-Prestige Cars, Pickups, and Vans are similar to one another, and Non-Prestige SUVs are priced a bit higher. There is a substantial gap between Prestige and Non-Prestige vehicles, with Prestige Cars and SUVs having similar prices from 1996-2001.

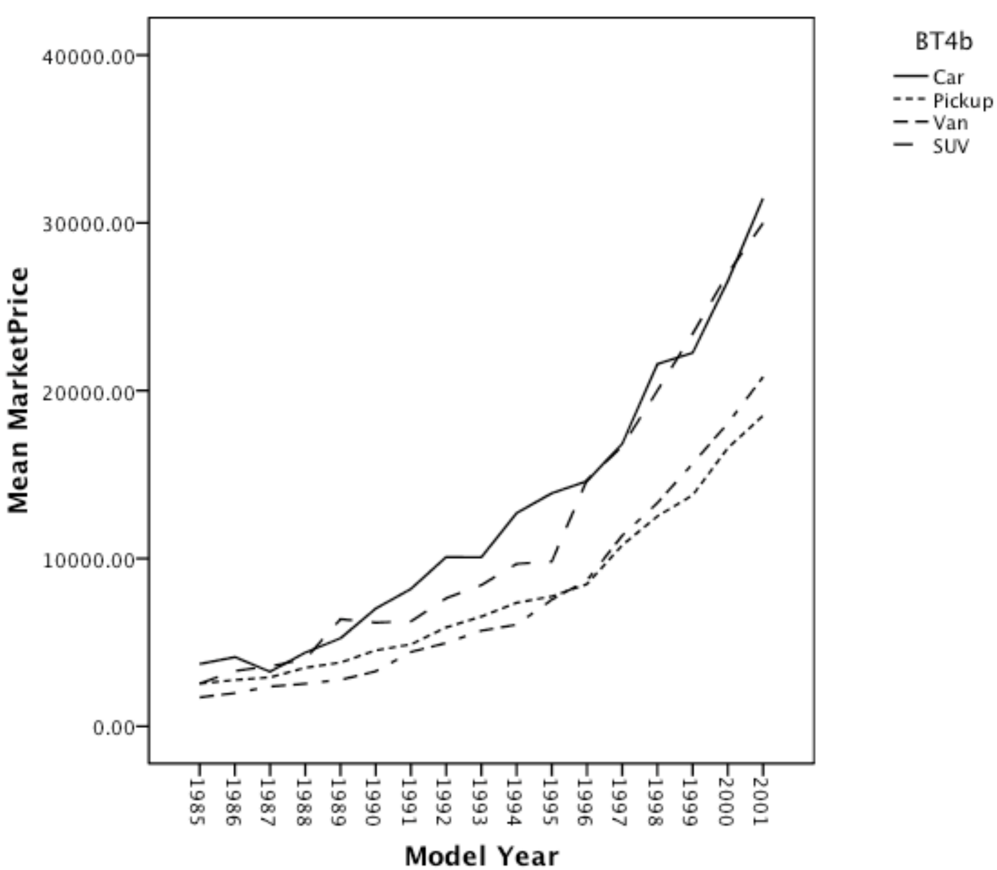

Figure 2.1 Ave. Market Prices by Body Type and Model Year (December 2001)

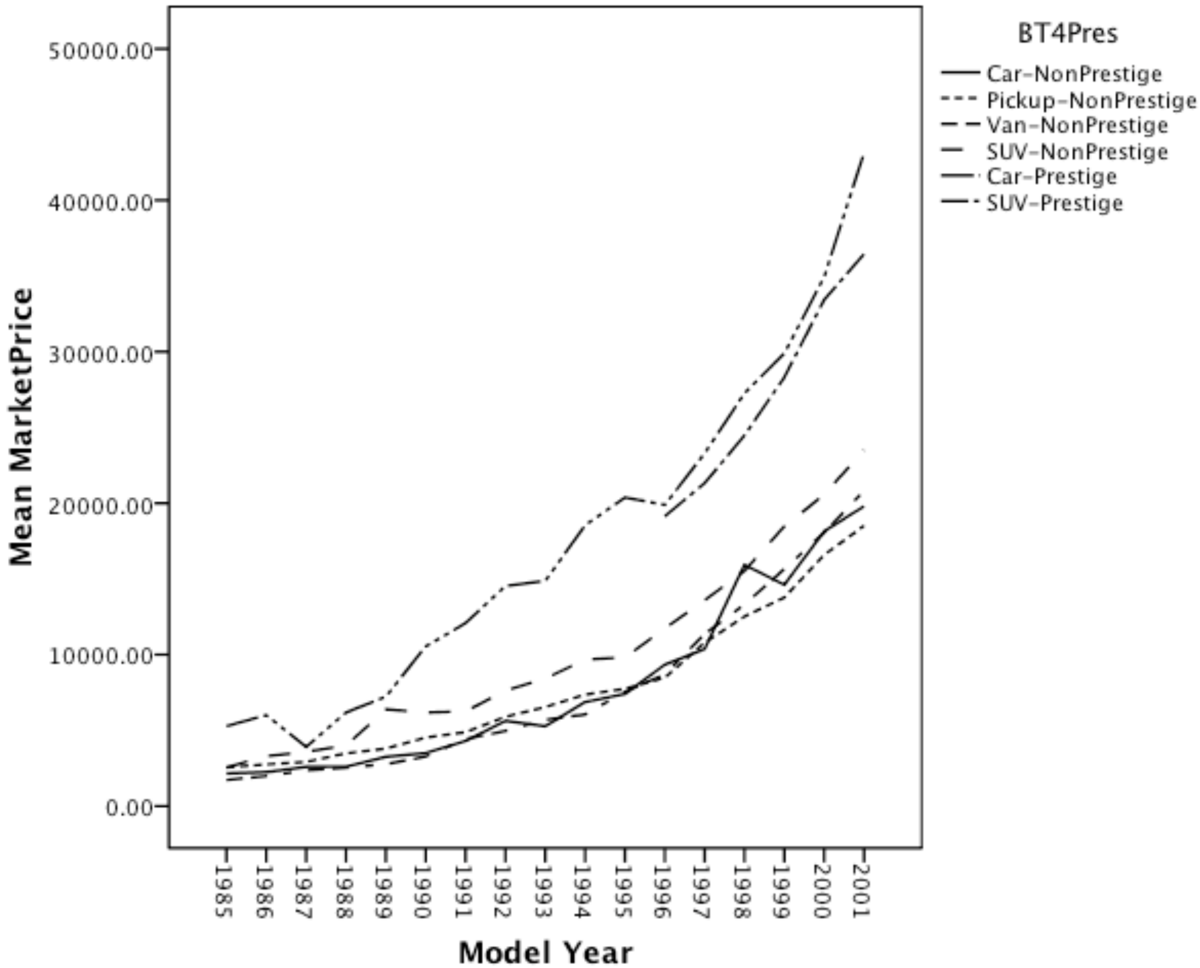

Figure 2.2 Ave. Market Prices by Body Type/Prestige Level and Model Year (December 2001)

2.2.2 Fuel Economy

Figure 2.3 shows average fuel economy for Body Type/Prestige level by model year. On average, the often-stated observation that fuel economy has remained relatively flat for a wide range of years is illustrated by this figure. The level of detail in Figure 2.3 also illustrates some other features of fleet fuel economy. For 1985, Non-Prestige Cars have the highest combined MPG, followed by Pickups, Prestige Cars, Vans, and Non-Prestige SUVs, respectively. In all years, the average fuel economy for Non-Prestige Cars is substantially higher than the light duty trucks, and also Prestige Cars. Prestige Car fuel economy lies below Pickups and above Vans until about 1995, when the steady downward trend in fuel economy for Pickups creates a crossover. The fuel economy of SUVs is well below the rest of the fleet.

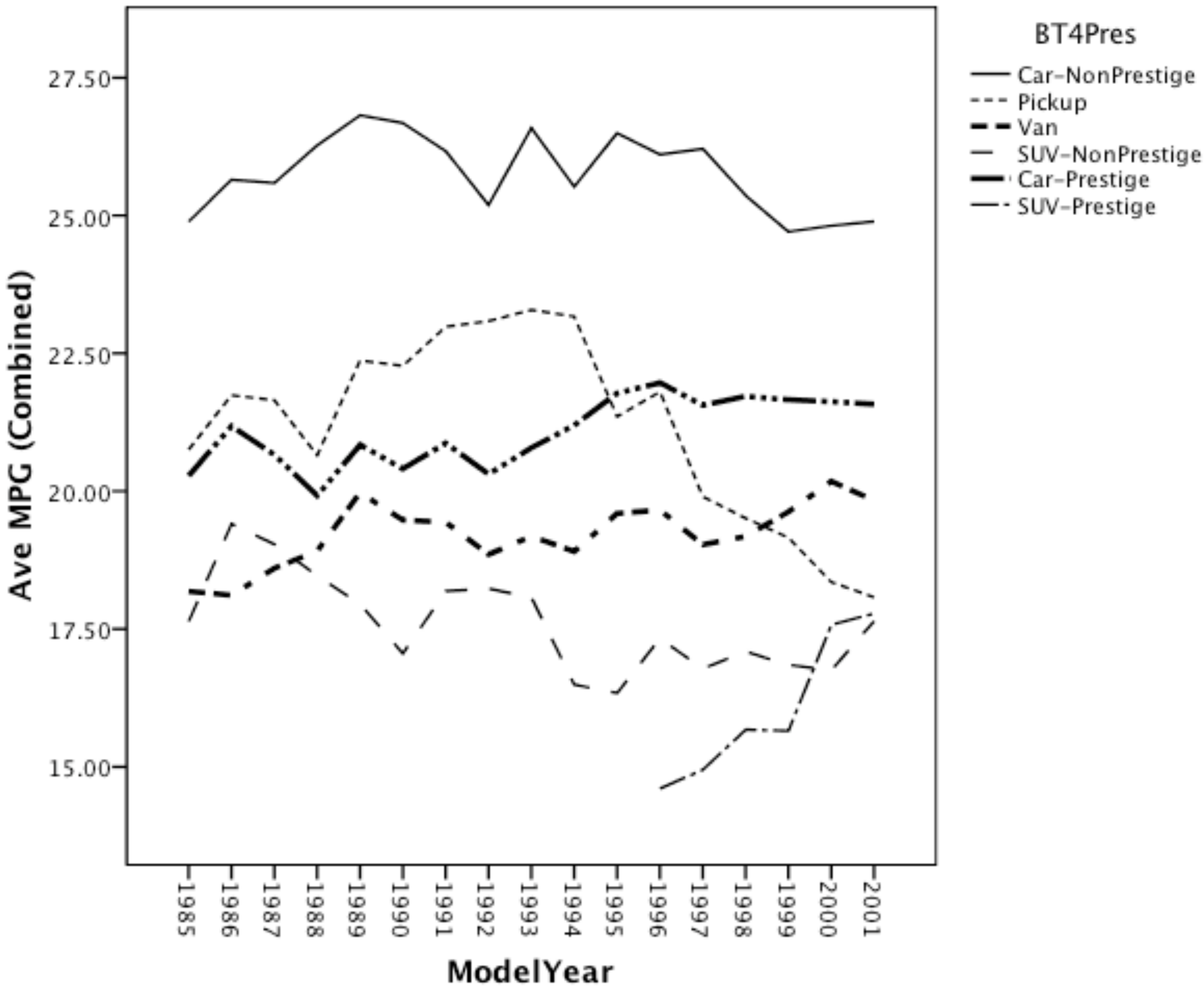

Figure 2.3 Average MPG (Combined) by Body Type/Prestige and Model Year

2.2.3 Performance

Average performance (measured by EPA_0_60) for Body Type/Prestige groupings by Model Year are given in Figure 2.4. In contrast to fuel economy, there is a noticeable upward trend in Performance (downward trend in 0-60 time) for most vehicle types, and a clear separation between Prestige Cars and all other vehicle types. Figures 2.3 and 2.4 illustrate an often-discussed issue in policy debates: Given available fuel technology, there is generally a tradeoff between fuel economy and performance, and in recent years advances in fuel technology are used primarily to improve performance while leaving fuel economy relatively flat.

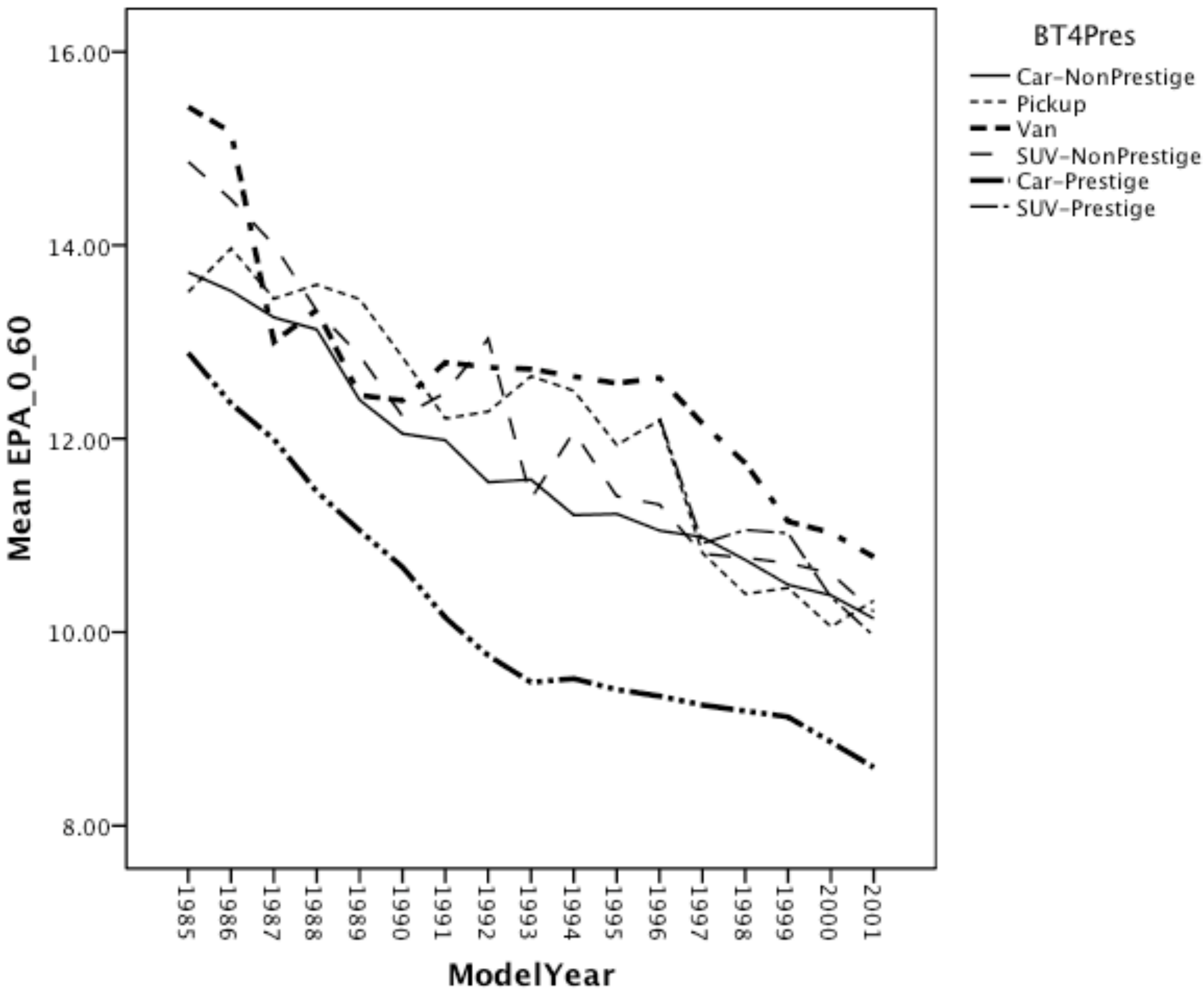

Figure 2.4 Average Performance by Body Type/Prestige and Model Year

3. Caltrans Travel Survey

The main household database used for updating CARBITS in this project is the **2000-2001 California Statewide Travel Survey**, which we will frequently refer to as the "Caltrans Travel Survey," or the "Caltrans Survey." The main reference is the survey's Final Report—see Bibliography. For purposes of background, the following is an excerpt from the Executive Summary of the Final Report:

> The California Department of Transportation (Caltrans) maintains a statewide database of household socioeconomic and travel information, which is used in regional and statewide travel demand forecasting. The most recent database, prior to this survey, contained data from the last statewide survey that was conducted in 1991. The 2000-2001 California Statewide Household Travel Survey was

conducted to update the database and will be used to help refine travel estimates, models, and forecasts throughout the State. The resultant data set will be used to estimate and forecast trip generation and distribution, mode choice, and assignments, as well as for vehicle emissions analyses and estimates.

The 2000-2001 **survey was conducted between October 2000 and December 2001** among households located in each of the 58 counties throughout the State. A total of **17,040 households** participated in the survey. Household socioeconomic data gathered in this survey includes information on **household size**, **income**, **vehicle ownership**, **employment status** of each household member, and housing unit type among other data. Travel information was also collected including trip times, mode, activity at location, origin and destination, and vehicle occupancy among other travel-related data. [Emphasis added.]

As discussed in previous parts of this report, the Caltrans survey has a large sample size, follows careful data collection procedures, and provides weight factors that make it an attractive option for our purposes. The items in bold above are the main elements required for vehicle choice modeling using "revealed preference" data. Table 3.1 reproduces key household statistics from the survey's final report.

Household Vehicles Available	21,448,770
Vehicles in Use on Average Weekday (71%)	15,252,463
Full-time Employees	10,130,359
Licensed Drivers	19,696,497
Occupied Housing Units	11,502,870
Single Housing Units	68%
Multiple and Other Housing Units	31%
Median Household Income	$54,946
Persons Per Household	2.8
Vehicles Per Household	1.9
No Vehicles	9.3%
One Vehicle	29.7%
Two Vehicles	37.7%
Three or More Vehicles	23.4%
Licensed Drivers Per Household	1.7

Table 3.1 Key Household Statistics from 2000-2001 California Statewide Household

The survey methodology includes the development of household weights that, when applied, provide a way to compute statistics (as in Table 3.1) that represent the entire California population. In particular, the weights are chosen so that certain statistics match those of the 2000 Census—see Chapter 6 of the Caltrans Survey Final Report.

3.1 Caltrans Survey Data Tables

Following standard database management practices, the data set is sub-divided into separate tables that correspond to three key entities: Households, Persons, and Vehicles. In this form, information is stored in a way that avoids inefficient replication of data elements. The three tables are linked together through a household id number (SAMPN). Documentation on selected variables from the Household and Vehicle tables is replicated in Tables 3.2 and 3.3, respectively. Important Household variables for choice modeling include income (INCOME), household size (HHSIZE), and number of workers (NWORK)—see Section 4. Identification of household ownership levels and characterization of vehicle holdings on the basis of body type, year, make, and model are also important, and present a number of practical challenges (to be discussed). The Persons table (not shown here) contains details for individual household members, including age, occupation, educational level, etc. The next sections explore data issues in more detail.

Var Name	Variable Description	Data Type	Width	Values
RECTYPE	Record Type	N	1	1=Household Data
SAMPN	HH ID Number	N	7	Assigned unique identifier
HHSIZE	Number of persons in household	N	2	Ordinal Variable
TOTVEH	Number of motorized vehicles available for use by HH members	N	2	Ordinal Variable
OWN	Owner/Renter Status	N	1	1=Own; 2=Rent; 7=Other, 8=DK, 9=RF
INCAT	Income Category	N	1	1=Above 50K; 2=Below 50K; 9=DK/RF
INCOME	Total 1999/2000 annual household income	N	2	1=<$10,000; 2=$10,000-$24,999; 3=$25,000-$34,999; 4=$35,000-$49,999; 5=$50,000-$74,999; 6=$75,000-$99,999; 7=$100,000-$149,999; 8=$150,000+; 9=DK/RF
NWORK	Number of HH Workers	N	2	Ordinal Variable
NSTUD	Number of HH Students	N	2	Ordinal Variable
WDWGT	Weekday Weight	N		

Table 3.2 Selected Household Variables from Caltrans Survey

Var Name	Variable Description	Data Type	Width	Values
RECTYPE	Record Type	N	1	3=Vehicle Data
SAMPN	HH ID Number	N	7	Assigned unique identifier
VEHNO	Vehicle Number	N	2	
MAKE	Vehicle X -Make	C	2	1=Acura; 2=Audi; 3=BMW; 4=Buick; 5=Cadillac; 6=Chevrolet; 7=Chrysler; 8=Dodge; 9=Ford; 10=Geo; 11=GMC; 12=Harley Davidson; 13=Honda; 14=Hyundai; 15=Infiniti; 16=Isuzu; 17=Jaguar; 18=Jeep; 19=Kawasaki; 20=Kia; 21=Lexus; 22=Lincoln; 23=Mazda; 24=Mercury; 25=Mercedes-Benz; 26=Mitsubishi; 27=Nissan; 28=Oldsmobile; 29=Plymouth; 30=Pontiac; 31=Porsche; 32=Range Rover; 33=Saab; 34=Saturn; 35=Subaru; 36=Suzuki; 37=Toyota; 38=Volkswagen; 39=Volvo; 40=Yamaha; 41=Daewoo; 42=Dotson; 43=International; 44=Winnebago; 45=MG; 97=Other, specify; 98=Don't know; 99=Refused
O_MAKE	Other make	C	60	
MODEL	Vehicle X-Model	C	60	
YEAR	Vehicle X - Year	F	4	8888=Don't know; 9999= Refused
BTYPE	Vehicle X -Body Type	N	2	1=Auto; 2=Van, 3=RV; 4=Sport utility vehicle; 5=Pick-up truck; 6=Other truck; 7=Motorcycle/Moped; 97=Other, specify; 99=DK/RF
WDWGT	Weekday Weight	N		

Table 3.3 Selected Vehicle Variables from Caltrans Survey

3.2 Caltrans Household Income Distributions

Household income distributions from the Caltrans Survey are presented in Table 3.4. The first columns of the table report distributions based on the un-weighted sample of 17,040 households. The final three columns show the same figures computed using the weights developed to match Census data to represent the 11.5 million households in California at that time. The table illustrates some common features of this type of survey work: Households at the lowest and highest income levels are frequently under-sampled, and many households (12-13% in this case) refuse to provide income information.

	Unweighted			Weighted		
	Freq	Percent	Valid Percent	Freq	Percent	Valid Percent
<$10,000	732	4.3	4.9	984705	8.6	9.7
$10,000-$24,999	2419	14.2	16.3	2003837	17.4	19.7
$25,000-$34,999	2244	13.2	15.1	1113007	9.7	11
$35,000-$49,999	2369	13.9	15.9	1297487	11.3	12.8
$50,000-$74,999	3389	19.9	22.8	1774103	15.4	17.5
$75,000-$99,999	1850	10.9	12.5	1103269	9.6	10.9
$100,000-$149,999	1268	7.4	8.5	1103019	9.6	10.9
$150,000+	583	3.4	3.9	775768	6.7	7.6
Total	14854	87.2	100	10155194	88.3	100
Don't Know/Refused	2186	12.8		1347671	11.7	
	17040	100		11502866	100	

Table 3.4 Household Income Distributions in the Caltrans Travel Survey

3.3 Vehicle Holdings

Another distribution of interest is the level of vehicle holdings by households. Despite the reference to "vehicle ownership" in the Executive Summary of the Caltrans Final Report, note that the survey generally relies a related measure termed "vehicle availability", i.e. the variable TOTVEH (Number of motorized vehicles available for use by HH members)—see Table 3.2. Using this variable in conjunction with weights yields the statistics in Table 3.1. An expanded distribution is given in Table 3.5. By this measure, fewer than 10% of California households have no motorized vehicles available (3.5 % of the sample). About 68% of households (73% of the sample) hold one or two vehicles. The mode in California is two-vehicle households.

	Unweighted		Weighted
No. of Vehicles	Frequency	Percent	Percent
0	601	3.5	9.3
1	5123	30.1	29.7
2	7343	43.1	37.7
3	2742	16.1	16
4	861	5.1	4.9
5	237	1.4	1.5
6	81	0.5	0.6
7	32	0.2	0.2
8	13	0.1	0.1
9	7	0	0
Total	17040	100	100

Table 3.5 "Vehicle Availability" Distribution for Caltrans Survey Households (see text for definition of vehicle availability)

However, one potential issue for this project is that "availability of motorized vehicles" is not necessarily equivalent the choice of "vehicle holdings" that we are concerned with, i.e., the household's light-duty vehicles. Specifically, in the Caltrans Survey "motorized vehicles" includes motorized vehicles of all types, as indicated in the text of the survey question:

> Question 19: "How many vehicles are presently available to members of your household? This includes all cars, vans, trucks, RVs, SUVs, motorcycles and mopeds, whether owned or leased or provided by an employer."

In contrast, consider the wording of the vehicle question used in the 2000 Census:

> Question #43: "How many automobiles, vans, and trucks of one-ton capacity or less are kept at home for use by members of your household?" There are seven possible responses to this question ranging from "none" to "6 or more." Note that this question does not ask about "vehicle ownership" per se, but about vehicles "kept at home" whether they are owned, leased, borrowed or company vehicles.]

The Census definition more closely matches the definition of vehicle holdings we are developing choice models for. However, comparing these two definitions raises a potential question about the validity of the weights in the Caltrans Survey, because it appears that the weights were constructed under the assumption that the two definitions are the same.

Another issue we faced in working with the Caltrans data was our discovery that the vehicle data were "dirty" in a number of ways, as can happen in surveys of this type. Relevant vehicle variables used in this project include body type, year, make, model, and fuel type of household vehicles—see Table 3.3. Problems we encountered included:

1. Item non-response, i.e., missing items (Don't Know or Refused) in variables for Year, Make, or Model of vehicle.
2. Limited information in Model variable (e.g., "Car" rather than the actual model name).
3. Errors in data entry, as evidenced by:
    a. Miss-matches between Make and Model (e.g., Nissan Camry).
    b. Miss-matches between stated body type and other variables. (For example, the body type could be listed as "Moped" for a 1999 Toyota Camry.)
    c. Miss-spelled model names, creating difficulties in vehicle matching.
    d. Miss-matches between year and model (e.g., a 1985 Toyota Prius does not exist, so there is a miss-match between year and make/model).

In addition, there were a relatively large number of very old vehicles in the data set. This can happen in a survey of this type due to sample response bias, e.g., individuals with a strong interest in cars might be "collectors," and would also be more likely to respond to the survey. For our work, we limited the "window" for vehicles to the 20-year period 1981-2001 for purposes of choice modeling (see Section 4). Constructing a data set to be used for choice model estimation requires that vehicles in the Caltrans Survey be 'identified' in enough detail to assign them to the vehicle classes discussed in Section 2. So, even though there were problems in exactly matching vehicles at the Year-Make-Model level, we established procedures to assign vehicle classes using available information. This is discussed in more detail in section 3.4. For now, we summarize some facts about the Caltrans vehicle data.

For a summary of vehicles successfully matched to vehicle technology data on the basis of Year, Make, and Model information for model years 1981-2002, see Table 3.6. The table is constructed using the Caltrans survey weights, indicating that vehicles representing 17.7M of the 21.4M (83%) are successfully matched. Data are presented in cross-tab form to highlight some of the data quality issues. Specifically, the "matched body type" is the body type from the vehicle technology database, whereas "btype" is the body type recorded in the survey data. Although they are highly correlated, they frequently disagree. In some cases the disagreements are significant, e.g., cases where Cars are assigned a body type (btype) of "moped/motorcycle" or "RV".

	Matched Body Type				
btype*	Car	Pickup	Van	SUV	Total
Auto	10,413,401	95,413	110,645	162,297	10,781,756
Pickup	24,167	2,831,608	14,356	39,922	2,910,053
Van	62,225	24,920	1,589,550	10,270	1,686,965
SUV	194,625	61,410	10,151	1,852,022	2,118,208
Other truck	14,584	61,112	8,111	67,316	151,123
RV	4,279	474	2,801	14,194	21,748
Moped/Motorcycle	20,740	161	1,724	1,702	24,327
Other	5,368			442	5,810
DK/Ref	4,533				4,533
Total	10,743,922	3,075,098	1,737,338	2,148,165	17,704,523

Table 3.6 Successfully Matched Caltrans Vehicles (1981-2002)
* btype variable from Caltrans Survey

Table 3.7 summarizes the status of unmatched Caltrans vehicles, and illustrates various data issues. There are a number of ways to look at these figures. First, if we omit concerns about the unreliability of the btype variable, this Table yields an estimate of 3M Autos, Pickups, Vans, and SUVs that are not included in Table 3.7, for a total of 20.7M light-duty vehicles out of the 21.4M "available vehicles," or about 97%. So, it may be using "available motorized vehicles" to represent "vehicle holdings of light-duty vehicles" is a reasonable approximation. About half of these 3M vehicles (1.5M, or 7% of the total) are excluded from Table 3.6 because they are older vehicles (model year < 1981). A relatively small number (500K, or 2%) are unmatched due to a missing model year. In all, the light duty vehicle fleet with model years 1981-2002 is estimated to lie in the range 18.6-19.1M vehicles, of which we have matched 17.7M (approx. 95%).

		YearFlag			
btype	DN/REF	1981-2002	1965-1980	< 1965	Total
Auto	331,794	508,128	706,063	153,028	1,699,013
Pickup	98,459	283,452	416,376	84,495	882,782
Van	48,643	98,566	100,040	3,415	250,664
SUV	28,938	70,011	69,232	10,973	179,154
Other truck	10,150	57,699	40,123	4,390	112,362
RV	4,641	104,881	51,861	892	162,275
Moped/Motorcycle	21,378	206,850	34,832	3,115	266,175
Other	2,650	4,724	2,051	1,152	10,577
DK/Ref	89,918	86,021	4,923	186	181,048
Total	636,571	1,420,332	1,425,501	261,646	3,744,050

Table 3.7 Summary of Unmatched Caltrans Vehicles

3.4 Vehicle Matching

This section provides additional details on the problem of "vehicle matching" using the Year-Make-Model variables from Table 3.2. Make information is collected in the form of a numerical code; however, the Model is typed in as a character string by an interviewer collecting the information from a respondent over the phone. Cleaning these data and performing the necessary steps to cross-reference these vehicles to entries in a Vehicle Technology Database can be a monumental task. In addition, this illustrates an important issue faced in vehicle choice modeling: the level of detail obtained in a household survey like this one is relatively coarse. Information on such things as trim levels, engine size, transmission, and drive train cannot be ascertained in a survey like this one.

To support the requirements of this project, Caltrans Vehicles were matched to vehicle records in the Chrome VINMatch database on the basis of Year-Make-Model (for more information on the Chrome database, see Appendix A). This is challenging because part of the matching process requires comparison of character string vehicle descriptions with no common standard. Vehicles were matched to the highest level of detail possible. In most cases, this resulted in multiple Chrome records being matched to each Caltrans Vehicle (since Chrome vehicle records are relatively detailed). This approach provided the maximum amount of flexibility for matching Caltrans Vehicles to vehicle technology data by using the more detailed Chrome records as the potential links. Specifically, this provided the flexibility to accommodate alternative Vehicle Class definitions should the need arise (now or in the future). To provide the data necessary for estimating the models discussed in Section 4, Caltrans Vehicles were linked to the appropriate Vehicle Classes from Section 3 to represent each household's vehicle holdings. Although there are usually multiple Chrome vehicles associated with each Caltrans Vehicle, the relative lack of detail at the Vehicle Class level can help simplify the process of matching a Caltrans Vehicle to a Vehicle Class. Specifically, in most cases all of the Chrome vehicles matched to a Caltrans Vehicle belong to the same Vehicle Class. (In those cases where this is not true, the assignment is made at random using weights created from processing the DMV data.) The next section discusses how the choice of vehicle holdings by households is modeled.

4. CARBITS 2.0 Vehicle Market Demand Models

This section describes development of a vehicle market demand model for CARBITS 2.0. Specifically, this is the model that performs the calculations in Step 6 ("Simulate Personal Vehicle Market Behavior for Current Year") of the CARBITS Vehicle Market Simulation Framework discussed in Section 1. CARBITS simulates the vehicle choice behavior for households in response to current market conditions. It uses a sample of households (with weights) to represent California in each time period. Although there are a number of additional details associated with simulating market behavior, the fundamental requirement is for some type of choice model to "simulate" each household's "vehicle demand" in response to a given set of market conditions.

There are a number of options for modeling household-level vehicle purchase/ownership behavior. At the household level, behavior is formulated in terms of (i) a universe of choice options, and (ii) choice probabilities for those options. These "choice options" can be characterized in various ways, e.g., the choice to purchase a vehicle, the choice to hold a vehicle portfolio, or, the choice to engage in a vehicle transaction (replacement, addition, or disposal of currently held vehicles). This section reviews background on vehicle choice models, describes the approach taken in CARBITS 2.0, and presents model estimation results.

4.1 Background on Vehicle Choice Models

There are many types of vehicle choice models in the literature, and choosing which type to use is based on a number of factors, including the purpose of the model. For example, many models of vehicle demand are exclusively focused on the new vehicle market. However, policy-related models like CARBITS are required to address the entire vehicle fleet (both new and used vehicles), which includes a much larger number of choice options than when considering the new vehicle market alone. Moreover, the decision-making unit in CARBITS is the Household (not an individual making a single purchase). In this section we briefly review some relevant background. For a more complete introduction, see Bunch and Chen (2008). There are two options that are generally available: Holdings models, and transactions models. For a holdings model, a household's decision-making process is described (informally) as follows:

1. For an entire one-year period, a household will own and use a specific portfolio of one or more vehicles (or, the household may own no vehicles).
2. Once per year, households revisit their entire set of vehicle ownership decisions.
3. At the annual "decision point," household's perform a "complete analysis" in which they make the following decisions for the *coming* year:
    a. How many vehicles to own (0, 1, 2 or more).
    b. Conditional on the number of vehicles, *which* vehicles to own.
4. A choice model estimates the probability of each "holdings outcome."

In contrast, a transactions model is described as follows:

1. A household starts in a "base period" with a set of vehicle holdings (including the possibility of "no vehicles").
2. At certain points in time (perhaps annually), a household makes the following sequence of decisions:
    a. Should we transact? (Yes or No)
    b. If YES, do we:
        i. Replace one of our current vehicles?
            1. If so, which vehicle is to be replaced?
            2. What vehicle will be purchased as the replacement vehicle?
        ii. Add a new vehicle to the household fleet? If so, which one?
        iii. Sell one of the currently held vehicle(s)? If so, which one?
3. A choice model estimates the probability of each "transaction outcome."

The argument for a transaction model is that it seems like a more "realistic" description of household vehicle purchase behavior. In particular, a household will go along for a period of time (perhaps years) until some event "triggers" the need for a transaction.

During this period vehicles are driven, they accumulate miles, get worn out, require repairs, etc. In this regard, transactions models are considered to be better able to capture "dynamic effects" such as inertia. In contrast, a simple holdings model would seem to be vulnerable to a much quicker market response to changes in market conditions.

Based on this discussion, a transactions model would appear to be a superior choice. However, transactions models:

1. Require detailed household level data on such transactions in order to support model estimation, i.e., panel data.
2. Are much more computationally intensive that holdings models (when implemented based on the above descriptions).
3. Have not been demonstrated to be superior in any published academic studies.

CARBITS 1.0 was implemented as a transactions model as part of a University of California research project in the mid-1990's. Choice models were estimated using a panel data set collected on California households as part of that project. The market simulation was implemented using a "pure microsimulation" approach, as implied by the above description. Specifically: In each period a household's choice probabilities are *conditional* on a specific set of vehicle holdings that a household has carried forward from the previous period. Then, based on these probabilities, a transaction is *simulated* for the current period. In most cases (as in the real world), a household will elect to retain its current set of vehicles for another year. A very large number of households, and many repeated replications of the simulation, are required in order to obtain an estimate of annual market vehicle distribution.

In contrast, a holdings model (as described above) can be estimated using the more usual version of household survey data in which households are interviewed at a single point in time, and are asked to report their current vehicle holdings. Choice models are estimated using the household sample. In the market simulation, the choice model produces a *probability* for each household's choice options. In this case, the market vehicle distribution can be computed by taking a weighted average of the choice probabilities over the sample of households. These numbers are *deterministically* computed, with no "simulation noise."

This discussion provides some additional background on why CARBITS 2.0 has been implemented as a holdings model. As noted previously, the major reason is the availability of the Caltrans Travel Survey Data. Specifically,

1. This survey contains a very large number of California households, and also includes weights developed by Caltrans so that the survey sample can be used to "represent" California.
2. This survey is a cross-sectional survey (not a panel survey) and contains the usual vehicle information, which is limited to vehicle holdings (not transactions).
3. In addition to the large sample size, the data in this survey are five years more recent the data used in CARBITS 1.0. Moreover, the panel survey data used in

> CARBITS 1.0 was a special-purpose survey that is highly unlikely to be replicated. In contrast, the Caltrans Survey is likely to be updated at regular intervals. Historically, it has been replicated every ten years or so, and a certain level of continuity and consistency in methodology has been maintained.

One final note: the above description of the two types of models is rather stylized, and designed to illustrate certain points. In reality, the two types of models can actually be more similar than they appear, depending on what features are included.

For example, some holdings models can be estimated with a "transactions dummy variable" *if* information on the household's vehicle portfolio from the previous period is available. This can be used to identify an "inertia" effect by representing the fact that, for a household to *switch* vehicle holdings requires a transaction to occur (at some cost to the household), so that the household's current portfolio has a much higher probability of being chosen than the other options. If this feature is added, the model results can be interpreted as being "transactions based" rather than "holdings based," even though the computations are very similar.

The key question in all of this: How much information about each vehicle's *holding time* is included? *If* the only information carried forward in the model is whether or not a vehicle was held during the *previous period*, then the two models are essentially the same. However, in CARBITS 1.0 the model kept track of exactly *how many* periods each vehicle was held by a household, and the probability of a transaction was computed conditional on how long the household had owned the vehicle. This feature created the requirement for a pure microsimulation approach, as indicated earlier.

4.2 Vehicle Holdings Models for Caltrans Travel Survey Data

This section summarizes vehicle holdings choice models estimated using the Caltrans Travel Survey Data. The models are of the conditional-multinomial-logit/nested-multinomial-logit type similar to those that have appeared elsewhere. A full discussion is beyond the scope of this report, but relevant references include Train (1986), Berkovec (1985), Hensher, et al. (1992), and Bunch and Chen (2008).

As discussed in the previous section, a complete vehicles holdings choice model includes both the choice of *how many* vehicles to own, and *which* vehicle(s). One model form that has been applied in these settings is the nested logit model. The top level has "branches" that correspond to the decision of how many vehicles to own (0, 1, 2, etc.). Under each (non-zero) branch are the options for vehicle portfolios that a household may chose to own. A typical nested logit model for vehicle holdings is illustrated in Figure 4.1.

One decision when developing a holdings model is how large the maximum vehicle portfolio size should be. Most models in the literature (e.g., Train 1986) stop with vehicle pairs, as depicted in Figure 4.1. A few references estimate models for three-vehicle households (e.g., Berkovec 1985). The vehicle holdings distribution for the Caltrans Survey households was provided in Table 3.3. Roughly 28% of households hold three or more vehicles. A practical issue is that the number of possible vehicle portfolios increases dramatically when the portfolio size increases. In Section 2 we developed 350 Vehicle Classes to represent the vehicle market in 2001. A one-vehicle household therefore has 350 options to choose from. A two-vehicle household could theoretically hold one of the possible pairs that can be constructed from the 350 vehicle classes, yielding 350*349/2 = 61,075 portfolio options. There are over 7 million possible vehicle portfolios of size 3. Even if the model is limited to pairs, some type of sampling procedure is typically employed to construct choice sets with a smaller number of options.

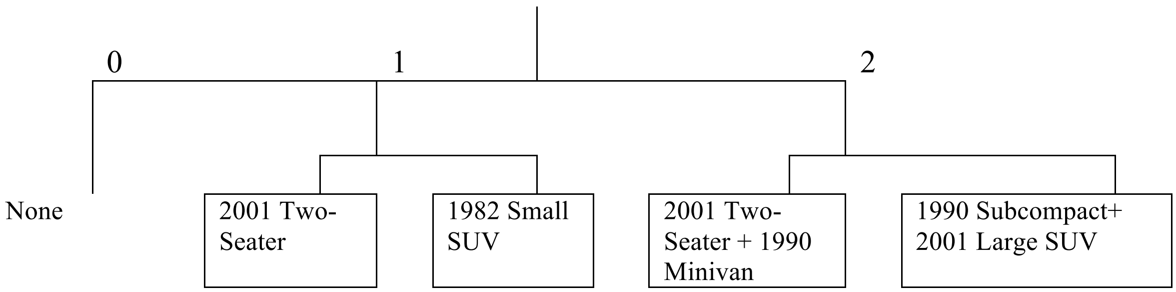

Figure 4.1 Nested-logit Structure for a Vehicle Holdings Model

Our main modeling concern is capturing the interaction effects that would occur when a household decides to hold more than one vehicle. Some combinations are more attractive than others, e.g., households frequently hold more than one body type so that their fleet can be used for multiple purposes. (The three-vehicle models estimated by Berkovec ignored such interaction effects in order to make the model estimation more tractable.) For this project, we followed the typical practice of estimating holdings models with 0, 1, and 2 vehicles. When simulating market behavior, a weighting procedure is employed so that the 2-vehicle model is used to represent the vehicle choices of households with more than two vehicles.

In a nested logit model, the "utility" of how many vehicles to own (one or two) is a function of the "expected maximum utility" conditional on the quantity choice. Consider the case of the choice of one vehicle, conditional on the assumption that one vehicle is being chosen. A household (n) will choose to hold one of the J Vehicle Classes that are available. Using a multinomial logit model (MNL), household n's choice probability for Vehicle Class c is given by

$$P_{cn,1} = \frac{e^{V_{cn}}}{\sum_{j=1}^{J} e^{Vjn}}$$

where V_{jn} is household n's preference index for Vehicle Class j. When choosing whether to own one or two vehicles, the expected maximum utility from the decision to purchase one of the J Vehicle Classes is given by the so-called Inclusive Value (IV):

$$IV_{n1} = \ln \sum_{j=1}^{J} e^{Vjn}.$$

An analogous expression can be derived for the conditional two-vehicle choice model. If these values were known, these and some additional factors (e.g., household income, size, etc.) would be expected to determine the probability of choosing one versus two vehicles. The vehicle quantity choice model for household *n* can be written as

$$Q_{nm} = \frac{e^{W_{nm}}}{e^{W_{n1}} + e^{W_{n2}}}$$

where Q_{nm} is the probability that household *n* holds *m* vehicles, W_{n1} and W_{n2} are the preference indexes for holding 1 and 2 vehicles, respectively, and each would include their respective inclusive values, as well as other factors, as explanatory variables. The full nested logit model can be directly estimated; however, a typical practice (following the above narrative) is to perform sequential estimation as follows:

1. Conditional one-vehicle household choice model.
2. Conditional two-vehicle household choice model.
3. Vehicle-quantity choice model.

This approach has been taken to estimate household-level vehicle holdings choice models using the Caltrans data. Results are presented in the next sections.

4.2.1 Conditional One-Vehicle Choice Model

Consider the case of a Caltrans Household that has already decided to hold one vehicle. A one-vehicle-household choice model can be estimated using the sample of one-vehicle households from the survey. Based on the discussion in section 3, the household has 350 Vehicle Classes from which to choose (summarized in Table 2.5). As noted above, the conditional choice probability of household *n* choosing Vehicle Class *c* can be modeled using a multinomial logit model, V_{jn} is household *n*'s preference index for Vehicle Class *j*, given by the linear-in-parameters form

$$V_{jn} = \sum_{k=1}^{K} \beta_k Z_{k,jn}.$$

The vector Z_{jn} contains explanatory variables that are a function of vehicle attributes for Vehicle Class *j* and household demographics from household *n*, and β is a K-dimensional vector of model parameters. Household demographics used in our models are:

1. Household income categories
    a. Income < $10K
    b. $10K ≤ Income < $25K
    c. $25K ≤ Income < $50K
    d. $50K ≤ Income < $75K
    e. Income ≥ $75K
    f. Income < $75K
2. Household size
    a. Household Size > 3
    b. Household Size ≤ 3
    c. Household Size > 2
    d. Household Size ≤ 2

Vehicle attributes include:

1. Dummy variables for Body-Type-Size classes
    a. TwoSeater [Car]
    b. Small [Car]
    c. Midsize [Car]
    d. Large [Car]
    e. Truck [Pickup]
    f. Van
    g. SUV
    h. LargeSUV
    i. SmallSUV
2. Price (vehicle market price, in year-2000 $)
3. OpCost (fuel operating cost, in cents per mile)
4. Accel (acceleration time, seconds for 0-60 mph)
5. LnMods (Log of number of vehicle models in the vehicle class)
6. LnVAge (Log of vehicle age when vehicle age is ≥ 1, 0 otherwise)
7. Prestige dummy variable

The vehicle attributes chosen for these models were based on a number of factors, including a careful review of the literature and past experience. Price, fuel operating cost, and acceleration cover three very important aspects of vehicle choice that are included in essentially all (household-level) choice models. There are a number of possible measures of performance that could be used (e.g., top speed, horsepower, horsepower to weight ratio, etc.). We chose to use acceleration time because it is a measure that consumers can relate to in terms of their direct experience (in contrast to the engineering characteristics). This measure is frequently used in choice experiments in which respondent are asked to indicate their most preferred alternative. This keeps open the possibility of, e.g., updating these choice models using stated choice data should the need arise.

The other important dimension of vehicle functionality and size are captured relatively well by dummy variables related to Vehicle Class. We considered using some alternative measures of size such as passenger volume and luggage space (and even did some testing), and also vehicle footprint. However, these measures (i) add to the vehicle data requirements, and (ii) are less amenable to issues related to model re-calibration. In particular, and vehicle characteristic included in the vehicle choice model must be *forecasted* for any scenario analysis being performed. The log(Number of Models) attribute always raises concerns, but it has been shown to be important in models of this type, i.e., those that estimate choice at the vehicle class level. (A full discussion is beyond the scope of this report; see, e.g., Train 1986 as a reference.) In addition to the variables listed above, some interaction effects are also included (e.g., interaction of income category with Price, interaction of household-size dummy variables with different body-type-size dummy variables).

Table 4.1 gives estimates of a multinomial logit model for 4,410 one-vehicle households. The full choice set of 350 alternatives was used for each household (yielding a data set with 1,543,500 rows). The estimator is maximum likelihood, and results were obtained using Stata (Version 10.1).

```
Conditional (fixed-effects) logistic regression   Number of obs   =    1543500
                                                  LR chi2(29)     =    6585.46
                                                  Prob > chi2     =     0.0000
Log likelihood = -22540.757                       Pseudo R2       =     0.1275
```

yij	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
PrLT10	-.0001891	.0000169	-11.22	0.000	-.0002222	-.0001561
Pr10_25	-.000164	.0000121	-13.54	0.000	-.0001878	-.0001403
Pr25_50	-.0000932	.0000103	-9.02	0.000	-.0001134	-.0000729
Pr50_75	-.0000499	.0000103	-4.83	0.000	-.0000701	-.0000296
PrGT75	-.000032	.0000109	-2.94	0.003	-.0000534	-.0000107
PrMiss	-.0000852	.0000118	-7.20	0.000	-.0001084	-.000062
OpCost	-.2528365	.0279414	-9.05	0.000	-.3076005	-.1980724
Accel	-.2880763	.0265671	-10.84	0.000	-.3401469	-.2360057
Pres_GT75	-.4308841	.213239	-2.02	0.043	-.8488249	-.0129433
Pres_LE75	-1.157587	.1504355	-7.69	0.000	-1.452436	-.8627394
Car_GT3	-.2989819	.1685597	-1.77	0.076	-.6293528	.031389
TwoSeat	-2.133135	.2545379	-8.38	0.000	-2.63202	-1.63425
TwoSGT2	-1.719884	1.016602	-1.69	0.091	-3.712388	.2726192
PresTS	.6419015	.616637	1.04	0.298	-.5666849	1.850488
Subcompact	-.5827533	.0507627	-11.48	0.000	-.6822463	-.4832603
Midsize	.2291438	.0563366	4.07	0.000	.1187262	.3395615
Large	-.6116656	.118392	-5.17	0.000	-.8437096	-.3796216
PresLCar	1.159216	.154017	7.53	0.000	.8573481	1.461084
Tr_GT2	-.3725133	.1770807	-2.10	0.035	-.7195851	-.0254414
Tr_LE2	.0407489	.1102985	0.37	0.712	-.1754322	.2569301
Van_GT3	.7609437	.2287049	3.33	0.001	.3126902	1.209197
Van_LE3	-.5785032	.138389	-4.18	0.000	-.8497406	-.3072658
SUV_GT75	-.3073329	.2273527	-1.35	0.176	-.752936	.1382702
SUV_LE75	-.8390846	.1814757	-4.62	0.000	-1.19477	-.4833989
LSUV	.4237661	.2498143	1.70	0.090	-.065861	.9133932
SmallSUV	1.014431	.1393791	7.28	0.000	.7412534	1.287609
New	-.9890594	.0755862	-13.09	0.000	-1.137206	-.8409132
LnVAge	-.8244201	.0716202	-11.51	0.000	-.9647932	-.684047
LnMods	.6877352	.0679447	10.12	0.000	.5545661	.8209043

Table 4.1 Estimates of One-Vehicle Choice Model using Caltrans Data

The coefficient estimates are highly significant, and all have interpretations that are consistent with theory. The Price coefficients (which are interacted with six income categories) are negative, and get smaller in magnitude with increasing income category, i.e., households become less price sensitive as income increases. Coefficients on OpCost and Accel are both negative, and are of similar magnitudes (similar to other models in the literature that use these same units).

The base body-type-size category is Compact Car, with a normalized utility of zero (not shown). In this sample, Midsize has a positive coefficient, whereas TwoSeater, Subcompact, and Large cars have negative coefficients. However, the PrestigeLarge-Car interaction is strongly positive, so that the total utility of a PrestigeLarge Car is 1.16 – 0.61 = 0.55, making it the largest Car coefficient. All sizes of Cars have less utility when households have more than 3 members, and specification testing revealed that this occurs in about the same amount so that a single coefficient can be used.

4.2.2 Conditional Two-Vehicle Choice Model

Coefficients for two-vehicle households are given in Table 4.2. Recall that there are 350 Vehicle Classes. If one were to use all possible vehicle portfolios consisting of pairs, the choice set size would be approximately 61,000. This model was estimated using choice sets that were generated by a procedure designed to yield 45 vehicle pairs per household (discussed below). Maximum likelihood estimates were obtained for a sample of 5,393 households.

In the two-vehicle model, we follow the frequently used practice of using the *sum* of attributes for the two vehicles in the portfolio, e.g., Price is the sum of the two market prices, OpCost and Accel are the sum of the values for the vehicle pair, etc. As in the one-vehicle case, most coefficient estimates are highly significant, and have signs that conform to theory. As before, households with progressively higher incomes become less price sensitive. The coefficients for OpCost and Accel are similar to those in the one-vehicle case.

This model includes many dummy variables that capture the relative desirability of different pairs of vehicle types, e.g., Car_Truck, Car_Van, Car_SUV, Truck_Van, etc. In addition, the sizes of Cars in the portfolio can play a role. In this specification, the "base" combination is a pair of Cars where one is "Small" (Subcompact or Compact), and the other is "Large" (Midsize or Large). In addition, some of these are also interacted with household size indicators ($>$ 3 versus $\leq$ 3), income level ($\geq$ \$75K versus not), and Prestige.

To illustrate, "SmSm_GT3" denotes two small cars, and a household with more than 3 members. Similarly, "SmSm_LE3" denotes two small cars, and a household with fewer than four members. The signs of both coefficients are negative, indicating that two small cars are less preferred than the base alternative ("Small Car-Large Car"). Moreover, the coefficient for SmSm_GT3 is more negative than SmSm_LE3, which seems logical.

Conditional (fixed-effects) logistic regression

Log likelihood = -13742.809	Number of obs	=	242685
	LR chi2(33)	=	13573.04
	Prob > chi2	=	0.0000
	Pseudo R2	=	0.3306

yij	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
PrLT10	-.0002038	.0000167	-12.20	0.000	-.0002366	-.0001711
Pr10_25	-.000224	8.12e-06	-27.60	0.000	-.0002399	-.0002081
Pr25_50	-.0001894	5.03e-06	-37.65	0.000	-.0001993	-.0001795
Pr50_75	-.0001559	4.87e-06	-32.03	0.000	-.0001654	-.0001463
PrGT75	-.0001116	4.14e-06	-26.96	0.000	-.0001197	-.0001035
PrMiss	-.000141	5.69e-06	-24.77	0.000	-.0001522	-.0001299
OpCost	-.3069833	.0129964	-23.62	0.000	-.3324558	-.2815107
Accel	-.3768886	.0156296	-24.11	0.000	-.4075221	-.3462552
SmSm_GT3	-.4179995	.1329962	-3.14	0.002	-.6786673	-.1573317
SmSm_LE3	-.178349	.07208	-2.47	0.013	-.3196233	-.0370748
MidL_MidL	.1055342	.0818664	1.29	0.197	-.0549211	.2659894
HasPr_GT75	1.065289	.0833516	12.78	0.000	.9019232	1.228655
HasPr_LE75	.0781934	.0776889	1.01	0.314	-.0740739	.2304608
Car_Truck	.9636605	.0651551	14.79	0.000	.8359588	1.091362
MidL_Truck	.5590861	.0578555	9.66	0.000	.4456914	.6724808
Pr_Tr_GT75	-.5405837	.147469	-3.67	0.000	-.8296176	-.2515498
Pr_Tr_LE75	.2162532	.1133964	1.91	0.057	-.0059995	.438506
Car_Van_GT3	1.589949	.0969137	16.41	0.000	1.400002	1.779896
Car_Van_LE3	.3159519	.0843007	3.75	0.000	.1507256	.4811783
Car_SUV	1.144865	.0796171	14.38	0.000	.9888181	1.300911
Car_SUV_GT75	.7329727	.0868222	8.44	0.000	.5628045	.903141
Truck_SUV	2.594509	.1045276	24.82	0.000	2.389638	2.799379
Van_SUV	1.805282	.1401018	12.89	0.000	1.530687	2.079876
TrVan_GT3	2.709814	.1334341	20.31	0.000	2.448288	2.97134
TrVan_LE3	1.339568	.1260801	10.62	0.000	1.092456	1.586681
Van_Van	.5157106	.1995178	2.58	0.010	.1246629	.9067583
SUV_SUV	2.397272	.146996	16.31	0.000	2.109165	2.685379
Truck_Truck	.7401467	.1232593	6.00	0.000	.4985629	.9817304
LnSMods	2.328702	.0635653	36.63	0.000	2.204117	2.453288
numVG1	-2.529077	.1073028	-23.57	0.000	-2.739387	-2.318768
LnTotAge	-1.074708	.055758	-19.27	0.000	-1.183992	-.9654246
numTS	-.8002239	.1239146	-6.46	0.000	-1.043092	-.5573558
nTSGT3	-1.014007	.3739025	-2.71	0.007	-1.746842	-.2811715

Table 4.2 Estimates of Two-Vehicle Choice Model Using Caltrans Data

Essentially all of the other vehicle type combinations are preferred to the base alternative (i.e., they have positive and statistically significant coefficients). Generally speaking, most of these involve *different* types and sizes of vehicles, and there is a clear preference for variety. For example, the smallest coefficients are for two "Large" cars (MidL_MidL), Van_Van, and Truck_Truck (an apparent exception is SUV_SUV, with a relatively large coefficient). Combinations such as Car_Van, Truck_Van are more strongly preferred by households with more than two members, as might be expected, due to the desirability for extra space.

There are also interactions involving Prestige and Income level. Households with more than $75K in income have a higher preference for Prestige Cars. Interestingly, households with this income have a negative coefficient for the case where a Prestige Car is combined with a Truck. Another interaction involves Car_SUV. High-income households prefer this pair type more strongly.

As in the one-vehicle case, TwoSeaters have disutility. The coefficients here are for the *number* of TwoSeaters, which are negative. In addition, there is more disutility for larger households (more than 3 members). Finally, as in the one-vehicle model, coefficients on Log(Number of Models) , Log(Sum of Vehicle Ages) and number of New vehicles (defined as model year 2000 and 2001) have the expected signs.

A final note on choice set generation: Because it is impractical to include the full choice set of all possible pairs, subsets of alternatives are used. We elected to use an approach with more slightly more structure than a simple random sample. We followed the following procedure:

1. Generate all possible pairs of the 350 Vehicle Classes.
2. Randomize their ordering of the pairs.
3. Going through the list of households, one household at a time, "deal" *P* (e.g., 45) pairs to each household from the full set. Continue until there are no more pairs left in the "deck". (In other words, pairs are randomly assigned to households from the set of all possible pairs, without replacement).
4. If all households in the database have *P* pairs, stop. If there are still households in the database without an assigned pair: Go to Step 1 and repeat the process for those households without assigned pairs. (If the last household in Step 3 received a partial set of pairs, those pairs are discarded and this household becomes the starting point for the next iteration.)

This approach ensures full coverage of the space of all possible vehicle pairs, and should lead to more efficient estimates. This procedure is used for both estimation and simulation. In the case of estimation, the set must include the household's actual held vehicles. If the randomly assigned choice set does not already include the household's actual holdings, one of the pairs is replaced (at random) with the actual holdings. Note: The results in this report are based on using choice sets with P = 45 (45 vehicle pairs). However, ongoing testing could lead to variations with, e.g., larger choice set sizes.

4.2.3 Vehicle Quantity Choice Model

Inclusive values can be computed using the results of the previous sections, and used as explanatory variables in a vehicle quantity choice model. In addition, the literature suggests that the following factors are useful for explaining vehicle quantity choice:

1. Household size
2. Number of workers
3. Household income
4. Availability of transit.

As in the more traditional form of multinomial logit, these factors can be interacted with the choice alternative (one or two vehicles) as they would be expected to have different effects. The estimated coefficients for a vehicle quantity model using Caltrans data are in

Table 4.3. In the current version, an index of transit availability is not available. For this model, we used the full sample of households (17,040), which includes some zero-vehicle households. The distribution of vehicle ownership was provided in Table 3.4. The coefficients from the conditional one- and two-vehicle choice models were used to compute inclusive values for the one- and two-vehicle choice options, respectively.

```
Conditional (fixed-effects) logistic regression   Number of obs   =      51120
                                                  LR chi2(11)     =   18007.96
                                                  Prob > chi2     =     0.0000
Log likelihood = -9716.3719                       Pseudo R2       =     0.4810
```

v1	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Workers-1v	.4175632	.0788392	5.30	0.000	.2630413	.5720851
Workers-2v	.9057387	.0789009	11.48	0.000	.7510958	1.060382
Ln(HHSize)-1v	-.0896374	.1052728	-0.85	0.395	-.2959683	.1166935
Ln(HHSize)-2v	1.802626	.1068367	16.87	0.000	1.59323	2.012022
IncLT10K-1v	-1.566434	.1280483	-12.23	0.000	-1.817404	-1.315464
IncLT10K-2v	-3.424094	.1464813	-23.38	0.000	-3.711192	-3.136996
Inc10-25K-1v	-.6500682	.1120722	-5.80	0.000	-.8697256	-.4304108
Inc10-25K-2v	-2.000327	.1164437	-17.18	0.000	-2.228553	-1.772102
One-Veh dummy	3.138727	.1196952	26.22	0.000	2.904129	3.373325
Two-Veh dummy	3.738774	.2697936	13.86	0.000	3.209989	4.26756
InclValue	.2567455	.0365396	7.03	0.000	.1851292	.3283618

Table 4.3. Estimates of Vehicle Quantity Choice Model Using Caltrans Data

The current specification is similar to Train (1986). All coefficients except one are statistically significant, and the signs are what might be expected. The alternative specific constant for two-plus vehicles is slightly larger than for one vehicle, and both are positive (versus a value of 0 for the base alternative of no vehicles), indicating a preference for more vehicles, all else equal. Coefficients for number of workers, and natural log of household size, are estimated as interactions with the one-vehicle and two-plus-vehicle alternatives, respectively. The coefficients for these two demographic factors are larger for the two-vehicle alternative than the one-vehicle alternative, as would be expected.

We also include interaction effects for the two lowest income groups. All of these coefficients are negative. The coefficients for the lowest income group (Less than $10K) are more negative than the next-lowest group ($10-25K), and the coefficients for the two-vehicle option are more negative than for the corresponding one-vehicle option. In other words, lower incomes result in a decrease in the expected number of vehicles per household. The coefficient for the Inclusive Value term is positive, indicating that any changes in vehicle features that yield increased utility will cause the probability of that branch to increase.

The vehicle holdings models estimated here specifically model household vehicle demand behavior, conditional on current market conditions (whatever they may be). These models are combined with other elements of CARBITS to simulate the total market "system."

5. Department of Motor Vehicle (DMV) Registrations Data

The models estimated in Section 4 are based on a specific sample of survey respondents. These household-level data are useful for identifying important behavioral effects when individual households make vehicle purchases. However, the sample sizes associated with survey data are not large enough to provide an accurate measure of *aggregate-level market statistics* (e.g., new vehicle sales of various vehicle types) that can be important when performing policy analysis. To address this issue, models estimated using survey data are typically recalibrated so that they match aggregate level statistics from other data sources. For example, in the case of CARBITS it would be desirable for the market demand model to "simulate" new vehicle sales in the base year that match actual vehicle sales. Moreover, because CARBITS also models the used vehicle market, it would be desirable to match vehicle count distributions by model year as well. Finally, if the model explicitly simulates vehicle exit/scrappage, it would be desirable to match known vehicle exit/scrappage rates (if such data are available).

For this project, procedures have been developed for processing California DMV registrations data to meet these needs. Specifically, the DMV has been producing regular biannual data "dumps" of all registrations for quite a number of years. Each data dump can be thought of as a snapshot of vehicle registrations at a particular point in time. The snapshots generally occur in October and April of each year. The practice of generating these data sets began as the result of joint effort by the California Energy Commission (CEC), ARB, and Caltrans to obtain data that could be used to meet needs of the various agencies. (A full history is beyond the scope of this project. The lead agency on this has been the CEC, with varying levels of participation from the other two agencies.)

In what follows, we look at registrations data from October 2001. October is an attractive month to consider because, by this time of the year, most sales of new vehicles with the *model* year corresponding to the current *calendar* year have occurred. For example, by October 2001 most sales of new 2001 model year vehicles have occurred. In addition, some sales of new model year 2002 vehicles have also occurred. However, in the DMV data there are very few of these vehicles, and our current practice is to drop them. For an illustration using the October 2001 DMV snapshot, see Figure 5.1.

The data in Figure 5.1 are limited to light-duty vehicles. Wherever possible, vehicles that are known to be part of government or commercial fleets have been excluded. The vehicle total for model years 1982-2002 is approximately 18.8 million. A few features of this figure are noteworthy. During this period there were economic recessions in 1980-1982, 1990-1991, and 2001-2003, with periods of steady growth in between. The downturns in Figure 5.1 correspond to these periods.

As a point of comparison, recall that the Caltrans Travel Survey data were collected from October 2000 to December 2001, and the sample is weighted so that 21.4 million vehicles are "available to households" (see Section 3). The number of light-duty vehicles with model years 1982-2001 using this weighted sample is estimated to be 18.5 M versus the

18.8M in the October 2001 DMV snapshot. For a comparison of the model year distributions from the two data sets, see Figure 5.2.

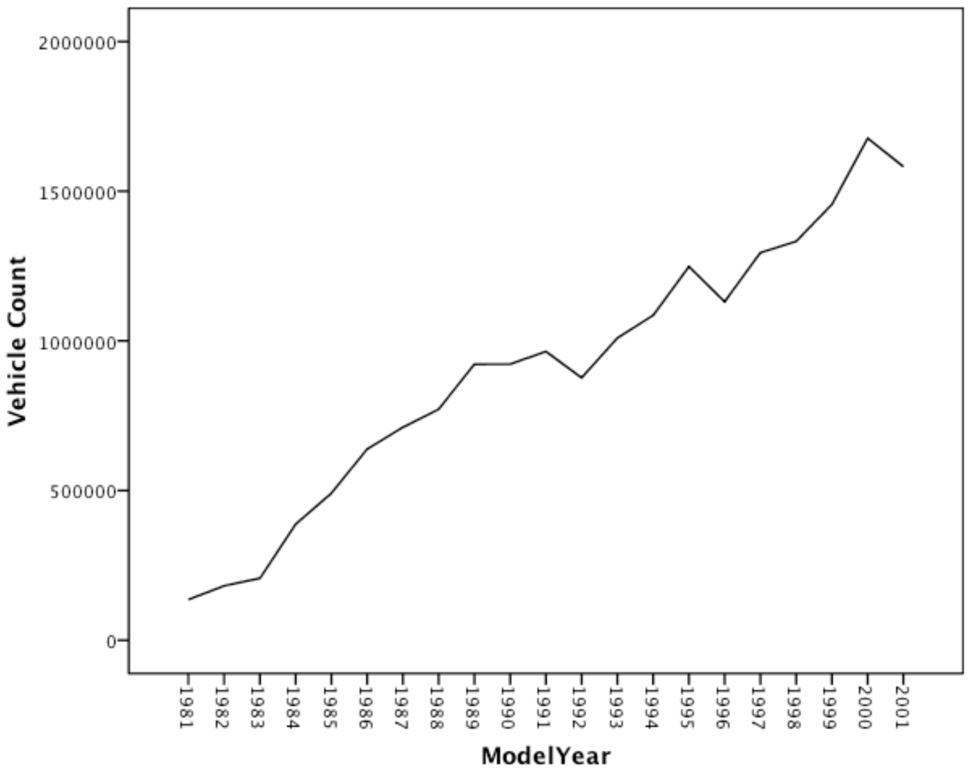

Figure 5.1 Model Year Distribution for October 2001 DMV Registrations (Light-Duty Vehicles)

Based on our past experience in comparing such distributions across different data sources, these are remarkably close. The DMV curve is much smoother than the Caltrans curve, as would be expected due to the issue of sample size. The main difference is that the vehicle counts for model year 2001 are substantially lower for the Caltrans data. This is easily explained: The Caltrans data were collected from households over an extended period of time starting in October 2000. Sales of model year 2001 vehicles accumulate over the entire calendar year and beyond into the following calendar year. The earlier a household was interviewed, the more likely it was that they could have purchased a 2001 model year vehicle after they were interviewed. More generally, this is a typical issue faced with choice model estimation: Households interviewed early in the process could have purchased a vehicle in the new vehicle market with model year 2000. In other words, it can be difficult to determine "new vehicle sales" on the bases of vehicle model year registrations. It is these phenomena that lead to the need for re-calibration of model constants for market simulation.

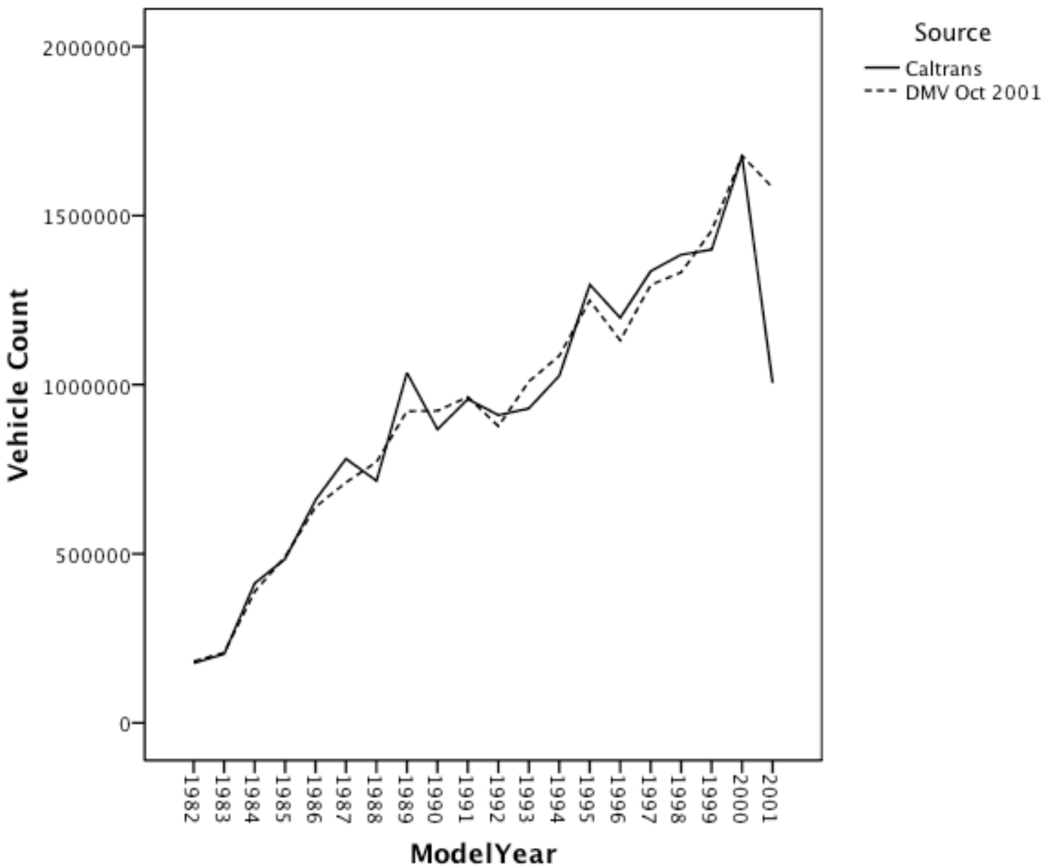

Figure 5.2 Model Year Distributions for DMV versus Caltrans Travel Survey

There are many details associated with processing DMV data that are not discussed in this section—see Appendix A. An important requirement is to be able to link vehicle counts for specific year-make-model vehicles in the California fleet to the corresponding vehicles in other data sets (e.g., the vehicle technology database) in order to perform various modeling tasks.

6. Vehicle Market Exit

As discussed in Section 1, one of the stated project goals is to explicitly model the exit of vehicles from the used vehicle fleet. In CARBITS 1.0, the exit of vehicles from the California fleet was an implicit outcome of household vehicle transaction choices for used vehicles over time. As vehicles continue to get older, their attractiveness diminishes so that more used vehicles of a particular class are sold than are purchased, leading to a net exit of vehicles from the market. An argument in favor of this approach is that the vehicle fleet distribution is determined by an internally consistent behavioral model of individual-level household vehicle preference and choice. A number of models (including the CalCars model of CEC) take this approach.

A potential vulnerability of this approach is that, combined with microsimulation, exit patterns of individual vehicle classes could appear noisy or inconsistent with typical scrappage patterns when compared to smoothed, well-behaved curves generated by models based on *aggregate* vehicle count data. The primary vulnerability is that it leaves the model open to criticism by hired consultants who use aggregate-level models, which are much simpler, and easier to both control and explain.

The literature contains examples of forecasting models in which household-level vehicle choice models are combined in the same system with scrappage models based on aggregate data—see, for example, Berkovec (1985) and Bento, at al. (2006). Although this approach is not based on a theoretical framework that is completely internally consistent, there is some behavioral theory that underlies the specification of the scrappage models, and this approach can be considered a way of incorporating additional information from aggregate data sources into the system. This project included a task to add this feature to CARBITS.

Before continuing, we make a few remarks about the general issue of modeling "vehicle scrappage." It will be noted that in this report we sometimes use the term "exit," and we sometimes use the term "scrappage." The main idea is that, when modeling the behavior of a vehicle market over time, older vehicles eventually "disappear" from the vehicle fleet by some process. At some point in time, most vehicles reach a state where they cease to exist and can never be "on the road" again. Vehicles that have been totaled in an accident, or simply become unusable, are scrapped for raw materials and spare parts. However, getting accurate data on this process is extremely difficult, and represents a challenge for modelers.

Another issue is that, when modeling a vehicle market over time, the market can ideally be treated as a "closed system" whereby all vehicles entering the market first do so through new vehicle sales, and they eventually exit by being scrapped. When modeling the domestic vehicle market for the entire United States, this may be a reasonable approximation. However, when modeling a submarket (e.g., California), the market is not really a "closed system." Vehicles of all vintages can both enter and leave the market through migration to and from other States. In this regard, there may be a net "exit" of vehicle classes from the market, but this process contains a mixture of immigration and scrappage processes. For this reason, we prefer to discuss vehicle "exit" rather than "scrappage."

Unless immigration processes are explicitly included in the model system, some modeling assumptions are required for simulating vehicle "exit" from the market. However, the more immediate issues are: What data should be used for estimating such a model, and what should a model look like? In this project, we use DMV registrations data for two consecutive years (October 2000 and October 2001) to estimate vehicle "exit rates" corresponding to the time frame of the Caltrans Travel Survey.

Our experiences mirror those reported in other research publications. Specifically, there is little or no vehicle exit during the first few years of most vehicle types. In fact, the data show a continued *increase* in vehicle counts for many vehicle models after the initial year of introduction. In our case, at least part of this effect can be attributed to immigration of vehicles into the State. However, researchers working with national-level registrations data also observe this effect, and have attributed it to continued new vehicle sales from the initial model year inventory for periods of up to four or more years—see Berkovec (1985). There are also issues with very old vehicles, where certain types of vehicles may be reconditioned and re-registered, leading to a net increase in vehicle counts that should theoretically not occur.

For our analysis, we computed vehicle exit rates for vehicles at the Year-Make-Model level. One useful piece of information contained in the DMV data is where the vehicle was originally sold as new: either in California, or Out of State (OS). This enabled us to confirm that there was a substantial amount of vehicle immigration for more recent model years, so that, on average, about 20% of the vehicle fleet will have originated from Out of State. We estimated vehicle exit rates by first removing net increase in OS vehicles over the period. This is a completely practical approach, and quite literally this is a vehicle "net exit" model, since, e.g., it is not possible for us to know if a vehicle originating in California left the fleet by leaving the State, through scrappage, etc. Moreover, vehicles could leave California and then return at a later time. The current analysis cannot separately identify this effect. (Later, we will remark on the possibility of future work that can be done in this area.)

We estimate a model using the same approach as earlier work in the literature—see Berkovec (1985). For each vehicle type $n = 1, \ldots, N$, the estimated exit rate is given by

$$R_n^{2001} = (Q_n^{2001} - Q_n^{2000}) / Q_n^{2000}$$

where Q_n^y is the vehicle count of vehicle type n in year y. We use a data set with N = 2,385 vehicle types (at the level of Year-Make-Model) using vehicles from model year 1982 to 1994. As noted earlier, it is typical in the literature to drop the first few years of data and treat the scrappage rate as zero (four is a typical number), for the reasons discussed. In our case, these are State-level (not national) data, and vehicle immigration seems to be a major effect. The "exit rate" figures for the first six years exhibited some unusual patterns, so these years have been dropped (this issue could be explored in more detail at a later time, if it seems warranted). Figure 6.1 provides plots of average exit rates as a function of Body Type/Prestige and Model Year.

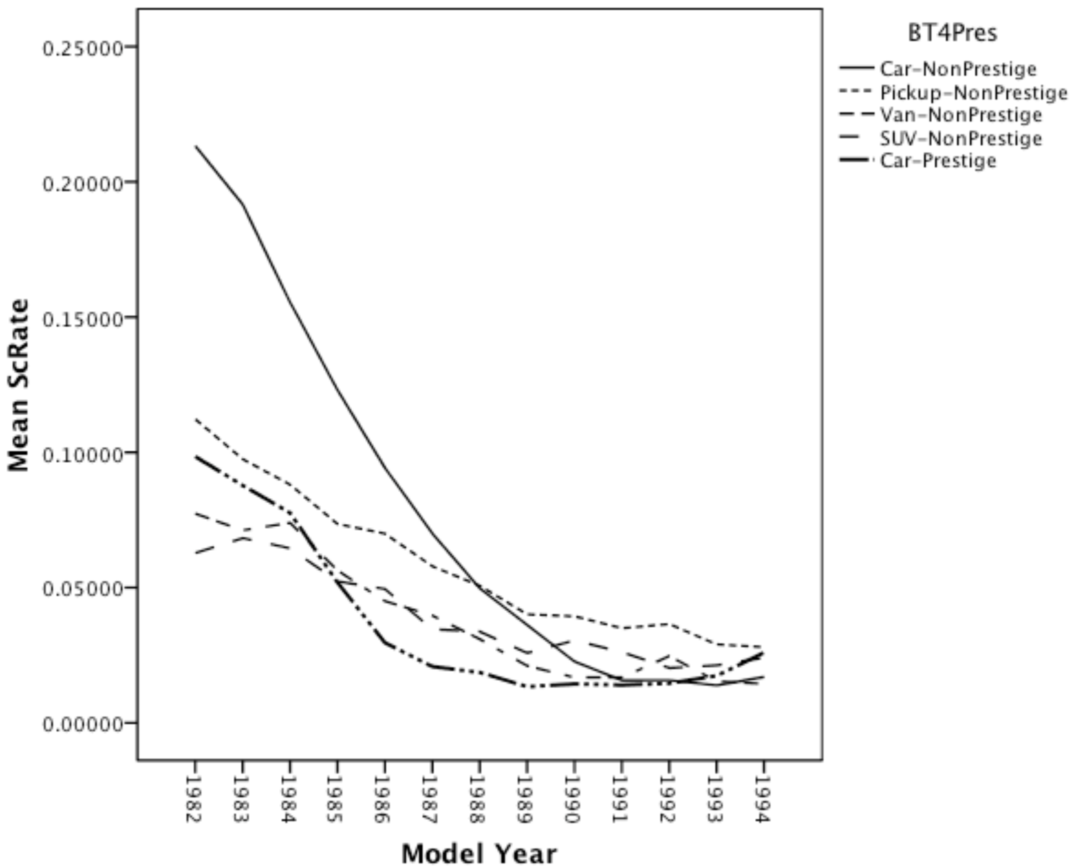

Figure 6.1 Mean Vehicle Exit Rate by Body Type/Prestige and Model Year

There are some noticeable patterns in this plot. The exit rates for Prestige versus Non-Prestige Cars are extremely different. Starting with the most recent model year and going back, the curve starts out relatively flat (below those of light duty trucks) for the first few years. Thereafter, there is a sharp increase in exit rates for Non-Prestige Cars, reaching a level of over 20% for the oldest vehicles. Prestige Cars have lower exit rates compared to all other Body Types for the newest 8-9 model years, and are comparable to the light duty trucks for the oldest model years. Exit rate curves for light duty trucks start out higher than cars, and have shapes that are (i) similar to each other, but (ii) different from either type of car. The curves for Vans and SUVs are similar to each other, and below the curve for Pickups.

What this plot does not include is the role that economic behavior might play in explaining the differences across the different vehicle types. We used these data to fit a standard scrappage model from the literature of the form:

$$\ln(R_n) = \beta_0 + X_n\beta_1 + \beta_2 P_n + \beta_3 P_n^2 + \varepsilon_n$$

where X_n includes vehicle characteristics such as age, body type, weight, etc., and P_n is the vehicle's price. This form is based on a model by Manski and Goldin (1982), and is derived from an economics-based theory of scrappage behavior. The model assumes that aging vehicles are continually subjected to random "events" that require repairs to occur in order to keep them operating. The distribution of repair costs is also random. As the value (market price) of the vehicle declines, at some point a random event generates a repair cost such that the scrap value of the vehicle exceeds the market value plus the repair cost, causing the owner to scrap rather than repair the vehicle.

The model above is estimated using linear regression—for results, see Table 6.1. The model is highly significant, with a rather high goodness-of-fit (adjusted R-square = 0.58). The signs of the coefficients are consistent with some of the conclusions from the plot in Figure 6.1. Vehicles with higher market prices tend to have lower exit rates. As prices drop, exit rates increase and do so at an increasing rate. One complicating factor is that prices (which are determined by economic market forces) are a function of many factors, including vehicle quality, age, and body type. However, price does not completely incorporate these effects so they are also included as explanatory variables. For example, the Prestige factor causes the exit rate to decrease (all else equal). The coefficients Y## denote model-year-specific dummy variables. After price and other effects are taken into account, newer vehicles have lower exit rates than average. Exit rates then systematically increase with increasing vehicle age. With regard to body type effects, the model uses Compact Car as the base vehicle type. All sizes of Cars except for Small Cars have insignificant coefficients (which are included here for completeness), indicating that Small Cars experience higher exit rates than all other types of cars after other factors are taken into account. Trucks have higher exit rates, and Vans have lower exit rates.

	B	t	Sig.
(Constant)	-3.238	-32.25	0.00
Price	-0.086	-5.797	0.00
PriceSqr	0.003	7.158	0.00
Y93	-0.251	-3.388	0.00
Y92	-0.332	-4.401	0.00
Y91	-0.316	-4.12	0.00
Y90	-0.179	-2.249	0.03
Y89	0.073	0.882	0.38
Y88	0.326	3.956	0.00
Y87	0.584	6.921	0.00
Y86	0.854	9.666	0.00
Y85	1.142	12.545	0.00
Y84	1.369	14.38	0.00
Y83	1.58	16.145	0.00
Y82	1.64	16.532	0.00
Weight	-1.03E-04	-3.051	0.00
Prestige	-0.372	-6.96	0.00
TwoSeat	-0.042	-0.408	0.68
SmallCar	0.161	3.355	0.00
MidCar	0.024	0.478	0.63
LargeCar	0.1	1.495	0.14
Truck	0.466	7.559	0.00
Van	-0.137	-1.934	0.05
SUV	0.067	1.032	0.30

Table 6.1 Vehicle Exit Model from DMV Data

7. Calibration Procedures

Other sections in this report have made reference to the issue of "calibration procedures." These procedures can be applied in multiple contexts, but the general idea is that there are circumstances under which *model* parameters can be adjusted so that certain distributions computed from the model's output are made to more closely match distributions obtained from an alternative (and presumably more reliable) *data source*. As an example, consider Figure 5.2 that compares the model year distribution of vehicle counts from the DMV and Caltrans data sets. Based on these data, choice models estimated from the Caltrans data will exhibit the same property of under-predicting the number of 2001 model year vehicles (when compared to the DMV vehicle count). However, the model includes parameters that reflect the household's utility for new versus used vehicles. These parameters can be adjusted to shift the vehicle count distribution produced by the model so that it matches the DMV distribution more closely.

As another example of the potential need for calibration, consider the following: Suppose that CARBITS 2.0 has been used to simulate vehicle market behavior starting in the base year of 2001, and has been run to 2006. Suppose DMV data are available for 2006, and the vehicle distributions from CARBITS are compared to the "actual" distributions from these data. These distributions will almost certainly be different due to factors that may have changed and/or are not taken into account in the original model estimation.

The vehicle choice models estimated in Section 4 contain coefficients that represent household preferences for such things as market price, fuel operating cost, and performance. These coefficients are regarded to be *generic* preference weights, and, in the absence of additional data, or original estimates represent the best source of this information. However, the models also include "alternative specific constants" that are associated with various vehicle classes (Cars, Pickups, Vans, SUVs, etc.). The interpretation of these coefficients is more challenging, and they are generally regarded to include factors that are not explicitly captured by the generic coefficients. Informally, these constants can be considered as capturing "all the other effects that we cannot see." Based on this interpretation, it is considered valid to "calibrate" these types of alternative-specific constants so that the model's vehicle distributions match those of some outside benchmark figures. The generic preference coefficients are left unaltered because (as noted) the household data still provide the most valid information on these quantities. If additional household data become available, then the models can be re-estimated to update these coefficients as well.

The procedures for performing recalibration in simple multinomial logit models are well known and relatively straightforward: See the book by Train (1986). The situation is slightly more complex for the new CARBITS model due to the existence of the various sub-models, but the basic principles are the same. Preliminary procedures for recalibration have been developed to make adjustments based on the following distributions:

1. Vehicle totals for four high-level vehicle types: Car, Pickup, Van, and SUV.
2. Vehicle total for new versus used vehicles.
3. For used vehicles, vehicle totals by model year.

Using the current approach, we would expect that these constants would be updated periodically based on the availability of DMV vehicle count data. Additional minor modifications to the MATLAB would also be required, e.g., updating an internal parameter to reflect a new base year, and re-initializing the historical vehicle attribute file to include records for vehicle classes up to and including the base year. Detailed instructions for this process will be provided.

8. Incorporation of Hybrid Electric Vehicles

8.1 Background

The proposal for this project was developed in early 2005 in close consultation with ARB staff. In considering possible improvements to CARBITS, there was a mutual desire to anticipate future needs for policy analysis involving the California vehicle market, and to consider ways in which CARBITS could be modified to address those needs. Such discussions are always highly speculative, since these needs are by their nature difficult to anticipate. Moreover, these issues typically relate to possible developments in areas with little (or non-existent) presence in the current marketplace. With regard to vehicles, the area of concern is typically the penetration of future fuel technologies, including alternative fuels such ethanol and natural gas, but more recently "clean diesel" and hydrogen. At that time, the new fuel technology that seemed to have the most obvious likelihood for successful penetration was hybrid electric. Modifying CARBITS to consider the impact of hybrid electric vehicles on the vehicle market seemed to be the least speculative in terms of likelihood, and potentially the most tractable since it did not require the consideration of requirements for new refueling infrastructure. Moreover, hybrid electric vehicles are more similar to existing market vehicles than are many of the other fuel technologies that are frequently discussed.

Based on these considerations, it was agreed to pursue the incorporation of hybrid electric vehicles into CARBITS as one of the project goals. This was done while recognizing that it was not immediately obvious how this could be achieved. It is always possible to modify a model to include choice alternatives that are not currently available. Such modifications inevitably depend on some type of judgment about modeling assumptions. The main issue is how much useful information and/or data are available to support these judgments. In this regard, we speculated that we might be able to draw on two different types of information: (i) potential access to data from stated choice experiments conducted by other researchers, and (ii) the potential for new data from additional penetration of hybrid electric vehicles into the existing marketplace. (Again, these discussions occurred in early 2005.) With this as background, we now describe the current status of modifying CARBITS to incorporate hybrid electric vehicles.

8.2 Hybrid Electric Vehicles in the Current Vehicle Market

To provide some perspective, we use the April 2008 DMV data to explore various aspects of the current market. Table 8.1 provides a cross-tabulation of vehicle counts by model year versus fuel (technology) type for the relevant range of years (2000-2008). Note: We also reviewed vehicle registrations data from DMV snapshots for the years 2000 through 2006. Although there is some variability in, e.g., registrations for a particular model year in successive calendar years, we concluded that the registration totals from April 2008 provided a reasonable representation of "new vehicle sales" for model years 2000-2007, and will be treated as such for discussion purposes (see also the remark in the next paragraph). Recall from Section 6 that vehicle exit rates are essentially flat for the first few years. For example, our model for vehicle exit between 2000 and 2001 uses data starting with model year 1995.

The numbers in Table 8.1 are highly illustrative of many of the general observations made about the hybrid electric vehicle market in, e.g., the popular press. First, taken in isolation, the penetration of hybrid electric vehicles has been rather fast during this period, increasing dramatically in percentage-increase terms during 2000 thru 2007. For example, sales increased from about 5.4 thousand in 2001 to 83 thousand in 2007, or about a 15-fold increase. (Remark: This DMV snapshot is from April 2008. Due to the timing of the snapshot, plus the disruption of the emerging recession in 2008, vehicle registrations for model year 2008 are clearly incomplete. We will consider registrations from model years 2000-2007 for purposes of comparison.) Still, despite the fast penetration and very high profile of these vehicles, hybrid electric vehicles represented only about 5.2% of new vehicle sales in 2007.

ModelYear * FuelType Crosstabulation					
		FuelType			
		Diesel	Gasoline	Hybrid	Total
2000	Count	18753	1440169	888	1459810
	% within ModelYear	1.30%	98.70%	0.10%	100.00%
2001	Count	26658	1469070	5379	1501107
	% within ModelYear	1.80%	97.90%	0.40%	100.00%
2002	Count	29530	1460431	6433	1496394
	% within ModelYear	2.00%	97.60%	0.40%	100.00%
2003	Count	34867	1510119	10616	1555602
	% within ModelYear	2.20%	97.10%	0.70%	100.00%
2004	Count	37652	1494933	17520	1550105
	% within ModelYear	2.40%	96.40%	1.10%	100.00%
2005	Count	33973	1509800	42493	1586266
	% within ModelYear	2.10%	95.20%	2.70%	100.00%
2006	Count	41368	1520379	49771	1611518
	% within ModelYear	2.60%	94.30%	3.10%	100.00%
2007	Count	22152	1500856	82718	1605726
	% within ModelYear	1.40%	93.50%	5.20%	100.00%
2008	Count	10246	640251	36383	686880
	% within ModelYear	1.50%	93.20%	5.30%	100.00%
Total	Count	255199	12546008	252201	13053408
	% within ModelYear	2.00%	96.10%	1.90%	100.00%

Table 8.1 Cross-Tabulation of Vehicle Counts by Model Year versus Fuel Type

In order to gain the additional level of understanding necessary for modifying the CARBITS vehicle choice models, these data must be explored in more detail with regard to factors such as: vehicle class, number of make/model offerings, market price, and fuel economy. Table 8.2 summarizes on Model Year 2007 vehicle counts for a selection of Vehicle Classes. For (model year) 2007, hybrid electric vehicle penetration was essentially limited to Subcompact and Midsize Cars, and SUVs. (The only Compact Car was the Saturn Aura, which sold 58 units in 2007).

Potentially noteworthy is that, although overall hybrid penetration is limited at 5.4 percent, the penetration is almost 17% for Midsize Cars (from 5 different models). Subcompact sales were due to one model (Honda Civic Hybrid). (Note: The Toyota Prius, which was originally introduced as a Subcompact, became a Midsize Car starting with model year 2004.) Hybrids comprised only 3.4% of SUV sales, primarily from three models (Ford Escape, Toyota Highlander, and Lexus RX 400h).

				FuelType	
			Gasoline	Hybrid	Total
Class	Subcompact	Count	188017	8817	196834
		% within class	95.50%	4.50%	100.00%
	Compact	Count	241969	58	242027
		% within class	100.00%	0.00%	100.00%
	Midsize	Count	303792	60733	364631
		% within class	83.30%	16.70%	100.00%
	SUV	Count	374498	13110	387682
		% within class	96.60%	3.40%	100.00%

Table 8.2 Cross-Tabulation of Vehicle Counts by Vehicle Class versus Fuel Type (2007)

Focusing on the Midsize Cars, see Table 8.3 for a summary of average vehicle attributes for four different types of Midsize Cars: Non-Prestige-Gasoline, Non-Prestige-Hybrid, Prestige-Gasoline and Prestige-Hybrid. In Section 2, we originally defined two Midsize Car Vehicle Classes: Prestige and Non-Prestige. Both of these are gasoline. Table 8.3 suggests the possibility of doubling the number of Midsize Car classes to four by adding hybrid electric versions.

Prestige	0	0	1	1
FuelType	Gasoline	Hybrid	Gasoline	Hybrid
MSRP	23013	26476	57525	54900
Horsepower	205	169	320	292
CurbWeight	3363	3414	3867	4134
HP/Wt	0.06079	0.04835	0.08186	0.07063
MPGCity	22	41	17	25
MPGHwy	30	39	25	28
Wheelbase	108	108	112	112
%Import	0.619	1	0.8276	1
NumModels	21	4	29	1
Sales	226,392	59,776	76,484	957

Table 8.3 Comparison of Vehicle Attributes for Four Classes of Midsize Cars

The information in Table 8.3 illustrates the practicality of introducing hybrid electric vehicles into CARBITS 2.0. The Vehicle Classes defined on the basis of historical data included two Midsize Car (gasoline) classes: Non-Prestige and Prestige. The obvious extension is to add two more Midsize Car classes: Hybrid-Non-Prestige and Hybrid-Prestige. With regard to vehicle size as measured by wheelbase, there is no difference between the two fuel types, once the prestige factor is taken into account. More importantly, the observed market demand for hybrids can generally be explained by the vehicle choice models in Section 4 due to the differences in vehicle attributes reported in Table 8.3.

We explore this argument in more detail. For Non-Prestige Midsize Cars, gasoline vehicles outsold hybrid vehicles by a 4-to-1 ratio. Hybrid vehicles clearly have better fuel efficiency than standard gasoline vehicles, which is their major positive feature. At the same time, the following are true: Hybrid vehicles are more expensive, have lower performance (as measured by horsepower-to-weight ratio), and are offered in many fewer models than are gasoline vehicles. All of these factors are consistent with a higher sales level of gasoline vehicles. In particular, the number of models offered is a very important variable in the vehicle choice models: the number of gasoline models outnumbers hybrids by a 5-to-1 ratio.

With regard to Prestige vehicles, gasoline vehicles had much larger sales than hybrids. At the same time, the ratio of model offerings is 29-to-1. On the other hand, due to the skewed price distribution, the hybrid price is a bit lower than the average price for prestige gasoline vehicles. In any case, these data suggest that Hybrid vehicles can be readily introduced into CARBITS by expanding existing Vehicle Classes to include Hybrid versions. Household vehicle preferences for generic attributes such as vehicle price, fuel operating cost, and performance would use the same coefficients as before. Any additional differences in preference would be addressed by calibrating a hybrid electric constant following the usual procedures discussed in Section 7.

8.3 Incorporation of New Hybrid Electric Vehicle Classes

Note that, based on Table 8.2, at this time we could consider adding five (5) new hybrid electric vehicle classes. In addition to the two Midsize classes just discussed, there would be a Non-Prestige Hybrid Small Car, a Non-Prestige Hybrid Midsize SUV, and a Prestige Hybrid Midsize SUV. The data to support this would come from the most recently available DMV snapshots. However, it should be recognized that the total vehicle counts for these classes would be rather small. At this time we have identified how these classes could be incorporated, but any final implementation will depend on additional discussions and collaboration with ARB staff. Similarly, more Hybrid classes could possibly be added for years beyond 2008, but would require more in the way of assumptions and judgments in an ongoing collaboration with ARB staff (which we are willing to pursue).

9. Concluding Remarks

This report includes background material relevant to, and descriptions of, the main outcomes of this project in a format designed to directly address the specific goals and objectives articulated in the original technical proposal and project plan. There are additional aspects of this project that lie outside the scope of this report that will be addressed separately. For example, technical details related to the MATLAB system and computer programs written to implement the CARBITS 2.0 platform, household behavioral models, market simulation, and calibration procedures will be documented as part of the final delivery to ARB staff. This report was primarily concerned with the development of different "parts" of CARBITS, but the operation and behavior of the system as a "whole" will be the subject to ongoing testing and refinement in collaboration between ARB staff and ITS researchers. This collaboration is expected to lead to fine-tuning and augmentation of the results reported here. Most of these changes would most likely address details related to computer program implementation, input and output interfaces, etc.

One exciting development with regard to the CARBITS model is that the conclusion of this project fortuitously coincides with another ARB-sponsored research project in which CARBITS will play a role in performing policy analysis related to a potential feebate program for California. The new project will essentially pick up where this one has left off, ensuring an uninterrupted continuation of CARBITS-related modeling activities that provide an opportunity for continued collaboration, testing, and development.

10. Bibliography

Bento, A. M., L. H. Goulder, M. R. Jacobsen, R. H. von Haefen, "Distributional and Efficiency Impacts of Increased U.S. Gasoline Taxes," Working Paper, July 2006.

Berkovec, J. (1985) "New car sales and used car stocks: a model of the automobile market", *Rand Journal of Economics*, 16(2): 195–214.

Bunch, D. and B. Chen (2008) "Automobile Demand and Type Choice," *Handbook of Transport I: Transport Modeling (Second Edition),* David A. Hensher and Kenneth J. Button, editors, Pergamon, pp. 463-479.

2000-2001 California Statewide Travel Survey, Final Report, California Department of Transportation (Caltrans), June 2002.

Hensher, D.A., N.C. Smith, F.W. Milthorpe and P.O. Barnard (1992) *Dimensions of automobile demand: A longitudinal study of household automobile ownership and use*. Amsterdam: North- Holland.

Sheng, Hongyan (1999). *A Dynamic Household Alternative-Fuel Vehicle Demand Model using Stated and Revealed Transaction Information*. Unpublished PhD dissertation, Department of Economics, University of California, Irvine.

Train, K. (1986) *Qualitative choice analysis: Theory, econometrics, and an application to automobile demand*. Cambridge: MIT Press.

Train, K. and C. Winston (2007) "Vehicle Choice Behavior and the Declining Market Share of U. S. Automakers," *International Economic Review*, Vol. 48, No. 4, November 2007.

Appendix A.

Development of Vehicle Technology, Price, and Market Demand Databases

As noted in various sections of this report, the data requirements for updating CARBITS were deceptively extensive. This Appendix provides additional detail on the data-related work performed in this project. Although they have been mentioned elsewhere, we first provide a brief overview of the basic issues. The most obvious requirement for updating household-level vehicle choice models is information on demographics and vehicle holdings (and/or transactions) from household survey data. This project used the 2000-2001 Caltrans Travel Survey and estimates vehicle holdings models—see Section 4 of Project Outcomes. However, these data must be supplemented with additional data on vehicle attributes (which are a function of vehicle technology) and market prices—see Sections 2 and 4. Moreover, aggregate market-level data on vehicle are required for multiple purposes. Our major data source is the Department of Motor Vehicle (DMV) registrations data—see Section 5.

Generally speaking, most issues arise from the fact that data about vehicles can vary widely in their level of detail. The most critical issue can be described as follows: (i) households make vehicle purchase decisions based on such factors as price, fuel economy and performance, (ii) these factors vary across vehicles due to differences in trim level, engine type, transmission, and drive train, (iii) information about household survey vehicles, and frequently aggregate vehicle demand, are typically not available at this level of detail. In the case of the models estimated in Section 4, vehicles are represented by Vehicle Classes, which contain even less detail than, e.g., Year-Make-Model. At the same time, models require representative values for vehicle attributes.

Based on the above description, we require the following: (i) data on vehicle prices and attributes at a very high level of detail, (ii) some way of generating representative values for vehicles at a lower level of detail (e.g., Year-Make-Model, or Vehicle Class). One approach is to find a way to estimate "sales-weighted averages" for vehicle attributes. Assuming the necessary data can be found, all of this requires the capability of matching vehicles across multiple data sources (frequently at different levels of detail).

This Appendix provides additional detail on this subject. The most logical starting point is to discuss how vehicle count data can be obtained from DMV registrations (see Section 5 of Project Outcomes), and how these counts can be linked to other databases containing information on vehicle technology and prices. The most important element of this process is the existence of the Vehicle Identification Number (VIN) system, and the concept of a VIN Prefix (the variable name we use is "VINPrefix"). The next section provides background, and succeeding sections provides further discussion on database development issues.

A1. Vehicle Identification Numbers (VINs) and VIN Prefixes

When working with vehicle data, a fundamental problem is that there is a wide variety of terminology in common use, and very limited standardization. Commercial sources of vehicle data each have their own way of identifying and describing individual vehicle offerings. These usually involve a character string description making it extremely difficult to combine data from multiple sources. One of the few standardized systems of coding vehicle information is the Vehicle Identification Number (VIN) system mandated by the U.S. federal government.

Although manufacturers were stamping some form of identification number on their vehicles as far back as the 1950's, the first standardized system was established at the behest of U. S. government in 1977 and underwent a major revision in 1983 (with subsequent revisions at later points in time). The official VIN requirements can be found in the Code of Federal Regulations [Title 49 (Transportation), Part 565]. The official documents are rather voluminous and can be tedious and difficult to read. One starting point for further exploration is the wiki page: http://en.wikipedia.org/wiki/Vehicle_Identification_Number. This page (as well as many other web pages) contains links to more detailed references.

Development of such a system requires specific legal definitions, terminology, etc. in order to rigorously specify the regulations (these are also included in the Code). Such terms include: Carline (or Line), Series, Body Type (or Type), Passenger Car, Light-Duty Vehicle, Light-Duty Truck, Model, Engine type, etc. The information requirements for VINs are established using well-defined terminology, with groups such as the International Organization for Standardization (ISO) and the Society of Automotive Engineers (SAE) playing a role.

Established regulations require that all newly registered and re-registered vehicles must have a 17-character VIN. The VIN can be:

- stamped into the vehicle structure (often the firewall) during manufacture, or
- stamped on a metal plate and fixed onto the vehicle body, or
- etched onto the rear window of the vehicle.

The use of digits is broken down into four sections:

1. WMI - World Manufacturer Identifier. [Positions 1-3 (with some exceptions)]
2. VDS - Vehicle Descriptor Section. [Positions 4-8]
3. Accuracy Check Digit [Position 9]
4. VIS - Vehicle Identifier Section. [Positions 10-17.]

The first ten digits include the data needed for vehicle technology identification purposes. Because position 9 is an accuracy check digit, the relevant data are coded as a 9-digit variable denoted "VINPrefix" (consisting of positions 1-8, plus 10).

A2. Vehicle Technology

A vehicle technology database is a fundamental requirement for the type of work performed under this project because vehicle demand is the result of the aggregation of household-level vehicle purchase decisions, and households purchase vehicles on the basis (either directly or indirectly) of a vehicle's physical characteristics. In addition, it is the vehicle's physical characteristics (along with vehicle usage) that determine the environmental and economic impacts of vehicles on society. Policies designed to mitigate these impacts must do so (directly or indirectly) by inducing change in vehicle characteristics.

To give additional background in terms of specifics, we note that during the course of this project we identified and purchased vehicle data from the Chrome company. Chrome offers a number of vehicle data-related products. We purchased two: Chrome VINMatch data, and Chrome New Vehicle Data (NVD). These datasets provide information on vehicles at a relatively high level of detail. The Chrome data are attractive because they have been implemented using modern relational database management (RDBM) techniques. Each data set has a table that uses a numeric variable called a Chrome StyleID that is the index variable to a database record containing a high level of detail for a particular vehicle type called a Chrome "Style."

In the NVD, the Style table is the central table in the database. In VINMatch, there are two "parallel" universes of vehicle definitions, one based on Chrome Style, and one based on VIN prefix. Although it is beyond the scope of this document to discuss these tables in detail, a summary of the variable names, types (N=numeric, S=string), and space allocations are given Table A1. Each table uses the same key variable (although they have slightly different names). The VINMatch table provides vehicle model and style information in a more user-friendly character string format. The NVD has more vehicle-attribute information directly included in the table, including MSRP, MktClassID (related to EPA vehicle classification), Consumer Friendly (CF) body type, Passenger Capacity, and information on whether certain transmission and drive train equipment are Standard, Optional, or not available. There is an indicator in the VINMatch table of whether the corresponding Chrome style exists in the NVD.

It is important to note the type of detail that Chrome Style is oriented toward: It is primarily defined on the basis of Year-Make-Model-Style information. In this sense, Chrome Style is the natural entry point to provide linkages to survey vehicles, which generally have this same type of information. At the same time, Chrome Style is generally much more detailed than the vehicle identifiers in survey data. As discussed in the main body of the report, the Caltrans Travel Survey vehicles were linked to vehicle technology data by first matching them to Chrome Styles.

However, as detailed as this information is, **a vehicle specification requires additional information on engine, transmission, and drive train**. We emphasize this because critical **vehicle attributes such as fuel economy and performance are a direct**

function of these vehicle characteristics. During the course of this project, we developed a detailed vehicle technology database on the basis of a construct called a "Chrome Vehicle," which can be uniquely identified on the basis of the following key variables:

Chrome StyleID [=> Year-Make-Model-Body Type]
VINPrefix
FuelSystem
EngineCategory
ForcedInduction
TransmissionType
TransmissionSpeed
DriveTrain

Style (NVD)		
StyleID	N	4
HistStyleID	N	10
ModelID	N	4
ModelYear	N	4
Sequence	N	2
StyleCode	S	8
FullStyleCode	S	9
StyleName	S	35
TrueBasePrice	S	1
Invoice	N	8
MSRP	N	8
Destination	N	6
StyleCVCList	S	71
MktClassID	N	2
StyleNameWOTrim	S	35
Trim	S	24
PassengerCapacity	N	2
PassengerDoors	N	1
ManualTrans	S	1
AutoTrans	S	1
FrontWD	S	1
RearWD	S	1
AllWD	S	1
FourWD	S	1
StepSide	S	1
Caption	N	1
AutoBuilderStyleID	S	14
PriceState	S	9
CFModelName	S	28
CFStyleName	S	40
CFDriveTrain	S	17
CFBodyType	S	31

YearMakeModelStyle (VINMatch)		
ChromeStyleID	N	4
Country	S	2
Year	N	4
DivisionName	S	13
SubdivisionName	S	21
ModelName	S	18
StyleName	S	35
TrimName	S	25
MfrStyleCode	S	7
FleetOnly	S	1
AvailableInNVD	S	1
DivisionID	N	2
SubdivisionID	N	4
ModelID	N	4
AutoBuilderStyleID	S	14
HistoricalStyleID	N	10

Common variables
(Chrome)StyleID
(Model)Year
StyleName
ModelID
AutoBuilderStyleID
Trim(Name)

Table A1. Variables contained in Chrome NVD and VINMatch "Style Tables"

In other words, taken together, these variables constitute a "composite key" that uniquely defines a "complete Chrome vehicle" at the level of detail required to determine key characteristics related to fuel economy and performance. Once vehicle technology data are developed at this level of detail, they can be used in various ways to meet various modeling requirements.

A3. Linking Vehicles from Different Databases

In the previous section, one of the variables is "VINPrefix." Although a complete discussion is beyond the scope of this report, the critical role of this variable must be described. A VINPrefix represents the first nine digits of the unique Vehicle Identification Number (VIN) assigned to all vehicles. Because it is required by law to include certain types of vehicle information, it has value as a vehicle identification variable that will be completely *consistent* across multiple databases. As such it is an aid to matching vehicles across databases. However, it is not perfect because it is not required to contain certain types of information. For example, it includes information on the Model Year, Make, Body Type, Fuel System, and Engine of every vehicle. Critical information on Transmission and Drive train are not required.

In any case, we briefly describe here the importance of VINPrefix. DMV data include the VINs of all registered vehicles. This means that *vehicle counts* can be obtained from the DMV data on the basis of VINPrefix. This means that vehicle count totals for makes and models can be determined from DMV data. Similarly, the National Automobile Dealer Association has a data product called a "VINPrefix Solution." This means that VINPrefix provides an entry point to our data source for used vehicle prices.

However, as already noted, VINPrefix does not define a complete vehicle at the level of detail discussed above. For example, suppose we know from Caltrans Survey Data that a household owns a 2001 Toyota Camry, and that it was purchased as a used vehicle in 2004. Consider the following questions:

1. What is the fuel economy for this vehicle?
2. What is the performance of this vehicle?
3. What was the likely market price paid by the household?

The answer to these questions cannot be accurately determined without more detailed information on the vehicle, e.g., engine type, transmission type, drive train, and trim level. (Note: It might be that there is only one drive train type available. However this may not be true for other types of vehicles.)

In order to construct typical values for these variables, we would like to compute weighted averages. However, VINPrefix does not provide enough detail to do this. In our work, we obtained BAR Smog Check data (which require the test operators to enter information on transmission and drive train). Specifically, the Smog Check database has the following information on individual vehicles: VIN, transmission type (automatic or manual), and drive train type (i.e., whether the vehicle has four/all wheel drive, or not). Data are aggregated over VINPrefix, yielding the percentage of automatic-versus-manual transmissions, and the percentage of four/all-wheel drive versus non- four/all-wheel drive for each VINPrefix. These percentages are then used as weights that can be used for averaging vehicle attributes.