ABSTRACT

FURLONG, KEVIN JOHN. The Competitive Effects of Variety. (Under the direction of Xiaoyong Zheng.)

This dissertation consists of two chapters and several technical appendices, all of which examine the welfare effects of new goods.

Chapter One examines the welfare effects from the hybrid vehicle innovation in the U.S. auto market between the years 2000 and 2008. Several structural discrete choice models are used to estimate the demand and supply curves for all vehicles in the U.S. auto market. Using the parameter estimates of the models, several counterfactual analyses are conducted examining the changes in producer surplus, the extent of first mover advantages, profit cannibalization, and business stealing effects from competitors. On the consumer side, price and variety effects are estimated which help identify the benefits from a productive and allocative efficiency standpoint, respectively. Furthermore, this chapter develops a simulation technique to recover these effects while allowing income to enter the indirect utility function nonlinearly. The results of this chapter suggest that, although hybrid vehicles represent a small share of the overall auto market, some households are willing to pay a premium for “green" and environmentally-friendly vehicles. Toyota was the first to recognize this market niche and responded by developing the Prius. Since the introduction, Toyota has enjoyed significant first mover advantages earning substantial economic profits and sustained growth in the hybrid vehicle market segment.

Chapter Two takes a more detailed approach to examine why consumers purchase hybrid vehicles. In particular, I show that in strictly financial terms, hybrid vehicles do not make sense. After factoring in vehicle depreciation, interest on financing, maintenance and repair costs, fees and taxes (including tax incentives), and fuel expenditures, I find that the majority of hybrid vehicle owners will not recover the higher premium through decreased fuel expenditures after five years of ownership. To explain why consumers continue to purchase them, a rich consumer demand model is developed using detailed household level data that explicitly links the year, make, and model of each vehicle a household owns to their specific demographic type. The results of this chapter suggest that, although hybrid vehicles may not make financial sense, for some households they are a better match to their preference type. In particular, one of the primary reasons consumers purchase hybrid vehicles is for conspicuous environmental effects; consumers not only want to be green, they want to look

green as well. As a result, some consumers are willing to pay a premium to send a signal to their community that they are “green" and living a low resource intensive lifestyle.

The Competitive Effects of Variety

by
Kevin John Furlong

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Economics

Raleigh, North Carolina

2012

APPROVED BY:

Roger von Haefen	Walter Thurman
Robert Hammond	Xiaoyong Zheng Chair of Advisory Committee

DEDICATION

To my parents.

BIOGRAPHY

Kevin Furlong was born in 1983 in New Jersey but was raised in the Northern Virginia area just outside of Washington, D.C. His family consists of his two parents, James and Nancy, and an older sister, Kelly. In 2001, he graduated from Paul VI Catholic High School and from there he went on to attend Virginia Tech in Blacksburg, VA as a mathematics major.

Kevin has been a lifelong competitive swimmer where he earned a full athletic scholarship to Virginia Tech, set numerous pool and school records, was named the 2003 Big East Conference Mens Most Outstanding Swimmer, and was named a 2005 NCAA Division I All American. Although he occasionally still swims from time to time, he has mostly taken up other activities including motocross, road biking, and running.

After graduating from Virginia Tech in 2005, Kevin decided to pursue a graduate degree in economics at North Carolina State University in Raleigh, NC. What started off as just a one and one-half year master's degree, turned into a five and one-half year PhD. Although he is not sure where he will be five years from now, he looks forward to overcoming more of life's challenges.

ACKNOWLEDGEMENTS

I would like to thank my graduate advisor, Dr. Xiaoyong Zheng, for sharing his knowledge, guidance, and patience throughout this entire process. This dissertation started off as a project in his class and he encouraged me to turn it into a final product. I would also like to thank the other members of my committee: Dr. Robert Hammond for his rooted knowledge of theoretical IO; Dr. Roger von Haefen for his expertise on discrete choice models and the vehicle demand literature; Dr. Walter Thurman for sharing his experience, level-headed, and rational approach to economics. I would also like to thank the following faculty members: Dr. Tom Vukina for everything I learned from the two graduate classes that I took from him as well as the numerous discussions we had in the hallway; Dr. Tamah Morant for constantly looking for new ways to improve the economics graduate program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

The introduction of new goods has been a reoccurring phenomenon that has existed throughout human history. New goods are continually being created, improved, modified, and destroyed ultimately being replaced by new and better products (Schumpeter, 1941). This process has grown exponentially since the industrial revolution with the advent of railways, telephones, electricity, automobiles, airplanes, and the internet, all of which have allowed information, goods, and services to be consumed across vast distances in shorter and shorter periods of time (Ridley, 2011). As a result, economies have grown, wealth has been created, and the well-being and living standards have risen across the developed world. Firms are the driving force behind this process as they seek transitory surpluses in economic profit. Firms can achieve this by either becoming the low-cost producer of a homogeneous good or differentiating their products from competitors (Hausman, 1997). Consumers, as a result, benefit from this process by having new and improved products available in their choice set. However, empirically quantifying the benefits of new products has been a long term challenge for economists primarily because the units come in dimensions other than price or quantity. This dissertation consists of two chapters and several technical appendices, all of which examine the welfare effects of new goods.

Chapter One examines the welfare effects from the hybrid vehicle innovation in the U.S. auto market between the years 2000 and 2008. Several structural discrete choice models are used to estimate the demand and supply curves for all vehicles in the U.S. auto market. Using the parameter estimates of the models, several counterfactual analyses are conducted examining the changes in producer surplus, the extent of first mover advantages, profit cannibalization, and business stealing effects from competitors. On the consumer side, compensating variation is decomposed into price and variety effects which help identify the benefits from a productive and allocative efficiency standpoint, respectively. The results of this chapter suggest that, although hybrid vehicles represent a relatively small share of the overall auto market, some households are willing to pay a premium for “green" and environmentally-friendly vehicles. Toyota was the first to recognize this market niche and responded by developing the Prius. Since the introduction, Toyota has enjoyed significant first mover advantages earning substantial economic profits and sustained growth in the hybrid vehicle market segment.

Chapter Two takes a more detailed approach to examine why consumers purchase hybrid vehicles. In particular, I show that in strictly financial terms, hybrid vehicles do not make sense. After factoring in vehicle depreciation, interest on financing, maintenance and repair costs, fees and taxes (including tax incentives), and fuel expenditures, I find that the majority of hybrid vehicle owners will not recover the higher premium through decreased fuel expenditures after five years of ownership. To explain why consumers continue to purchase them, a rich consumer demand model is developed using detailed household level data that explicitly links the year, make, and model of each vehicle a household owns to their specific demographic type. The results of this chapter suggest that, although hybrid vehicles may not make financial sense, for some households they are a better match to their preference type. In particular, one of the primary reasons consumers purchase hybrid vehicles is for conspicuous environmental effects; consumers not only want to be green, they want to look green as well. As a result, some consumers are willing to pay a premium to send a signal to their community that they are “green" and socially responsible.

A common theme throughout this dissertation is recognizing that consumers have heterogeneous preferences that vary both vertically and horizontally - what one consumer likes another may dislike. Selecting an appropriate demand model and specifying it correctly is crucial for capturing heterogeneous preferences for differentiated products. Below, I briefly discuss the methodologies used in both chapters of this dissertation to help the reader understand the advantages, similarities, differences, and caveats of each approach. I also briefly discuss the contributions that I make to the literature.

Lancaster (1966, 1979, and 1990) proposed the characteristics approach for which products are treated as bundles of various attributes that are packaged together in some combination and consumed. For example, a Ford Mustang is not just a Ford Mustang; it is a 2-door sports car with a 4.6 liter engine, 315 horsepower, dual exhaust, aluminum alloy wheels, power steering, windows and locks, and bucket seats. This same concept can be used to define a person. For example, a person has an age, gender, education, family size, income, and a household location that when bundled together define a person. Lancaster's approach assumes that each consumer, based on their preference type, has some “ideal" product bundle in mind. However, due to economies of scale and production constraints, this ideal product rarely exists in reality. As Chamberlin explains “If all products were ‘made to order,’ production would clearly be very much less efficient and fewer wants would be satisfied. But they would be satisfied with greater precision." As a result, when a consumer enters the market, their utility maximization problem essentially becomes a minimization problem whereby consumers search for the product that minimizes the distance between their “ideal" product and what is currently available on the market. The product that is the best match (i.e. shortest distance) is therefore the product that yields the highest utility.

However, economists rarely have information on all product attributes or consumer demographics because 1) the data simply does not exist or 2) the attributes may be difficult to quantify. For example, how do you measure quality, style, or prestige? Furthermore, even though two consumers may be of the same demographic type, they still may have different preferences due to unobservable (to the econometrician) reasons. Without fully accounting for these “unobservables," the demand model may yield biased parameter estimates or fail to fully capture heterogeneous preferences. The literature, as described below and in the chapters and appendices that follow, has developed several techniques to overcome these difficulties.

Ideally a demand model should accurately predict and forecast consumer behavior. However, regardless of how well a model is specified, it is impossible to predict exactly which product a particular consumer will choose due to their idiosyncratic preferences. For example, although I know everything about myself, I still do not know what I am going to do tonight - do I work on my dissertation or do I drink beer and watch TV? Thurstone (1927), which was later refined by Marchak (1960) and McFadden (1974), developed the idea of a Random Utility Model (RUM) which adds an element of randomness to a consumer’s utility function. As a result, this specification only allows researchers to infer the probability that a particular consumer will choose a specific product.

Discrete choice models are a type of random utility model that have grown in popularity among researchers to estimate demand when consumers are faced with a discrete (countable) set of alternatives. These models are based on the assumption that consumers are rational, have perfect information, and will only purchase the product that maximizes their utility. In the words of Herbert Simon (1978), “The rational man of economics is a maximizer, who will settle for nothing less than the best." By observing which product a consumer chooses, they are, in essence, revealing their preference for that particular product. This is important when analyzing demand for differentiated products, such as automobiles, because the “best" product is subjective and varies from person to person. By observing a large enough sample of consumers with information on their product choices, product characteristics, and consumer demographics, it is possible, with the correct specification, to develop a demand model that can fully characterize consumer behavior.

The logit discrete choice model is currently one of the most widely used among researchers as a result of its simplicity, closed form solution, and easily interpretable results (Train, 2003). However, there are several critiques to this model, most notably the Independence of Irrelevant Alternatives (IIA) property which, by construction, assumes that unobservable (to the econometrician) utility components are Independent and Identically Distributed (IID) extreme value type I. As a result, cross-price elasticities between products are driven completely by market shares and not by how similar the products are. For example, if a Ford Mustang (sports car) and a Chrysler Town and Country (minivan) have equal market shares, and the price of a Chevrolet Camero (sports car) increased, the logit model would imply that consumers would substitute away from the Camero and switch to both the Mustang and Town and Country in equal proportions, which clearly is unrealistic. To overcome this problem, several alternative discrete choice models have been developed including the Generalized Extreme Value (GEV), nested logit, probit, and mixed logit model. However, the flexibility of the mixed logit model allows it to approximate any discrete choice model with an appropriate specification of variables and mixing distributions. As a result, the mixed logit model has become regarded as the gold standard of discrete choice models (see Hensher and Greene, 2003).

The mixing distribution of the mixed logit model essentially captures how consumers' unobservable (to the econometrician) heterogeneous preferences are distributed across the population for various product characteristics. Although these preferences are technically "unobservable," the researcher may have an idea as to how they are distributed across the population. Common distributions include the normal and log-normal at various truncations (e.g. 99%, 95%, etc.), however, the researcher may specify any distribution that he/she deems fit. Empirically, the researcher is able to recover the mean and variance of these distributions which helps characterize heterogeneous preferences for various product characteristics. Though useful, this additional information says nothing about where a specific individual's preferences lie in the distribution. In Chapter Two below, I develop an estimation procedure, similar to Revelt and Train (2000), to infer tighter bounds on the mean and variance of consumers' unobservable (to the econometrician) preferences conditional on their product choice and demographic type.

The advantages of the mixed logit model are threefold. First, the model allows preferences to vary randomly across the population. Second, the model produces rich substitution patterns. Third, the model overcomes the IIA critique by allowing products to be correlated for unobservable (to the econometrician) reasons. However, the mixed logit model, as is true with the logit model, is susceptible to endogeneity problems if the researcher has limited information on product characteristics. This, of course, assumes that unobservable (to the econometrician) product characteristics are costly to produce and, as a result, would be correlated with price.

The methodologies of Berry (1994) and Berry, Levinsohn, and Pakes (1995) (henceforth BLP) are essentially analogous to the logit and mixed logit model, respectively, except that they both explicitly account for unobservable (to the econometrician) product characteristics via a structural error component which helps mitigate endogeneity problems. Furthermore, only aggregate data is required for estimation which may be advantageous when detailed micro data is expensive or proprietary in nature. However, like the logit model, the methodology of Berry (1994) does not allow unobservable (to the econometrician) utility components to be correlated across alternatives. As a result, the method produces unrealistic substitution patterns and suffers from the IIA critique. The BLP methodology, like the mixed logit model, overcomes these caveats by allowing the researcher to specify an appropriate mixing distribution that characterizes how consumers' unobservable (to the econometrician) preferences are distributed across the population. As a result, the BLP methodology produces rich substitution patterns and overcomes the IIA critique (Nevo, 2000).

The additional benefits of the BLP methodology, however, come with significant computational costs for several reasons. First, the model has no closed form solution and must be estimated via simulation. Second, the parameters are nonlinear which require parameter estimates to be iteratively search over. Third, the objective function is not guaranteed to be globally concave which may result in multiple local solutions. Furthermore, since only aggregate data is required for estimation, inference at the micro-level is somewhat limited as the researcher must rely on aggregate demographic data to infer correlations, rather than a direct link, between product choice and heterogeneous preferences which can lead to imprecise parameter estimates. Petrin (2002) developed a methodology to help mitigate the latter of these caveats by incorporating information from micro-data that directly links a consumer's product choice to their demographic type into the BLP methodology. In Chapter One below, I use all three structural discrete choice models of Berry (1994), BLP (1995) and Petrin (2002) and compare the results.

Compensating Variation (CV) and Willingness to Pay (WTP) are two similar, yet different, approaches to measuring the benefits of new products. Both measure how much a consumer would need to be paid, in dollar terms, to be compensated for a “change" in an economic variable between the original and counterfactual market. However, when measuring the benefits of new products, a “change" can either be 1) the removal of a new product(s) as a whole from the counterfactual market or 2) turning a product attribute on or off in the counterfactual market. In the discrete choice framework, CV allows consumers to re-sort to the next best alternative after the “change," whereas WTP does not. Therefore, if the researcher is interested in measuring the benefits of a new product(s) as a whole, CV would be the best approach. However, if the researcher is only interested in measuring the benefits of a specific product attribute, then WTP is the best approach. This dissertation uses both approaches.

In Chapter One, I am interested in measuring the benefits of hybrid vehicles as a whole. Therefore, I remove all hybrids from the market and allow consumers to re-sort to the next best alternative. For example, if a consumer originally purchased a Toyota Prius then, in the counterfactual market with all hybrids removed, they may, as a result, purchase a Toyota Corolla instead. However, because the Corolla was not the consumer's first choice, they will suffer a welfare loss and will therefore need to be compensated.

In Chapter Two, I am only interested in measuring how much consumers value hybrid technology. Therefore, only the hybrid attribute is turned off and, unlike Chapter One, consumers are *not* allowed to re-sort to the next best alternative. If consumers *were* allowed to re-sort, the welfare measure would be misleading. For example, suppose a consumer originally purchased a Toyota Prius. If I just turned the hybrid attribute off and allowed consumers to re-sort, they would most likely purchase an entirely different vehicle with entirely different attributes (different horsepower, different price, different size, etc.). Therefore, if CV was used rather than WTP, it will not only capture the loss from no longer having hybrid technology, but it will also capture the additional benefit or cost from having different attributes. For example, if the consumer purchased a Corolla in the counterfactual market, they may suffer a loss from it not being a hybrid, but they may also benefit by having more horsepower.

In addition to Chapters One and Two, several technical appendices are included at the end of this dissertation. Appendix I discusses the advantages, the theory, and how to empirically incorporate micro-moments into the BLP (1995) methodology following the work of Petrin (2002). Appendix II describes, in detail, the simulation technique that was developed in Chapter One of this dissertation to isolate both price and variety effect with nonlinear marginal utility of income. Appendix III discusses various marginal utility of income specifications that are commonly used in discrete choice models. Choosing the correct specification is crucial for accurately measuring compensating variation and willingness to pay estimates. Finally, Appendix IV describes the supplemental data used in Chapter Two.

Chapter 1

Quantifying the Benefits of New Products: Hybrid Vehicles

1.1 Introduction

In monopolistically competitive markets with low barriers to entry, firms are the driving force behind the research, development, and production of new goods and services as they seek transitory surpluses in economic profit. Firms can achieve this by either becoming the low-cost producer of a homogeneous good or differentiating their products from competitors (Hausman, 1997). Consumers, as a result, benefit from this process by having new and improved products available in their choice set. However, empirically quantifying the benefits of new products has been a long-term challenge for economists primarily because the units come in dimensions other than price and quantity.

This chapter measures the welfare effects from the hybrid vehicle innovation in the U.S. auto market between the years 2000 and 2008. Using several counterfactual analyses, I measure the changes in producer surplus, the extent of first mover advantages, and business stealing effects from competitors. Consumer welfare is examined from a productive and allocative efficiency standpoint by decomposing compensating variation into price and variety effects.

I follow the discrete choice differentiated products literature of Berry (1994), Berry, Levinsohn, and Pakes (1995) (henceforth BLP), and Petrin (2002) to estimate the demand and supply curves for all vehicles in the U.S. auto market. These models are ideal for estimating demand for large systems of differentiated products that capture heterogeneous preferences and control for unobservable product characteristics. Furthermore, these models are structural which allows various counterfactual market scenarios to be simulated. Micro-moments, following the work of Petrin (2002), are also included in the model to help identify and improve the significance of the parameter estimates.

Trajtenberg (1989) suggested that the best metric economists have available to quantify the benefits of new products is the amount of consumer and producer surplus generated by them. Bresnahan et al. (1997) added to the new products literature by examining the competitive effects of product segmentation with an application to the differentiated personal computer market. Specifically, they show that firms that can successfully differentiate their products are rewarded by a degree of transitory market power enabling them to enjoy first mover advantages. Hausman and Leonard (2002) examine the welfare effects of new products by decomposing consumer welfare into price and variety effects, thus helping to identify the benefits from a productive and allocative efficiency standpoint, respectively. Petrin (2002) quantified the benefits of the minivan innovation in the U.S. auto market using a structural discrete choice model combining both micro and macro level data. He finds that the minivan generated substantial benefits by offering would-be station wagon owners a better alternative. Furthermore, he found that non-minivan owners also benefited due to the increased price competition.

This chapter finds that hybrid vehicles generated substantial welfare effects totaling $14.7 billion between 2000 and 2008. Hybrid vehicle owners captured the majority of this, namely $12.0 billion, due almost entirely to the additional product variety. The median hybrid vehicle owner captured $4,280 in surplus in 2000 which grew to $5,475 by 2008. Price effects were found to be relatively modest generating a total of $780 million in compensating variation for all new vehicle purchasers combined over the course of 9 years. Toyota, the largest producer of hybrid vehicles, captured significant first mover advantages earning a total of $4.70 billion in variable profits. Honda also benefited, although substantially less, with variable profits totaling $490.2 million. All other producers were hurt from the hybrid innovation and would have been better off had hybrids never entered the market. Over the course of nine years, Toyota's competitors namely Honda, Ford, GM, and Nissan were relatively unsuccessful at stealing business away from Toyota, capturing $200.8 million, corresponding to 0.39%, of Toyota's total variable profits.

This chapter contributes to the literature in two ways. First, it provides an empirical estimation strategy to isolate both price and variety effects while allowing income to enter the indirect utility function nonlinearly. These effects are estimated conditional on both a consumer's original product choice and preference type, thus giving more precise and realistic welfare measures. Since no closed-form solution exists, simulation techniques are used. Second, this chapter adds to the new products literature by being the first to quantify the welfare effects from hybrid vehicles entering the U.S. auto market.

This chapter is organized as follows. Section Two gives a brief history of hybrid vehicles. Section Three explains the data. Section Four explains the demand side of the model. Section Five explains the estimation procedure. Section Six explains the supply side of the model. Section Seven describes the welfare calculations. The results are presented in Section Eight. Finally, Section Nine concludes the chapter.

1.2 History of Hybrid Vehicles

The Toyota Prius and Honda Insight were the first two hybrid vehicles to enter U.S. auto market, first appearing in 2000. Hybrid sales have steadily grown from 9,350 vehicles in 2000 to 306,080 in 2008, corresponding to market shares of 0.05% and 2.24% for all new vehicle sales, respectively. See Table 1.1. By 2008, several firms had entered the hybrid vehicle market including Ford, GM, Honda, Nissan, and Toyota offering a total of 12 hybrid models to choose from. See Table 1.2. The growth in sales is attributed to rising gasoline prices, increased environmental awareness, government tax incentives, and the consumer adoption process of new products. See Beresteanu and Li (2009), Diamond (2009), Erdem et al. (2010), Gallagher and Muehlegger (2011), Kahn (2007), Potoglou and Kanaroglou (2007), Sangkapicahi and Saphores (2008), and Turrentine and Kurani (2006) for related articles.

Toyota is viewed as the innovator and leader of the hybrid vehicle market with hybrid market shares consistently staying well above 50%. Nine years after the first Prius entered the market, competitors have still struggled to capture market shares. In fact, Toyota's position has grown stronger, despite the increased competition, with hybrid market shares growing to 78% by 2008. Toyota's success can be attributed to the strong branding and unique design of the Prius which gives consumers a vehicle with both high fuel economy and one that sends a signal to the community that they are green and living a low resource intensive lifestyle. Table 1.3 shows that each year, Prius sales have represented over half of all hybrid vehicle sales.[1]

[1]The one exception is 2006 in which Prius sales only represented 42.8% of hybrid sales.

Honda is the second largest producer of hybrid vehicles with the Civic Hybrid, released in 2003, being their most successful model. Honda's market shares, however, have steadily declined from 47% in 2003 to only 10% in 2008. During this time, Ford, GM, and Nissan entered the hybrid vehicle market potentially stealing business away from them. Furthermore, Toyota's market shares also increased from 53% in 2003 to 78% in 2008 further exacerbating Honda's decline. Although the Insight was Honda's first hybrid model released into the market, it was never widely accepted in the market with sales gradually diminishing to a mere 722 units in 2006 before being discontinued in 2007.[2]

Hybrids offer increased fuel efficiency compared to traditional 100 percent gasoline vehicles. They capture the energy lost from breaking and idling by using that energy, which would otherwise be lost, to charge a battery. The battery is then used interchangeably with the internal combustion engine to power the vehicle forward. A growing trend appears to be producing vehicles with hybrid technology available as an optional feature giving consumers the choice to purchase the technology or not. In fact, all but two hybrids available today also come in a non-hybrid form. The only two vehicles which do not are the Toyota Prius and Honda Insight.

From a social perspective, the increased fuel efficiency from hybrid technology has several substantial benefits. Environmentally, hybrids emit less tail-pipe emissions which helps mitigate global warming concerns and decrease air pollution. The increased fuel efficiency also helps decrease our dependence on foreign oil which has been a long term national objective since the 1970s oil price shock (EPA and NHTSA National Program, 2010).[3]

The Federal government recognizes the external benefits from hybrid vehicles and therefore offers tax incentives to help encourage more consumers to purchase them. Prior to 2006, there was a $2,000 tax deduction for all hybrids. Then, starting January 1, 2006 as part of the Energy Policy Act of 2005, tax credits were given to hybrid purchasers that varied in amount depending on the emissions and fuel efficiency rating of the specific model. The tax credits were designed to phase out once a manufacturer sold 60,000 hybrid vehicles. The first two calendar quarters after the 60,000th hybrid was sold, the credit would reduce to 50% of the original amount. The last two quarters of the calendar, the credit would reduce further to 25% of the original amount before reducing to zero for all subsequent quarters. In addition to federal incentives, several states also offer incentives to encourage the adoption of hybrid

[2]Honda has subsequently redesigned the Insight and has re-released it back into the market with sales in 2009 totaling 20,572. Remarkably, the next generation Insight looks almost identical to the Prius.

[3]As of 2010, the U.S. imported 60% of its oil.

vehicles including HOV lane access for solo drivers and further monetary incentives.

1.3 Data

Two main data sources are used in this chapter. The first is Ward's Automotive Yearbook which contains information on product characteristics, Manufacturer Suggested Retail Price (msrp) and quantities for all vehicles sold in the U.S. between the years 1999 and 2008. Product characteristics include the size (length x width), horsepower, weight, miles per gallon (mpg), country of origin, and categorizes vehicles into 26 segment classes. Quantities and msrp are linked to the baseline models since information on the trim level was unavailable. Vehicles for which fewer than 2,000 were sold are excluded from the data giving a total of 2,427 vehicle observations used in estimation. All prices are inflated to 2008 levels. See Tables 1.4 and 1.5 for summary statistics.

The second main data source is the 2009 National Household Travel Survey (NHTS) published by the U.S. Department of Transportation: Federal Highway Administration. The survey was conducted between March 2008 and May 2009, containing 150,147 household level observations. Information in this data set include the year, make, and model for each vehicle a household owns in addition to many demographical variables including household income, race, if they have children, and education level. Out of the full sample, 15,545 households owned a 2008 year vehicle which I assumed was a new vehicle purchase. The data also contained a total of 3,356 hybrid vehicle observations of various makes, models and years.[4]

The market size, M, is assumed to be the number of households in the U.S. and is obtained from the U.S. Census Bureau. The market share for each vehicle is the quantity sold per market size, $S_j = \frac{q_j}{M}$. The average price for all grades of gasoline is obtained from the U.S Energy Information Agency for years 1999-2008.

[4]The 2009 NHTS actually indicated that there are 7,145 “hybrid or alternative fuel vehicles." However, after inspecting the makes and models of these vehicles, it was concluded that many were actually “alternative fuel" rather than “hybrid" vehicles.

1.4 The Model: Demand Side

In this section I explain the methodology used to model the demand curves for all vehicles in the U.S. auto market. I follow the discrete choice literature for differentiated products which allows tastes to vary both vertically and horizontally across households.

1.4.1 Demand

In any given year, a household i faces the decision to either purchase a new vehicle, $j \in 1, 2, ..., J$, or not purchase a new vehicle, resulting in $J+1$ choices. A household's utility is assumed to be a function of product characteristics, $\{p_j, X_j, \xi_j\}$, household specific preferences and income, $\{y_i, V_i, \epsilon_i\}$, and parameters of the model, $\{\alpha^*, \bar{\beta}, \sigma, \xi_j\}$,

$$U_{i,j}(X_j, y_i, V_i, \epsilon_i; \theta_1, \theta_2) = \alpha^* \log(y_i - p_j) + \bar{\beta} X_j + \sum_k^K \sigma_k x_j^k v_i^k + \xi_j + \epsilon_{i,j}.$$

Households are assumed to choose one product j which yields them the highest utility among the $J+1$ alternatives subject to their financial constraint, $y_i = p_j + z_i$, with z_i representing the amount of the numeraire good consumed. In the BLP(1995) framework, product characteristics are separated into observable, X_j, and unobservable (to the econometrician), ξ_j, characteristics. The latter of which is estimated via a contraction mapping step.

The utility function is decomposed into a mean utility component,

$$\delta_j(X_j; \theta_1) = \bar{\beta} X_j + \xi_j,$$

and a heterogeneous utility component unique to each household,

$$\mu_{i,j}(X_j, y_i, V_i, \epsilon_{i,j}; \theta_2) = \alpha^* \log(y_i - p_j) + \sum_k^K \sigma_k x_j^k v_i^k + \epsilon_{i,j},$$

which allows tastes to vary across households. Here, $x_j^k \in X_j$ is the k^{th} observable characteristic for product j, p_j is the price for product j, y_i represents household i's income, $v_i^k \in V_i$ represents household i's deviation in taste for product characteristic k, and $\epsilon_{i,j}$ represents household i's idiosyncratic taste for product j. The parameters to be estimated are $\theta_1 = \{\bar{\beta}, \xi\}$ and $\theta_2 = \{\alpha^*, \sigma\}$, which correspond to the linear and nonlinear parameters, respectively.

The distribution of tastes for product characteristics are assumed to be normally distributed such that,

$$\beta_i^k = \bar{\beta}^k + \sigma_k v_i^k,$$

where $\bar{\beta}^k$ represents the average taste common across all households, σ_k is a scale parameter to be estimated, and $v_i^k \in V_i \sim N(0, I^K)$. As a result, the distribution of tastes for product characteristic k can be represented by the distribution, $\beta^k \sim N(\bar{\beta}^k, \sigma_k^2)$.

Income is allowed to enter the utility function nonlinearly, $\alpha^* \log(y_i - p_j)$, making high income households less sensitive to prices relative to low income households.[5] The marginal utility of income, as a result, is

$$MU_I = \frac{\partial U_{i,j}}{\partial y_i} = \frac{\alpha^*}{(y_i - p_j)},$$

which increases at a decreasing rate with respect to income while holding prices constant. To help capture income effects, the coefficients are allowed to vary according to income groups:

$$\alpha^* = \begin{cases} \alpha_1, & \text{if } y_i < \bar{y}_1 \\ \alpha_2, & \text{if } \bar{y}_1 \leq y_i < \bar{y}_2 \\ \alpha_3, & \text{if } y_i > \bar{y}_2, \end{cases}$$

where $\bar{y}_1$ and $\bar{y}_2$ divide the population into three equally sized groups ordered by income. Household income is assumed to follow a log-normal distribution, $y \sim \text{Ln N}(\mu_y, \sigma_y)$, where the parameters, μ_y and σ_y, are estimated using a maximum likelihood procedure and a sample of U.S. household incomes obtained from the 2009 Current Population Survey (CPS) published by the U.S. Census Bureau. Table 1.6 reports the income tiers used in estimation.

The demand system is complete by defining the utility from consuming the outside good,

$$u_{i,0} = \alpha^* \log(y_i) + \delta_0 + \epsilon_{i,0},$$

which is identified by normalizing the mean utility from the outside good to zero, $\delta_0 = 0$, and by including a constant and a random coefficient around the constant in the utility specification for the inside goods. The constant, as a result, measures a household's mean utility from simply purchasing a new vehicle.

By specifying a household's idiosyncratic tastes, $\epsilon_{i,j}$, as having an extreme value distribu-

[5]For those households whose simulated income was less than the price for product j, utility was set to zero thus avoiding the problem of taking the log of a negative number.

tion, it can subsequently be integrated out resulting in the logistic distribution,

$$P_j(X,i;\theta_1,\theta_2) = \frac{e^{\delta_j+\mu_{i,j}}}{\sum_{r=0}^{J} e^{\delta_r+\mu_{i,r}}},$$

which is interpreted as the probability a household of type i purchases product j. Furthermore, integrating over the distribution of household types, $i \in M$, yields the probability that any household purchases product j,

$$s_j(X;\theta_1,\theta_2) = \int_i P_j(X,i;\theta_1,\theta_2)di = \int_i \frac{e^{\delta_j+\mu_{i,j}}}{\sum_{r=0}^{J} e^{\delta_r+\mu_{i,r}}}di,$$

which is interpreted as the market share of product j.

1.5 Estimation

In this section I discuss the estimation procedure for the demand side. Parameters are estimated via the BLP (1995) and Petrin (2002) approach using three sets of moment conditions.[6]

1.5.1 BLP Moments

The first set of moment conditions match the share equation above, $s_j(X;\theta_1,\theta_2)$, to the actual product's market shares observed in the data, S_j,

$$||S_j - s_j(\delta(X;\theta_1);\theta_2)|| = 0.$$

The mean utility component, δ_j, is treated as a first stage parameter vector identified by the J moments above using the contraction mapping step defined in BLP(1995) such that the distance between the moments is exactly zero. The parameters, $\theta_1 = \{\beta,\xi\}$, are estimated by a second stage regression using Generalized Least Squares (GLS),

$$\hat{\beta} = (X'W^{-1}X)^{-1}X'W^{-1}\hat{\delta}(S;\theta_2),$$

[6]BLP(1995) and Petrin (2002) estimate the demand and supply side jointly. This helps identify additional random coefficients by providing additional moment conditions in the GMM objective function. However, since micro-moments are use in this chapter, fewer additional moment conditions are required which allows me to estimate the demand and supply side non-jointly following Beresteanu and Li (2009).

where X is the matrix of observable product characteristics and W is a weight matrix defined in a later subsection. The residual is interpreted as the unobservable (to the econometrician) demand characteristics,

$$\hat{\xi}_j(\theta_2) = \hat{\delta}_j(S_j;\theta_2) - \hat{\beta}X_j,$$

which is implicitly a function of θ_2.

The second set of moment conditions relies on the exogeneity assumption between the structural error term above, $\xi(\theta_2)$, and the observable demand side variables, X. To overcome endogeneity problems, a set of instrumental variables, Z, is constructed such that

$$E[Z\cdot\xi(\theta_2)] = 0,$$

which I explain in a later section.

1.5.2 Micro-moments

The third set of moment conditions are based on the work of Petrin (2002) in which additional household level data from the NHTS is incorporated into the model. These additional moments are based on transformations of the share equation forming probabilities (expectations) of demographic variables conditional on a household's new vehicle choice. The model's predicted probability is matched to the micro-based conditional probability by adjusting the parameters of the model.

In this chapter, I use three sets of micro-moments matching the probability of a household's income level conditional on a new vehicle purchase, thus helping identify the income effect parameters $(\alpha_1, \alpha_2, \alpha_3)$ which are critical in estimating marginal costs and markups. These moments are given by,

$$E\Big[I^i_{\text{NHTS}}\{y_i \le \bar{y}_1 \mid i \text{ purchased new vehicle}\} - \bar{P}_{\text{model}}(y \le \bar{y}_1 \mid \text{new vehicle purchase};\theta)\Big] = 0$$
$$E\Big[I^i_{\text{NHTS}}\{\bar{y}_1 \le y_i \le \bar{y}_2 \mid i \text{ purchased new vehicle}\} - \bar{P}_{\text{model}}(\bar{y}_1 \le y \le \bar{y}_2 \mid \text{new vehicle purchase};\theta)\Big] = 0$$
$$E\Big[I^i_{\text{NHTS}}\{y_i > \bar{y}_2 \mid i \text{ purchased new vehicle}\} - \bar{P}_{\text{model}}(y > \bar{y}_2 \mid \text{new vehicle purchase};\theta)\Big] = 0,$$

where $I^i_{\text{NHTS}}\{\cdot\}$ is an indicator variable constructed from the NHTS and $\bar{P}_{\text{model}}$ is the model's predicted conditional mean. The expectation is taken with respect to the number of observations in the NHTS that purchased a new vehicle in 2008. Table 1.6 provides the income levels, $\bar{y}_1$ and $\bar{y}_2$, as well as the model's prediction.

1.5.3 Instrumental Variables

Instrumental Variables (IVs) are critical in identifying BLP models for two reasons. First, price is most likely correlated with unobservable product characteristics, ξ, creating an endogeneity problem. Failure to correct for this will bias the marginal utility of income towards zeros making households appear less sensitive to prices than they actually are. Second, each random coefficient included in the model requires a corresponding IV for identification.

I follow the differentiated products literature of BLP (1995), Breshnahan et al. (1997), Sudhir (2001) and Petrin (2002) and construct optimal instruments. Product characteristics, X_j, are valid instruments for themselves, $Z_j^1 = X_j$. For price, IVs are based on the amount of competition in the market which is measured in two forms. First, the number (count) of competing products that exists for product j. Second, the proximity other products are in characteristic space to product j. Products with closer substitutes are expected to have lower, more competitive, prices.

To improve the measure of competition for product j, three sets of product clusters are defined based on the presence or absence of a distinct characteristic. The variation of product characteristics within and across product clusters are used as IVs. The first two sets of IVs are based on products produced by the same firm, $Z_j^2 = \sum_{r\neq j, r\in J_f}^J X_r$, and across competing firms, $Z_j^3 = \sum_{r\neq j, r\notin J_f}^J X_r$, where J_f is the set of products produced by firm f. The second two sets of IVs are based on products produced by the same firm and of the same type, $Z_j^4 = \sum_{r\neq j, r\in J_f, r\in C_j^t}^J X_r$, and across competing firms and across types, $Z_j^5 = \sum_{r\neq j, r\notin J_f, r\notin C_j^t}^J X_r$, where C^t is the cluster containing all products of type t, (e.g. car, truck, etc.). The last set of IVs are based on on all products in the same segment class, $Z_j^6 = \sum_{r\neq j, r\in C_j^s}^J X_r$, where C^s are all products in segment s (e.g. upper-small, small specialty, etc.).[7]

These IVs vary across products by observing i) they exclude different own-products j, ii) different firms produce different sets of products, iii) different clusters contain different product characteristics, and iv) there is variation across time as the choice set of available products changes each year. Furthermore, by including a constant in X, then $Z_j^2, Z_j^3, Z_j^4, Z_j^5$ and Z_j^6 will contain the number (count) of own-firm products, rival-firm products, and the number of competing products within and across each product cluster.

In all, there are 11 exogenous demand variables, 25 exogenous vehicle class segment

[7]Ward's Automotive Yearbook classifies vehicles into 26 segments based on vehicle size, price, and dominant feature.

dummies, and 7 nonlinear parameters resulting in a total of 43 parameters. Out of the IVs defined above, only 9 are statistically correlated with price and are therefore used in the model, resulting in a total of 2 (= 9 - 7) degrees of freedom. Including micro-moments increased the number of moments by 3, resulting in 5 (= 12 - 7) degrees of freedom.

1.5.4 The Objective Function

The GMM objective function contains two sets of moment conditions. The first is the BLP-like moments and the second is the micro-moments,

$$\begin{aligned} g_1 &= E[Z \cdot \xi(\theta_2)] \\ g_2 &= E[I^i_{\text{NHTS}}\{\cdot\} - \bar{P}_{\text{model}}(\theta_2)], \end{aligned}$$

which are stacked and jointed estimated,

$$G(\theta_2) = \begin{bmatrix} g_1(\theta_2) \\ g_2(\theta_2) \end{bmatrix}.$$

Using a numerical optimization routine, the parameters are adjusted such that the GMM objective function is globally minimized,

$$\hat{\theta}_2 = \min_{\theta_2} G'WG,$$

where the weight matrix, W, equals to the inverse of the variance-covariance of the moment conditions.[8] The asymptotic distribution of the parameter estimates follow the nonlinear GMM literature which has the general form,

$$\sqrt{J}(\hat{\theta} - \theta_0) \sim N\Big(0, (\Gamma' W \Gamma)^{-1} \Gamma' W \Sigma W \Gamma (\Gamma' W \Gamma)^{-1}\Big),$$

where $\Gamma = E[\frac{\partial G(\cdot)}{\partial \theta}]$ and $\Sigma = E[G(\cdot)G(\cdot)']$.[9]

[8]Two-step GMM is used yielding the smallest (i.e. efficient) variances for the parameter estimates. The variance, as a result, reduces to the simplified form, $\sqrt{J}(\hat{\theta} - \theta_0) \sim N(0, \Gamma' W \Gamma)^{-1}$.

[9]See the technical appendix "Micro Moment Methodology" for more details.

1.6 The Model: Supply Side

In this section I discuss the supply side of the model. I follow the approach used by BLP (1995) and Petrin (2002) in which economic theory is used to recover marginal cost estimates. Their approach also allows new equilibrium price and quantity vectors to be solved under various counterfactual scenarios which can then be used to conduct welfare analyses.

1.6.1 Supply

Each multi-product firm $f \in F$ produces a subset of products in the market, $J_f \subset J$. The profit function for each firm is,

$$\pi_f = \sum_{j \in J_f} (p_j - mc_j) M s_j(X; \theta_1, \theta_2) - FC_f,$$

where mc_j is the marginal cost of producing product j, M is the market size, $s_j(\cdot)$ is the share equation recovered from the demand side of the model for product j, and FC_f is the fixed cost for firm f.[10]

Firms are assumed to compete in a Bertrand-Nash market structure. As such, firms simultaneously set prices to maximize profits yielding first order conditions,

$$s_j(\cdot) + \sum_{r \in J_f} (p_r - mc_r) \frac{\partial s_r(\cdot)}{\partial p_j} = 0 \qquad \forall j \in J_f.$$

In matrix form, this is written as,

$$s(X; \theta_1, \theta_2) - \Delta(X; \theta_1, \theta_2)(p - mc) = 0,$$

where

$$\Delta_{jr} = \begin{cases} \frac{\partial s_r}{\partial p_j}, & \text{if j and r are produced by the same firm} \\ 0, & \text{otherwise.} \end{cases}$$

The system of equations is solved recovering price-cost markups,

$$(p - mc) = \Delta(X; \theta_1, \theta_2)^{-1} s(X; \theta_1, \theta_2).$$

[10] The total quantity of product j is equal to the market size multiplied by the share equation, $q_j = M s_j(\cdot)$.

Finally, marginal cost estimates are backed out,

$$mc = p - \Delta(X;\theta_1,\theta_2)^{-1}s(X;\theta_1,\theta_2).$$

1.6.2 Cost Parameters

The marginal cost of producing product j is assumed to be a log-linear function of observable cost shifters. Using the marginal cost estimates from the previous step, cost parameters are estimated using Ordinary Least Squares (OLS),

$$\ln(mc) = \ln\left(p - \Delta(\cdot)^{-1}s(\cdot)\right) = X_c\beta_c + \omega,$$

where X_c and β_c are observable cost characteristics and parameters, respectively. The error term, ω, is interpreted as unobservable product characteristics that are independent and identically distributed across products.

1.6.3 New Equilibrium Prices and Quantities

In the counterfactual market, new equilibrium price and quantities are solved using firms' profit maximizing first order condition defined above,

$$p^{cf} = \hat{m}c + \Delta(p^{cf};\hat{\theta})^{-1}s(p^{cf};\hat{\theta}),$$

where $\hat{\theta}$ are the parameter estimates from the original model, $\hat{m}c$ are the marginal cost estimates from the original model, and p^{cf} is the new equilibrium price vector that solves this system of equations.

The new prices and quantities are used to calculate the change in a firm's variable profits,

$$\Delta\pi_f = \Big(\sum_{j\in J_f^{cf}} \pi_f(p^{cf},\hat{m}c;\theta)\Big) - \Big(\sum_{j\in J_f} \pi_f(p,\hat{m}c;\theta)\Big),$$

where $\pi_f(p^{cf},\hat{m}c;\theta)$ is firm $f's$ profit in the counterfactual market, $\pi_f(p,\hat{m}c;\theta)$ is firm $f's$ profit in the original market, and J_f^{cf} and J_f are the set of products produced by firm f in the counterfactual and original market, respectively. The total change in variable profits for the auto industry as a whole is the sum of each firm's change in variable profits,

$$\Delta\pi_{\text{Industry}} = \sum_{f}^{F} \Delta\pi_f.$$

1.7 Welfare Effects

This section discusses the estimation procedure used to calculate the welfare effects generated from hybrid vehicles entering the U.S. auto market. Several counterfactual simulations are conducted in which hybrid vehicles are removed from the choice set allowing households to re-sort to the next best alternative. When quantifying the benefits of new products, two important economic questions to ask are 1) how much of the benefit was generated from the increased product variety and 2) how much was generated from the increased price competition. Following Hausman and Leonard (2002), I estimate both price and variety effects and add to their work by allowing income to enter the indirect utility function nonlinearly.

1.7.1 Compensating Variation

The total consumer welfare generated from hybrid vehicles is measured using compensating variation (CV); the dollar amount that an individual would need to be paid, in dollar terms, to adjust his/her counterfactual utility level back to its original state. Compensating variation lends itself easily to the indirect utility function,

$$\begin{aligned} &\max_{j\in 0,1,\ldots,J} \left\{\alpha^* \log(y_i - p_j) + \delta_j + \mu_{i,j} + \epsilon_{i,j}\right\} \\ = &\max_{r\in 0,1,\ldots,J^{cf}} \left\{\alpha^* \log(y_i - p_r^{cf} - cv_i) + \delta_r + \mu_{i,r} + \epsilon_{i,r}\right\}, \end{aligned}$$

where p_j is the original market price, p_r^{cf} is the counterfactual price, $J^{cf} \subset J$ is the counterfactual choice set, and cv_i is the amount individual i would need to be compensated to match the two utility levels. In this analysis, all hybrid vehicles are removed from the counterfactual market. Since income enters the indirect utility function nonlinearly, the log-sum formula proposed by Small and Rosen (1981) cannot be used. I therefore rely on simulation techniques.[11]

[11] See Appendix II "Decomposing Compensating Variation into Price and Variety Effects with Nonlinear Marginal Utility of Income" for more details.

1.7.2 Price and Variety Effects

Following Hausmand and Leonard (2002), both Price Effects (PE) and Variety Effects (VE) are estimated. The PE measures the consumer welfare generated from the increased price competition as a result of hybrid vehicles entering the market. Similarly, assuming hybrid vehicles are in greater accordance to some consumer's preferences, the VE measures the allocative efficiency gain. To estimate the PE, a household's utility is simulated using two market scenarios where the choice set (i.e. variety) remains fixed across both markets. Only prices are allowed to vary,

$$\begin{aligned}&\max_{j \in 0,1,\ldots,J^{cf}} \left\{ \alpha^* \log(y_i - p_j) + \delta_j + \mu_{i,j} + \epsilon_{i,j} \right\} \\ = &\max_{r \in 0,1,\ldots,J^{cf}} \left\{ \alpha^* \log(y_i - p_r^{cf} - cv_i^{pe}) + \delta_r + \mu_{i,r} + \epsilon_{i,r} \right\},\end{aligned}$$

where J^{cf} is the counterfactual choice set with all hybrid vehicles removed and p^{cf} is the corresponding vector of counterfactual equilibrium prices. The PE, cv_i^{pe}, is therefore the amount of money that household i would need to be compensated as a result of the price changes generated by the increased competition from hybrid vehicles entering the market.

Similarly, the VE is estimated by simulating a household's utility, again using two market scenarios. For this effect, prices remain fixed at original market levels across both markets, while only the choice set (i.e. variety) is allowed to vary,

$$\begin{aligned}&\max_{j \in 0,1,\ldots,J} \left\{ \alpha^* \log(y_i - p_j) + \delta_j + \mu_{i,j} + \epsilon_{i,j} \right\} \\ = &\max_{r \in 0,1,\ldots,J^{cf}} \left\{ \alpha^* \log(y_i - p_r - cv_i^{ve}) + \delta_r + \mu_{i,r} + \epsilon_{i,r} \right\}.\end{aligned}$$

The VE, cv_i^{ve}, is therefore the amount of money that household i would need to be compensated as a result of households having hybrid vehicles (i.e. additional variety) available in their choice set.

1.8 Results

In this section I discuss the results organized into 6 subsections: descriptive statistics, parameter estimates, markups and producer welfare, elasticities and consumer welfare, total welfare, and specification and robustness checks.

1.8.1 Descriptive Statistics

Vehicles are classified into 26 segment classes according to their size, price, and dominant feature. Each year the three most competitive vehicle classes, based on the number of competing models in each segment, are the “upper middle" class (e.g. Chevy Malibu, Honda Accord, Toyota Camry, and Toyota Prius), “upper small" (e.g. Ford Fusion, Honda Civic, and Toyota Corolla) and the “lower luxury" class (e.g. Acura TL, BMW 3-series, Lexus ES, and Mercedes C-Class). Table 1.7 provides the number of vehicles in each segment.

The sales-weighted average msrp (minus federal tax incentives) for hybrid vehicles is $23,782 with the lowest priced hybrid being the 2005 Honda Civic, at $20,530, and the highest priced hybrid being the 2005 Lexus RX 400, at $52,085. The median quantity of hybrids sold each year is 19,119 with the Honda Insight selling the fewest in 2002 at 2,216, and the Toyota Prius selling the most at 181,221 in 2007.[12] The sales-weighted average horsepower for hybrid cars is 93 which is substantially less than that for all other cars, at 167. Although not as powerful, hybrid cars offer increased fuel efficiency with a sales-weighted average mpg of 51.6 and 47.4, for city and highway driving, respectively, while the sales-weighted average for non-hybrid cars is 23.4 and 31.4 mpgs, respectively.[13] See Tables 1.4 and 1.5 for a complete summary.

1.8.2 Parameter Estimates

In total, four demand models are estimated: OLS Logit, TSLS Logit with IV correction, random coefficients with IV correction, and random coefficients with IV correction and micro-moments. The first two models follow Berry (1994) in which no heterogeneous taste coefficients are included in the model.[14] These two models are intended to serve as a baseline to compare the results from the full models to. The need to instrument for price is evident by comparing the price coefficients for the OLS and TSLS models. Table 1.8 shows that without IVs, the price coefficient is severely biased downward, making consumers appear relatively

[12]The Honda Insight actually had fewer sales in 2006, with only 722 units being sold. However, to improve the efficiency of the estimation routine, I only examine vehicles that had sales greater than 2,000 per year.

[13]MPG's are based on manufacturer sticker prices reported at the time of sale rather than the adjusted MPG published by the EPA. This seems to be a reasonable assumption because at the time of purchase, this is the only information the consumer had available to make his/her purchase decision.

[14]In the OLS and TSLS models, income is not interacted with price. Therefore, αp_j is included in the model specification rather than $\alpha^* \log(y_i - p_j)$. As a result, the α coefficients between the baseline and full models are not directly comparable.

unresponsive to prices, which in turn cause markups to be unrealistically high. Including IVs helped correct for this, more than tripling the price coefficient, resulting in significantly lower, yet still unreasonable, markups.

The last two models correct for the endogeneity problem and also include random coefficients which allows tastes to vary across households. The first of these models follow BLP (1995) in which only market-level data is used. The second model follows Petrin (2002) in which additional micro-level data is incorporated into the model primarily intended to help identify the price coefficients. With micro data, two of the three price coefficients are significant at the 5% level, whereas only one is significant at the 5% level without micro data. See Table 1.8 for all price coefficients.

For the two models with random coefficients, preferences are assumed to vary horizontally across vehicle types (car, suv, and trucks) whereas preferences for vehicle characteristics (size, acceleration, fuel efficiency, and luxury) are assumed to vary vertically.[15] This way, more of a characteristic is universally preferred to less within each vehicle type. For example, regardless of whether a household owns a Hummer H2 or a Toyota Prius, it is reasonable to assume that both households would prefer more fuel efficiency to less. Furthermore, this specification also suggests that a Hummer H2 owner presumably really likes SUVs whereas a Prius owner really likes cars; perhaps even disliking SUVs. See Table 1.9 for all random coefficient estimates.

In all four models, the mean taste coefficient for hybrid vehicles is negative indicating that the average household dislikes hybrid vehicles. However, a quadratic time trend is also interacted with the hybrid variable intended to capture the consumer adoption and diffusion process of the hybrid innovation. Table 1.8 shows that in all four models, the time trend is positive indicating that over time, the average household actually began to prefer hybrid vehicles.[16]

Several control variables are included in the model to account for macroeconomic effects that may have influenced a consumer's decision to purchase a new vehicle. These variables are reported in Table 1.9 and include a quadratic time trend, GDP percentage change, and a 2008 dummy variable intended to capture the economic downturn. Finally, 25 vehicle segment dummies are included in the model to control for as many unobservable product characteristics as possible.

---

[15]A random coefficient was originally included for hybrid vehicles, but was later removed due to identification issues.

[16]According to the model, this transition took place in 2005.

The cost side parameters are estimated using the marginal cost estimates from the full model with micro-moments. All parameters, except for the small truck segment dummy, are significant at the 5% level and are of the expected sign. It costs more to produce larger, more powerful, luxury vehicles. The hybrid cost coefficient, $\beta_{\text{hybrid}} = 0.22$, indicates that, on average, hybrid vehicles cost 22% more than non-hybrid vehicles to produce. For example, if a hybrid vehicle were to cost $15,000 to produce, $3,300 of that would be the costs associated with the hybrid technology (battery, hybrid power-train, etc.).[17] See Tables 1.10 for all cost parameter estimates.

1.8.3 Mark-ups, Producer Welfare, and Profit Dissipation

In this section I summarize marginal cost and markup estimates for all vehicles and separately summarize them for hybrid vehicles. The estimated markups are compared across all four demand specifications to demonstrate the importance of both instrumenting for endogenous price and allowing tastes to vary across households. I also explore, through several counterfactual scenarios, the change in each manufacturer's variable profits with and without all hybrid vehicles in the market, the change in Toyota's profits with and without the Prius in the market, and the business stealing effect from Toyota's competitors.

Marginal costs and markups are reported in Table 1.11 for all vehicles and compared across each of the four demand specifications. In the first two models, OLS Logit and TSLS Logit, the average markup is $40,548 and $11,807, corresponding to 78% and 1% negative marginal cost estimates, respectively, which is clearly unrealistic. As Petrin points out, the difference between these two estimates is driven almost entirely by the price coefficient, α, and thus demonstrates the importance of instrumenting for endogenous price.

To improve the marginal cost and markup estimates, random coefficients are included in the model. The average markup without micro-moments range from 17% to 50% with an average of 27%. With micro-moments, the markups are slightly less ranging from 17% to 38% with an average of 25%. These estimates are within the range of what previous papers have found, namely BLP (1995), Goldberg (1995), and Petrin (2002).[18] Furthermore, all marginal

[17]This is consistent with Auto USA which reports that hybrid battery replacement costs range from $3,200 to $3,400.

[18]Goldberg (1995), using a nested logit model, estimated markups to have an average of 38%. BLP(1995), using various model specifications and only market-level data, estimated average markups as low as 15% to as high as 40%. Petrin (2002) estimated average markups with and without micro-moments to be 16.7% and 40.7%, respectively.

cost estimates are positive in both models.

Table 1.12 reports the marginal cost and markups for hybrid vehicles only. Markups range from 18% to 28% with an average of 22% which is slightly less than that for all other vehicles. To help the reader visualize how the markups vary across each hybrid model, Table 1.13 reports the marginal cost and markup estimates for all hybrid vehicles sold in 2007. The Honda Civic hybrid has the smallest markup at $4,496 corresponding to a 20% markup, while the Lexus RX 400 hybrid has the largest markup at $11,435 corresponding to a 27% markup.

In the first counterfactual analysis, all hybrid vehicles are removed from the market allowing new equilibrium price and quantities to be solved, which are then used to construct counterfactual variable profits for each manufacturer. Marginal cost estimates are assumed to remain constant. Table 1.14 shows that Toyota and Honda, the two largest producers of hybrid vehicles, were the only two manufacturers to benefit from the hybrid innovation whereas all other manufacturers were hurt. Toyota benefited the most, with variable profits consistently growing from $20.7 million in 2000 to an impressive $1.35 billion in 2007, representing 0.23% and 8.33% of their total variable profits, respectively. Honda's variable profits were relatively modest growing from $12.3 million in 2000 to $148.0 million in 2005, corresponding to 0.22% and 1.73% of their total variable profits, respectively, before declining to just $34.6 million in 2007, corresponding to 0.38% of their total variable profits.[19] GM suffered the largest loss from the hybrid innovation, losing a total of $1.22 billion in variable profits over the course of nine years.

With all hybrid vehicles removed from the market, vehicle prices both increased and decreased depending on the manufacturer. Of the 227 vehicles that experienced price decreases, 75% were produced by Toyota and Honda, the two largest producers of hybrid vehicles, and 9% were produced by other manufacturers of hybrid vehicles, namely Ford, GM, and Nissan. These findings suggests that firms that had previously relied on hybrid technology to differentiate their products, now had to rely on lower prices, rather than hybrid technology, to capture market shares. Similarly, vehicles that experienced price increases were typically produced by manufacturers who either did not produce hybrid vehicles or only represented a small share in the hybrid vehicle market. As a result, these firms had less competition with hybrid vehicles removed from the market and could, as a result, increase prices.

The first counterfactual market also suggests that, had hybrid vehicles never been intro-

[19] I ignore 2008 simply because it was an off year for the auto market as a whole due the the economic downturn.

duced, the Toyota Camry would have sold an additional 18,651 units in 2007 representing the largest increase in sales among all other vehicles. The Toyota Matrix and Honda Accord saw the next largest increase in sales, selling an additional 12,176 and 9,356 units, respectively. See Table 1.15 for a ranking of the top 20 vehicles with the greatest change in sales.

In the second counterfactual analysis, only the Toyota Prius is removed from the market. Counterfactual equilibrium prices and quantities are solved which are then used to construct Toyota's variable profits with and without the Prius in the market, again using marginal cost estimates recovered from the original market. Table 1.16 suggests that had the Prius never been introduced, Toyota would have lost a total of $2.62 billion, corresponding to 5.12%, of their total variable profits over the course of nine years. This large loss suggests that the Prius did not cannibalize any of Toyota's profits that would have been generated from, say, the Toyota Camry.

In the final counterfactual analysis, the business stealing effect is estimated to determine how much larger Toyota's profits would have been had competitors never entered the hybrid vehicle market. To estimate this effect, all hybrid vehicles are removed from the market except for Toyota's, again allowing new equilibrium price and quantities to be solved. Counterfactual variable profits are constructed using counterfacutal equilibrium prices and quantities and marginal cost estimates recovered from the original market. The difference between Toyota's original and counterfactual variable profit is interpreted as the "Business Stealing Effect." Table 1.17 shows that from 2000 to 2008, competitors in the hybrid vehicle market, namely Honda, Ford, GM and Nissan, were relatively unsuccessful in stealing business away from Toyota, capturing $200.8 million, corresponding to 0.39%, of Toyota's total variable profits during this time. Though significant, few close substitutes appear to exist for Toyota's hybrid vehicles, most notably the Prius, namely because of its conspicuous green look.

1.8.4 Elasticities and Consumer Welfare

In this section I report the cross-price elasticities for all hybrid vehicles. This helps provide a picture of how many close substitutes exist for each model. Hybrid vehicles with fewer substitutes should generate larger welfare gains, and vice versa. I then summarize the total gain in consumer welfare, the price effect, and the variety effect generated from hybrid vehicles.

Cross-price elasticities for all 2007 hybrid truck and car models are reported in Tables 1.18 and 1.19, respectively. The Honda CR-V and Toyota Rav4 are the top two substitutes for the Ford Escape, Mercury Mariner and Saturn Vue hybrid models. The cross-price elasticity estimates indicate, for example, that a 1% increase in the price of the Ford Escape Hybrid will result in market shares for the Honda CR-V to increase by 0.114% as consumers substitute between the two vehicles. In the hybrid car segment, the Toyota Camry, Honda Accord, Toyota Matrix and Chevrolet Impala are all top substitutes for the various hybrid car models. For example, a 1% increase in the price of the Toyota Prius will result in market shares for the Toyota Camry to increase by 0.179%.

Table 1.20 and 1.21 report the Total Compensating Variation (CV) and Variety Effect (VE), respectively, conditional on purchasing a hybrid vehicle. The similarity between the VE and total CV estimates indicate that the majority of the welfare generated from hybrid vehicles is driven by the increased variety, rather than price effects. Each year, the total CV conditional on purchasing a hybrid vehicle grew from a median of $4,280 in 2000 to $5,475 by 2008. A relatively small number of hybrid vehicle owners, due to their extreme tastes, generated substantial welfare gains with VE estimates as high as $153,545 in 2007. Although high, only 2% of hybrid owners needed compensation greater than $20,000 and less than 1% needed compensation greater than $30,000.[20] These extreme welfare measures are also consistent with what Petrin (2002) found. Table 1.22 reports Price Effects (PE) which are found to be relatively modest, generating between $0.01 and $13 in benefits for new vehicle purchasers between the years 2000 and 2008. These small price effects are not surprising given the small market share of hybrid vehicles in the U.S. auto market.

1.8.5 Total Welfare Change

Table 1.23 shows that by 2008, hybrid vehicles had generated a total of $12.8 billion in welfare for all new vehicle purchasers combined. Hybrid vehicle owners captured the majority of this, with compensating variation totaling $12.0 billion. Consumers of non-hybrid vehicles also benefited due to the increased price competition, though substantially less, with price effects totaling $780.3 million. On the producer side, Toyota and Honda were the only two manufacturers to benefit from the hybrid innovation, with variable profits totaling $4.7 billion and $490 million, respectively. All other manufacturers were hurt, with GM losing the most at $1.2 billion in variable profits over the course of nine years. Since there were both winners and losers from the hybrid innovation, the auto industry as a whole only benefited a

[20]Compensating variation, variety effects, and price effects are estimated using 1 million simulations per year.

total \$1.9 billion in variable profits. In summary, hybrid vehicles generated a total of \$14.7 billion in welfare for the U.S. economy as a whole, with hybrid owners, non-hybrid owners, and producers capturing 68.7%, 4.5%, and 26.8% of the benefits, respectively.

1.8.6 Specification and Robustness Checks

In this subsection I discuss various model specifications and robustness checks. I test how the inclusion of various random coefficients affect the model in addition to various distributional assumptions for unobserved tastes. I also examine how sensitive the model is, in particular markups, to simulated household income distributions. Finally, I test the estimation procedure to ensure that the GMM objective function is globally minimized.

Two possible sets of random coefficients are included in the model. The first set is based on vehicle characteristics (size, acceleration, and fuel efficiency) and the second set is based on vehicle type (car, suv, truck). To determine which set of random coefficients to use, I relied on the matrix of cross-price elasticities and the significance of parameter estimates. I found that the model with random coefficients for vehicle characteristics yielded unreasonable substitution patters. For example, the Toyota Tacoma was found to be a close substitute to the Toyota Prius, which is clearly unrealistic. The model with random coefficients for vehicle types, however, yielded vary reasonable substitution patterns and was therefore used in the final model. A random coefficient was also included for the hybrid variable but was subsequently removed because of difficulties identifying the parameter.[21]

Various distributional assumptions for unobserved tastes were also experimented with including the normal, log-normal, and chi-squared distributions at various truncations (95% and 99%). Although ad-hoc, the 95% truncated normal distribution tended to yield the most realistic cross-price elasticities and was therefore used in the final model. The model was also found to be somewhat sensitive to the distribution of simulated U.S. household incomes. Occasionally the random set of simulated incomes contained an extreme household income draw (e.g. income > \$2 million) resulting in unreasonably large markups. This, of course, is driven by wealthier households being less sensitive to prices. To avoid this problem, I truncated the simulated U.S. household income distribution at \$350,000. In total, 400 simulated random draws are used to approximate the distribution of household incomes

[21]When a random coefficient was included for the hybrid vehicle dummy variable, the GMM objective function appeared to be flat, making it difficult for the numerical optimization routine to move in the right direction. This is presumably due to a lack of variation as a result of hybrid vehicles' small market share.

and unobservable tastes.

To guarantee the global minimization of the GMM objective function, I follow Knittle and Metaxoglou's (2009) suggestion and use multiple starting values and experiment with three different optimization packages. As numerous papers have pointed out, the BLP methodology is prone to numerical errors driven primarily by the estimation of nonlinear parameters and a non-globally concave objective function. The first two optimization packages I use are MATLAB's fminsearch and fminunc. The first of which is a non-gradient direct line search based on the Nedler-Mead Simplex method, while the second method uses a gradient-based trust region method. The first of these, fminsearch, did a poor job finding the minimum while the second, fminunc, did reasonably well, but was found to be inefficient (i.e. slow). The last optimzation package used was SolvOpt, which proved to be the best, consistently converging to the global minimum relatively quickly.[22] Knittle and Metaxoglou also found this optimization package to be the best when compared to nine other optimization packages examined in their paper. I also experimented with multiple starting values. However, when using the SolvOpt optimization package this proved to be unnecessary. A total of 10 starting values were used, all converging to the same GMM function value and same parameter estimates.

1.9 Conclusion

This chapter examined the welfare effects from hybrid vehicles entering the U.S. auto market between the years 2000 and 2008. Several structural discrete choice models are used to estimate the demand and supply curves for all vehicles in the U.S auto market. Using several counterfactual analyses, I measure the changes in producer surplus, the extent of first mover advantages, and business stealing effects from competitors. Consumer welfare is examined from a productive and allocative efficiency standpoint by decomposing compensating variation into price and variety effects, respectively. These effects are estimated conditional on both a consumer's original product choice and preference type, thus giving more precise and realistic welfare measures. Furthermore, this chapter developed a simulation technique to recover these effects while allowing income to enter the indirect utility function nonlinearly.

Although hybrid vehicles represent a relatively small share of the total U.S. auto market, there is strong evidence that some households are willing to pay a premium for "green" and

[22] SolvOpt is available for free at http://www.kfunigraz.ac.at/imawww/kuntsevich/solvopt/ .

environmentally-friendly vehicles. Toyota was the first to recognize this market niche and responded by developing the Prius. Since the introduction, Toyota has enjoyed significant first mover advantages earning substantial economic profits and sustained growth in the hybrid vehicle market segment. Their success can be attributed to the strong branding and unique design of the Prius which gives consumers a vehicle with both high fuel economy and one that sends a signal to the community that they are green and living a low resource intensive lifestyle. Competing hybrid vehicle models merely have an inconspicuous decal on their back bumper making them virtually indistinguishable from their all gas equivalent.

The new goods literature has thus far done a fairly adequate job in developing structural models to estimate consumer surplus. Producer surplus, however, has been limited to variable profit estimates which, consequentially, ignore fixed costs. This additional information may be useful in explaining firm entry decisions in monopolistically competitive market structures. This remains an area for future research.

Table 1.1: U.S. Vehicles Sold Per Year

Year	Number of Hybrid Models	New Vehicles Sold	Total Number of Hybrids Sold	Percent Sold Hybrid
1999	0	16,817,612	-	0.00%
2000	2	17,375,172	9,350	0.05%
2001	2	17,379,274	20,282	0.12%
2002	3	16,840,112	36,035	0.21%
2003	3	16,594,265	46,427	0.28%
2004	4	16,923,901	82,555	0.49%
2005	7	17,061,212	208,047	1.22%
2006	9	16,551,851	250,130	1.51%
2007	12	16,167,174	348,920	2.16%
2008	12	13,127,201	306,080	2.33%

[a] Data obtained from Ward's Automotive Yearbook (1999-2000).

Table 1.2: Hybrids Sold Per Manufacturer

(Totals)

Year	Honda	Toyota	Ford	GM	Nissan	Total
2000	3,788	5,562	-	-	-	9,350
2001	4,726	15,556	-	-	-	20,282
2002	15,916	20,119	-	-	-	36,035
2003	21,800	24,627	-	-	-	46,427
2004	25,571	53,991	2,993	-	-	82,555
2005	42,690	146,560	18,797	-	-	208,047
2006	36,849	189,958	23,323	-	-	250,130
2007	35,980	275,041	25,108	4,403	8,388	348,920
2008	31,297	239,797	19,502	6,665	8,819	306,080

(Percents)

Year	Honda	Toyota	Ford	GM	Nissan	Total
2000	41%	59%	-	-	-	100%
2001	23%	77%	-	-	-	100%
2002	44%	56%	-	-	-	100%
2003	47%	53%	-	-	-	100%
2004	31%	65%	4%	-	-	100%
2005	21%	70%	9%	-	-	100%
2006	15%	76%	9%	-	-	100%
2007	10%	79%	7%	1%	2%	100%
2008	10%	78%	6%	2%	3%	100%

[a] Data obtained from Ward's Automotive Yearbook (2000-2008).

Table 1.3: Toyota Prius Sales

Year	Total Number of Hybrids Sold	Total Number of Priuses Sold	Percent Sold Prius
1999	-	-	-
2000	9,350	5,562	59.5%
2001	20,282	15,556	76.7%
2002	36,035	20,119	55.8%
2003	46,427	24,627	53.0%
2004	82,555	53,991	65.4%
2005	208,047	107,897	51.9%
2006	250,130	106,971	42.8%
2007	348,920	181,221	51.9%
2008	306,080	158,884	51.9%

[a] Data obtained from Ward's Automotive Yearbook (1999-2008).

Table 1.4: Summary Statistics: Vehicle Characteristics (Cars)

All Cars

	Quantity	MSRP[b]	Size[c]	MPG Avg[d]	Horsepower	Weight	Gas Price
Average[e]		$ 24,441	1.31	27.1	167	3,087	
Average	57,828	$ 32,663	1.30	26.0	193	3,218	2.32
Min	2,004	$ 10,674	0.65	15.3	55	1,288	1.44
Q1	12,913	$ 20,044	1.20	22.0	140	2,855	1.60
Q2	30,290	$ 26,651	1.32	24.5	184	3,245	2.34
Q3	71,177	$ 39,591	1.40	27.5	227	3,590	2.85
Max	448,162	$ 109,485	1.73	65.6	500	4,938	3.32

Hybrid Cars

	Quantity	MSRP[b]	Size[c]	MPG Avg[d]	Horsepower	Weight	Gas Price
Average[e]		$ 22,403	1.21	49.1	93	2,998	
Average	38,620	$ 23,521	1.21	47.6	113	2,994	2.32
Min	2,216	$ 20,530	0.77	32.9	67	1,856	1.44
Q1	8,604	$ 21,259	1.16	38.9	73	2,765	1.60
Q2	24,627	$ 21,986	1.19	48.6	85	2,890	2.34
Q3	39,424	$ 23,841	1.32	52.2	147	3,492	2.85
Max	181,221	$ 33,034	1.37	65.6	255	3,680	3.32

[a] Data obtained from Ward's Automotive Yearbook (1999-2008).
[b] Federal taxes are subtracted from MSRP.
[c] Size =(Length x Width)/10,000.
[d] MPG Avg. assumes 55% city driving and 45% highway driving.
[e] Sales-weighted average.

Table 1.5: Summary Statistics: Vehicle Characteristics (Trucks)

All Light Trucks

	Quantity	MSRP^b	Size^c	MPG Avg^d	Horsepower	Weight	Gas Price
Average^e		$ 26,677	1.46	20.0	203	4,044	
Average	77,079	$ 31,660	1.43	19.9	216	4,177	2.15
Min	2,054	$ 14,589	0.98	12.5	97	2,269	1.22
Q1	16,825	$ 22,450	1.30	17.4	170	3,639	1.56
Q2	41,139	$ 28,414	1.42	19.5	210	4,081	1.92
Q3	89,419	$ 38,688	1.56	22.3	256	4,700	2.64
Max	891,482	$ 88,964	2.13	33.1	403	7,819	3.32

Hybrid Light Trucks

	Quantity	MSRP^b	Size^c	MPG Avg^d	Horsepower	Weight	Gas Price
Average^e		$ 35,108	1.31	29.5	183	4,049	
Average	13,952	$ 33,387	1.31	29.4	179	4,019	2.15
Min	2,329	$ 23,233	1.22	21.5	133	3,466	1.22
Q1	3,734	$ 27,093	1.24	28.7	133	3,717	1.56
Q2	17,291	$ 29,918	1.31	30.0	172	3,789	1.92
Q3	20,155	$ 38,291	1.35	30.7	208	4,190	2.64
Max	31,485	$ 52,085	1.60	33.1	332	5,617	3.32

[a] Data obtained from Ward's Automotive Yearbook (1999-2008).
[b] Federal taxes are subtracted from MSRP.
[c] Size =(Length x Width)/10,000.
[d] MPG Avg. assumes 55% city driving and 45% highway driving.
[e] Sales-weighted average.

Table 1.6: Predicted Income Probability Conditional on New Vehicle Purchase

Income Tier[a]	"True" Probability[b]	Model's Prediction	
		$\mu_{i,j} \neq 0$	$\mu_{i,j} \neq 0$
		(1)	(2)
P(Income $\leq$ \$36,494 | New)	13.42%	13.34%	32.76%
P(\$36,494 < Income $\leq$ \$71,167 | New)	27.44%	27.43%	29.75%
P(Income > \$71,167 | New)	59.15%	59.04%	37.46%
Uses NHTS Micro Data		Yes	No

[a] Income tiers divide households into three equally sized groups based on income.
[b] The "true" probability is based on 11,915 observations from the 2009 National Household Travel Survey (NHTS).

Table 1.7: U.S. Vehicle Segment Class Summary

	Year									
Segment	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Lower Small	4	6	6	4	4	5	4	7	7	6
Upper Small	23	23	23	25	22	23	22	23	22	22
Small Specialty	3	4	4	5	5	6	6	6	6	8
Lower Middle	5	6	6	6	5	7	7	4	3	3
Upper Middle	22	26	24	24	22	25	27	26	26	25
Middle Specialty	10	10	10	11	9	8	8	9	10	11
Large Car	10	9	9	7	7	11	10	10	9	8
Lower Luxury	14	16	18	18	19	20	21	22	20	16
Middle Luxury	14	14	13	13	13	12	14	12	11	10
Upper Luxury	9	10	9	8	9	8	8	7	8	6
Luxury Specialty	3	4	6	6	5	6	5	4	6	4
Luxury Sport	9	10	9	10	11	11	11	11	11	11
Small Cross	2	3	3	4	5	6	8	9	10	9
Middle Cross	4	7	10	12	15	18	19	21	23	21
Middle Luxury Cross	1	2	4	5	11	11	11	11	13	14
Large Cross	0	0	1	2	3	3	3	3	9	9
Large Luxury Cross	0	1	1	1	2	2	3	5	5	5
Small SUV	3	3	4	4	4	4	3	4	4	4
Middle SUV	11	11	14	15	15	15	16	14	14	12
Middle Luxury SUV	7	7	7	8	8	8	8	7	7	7
Large SUV	6	7	6	6	6	7	7	7	7	8
Large Luxury SUV	5	6	6	7	9	10	11	10	10	9
Small Van	11	12	12	11	11	12	12	11	7	6
Large Van	2	3	3	3	3	2	1	1	1	1
Small Truck	8	8	7	7	8	10	8	10	9	8
Large Truck	8	9	10	11	10	10	12	12	12	13
Total	194	217	225	233	241	260	265	266	270	256

[a] Data obtained from Ward's Automotive Yearbook (1999-2008).

Table 1.8: Parameter Estimates for the Demand Side: Base Coefficients

Variable	Random Coefficient $\mu_{i,j} \neq 0$ (1)	Random Coefficient $\mu_{i,j} \neq 0$ (2)	IV Logit $\mu_{i,j} = 0$ (3)	OLS Logit $\mu_{i,j} = 0$ (4)
Constant	-17.01^{*}	-18.69^{*}	-9.03^{**}	-8.25^{**}
	(9.25)	(10.83)	(0.40)	(0.30)
Size	3.85^{*}	4.03	2.42^{**}	1.52^{**}
	(2.36)	(9.24)	(0.37)	(0.24)
HPWT	0.08^{*}	0.09	0.04^{*}	-0.31^{**}
	(0.045)	(0.15)	(0.032)	(0.021)
DPM	-0.03^{**}	-0.02^{*}	-0.08^{**}	-0.18^{**}
	(0.01)	(0.01)	(0.038)	(0.019)
Hybrid	-3.07	-3.09	-3.89^{**}	-4.40^{**}
	(3.08)	(6.77)	(1.25)	(1.13)
Hybrid*Time	0.88	0.82	0.96^{**}	0.96^{**}
	(0.92)	(2.20)	(0.41)	(0.370)
Hybrid*Time^2	-0.06^{**}	-0.05	-0.07^{**}	-0.07^{**}
	(-0.03)	(-0.09)	(0.030)	(0.028)
GDP	0.01^{**}	0.00^{**}	0.01	0.015
	(0.002)	(0.000)	(0.024)	(0.023)
2008	-0.48^{*}	-0.54^{*}	-0.18^{*}	-0.069
	(0.31)	(0.34)	(0.14)	(0.12)
Time	-0.05	-0.09	-0.67^{*}	-0.26^{**}
	(0.076)	(0.16)	(0.42)	(0.052)
Time^2	-0.01	0.00	0.37	0.033^{**}
	(0.00)	(0.00)	(0.39)	(0.005)
Price Coefficients				
α_1	4.49	1.013	-0.086^{**}	-0.025^{**}
	(6.83)	(3.40)	(0.002)	(0.003)
α_2	7.94^{**}	4.56^{*}		
	(2.82)	(4.31)		
α_3	13.10^{**}	11.56^{**}		
	(1.27)	(1.48)		
J-stat (degrees of freedom)	8.00(5)	13.34(2)	-	-
R-squared	-	-	0.35	0.19
Uses NHTS Micro Data	Yes	No	No	No

[a] * t-statistics > 1, ** t-statistics > 2.
[b] Demand parameters for the 25 vehicle segment variables were also included in the model but were excluded from the table to save space.
[c] Number of observations used in estimation = 2,427.

Table 1.9: Parameter Estimates for the Demand Side: Random Coefficients

Variable	Random Coefficient $\mu_{i,j} \neq 0$ (1)	Random Coefficient $\mu_{i,j} \neq 0$ (2)
Constant	0.125	0.255
	(15.35)	(29.29)
Car	6.864**	7.220**
	(3.318)	(1.842)
SUV	5.353**	5.979**
	(1.736)	(1.781)
Truck	6.207	13.554
	(9.962)	(15.94)
J-stat (degrees of freedom)	8.00(5)	13.34(2)
Uses NHTS Micro Data	Yes	No

[a] * t-statistic > 1, ** t-statistic > 2.
[b] Number of observations used in estimation 2,427.

Table 1.10: Parameter Estimates for the Cost Side

Variable	Parameter Estimate	Standard Error	T-Stat
Constant	8.665**	0.047	182.622
Log(Size)	0.578**	0.057	10.123
Log(HPWT)	0.248**	0.027	9.268
Hybrid	0.224**	0.031	7.247

[a] Cost parameter estimates are based on marginal cost (MC) estimates recovered from the full model with random coefficients and micro-data.
[b] * t-statistic > 1, ** t-statistic > 2.
[c] Cost parameters for the 26 vehicle segment variables were also included in the model but were excluded from the table to save on space.
[d] Number of observations used in estimation = 2,427.

Table 1.11: Implied Markups Derived From Demand-Side Estimates and Bertrand-Nash Pricing Assumption: All Vehicles (1999-2000)

	Random Coefficients $\mu_{i,j} \neq 0$ (1)			Random Coefficients $\mu_{i,j} \neq 0$ (2)			IV Logit $\mu_{i,j} = 0$ (3)		OLS Logit $\mu_{i,j} = 0$ (4)	
Statistic	MC	MrkUp	% markup	MC	MrkUp	% markup	MC	MrkUp	MC	MrkUp
Average	\$ 24,604	\$ 7,620	25	\$ 24,399	\$ 7,825	27	\$ 20,417	\$ 11,807	\$ -8,324	\$ 40,548
5th Percentile	\$ 10,658	\$ 3,310	17	\$ 8,395	\$ 4,589	17	\$ 3,390	\$ 11,587	\$ -25,378	\$ 39,793
Q1	\$ 16,161	\$ 4,721	21	\$ 14,954	\$ 5,711	21	\$ 9,400	\$ 11,615	\$ -19,392	\$ 39,890
Q2	\$ 20,647	\$ 7,228	23	\$ 22,003	\$ 6,712	23	\$ 15,927	\$ 11,797	\$ -12,941	\$ 40,513
Q3	\$ 29,054	\$ 10,364	26	\$ 30,025	\$ 9,495	29	\$ 27,167	\$ 11,965	\$ -1,581	\$ 41,092
95th Percentile	\$ 54,922	\$ 12,957	38	\$ 53,564	\$ 14,162	50	\$ 56,239	\$ 12,119	\$ 27,348	\$ 41,619
% Negative MC Estimates	0			0			1%		78%	
Uses NHTS Micro Data	Yes			No			No		No	

[a] Markups are not reported for IV and OLS logit because of the presence of negative marginal cost estimates.
[b] $\mu_{i,j} \neq 0$ indicates that random coefficients were included in the model.

Table 1.12: Implied Marginal Cost and Markup Estimates: Hybrid Vehicles (2000-2008)

	Random Coefficients $\mu_{i,j} \neq 0$ (1)			Random Coefficients $\mu_{i,j} \neq 0$ (2)		
	MC	MrkUp	% markup	MC	MrkUp	% markup
Average	$ 21,270	$ 6,326	22	$ 20,949	$ 6,647	24
5th Percentile	$ 16,693	$ 3,804	18	$ 14,633	$ 4,755	18
Q1	$ 17,531	$ 4,219	19	$ 15,947	$ 5,618	22
Q2	$ 18,807	$ 5,936	23	$ 18,064	$ 6,112	25
Q3	$ 24,067	$ 7,363	25	$ 25,490	$ 6,606	27
95th Percentile	$ 33,934	$ 12,191	28	$ 34,369	$ 12,070	31
Uses NHTS Micro Data	Yes			No		

[a] $\mu_{i,j} \neq 0$ indicates that random coefficients were included in the model.

Table 1.13: Implied Marginal Cost and Markup Estimates: 2007 Model Year Hybrid Vehicles

			Random Coefficient $\mu_{i,j} \neq 0$ (1)			Random Coefficient $\mu_{i,j} \neq 0$ (2)		
Make	Model	Price	MC	MrkUp	% markup	MC	MrkUp	% markup
Ford	Escape Hybrid	$ 26,728	$ 20,342	$ 6,386	24	$ 19,802	$ 6,926	26
Mercury	Mariner Hybrid	$ 27,770	$ 21,021	$ 6,749	24	$ 20,996	$ 6,774	24
Saturn	Vue Hybrid	$ 23,233	$ 17,517	$ 5,716	25	$ 16,612	$ 6,621	28
Honda	Accord Hybrid	$ 32,258	$ 24,544	$ 7,714	24	$ 26,137	$ 6,121	19
Honda	Civic Hybrid	$ 21,990	$ 17,495	$ 4,496	20	$ 16,271	$ 5,719	26
Nissan	Altima Hybrid	$ 24,337	$ 19,154	$ 5,183	21	$ 19,665	$ 4,672	19
Lexus	RX 400 hybrid	$ 42,687	$ 31,252	$ 11,435	27	$ 32,289	$ 10,399	24
Toyota	Camry Hybrid	$ 26,527	$ 20,292	$ 6,235	24	$ 21,803	$ 4,724	18
Toyota	Highlander Hybrid	$ 33,398	$ 24,658	$ 8,740	26	$ 27,330	$ 6,068	18
Toyota	Prius	$ 21,986	$ 17,239	$ 4,747	22	$ 15,853	$ 6,133	28
Uses NHTS Micro Data				Yes			No	

[a] $\mu_{i,j} \neq 0$ indicates that random coefficients were included in the model.

Table 1.14: Change In Manufacturer Variable Profits: (Counterfactual - Original)

Manufacturer	Year										Total
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	
Chrysler	$0	$292	$536	$886	$960	$1,899	$8,295	$9,957	$13,070	$9,171	$45,068
Ford	$0	$745	$1,454	$2,326	$2,446	$2,159	$4,294	$740	$2,713	$2,790	$19,667
GM	$0	$1,068	$2,041	$3,532	$4,205	$6,991	$25,173	$27,609	$30,948	$20,634	$122,201
Daewoo	$0	$13	$21	$28	$0	$0	$0	$0	$0	$0	$61
Honda	$0	-$1,231	-$1,127	-$5,734	-$6,882	-$7,233	-$14,802	-$8,240	-$3,466	-$305	-$49,019
Hyundai	$0	$48	$144	$258	$359	$675	$1,692	$2,173	$3,080	$2,754	$11,184
Kia	$0	$18	$68	$115	$157	$359	$858	$1,076	$1,509	$1,420	$5,580
Mazda	$0	$56	$113	$190	$237	$487	$1,037	$1,251	$2,179	$2,113	$7,662
Mitsubishi	$0	$72	$160	$305	$235	$272	$589	$636	$1,018	$834	$4,122
Nissan	$0	$150	$292	$666	$880	$1,626	$5,668	$6,934	$6,279	$4,954	$27,448
Subaru	$0	$41	$98	$175	$198	$353	$1,062	$1,337	$1,732	$1,624	$6,621
Suzuki	$0	$4	$8	$20	$23	$94	$280	$398	$565	$502	$1,895
Toyota	$0	-$2,073	-$5,682	-$6,571	-$7,759	-$17,377	-$72,330	-$97,287	-$135,596	-$125,123	-$469,798
BMW	$0	$87	$187	$426	$539	$857	$3,234	$6,012	$8,397	$7,512	$27,251
Mercedes	$0	$87	$182	$365	$459	$751	$2,611	$4,183	$6,708	$6,603	$21,950
Porsche	$0	$12	$25	$47	$40	$58	$430	$728	$937	$791	$3,068
Volkswagon	$0	$157	$328	$573	$627	$898	$2,166	$2,787	$5,603	$4,725	$17,864
Isuzu	$0	$2	$3	$3	$2	$7	$54	$40	$28	$4	$143
Land Rover	$0	$1	$2	$4	$4	$18	$797	$979	$933	$665	$3,404
Total	$0	-$450	-$1,149	-$2,386	-$3,269	-$7,103	-$28,891	-$38,686	-$53,362	-$58,330	-$193,627

[a] Profits are in $10,000s.
[b] Counterfactual equilibrium price and quantities are solved with all hybrid vehicles removed from the market.

Table 1.15: Counterfactual Price and Quantity Comparison: (2007 Market - Sorted by Changes in Quantity)

Make	Model	Price	CF Price[b]	Δ	% Δ	Quantity	CF Quantity[b]	Δ	%Δ
Toyota	Camry	$ 19,578	$ 19,513	-$64	-0.33%	418,631	437,282	18,651	4.5%
Toyota	Matrix	$ 16,452	$ 16,407	-$45	-0.27%	348,016	360,192	12,176	3.5%
Honda	Accord	$ 21,883	$ 21,904	$20	0.09%	388,826	398,182	9,356	2.4%
Nissan	Altima	$ 19,282	$ 19,308	$27	0.14%	276,374	281,334	4,960	1.8%
Chevrolet	Impala	$ 22,346	$ 22,403	$58	0.26%	311,128	316,055	4,927	1.6%
Honda	Civic	$ 14,961	$ 14,971	$10	0.07%	298,520	303,210	4,690	1.6%
BMW	3 Series	$ 37,385	$ 37,391	$6	0.02%	142,490	146,752	4,262	3.0%
Lexus	ES 350	$ 35,193	$ 35,051	-$142	-0.40%	82,867	86,938	4,071	4.9%
Toyota	Avalon	$ 28,515	$ 28,384	-$131	-0.46%	72,945	76,860	3,915	5.4%
Chrysler	300M	$ 25,425	$ 25,451	$26	0.10%	120,636	123,783	3,147	2.6%
Lexus	IS 250	$ 32,166	$ 32,024	-$142	-0.44%	54,933	57,773	2,840	5.2%
Dodge	Charger	$ 23,343	$ 23,379	$36	0.16%	119,289	121,942	2,653	2.2%
Hyundai	Sonata	$ 18,690	$ 18,714	$24	0.13%	145,568	148,138	2,570	1.8%
Scion	tC	$ 17,594	$ 17,542	-$52	-0.29%	63,852	66,329	2,477	3.9%
Ford	Mustang	$ 20,767	$ 20,807	$40	0.20%	134,626	136,920	2,294	1.7%
Ford	Fusion	$ 18,690	$ 18,722	$32	0.17%	149,552	151,824	2,272	1.5%
Infiniti	G35	$ 35,468	$ 35,466	-$2	-0.01%	71,811	74,054	2,243	3.1%
Chevrolet	Cobalt	$ 14,193	$ 14,216	$23	0.16%	200,620	202,740	2,120	1.1%
Ford	Focus	$ 14,618	$ 14,639	$21	0.14%	173,213	175,280	2,067	1.2%
Pontiac	G6	$ 18,394	$ 18,431	$38	0.20%	150,001	151,965	1,964	1.3%

[a] All hybrid vehicles are removed from the counterfactual market.
[b] CF: Counterfactual market equilibrium.
[c] A total of three starting values were used, all of which converged to the same counterfactual price: 1) original equilibrium prices 2) original equilibrium prices +/- a $10 random deviation and 3) original equilibrium prices +/- a $20 deviation.

Table 1.16: Toyota's Lost Variable Profits: Prius Removed from Market

Year	Lost Variable Profits	Change
2000	\$220	0.23%
2001	\$584	0.54%
2002	\$72,5	0.67%
2003	\$856	0.83%
2004	\$1,855	1.49%
2005	\$3,885	2.80%
2006	\$3,899	2.63%
2007	\$7,418	4.56%
2008	\$6,797	4.80%
Total	\$26,243	2.06%

[a] Profits in \$10,000s.
[b] Counterfactual equilibrium price and quantities solved with the Prius removed from the market.

Table 1.17: Toyota's Profit Dissipation With Entry of Hybrid Competitors

Year	Lost Variable Profits	Change
1999	\$0	0.00%
2000	\$13	0.03%
2001	\$16	0.03%
2002	\$69	0.14%
2003	\$81	0.18%
2004	\$121	0.22%
2005	\$362	0.56%
2006	\$401	0.58%
2007	\$497	0.69%
2008	\$443	0.69%
Total	\$2,008	0.39%

[a] Profits in \$10,000s.
[b] Counterfactual equilibrium price and quantities solved with all hybrids, except for Toyota's, removed from the market.

Table 1.18: Top 20 Hybrid Cross-Price Elasticities: Light Trucks (2007 Market)

Ford Escape Hybrid		Mercury Mariner Hybrid		Saturn Vue Hybrid		Lexus RX400 Hybrid		Toyota Highlander Hybrid	
	elast.		elast.		elast.		elast.		elast.
Honda CR-V	0.114	Honda CR-V	0.114	Honda CR-V	0.128	Chevrolet Tahoe	0.108	Chevrolet Tahoe	0.080
Toyota RAV4	0.090	Toyota RAV4	0.090	Toyota RAV4	0.099	Lexus RX 350	0.070	Honda CR-V	0.073
Ford Escape	0.065	Chevrolet Trailblazer	0.065	Ford Escape	0.081	Chevrolet Suburban	0.067	Toyota RAV4	0.072
Ford Explorer XLT	0.064	Ford Explorer XLT	0.064	Jeep Wrangler	0.068	Jeep Grand Cherokee	0.067	Ford Explorer XLT	0.063
Chevrolet Trailblazer	0.061	Ford Edge	0.061	Chevrolet Trailblazer	0.063	Ford Explorer XLT	0.063	Jeep Grand Cherokee	0.060
Ford Edge	0.060	Ford Escape	0.060	Ford Explorer XLT	0.062	Honda Pilot	0.062	Chevrolet Trailblazer	0.060
Chevrolet HHR	0.055	Jeep Grand Cherokee	0.055	Ford Edge	0.059	Honda CR-V	0.060	Ford Edge	0.059
Jeep Grand Cherokee	0.054	Chevrolet HHR	0.054	Chevrolet HHR	0.058	Ford Edge	0.060	Honda Pilot	0.057
Jeep Wrangler	0.054	Chevrolet Tahoe	0.054	Hyundai Santa Fe	0.053	Toyota RAV4	0.059	Lexus RX 350	0.048
Honda Pilot	0.053	Honda Pilot	0.053	Jeep Liberty	0.050	Chevrolet Trailblazer	0.059	Toyota Highlander	0.047
Toyota Highlander	0.051	Toyota Highlander	0.051	Chevrolet Equinox	0.048	Ford Expedition	0.055	Chevrolet Suburban	0.047
Hyundai Santa Fe	0.048	Hyundai Santa Fe	0.048	Chrysler PT Cruiser	0.047	Toyota 4Runner	0.053	Ford Expedition	0.047
Chevrolet Tahoe	0.047	Jeep Wrangler	0.047	Jeep Grand Cherokee	0.046	Toyota Highlander	0.053	Ford Escape	0.044
Chevrolet Equinox	0.046	Chevrolet Equinox	0.046	Saturn Vue	0.045	Acura MDX	0.050	Toyota 4Runner	0.044
Saturn Vue	0.042	Saturn Vue	0.042	Toyota Highlander	0.044	GMC Yukon	0.048	Nissan Murano	0.038
Jeep Liberty	0.041	Jeep Liberty	0.041	Dodge Nitro	0.042	GMC Acadia	0.044	GMC Acadia	0.038
Toyota 4Runner	0.040	Ford Expedition	0.040	Honda Pilot	0.041	Nissan Murano	0.042	GMC Yukon	0.035
Ford Expedition	0.039	Toyota 4Runner	0.039	Chevrolet Tahoe	0.039	Cadillac Escalade	0.037	Chevrolet Equinox	0.034
Dodge Nitro	0.034	Nissan Murano	0.034	Kia Sportage	0.035	GMC Yukon XL	0.037	Acura MDX	0.033
Chrysler PT Cruiser	0.031	Dodge Nitro	0.031	Ford Expedition	0.031	Jeep Commander	0.035	Jeep Liberty	0.033

[a] Interpretation: ex) A 1% increase in the price of the Honda Accord Hybrid causes market shares for the Toyota Camry to increase by 0.142%.

Table 1.19: Top 20 Hybrid Cross-Price Elasticities: Cars (2007 Market)

Honda Accord Hybrid	elast.	Honda Civic Hybrid	elast.	Nissan Altima Hybrid	elast.	Toyota Camry Hybrid	elast.	Toyota Prius	elast.
Toyota Camry	0.142	Toyota Camry	0.203	Toyota Camry	0.187	Honda Accord	0.159	Toyota Camry	0.179
Honda Accord	0.141	Toyota Matrix	0.173	Honda Accord	0.166	Toyota Camry	0.141	Honda Accord	0.171
Chevrolet Impala	0.118	Honda Accord	0.171	Chevrolet Impala	0.135	Chevrolet Impala	0.131	Chevrolet Impala	0.137
Nissan Altima	0.091	Chevrolet Impala	0.137	Toyota Matrix	0.133	Toyota Matrix	0.110	Toyota Matrix	0.124
BMW3 Series	0.086	Nissan Altima	0.135	Nissan Altima	0.100	Nissan Altima	0.090	Nissan Altima	0.117
Toyota Prius	0.079	Honda Civic	0.094	Toyota Prius	0.084	Toyota Prius	0.085	Honda Civic	0.094
Toyota Matrix	0.077	Toyota Prius	0.080	Honda Civic	0.069	Chrysler 300M	0.057	Ford Fusion	0.062
Chrysler 300M	0.058	Hyundai Sonata	0.072	Ford Mustang	0.054	Honda Civic	0.056	Pontiac G6	0.061
Honda Civic	0.053	Ford Fusion	0.062	Chrysler 300M	0.054	BMW 3 Series	0.053	Hyundai Sonata	0.060
Lexus ES 350	0.051	Pontiac G6	0.061	Dodge Charger	0.053	Dodge Charger	0.053	Chevrolet Cobalt	0.059
Dodge Charger	0.050	Chevrolet Cobalt	0.059	Ford Fusion	0.052	Ford Mustang	0.051	Ford Mustang	0.059
Hyundai Sonata	0.044	Ford Mustang	0.059	Pontiac G6	0.050	Ford Fusion	0.046	Ford Focus	0.053
Infiniti G35	0.044	Mazda Mazda3	0.058	Hyundai Sonata	0.050	Hyundai Sonata	0.044	Chevrolet Malibu	0.052
Ford Mustang	0.043	Ford Focus	0.053	Chevrolet Malibu	0.043	Pontiac G6	0.044	Dodge Charger	0.051
Toyota Avalon	0.043	Nissan Sentra	0.053	Chevrolet Cobalt	0.042	Pontiac Grand Prix	0.038	Volkswagen Jetta	0.049
Mercedes C Class	0.039	Chevrolet Malibu	0.052	Volkswagen Jetta	0.041	Chevrolet Malibu	0.038	Chrysler 300M	0.049
Acura TL	0.037	Dodge Charger	0.051	BMW 3 Series	0.041	Buick LuCerne	0.037	Mazda Mazda3	0.038
Pontiac Grand Prix	0.036	Volkswagen Jetta	0.049	Nissan Sentra	0.039	Lexus ES 350	0.036	Pontiac Grand Prix	0.038
Buick LuCerne	0.036	Chrysler 300M	0.049	Pontiac Grand Prix	0.039	Volkswagen Jetta	0.035	Toyota Yaris	0.037
Ford Fusion	0.035	Pontiac Grand Prix	0.038	Ford Focus	0.038	Toyota Avalon	0.035	Nissan Sentra	0.037

[a] Interpretation: ex) A 1% increase in the price of the Honda Accord Hybrid causes market shares for the Toyota Camry to increase by 0.142%.

Table 1.20: Compensating Variation: Full Effect

Conditional on Hybrid Vehicle Purchase									
	2000	2001	2002	2003	2004	2005	2006	2007	2008
Average	$6,908.4	$7,214.3	$7,106.7	$7,553.7	$7,327.6	$9,507.7	$9,480.1	$9,042.3	$10,144.8
Min	$13.6	$1.8	$5.0	$4.6	$0.8	-$22.9	-$25.9	-$69.2	-$10.0
Q1	$2,247.5	$1,303.7	$1,477.5	$1,528.5	$1,567.4	$1,798.9	$1,938.5	$1,890.2	$2,139.7
Q2	$4,280.3	$3,797.9	$3,854.6	$4,274.1	$3,944.9	$5,049.4	$5,213.4	$5,075.6	$5,475.7
Q3	$7,388.3	$9,831.1	$8,677.1	$9,736.2	$9,381.7	$12,052.1	$12,053.5	$11,734.6	$12,590.1
Max	$51,299.5	$47,545.3	$77,102.8	$59,794.5	$106,296.4	$123,604.4	$129,034.1	$153,545.6	$132,571.6

Conditional on Any New Vehicle Purchase									
	2000	2001	2002	2003	2004	2005	2006	2007	2008
Average	$4.4	$9.5	$13.5	$20.3	$34.3	$126.0	$158.0	$204.0	$256.1
Min	-$3.3	-$8.3	-$11.5	-$12.7	-$25.3	-$89.6	-$102.3	-$142.1	-$135.0
Q1	-$0.1	-$0.1	-$0.2	-$0.4	-$1.8	$0.0	$0.5	$0.1	$0.1
Q2	$0.0	$0.1	$0.1	$0.0	$0.0	$6.2	$8.4	$13.4	$19.4
Q3	$0.5	$0.8	$1.5	$1.2	$1.2	$15.3	$19.0	$26.8	$32.4
Max	$51,299.5	$47,545.3	$77,102.8	$59,794.5	$106,296.4	$123,604.4	$129,034.1	$153,545.6	$132,571.6

[a] Estimates are based on one million simulations per year.

Table 1.21: Compensating Variation: Variety Effect

Conditional on Hybrid Vehicle Purchase									
	2000	2001	2002	2003	2004	2005	2006	2007	2008
Average	\$6,908.2	\$7,214.3	\$7,106.5	\$7,568.2	\$7,328.3	\$9,492.1	\$9,470.0	\$9,035.0	\$10,130.8
Min	\$12.6	\$1.8	\$6.2	\$1.4	\$5.5	\$2.8	\$0.5	\$0.4	\$4.0
Q1	\$2,247.0	\$1,303.7	\$1,482.3	\$1,528.5	\$1,568.6	\$1,780.4	\$1,928.2	\$1,890.2	\$2,144.0
Q2	\$4,279.6	\$3,798.4	\$3,854.6	\$4,283.5	\$3,951.2	\$5,029.9	\$5,210.1	\$5,052.4	\$5,472.4
Q3	\$7,388.3	\$9,831.1	\$8,679.4	\$9,774.0	\$9,382.4	\$12,052.1	\$12,046.7	\$11,712.2	\$12,590.1
Max	\$51,299.5	\$47,547.4	\$77,102.2	\$59,794.5	\$106,320.5	\$123,578.5	\$128,964.1	\$153,530.8	\$132,571.6

[a] Estimates are based on one million simulations per year.

Table 1.22: Compensating Variation: Price Effect

Conditional on Any New Vehicle Purchase									
	2000	2001	2002	2003	2004	2005	2006	2007	2008
Average	$0.1	$0.0	$0.0	-$0.5	-$1.4	$6.6	$7.8	$9.8	$13.0
Min	-$3.3	-$8.3	-$11.5	-$12.7	-$25.3	-$89.6	-$102.3	-$142.1	-$135.0
Q1	-$0.1	-$0.1	-$0.2	-$0.4	-$1.8	-$0.3	$0.0	$0.1	$0.1
Q2	$0.0	$0.1	$0.1	$0.0	$0.0	$5.7	$8.3	$12.6	$18.7
Q3	$0.5	$0.8	$1.5	$1.0	$1.2	$13.9	$18.9	$26.7	$31.2
Max	$8.9	$5.3	$9.1	$8.9	$25.9	$224.4	$213.6	$345.7	$309.0

[a] Estimates are based on one million simulations per year.

Table 1.23: The Total Change in U.S. Welfare from the Hybrid Innovation: (2000-2008)

Year	Compensating Variation for all Hybrid Owners	Compensating Variation for all Households	Change in Producer Profits	Total Welfare Change
2000	$645	$761	$45	$806
2001	$1,463	$1,654	$114	$1,769
2002	$2,560	$2,271	$238	$2,510
2003	$3,506	$3,372	$326	$3,699
2004	$6,049	$5,805	$710	$6,515
2005	$19,780	$21,502	$2,889	$24,392
2006	$23,712	$26,146	$3,868	$30,014
2007	$31,550	$32,983	$5,336	$38,320
2008	$31,051	$33,624	$5,833	$39,457
Total	$120,321	$128,124	$19,362	$147,487

[a] Profits in $10,000s.
[b] Estimates are based on one million simulations per year.

Chapter 2

Financial Loss, Welfare Gain: A Story of Hybrid Vehicle Consumption

2.1 Introduction

Since hybrid vehicles were introduced into the U.S. auto market in 2000, consumers have been asking the question of whether buying one makes financial sense. Hybrids cost several thousand dollars more than a comparable non-hybrid vehicle but promise increased fuel savings which may or may not offset the initial cost after several years of driving. Several papers, including this one, show that the majority of hybrid owners will not recover this premium. Despite this fact, hybrid sales have continued to see growth throughout this past decade. This finding suggests consumers value hybrids for reasons other than just the gain in fuel savings.

Consumer reports (2006) found that the Honda Civic Hybrid and Toyota Prius will cost \$3,700 and \$5,250 more than their all-gas equivalent after five years of ownership. For the Honda Accord Hybrid, Lexus RX400h, and Toyota Highlander Hybrid the costs are even higher, ranging from \$10,250 to \$13,300.[1] The Wall Street Journal (2007) found similar results by examining the length of time required to recoup the premium of owning a hybrid vehicle. They find that the Toyota Prius, when compared to the Toyota Corolla, will take consumers 17.9 years on average. Even with a more favorable comparison to the Toyota Camry, they find

[1] These results assume the household lives in California (the number one hybrid vehicle market), drives 15,000 miles per year, \$4.00 per gallon gasoline and incorporated vehicle depreciation costs, fees and taxes, maintenance and repair costs, and federal tax incentives.

that the Prius will take 6.5 years.

However, evidence suggests that consumers do not factor ownership costs correctly, nor do they seem to care that much about them. Turrentine and Kurani (2007) conducted a survey across nine lifestyle sectors and found that no household analyzed their fuel costs in a systematic way, almost none of the households kept track of their fuel expenditures over time, and most households could not recall how much they spent on fuel just a few weeks prior. Their survey also contained several households that actually owned a hybrid vehicle. For these households, not only did they not track their fuel expenditures over time, they also said that saving money on gasoline was not the primary reason they purchased a hybrid. Rather, it was because they wanted to “set an example," “be a pioneer," “live lighter," or “be able to talk to other people about their car." The Wall Street Journal (2007) reported similar findings citing that “the majority of people just want to feel better about themselves and that they are trying to make some effort to be more green, be more socially responsible, and reduce their carbon footprint."

Conspicuous environmentalism is a growing phenomenon that has gained attention in the literature. Consumers not only want to be green, they want to look green as well. Kahn (2007) showed that consumers living in “green" neighborhoods based on the number of Green Party voter registrations are more likely to drive a Toyota Prius over other equally fuel efficient vehicles because of its distinguished green look. He coins this the Prius effect. Sexton and Sexton (2011) found similar results by comparing vehicle registration data for the Toyota Prius and Honda Civic Hybrid in Colorado and Washington state. They find that consumers living in communities with a high percentage of Democrats are willing to pay up to $7,000 more for a Prius than a Civic Hybrid because of the stronger signal it sends to the community that they are an environmentally conscious consumer.

Several surveys have examined the types of people that purchase hybrid vehicles. J.D. Power and Associates (2008) find that females, college educated, wealthy, and older consumers are more likely to purchase hybrid vehicles than the average new car buyer. Scarborough Research (2007) conducted a similar survey and found that hybrid owners are more likely to be Democrats, wealthy, educated and are more likely to engage in outdoor activities such as biking, hiking, gardening, snow skiing, and eat organic foods.

The purpose of this chapter is to develop a rich consumer demand model to explain why, despite the higher ownership costs, consumers continue to purchase hybrid vehicles. Economic theory would suggest that households will only purchase a hybrid vehicle if they are willing to pay more than it costs to own. This chapter determines what these reasons are and, more importantly, estimates how much consumers are willing to pay for hybrid vehicles conditional on their preference type.

This chapter complements the growing literature of hybrid vehicle demand by using detailed household level data at the U.S. national level that explicitly links the year, make, and model of each vehicle a household owns to their actual demographic type. Previous papers have used vehicle registration data at the aggregate level, however, they lacked information on specific household demographics (Kahn (2007), Diamond (2009), Beresteanu and Li (2010), Chandra et al. (2010), Gallagher and Muehlegger (2011), and Sexton and Sexton (2011)). Therefore they had to rely on aggregate demographic data to form correlations, rather than a direct link, between vehicle choice and heterogeneous preferences. Other papers have used detailed household level data directly linking vehicle choice to a household's specific demographic type but were only able to cover small geographic areas, had relatively small sample sizes, and were based on stated preference rather than revealed preference surveys (Sangkapichai and Saphores (2009), Erdem et al. (2010), Potoglou and Kanaroglou (2007)). This chapter overcomes these previous data limitations and, as a result, is able to identify heterogeneous preferences for hybrid vehicles with a high degree of statistical precision.

Furthermore, this is the first research to directly link a household's willingness to pay for hybrid vehicles to the true cost of ownership by factoring in vehicle depreciation, insurance premiums, interest on financing, fees and taxes (including tax incentives), maintenance and repair costs, and fuel expenditures. As a result, I am able to show that, although hybrid vehicles may not make financial sense, for some households they are a better match to their preference type.

This research also complements the conspicuous environmental literature of Kahn (2007) and Sexton and Sexton (2011). Kahn (2007) was the first to identify the conspicuous environmental effect from owning a Prius over other almost equally green vehicles by examining the relationship between vehicle registration data and the percentage of Green Party voter registrations at the Census Tract level in Los Angeles county. However, Kahn never quantified how much households were willing to pay to look conspicuously green from owning a Prius, nor did he examine the cost of ownership differential between hybrid vehicles and their closest all gas equivalent. Sexton and Sexton (2011) attempted to quantify how much households are willing to pay to look conspicuous green from owning a Prius by examining the relationship between vehicle registration data and the percentage of Democrats at the

ZIP code level in Colorado and Washington state. However, their willingness-to-pay (WTP) estimates for the conspicuous Prius effect are calculated using the price elasticity of demand for the Toyota Corolla, rather than for the Prius. Furthermore, the Corolla price elasticity is obtained from the literature, rather than their own demand model, with BLP (1995) being the most recent paper used.[2] This research overcomes the short-comings of Sexton and Sexton (2011) and builds upon the work of Kahn (2007) by estimating how much more households living in green communities, based on the percentage of Democrats, are willing to pay for the Toyota Prius than any other hybrid vehicle.

The results of this chapter suggest that an average household living in a county that voted 75% for the Democratic Party would be willing to pay $11,064 more for a Toyota Prius than any other hybrid vehicle. On the other end of the political spectrum, the same household living in a county that only voted 25% for the Democrat Party would only be willing to pay $4,618 more. To help the reader understand this welfare measure, consider the following thought experiment. In the first scenario, suppose you own a Prius that gets a combined fuel economy of 46 miles-per-gallon. Also, because of the Prius's conspicuous green look, all of your neighbors, co-workers, and other members of the community recognize you as "green" and socially responsible. Other hybrid vehicles simply have an inconspicuous decal on the back bumper making them virtually indistinguishable from their all gas equivalent. In the second scenario, suppose you also own a Prius and that all the physical attributes (horsepower, fuel economy, size, style, comfort, performance, etc.) remain the same. The only thing that is different, is that now the Prius is no longer conspicuously green. In other words, your community no longer recognizes you as "green" and socially responsible, which effectively makes your Prius no different than any other hybrid vehicle. As a result, you suffer a welfare loss from lack of recognition. Therefore, this welfare measurement should be interpreted as the the dollar amount that an average household would be willing to pay to be recognized by their community as "green" and socially responsible (i.e. as a conspicuous environmentalist).

Finally, this research complements the applied welfare literature by developing a simulation procedure to infer where a consumer's unobservable preferences lie relative to that of the population as a whole. My method is similar to Revelt and Train (2000), but allows additional demographic data to be incorporated into the conditional distributions of unobservable preferences. This additional information can be used to construct tighter bounds on WTP

[2]BLP (1995) examined the U.S. auto market between the years 1971 and 1990.

estimates by reducing the variance in unobservable factors.

The rest of this chapter is as follows: Section Two provides a brief literature review, Section Three discusses the data, Section Four describes the methodology, Section Five describes the model specification, results are presented in Section Six, finally Section Seven concludes the chapter.

2.2 Literature Review

Several papers have examined the relationship between rising gasoline prices and government incentives to the recent growth in hybrid vehicle sales. Beresteanu and Li (2010) use a structural discrete choice model and new vehicle sales data at the MSA level between 1999 and 2006. They find that recent increases in gasoline prices and federal tax incentives all had a significant impact on the growth in hybrid vehicle sales. Specifically, they show that had gasoline prices stayed at 1999 levels ($1.53) rather than 2006 levels ($2.60) that hybrid sales would have been about 37% less. Federal income tax deductions between 2001 and 2005 were found to be attributed to 5% of the growth in hybrid vehicles sales while federal income tax credits in 2006 were attributed to 20% of the growth.

Diamond (2009) examine quarterly state-level hybrid vehicle registration data between 2001 and 2008. He finds that high gasoline prices were the primary driver of the recent growth in sales while government incentives, though significant, had a much weaker impact. Gallagher and Muehlegger (2011) conducted a similar study examining quarterly state-level sales data for eleven hybrid models between 2000 and 2006. After controlling for gasoline prices, non-monetary government incentives, and socioeconomic variables, they find that state sales tax waivers are ten times more effective than income tax credits in promoting the adoption of hybrid vehicles. They also find that rising gasoline prices and access to HOV lanes are significant contributing factors. Chandra et al. (2010) use a discrete choice model to analyze the effectiveness of hybrid vehicle tax rebates across several Canadian Provinces. They find that 26% of hybrid vehicles sold during the time of the rebate programs can be attributed to the rebates. They also conclude that hybrid vehicles, in terms of reducing carbon dioxide, are “strikingly more expensive than other means of reducing carbon emissions."

Several papers have used surveys to examine households’ stated preferences for hybrid vehicles. Erdem et al. (2010) conducted a phone survey in Turkey and used an ordered probit model to show that income, gender, education, and concerns for global warming all had a significant impact.[3] Sangkapichai and Saphores (2009) conducted a 2004 statewide phone survey in California asking respondents questions regarding their preferences towards various State and National environmental policies as well as their socioeconomic background. They were then asked if they were 1) interested in hybrids 2) indifferent or 3) not interested. Using an ordered logit model they found that education, age, rising gasoline prices, environmental concerns, and access to HOV lanes were all significant factors affecting the likelihood of purchasing a hybrid vehicle.

2.3 Data

Several data sets are used in this chapter. The first is the 2009 National Household Travel Survey (NHTS) conducted by the U.S. Department of Transportation Federal Highway Administration. The survey was conducted between March 2008 and May 2009, containing 150,147 household level observations. This is a rich data source providing information on the year, make, and model of each vehicle a household owns, in additional to many demographical variables including household income, race, education, household size, children, and age. The survey also indicated whether the vehicle is a hybrid or not.[4] A confidential version of this data was obtained providing the Census Tract of each household's location which was then used to link to additional socioeconomic variables.

Vehicle attributes are obtained from Ward's Automotive Yearbook and are linked to the NHTS. These attributes include the manufacturer's suggested retail price (MSRP), horsepower, weight, vehicle dimensions, miles-per-gallon (MPG), and classifies vehicles into one of 26 segment classes (e.g. small car, luxury suv, van, etc.). Since the NHTS did not provide information on the vehicle's trim level, attributes for the base model are used instead. EPA adjusted fuel economy estimates, obtained from fueleconomy.gov, are also used which are intended to represent the actual fuel economy that a typical driver can expect under regular driving conditions.

Gasoline prices and a regional cost of living index (COLI) are obtained from the American Chamber of Commerce's ACCRA data base for 322 Core Based Statistical Areas (CBSAs).

[3]Due to the small market share of hybrid vehicles in Turkey, the authors acknowledged that consumers lacked basic information about what a hybrid vehicle actually was and were therefore offered a brief explanation.

[4]The survey indicated that there were 7,145 "hybrid or alternative use vehicles." However, after inspecting the makes and models of these vehicles, it was concluded that only 3,356 were actually hybrids. The remaining models were assumed to be "flex fuel."

Gasoline prices are for one gallon of regular unleaded and includes all taxes (federal, state, and local). The COLI is used to scale household income to control for purchasing power differences across regions. For households not located in one of the 322 CBSAs, the state average gasoline price and COLI is used instead.

Information on state High Occupancy Vehicle (HOV) lane laws are obtained from various state Department of Motor Vehicle (DMV) websites as well as the U.S. Department of Transportation.[5] To estimate the relationship between hybrid vehicle demand and HOV lane access, only households that live in both 1) a state that exempts hybrid vehicles from the HOV lane minimum occupancy requirement and 2) a MSA with HOV lanes, are included in the hybrid vehicle interaction term described below. This prevents a household that lives, for example, 100 miles from an HOV lane to be influenced by a state law exempting hybrid vehicles from the minimum occupancy requirement.

Cost-of-ownership estimates for hybrid vehicles and their all gas equivalent are obtained from driverside.com. These estimates are for the national average and assumes a household drives 15,000 miles per year. These estimates include the vehicle's expected depreciation, insurance rate premiums, maintenance and repair costs, fees and taxes (including tax incentives), and interest on financing. Insurance costs are for a middle-aged driver with a clean driving record. Financing assumes a 60-month loan with a 6.46% interest rate and a 15% down payment. Even if the household does not finance and purchases their vehicle outright, the financing cost should be close to the opportunity cost from lost investment.

The percentage of Democrats per county is obtained from USA Today. I use the 2008 presidential results and link the percentage of votes per county that went to Obama (Democrat) and McCain (Republican) to each household. This variable is used to proxy for how green a community is. Counties with a higher percentage of Democrats are expected to be more pro-environment than counties with a low percentage of Democrats.

The final two data sources are used to control for unobservable product characteristics. The first is the Consumer Report's Road Test Score which rates vehicles on a scale of 1 to 100, with 100 being the best. The Score is based on things such as performance, handling, and reliability. The second is the Kelley Blue Book: Residual Guide which reports the expected retained value for each vehicle 24 months, 36 months, 48 months, and 60 months into the future.

[5] See Appendix IV "Supplemental Data: Chapter 2" for more details on each state's HOV lane laws.

2.4 Method

This chapter follows a similar approach taken by Train and Winston (2007) and develops a discrete choice model conditional on buying a new car. This is different than an unconditional choice model in which consumers have the choice of purchasing a new car, used car, or no car at all. As a result, there is no outside good. This limits me to only being able to infer preferences for “new car purchasers" rather than the population as a whole.

A random utility function is specified that characterizes consumers’ choice of new vehicles by make and model as a function of vehicle attributes, demographics, and unobservable heterogeneous tastes. Specifically, let new car purchasers be indexed by $i = 1, ..., N$ and the available new vehicles by $j = 1, ..., J$. The utility, $U_{i,j}$, that consumer i derives from vehicle j is given by

$$U_{i,j} = \alpha^* h(y_i, p_j) + \beta' x_j + \gamma' z_{i,j} + \mu_i' x_j + \epsilon_{i,j}.$$

The first term, $\alpha^* h(y_i, p_j)$, captures the price and income effects, where the function $h(y_i, p_j)$ specifies how income, y_i, and price, p_j, are related. The second term, $\beta' x_j$, captures the average utility component where x_j is a vector of vehicle attributes and β is a vector of average marginal utility coefficients. The third component, $\gamma' z_{i,j}$, captures observed heterogeneity where $z_{i,j}$ is a vector of consumer i's demographics interacted with vehicle characteristics and γ is a vector of average demographical taste shifters. The fourth component, $\mu_i' x_j$, captures unobservable heterogeneity where μ_i is vector of random unobservable tastes with a variance of σ and a mean of zero. The last term, $\epsilon_{i,j}$, represents consumer i's idiosyncratic tastes for vehicle j, which captures all remaining elements of utility. Consumers are assumed to purchase the vehicle that yields the highest utility subject to their financial constraint.

Assuming $\epsilon_{i,j}$ is i.i.d. extreme value, the probability that consumer i chooses vehicle j conditional on knowing $\{\alpha^*, \beta, \gamma, \mu_i\}$ is represented by the logit probability function,

$$L_{i,j}(\alpha^*, \beta, \gamma, \mu_i) = \frac{e^{\alpha^* h(y_i, p_j) + \beta' x_j + \gamma' z_{i,j} + \mu_i' x_j}}{\sum_{r=1}^{J} e^{\alpha^* h(y_i, p_r) + \beta' x_r + \gamma' z_{i,r} + \mu_i' x_r}}.$$

However, since μ_i is only observable to the consumer and not the researcher, I am only able to infer the density of the random variable, $f(\mu|\sigma)$, which has a variance σ centered around zero. The goal of the researcher is to estimate the parameters $\{\alpha^*, \beta, \gamma, \sigma\}$ that fully characterize the distribution of consumers’ preferences. By integrating over all values of μ, the unconditional probability that consumer i chooses vehicle j is represented by the mixed logit probability function,

$$P_{i,j}(\alpha^*,\beta,\gamma,\sigma) = \int L_{i,j}(\alpha^*,\beta,\gamma,\mu) f(\mu|\sigma) d\mu.$$

The integral above is estimated via simulation since no closed form solution exists using draws of μ^r taken from the distribution $f(\mu|\sigma)$. The Logit probability function for each μ^r is calculated and the results are averaged over resulting in the simulated probability function,

$$\tilde{P}_{i,j}(\alpha^*,\beta,\gamma,\sigma) = \frac{1}{R}\sum_{r=1}^{R} L_{i,j}(\alpha^*,\beta,\gamma,\mu^r),$$

where R is the number of simulations. The simulated probability function is inserted into the log-likelihood function resulting in the simulated log-likelihood function,

$$SLL(\alpha^*,\beta,\gamma,\sigma) = \sum_{i=1}^{N}\sum_{j=1}^{J} d_{i,j} \ln \tilde{P}_{i,j}(\alpha^*,\beta,\gamma,\sigma),$$

where $d_{i,j} = 1$ if consumer i choses vehicle j and zero otherwise. Values of $\{\alpha^*,\beta,\gamma,\sigma\}$ are searched over such that the simulated log-likelihood function is maximized using a numerical optimization routine.

2.5 Model Specification

This section describes the variables and specification of the model. For computational reasons, I limit the analysis to the 2008 new vehicle market, resulting in $N = 11,915$ household observations and $J = 204$ new vehicles in each consumer's choice set. Since the 2009 NHTS was conducted between March 2008 and May 2009, I assume all households with a 2008 vehicle purchased it new.

2.5.1 Vehicle Attributes

I follow the vehicle demand literature and include the following vehicle attributes: horsepower to weight ratio (HPWT), dollars-per-mile (DPM), and size = (length x width). The HPWT variable is a proxy for vehicle acceleration and the DPM variable measures how much it costs to drive each vehicle in terms of fuel expenditures per mile. For example, if the price of gasoline is \$3.00 per gallon, it will cost a Toyota Prius owner $\frac{\$3.00/\text{gallon}}{46\text{miles/gallon}} = \frac{\$0.07}{\text{mile}}$ to drive their vehicle. The manufacturer's suggested retail price (MSRP) is used since transaction prices are unavailable. Dummy variables are constructed indicating if the vehicle is a car, suv, van, or truck and also for luxury and hybrid vehicles. Finally, a dummy variable is constructed if the vehicle is produced by a USA manufacturer (Ford, GM, and Chrysler). See Table 2.1 for summary statistics describing vehicle attributes.

2.5.2 Heterogeneous Tastes

To capture observable heterogeneous tastes, several demographic variables are interacted with the vehicle attributes listed above. These variables include household income, age, the percentage of Democrats that live in each household's county, a dummy variable if the household has a college degree, a dummy variable if the household has a child, and a dummy variable if the household lives in a Metropolitan Statistical Area (MSA) that exempts hybrid vehicles from the HOV lane minimum occupancy requirement.[6] See Table 2.2 for summary statistics describing household demographics.

Unobservable (to the econometrician) tastes are assumed to be normally distributed, $\mu \sim N(0, \sigma)$, with a mean of zero and variance σ. I use 200 Halton draws rather than pseudo-random draws to improve the coverage of the simulation (see Train 2003).

2.5.3 Endogenous Price

Unobservable vehicle characteristics such as quality, performance, and reliability are expected to be correlated with price, which could create an endogeneity problem. Failing to control for these effects will result in the price variables being positively correlated with the error term. This, in turn, will bias the price variables towards zero making households appear less sensitive to prices than they actually are. To overcome this problem, I include vehicles' 24 month expected retained value as an additional explanatory variable. Train and Winston (2007) argue that if a manufacturer were to raise the price of their vehicle without also improving its attributes, the retained value would not rise proportionally and may not rise at

[6]The one exception is for the Washington, D.C. MSA because it encompasses Virginia, Maryland and D.C. Because only Virginia exempts hybrid vehicles from the HOV lane minimum occupancy requirement, I only construct a dummy variable if the household lives in the Northern Virginia area, rather than D.C. or Maryland. It should also be noted that hybrid vehicles are only permitted to drive on the I-66 HOV lanes and not I-395 HOV lanes in Northern Virginia. See Appendix IV for more details.

all. Therefore, higher retained values should proxy for higher quality vehicles. Consumer Report's Road Test Score is also included in the model to control for additional unobservable vehicle characteristics. The score is based on over 50 criteria and includes things such as performance, handling, and reliability.

2.5.4 Price Effects, Income Effects, and the MU of Income

Capturing price and income effects is crucial for obtaining accurate marginal utility of income estimates. The literature has used several specifications in the past including a constant marginal utility of income specification,

$$\alpha h(y_i, p_j) = \alpha(y_i - p_j),$$

diminishing marginal utility of income,

$$\alpha h(y_i, p_j) = \alpha \log(y_i - p_j),$$

income as a taste parameter,

$$\alpha h(y_i, p_j) = \alpha \frac{p_j}{y_i},$$

and a constant marginal utility of income within income groups, allowing it to vary across groups,

$$\alpha h(y_i, p_j) = \alpha^*(y_i - p_j),$$

where

$$\alpha^* = \begin{cases} \alpha_1 & \text{if income} \leq y_1 \\ \alpha_2 & \text{if } y_1 < \text{income} \leq y_2 \\ & \vdots \\ \alpha_m & \text{if income} > y_m. \end{cases}$$

In the expression above, m is the number of income groups specified by the researcher. See Appendix III for an in-depth discussion of each specification.

In this chapter, I use a constant (fixed) marginal utility of income within income groups, allowing it to vary across groups. Specifically, I partition households into three groups with $y_1 = \$50,000$ and $y_2 = \$100,000$. This is less restrictive than the constant marginal utility of income specification for all households and also simplifies the welfare analysis procedure.

Furthermore, since my model is conditional on purchasing a new vehicle rather than the population as a whole, households should be more homogeneous in regards to their valuation of a marginal dollar. I also use “taste income" to interact with the hybrid and luxury dummy variables to capture wealthier households' preferences towards these vehicles. However, this will not affect the marginal utility of income since it is “taste income" rather than “financial income" that is interacted with these terms.

2.5.5 WTP for Hybrid Vehicles

Households' WTP for hybrid vehicles is estimated as a function of both observable and unobservable (to the econometrician) heterogeneous tastes. For simplicity, let households' conditional indirect utility function be specified as a function of the price and income effect component, $h(y,p)$, hybrid vehicle terms, X_h, and control variables,

$$V = \alpha^* h(y,p) + \beta_h X_h + \text{controls},$$

where α^* is the marginal utility of income, β_h is the marginal utility from hybrid vehicles, and X_h is a hybrid dummy variable. By setting the total derivative of the indirect utility function to zero, the marginal rate of substitution (MRS) between the price and income effect component, $h(\cdot)$, and hybrid vehicles, X_h can be used to calculate households' WTP,

$$WTP_{\text{hybrid}} \equiv \frac{\partial h(y,p)}{\partial X_h} = -\frac{\beta_h}{\alpha^*}.$$

This is interpreted as the amount of money that would have to be added (or subtracted) from the price of the vehicle (which will provide negative utility) if the vehicle has hybrid technology (which may or may not provide positive utility given a household's observable and unobservable preferences) that will keep the household's utility level constant.

Since the marginal utility of income is allowed to vary across income groups, WTP estimates will also vary,

$$WTP_{\text{hybrid}} = \begin{cases} -\frac{\beta_h}{\alpha_1} & \text{if income} \le y_1 \\ -\frac{\beta_h}{\alpha_2} & \text{if } y_1 < \text{income} \le y_2 \\ -\frac{\beta_h}{\alpha_3} & \text{income} > y_2. \end{cases}$$

Furthermore, since the hybrid effect consists of an average component, demographical taste shifters, and a random component, household i's marginal utility from hybrid vehicles is

$$\beta_h^i = \bar{\beta}_h + \gamma_h D_h^i + \sigma_h \mu_h^i,$$

which will vary across households, where D_h^i is a vector of household i's demographics and μ_h^i is a scalar of household i's unobservable tastes for hybrid vehicles.

2.5.6 WTP Conditional on Vehicle Choice

Unobservable (to the econometrician) heterogeneous preferences, μ_i, are assumed to vary normally across households, $f(\mu|\sigma)$. The variance of which, σ, is estimated from the maximum likelihood procedure outlined above. However, the distribution, $f(\mu|\sigma)$, only characterizes preferences for the population as a whole and provides no information in regards to where a specific household's preference may lie in this distribution. For example, do hybrid vehicle owners prefer fuel-savings more than the average person? Do Prius owners dislike trucks? Do Corvette owners really like horsepower? It may be important to market researchers to know this information. Revelt and Train (2000) suggest that a household's purchase history can be used in a post-estimation procedure to help reveal this.

I propose a similar method that relies on a simulation technique which is, in some ways, easier to implement. My method can also be extended to yield distributions of unobservable preferences conditional on both 1) a particular vehicle and 2) specific observable demographics. For example, I may want to know what the distribution of unobservable preferences looks like given that the household purchased a Toyota Prius and makes less than $50,000 per year, lives in an area where the price of gasoline is less than $3.50, and lives in a county where a percentage of Democrats is less than 50%. This additional information can be used to construct bounds on WTP estimates by reducing the variance of unobservable factors.

The idea is simple - using the parameter estimates from the model and random draws from both observable and unobservable heterogeneous tastes, I simulate a household's utility for each $j \in J$ product in the market. If product j yields the highest utility, and is also the product of interest (i.e. Toyota Prius), I keep that household's simulated unobservable preferences. After repeating this several million times, I am left with a distribution of unobservable preferences conditional on a household's vehicle choice and observable demographics.

2.6 Results

In this section I discuss the results organized into four subsections: cost-of-ownership estimates, parameter estimates, WTP estimates, and conditional distributions of unobservable preferences.

2.6.1 Cost-of-Ownership

To calculate vehicle cost-of-ownership, I follow Consumer Report's approach and incorporate vehicle depreciation, interest on financing, insurance premiums, maintenance and repair costs, fees and taxes (including tax incentives), and fuel costs over a five year period. Because hybrid models typically come with more features than the standard base model, I compare hybrid vehicles to their all gas equivalent with the closest trim level. For example, the Honda Civic Hybrid is compared to the Honda Civic LX rather than the Civic DX.[7] Since it would be difficult to present the results for each state, ownership costs are based on the National average. I temporarily exclude fuel costs so the reader can disentangle fuel savings from the higher price premium. See Table 2.3 for a complete five year cost-of-ownership comparison for all 2008 hybrid vehicles (in my data) to their closest all gas equivalent. For the readers convenience, Table 2.4 provides a comparison of EPA adjusted fuel economy estimates and vehicle prices and Table 2.5 provides a comparison of expected fuel savings.

Excluding fuel costs, the Toyota Prius will cost the average owner $27,106 after five years whereas it would only cost $21,025 to own a Toyota Corolla, for a difference of $6,081. The Prius, however, has better fuel economy saving households between $1,912 to $3,823 after 5 years with gasoline prices at $2.00 and $4.00, respectively, when compared to the Corolla.[8] See Table 2.5. Therefore, at best a Toyota Prius owner will lose $2,258 and at worst lose $4,196. The Honda Civic Hybrid has less favorable results. At best, a Honda Civic Hybrid owner will lose $5,099 and at worse lose $6,700. There are, however, several hybrid models for which the owner will actually save money. If gasoline prices are greater than $3.50, owners of the Nissan Altima Hybrid and Ford Escape Hybrid will recover their costs. If gasoline prices are greater than $2.50, owners of the Lexus RX400 hybrid will recover their costs. The Saturn Vue Hybrid is the only vehicle for which all costs are recovered after 5 years for all gasoline

[7]These features include cruise control, power locks and mirrors, remote keyless entry, and a compact disc player.

[8]Fuel expenditures are calculated using the EPA adjusted fuel economy estimates rather than the manufacturer's estimates to reflect more realistic driving habits.

prices examined in this chapter. Interestingly, the hybrid vehicles with the greatest chance of recovering the higher ownership costs represent a small share of the total hybrid vehicle market.[9] See Table 2.6 for a complete list of total cost-of-ownership estimates.

2.6.2 Parameter Estimates

The price coefficients, as indicated in Table 2.7, are all of the correct sign and are statistically significant at the 1% level. As expected, households that make above $100,000 per year are less sensitive to prices than households that make between $50,000 and $100,000 per year, and those households are in turn are less sensitive to prices than households that make less than $50,000 per year. Vehicles' 24 month expected retained value is also included in the model to help control for unobservable product characteristics. As expected, vehicles with higher retained values are preferred to those with less. Also, the magnitude of the retained value coefficient, when compared to the price coefficients, suggests that households are more concerned about the vehicle's resale value rather than the initial upfront cost of the vehicle.

Several vehicle characteristics are included in the model and are presented in Table 2.7. Some of these characteristics are interacted with household demographics and unobservable heterogeneous tastes. The first variable is Consumer Report's Road Test Score which is intended to capture additional unobservable product characteristics to further minimize any potential endogeneity problem. The coefficient is significant at the 1% level and is positive indicating that consumers, on average, prefer vehicles with better performance, handling, and reliability. Fuel expenditures are accounted for in the model by including a dollars-per-mile (DPM) variable which is calculated using local gasoline prices and the fuel economy (mpg) for each vehicle. A random coefficient is also included for this variable, but it is not statistically significant indicating that households universally dislike fuel costs equally as much. The coefficient for vehicle size (length x width) is positive and significant at the 1% level indicating that consumers prefer larger vehicles on average. The random coefficient, however, is small in magnitude and is only significant at the 10% level indicating that there is very little variation in unobservable tastes for size. To capture vehicle acceleration, I follow the literature and include the horsepower to weight ratio (HPWT) for each vehicle. I take the log of this variable, Log(HPWT), to capture the diminishing returns from additional

[9] Market shares for the 2008 hybrid vehicle market are the following: Prius = 50.1%, Camry = 14.8%, Civic = 10%, Highlander = 6.2%, Escape = 5.5%, RX400h = 4.9%, Altima = 2.8%, Vue = 0.9%, Mariner = 0.7%.

acceleration which also appeared to fit the data better. Surprisingly, the base coefficient is negative indicating that the average person dislikes acceleration. However, the random coefficient is large in magnitude and is statistically significant at the 1% level indicating that there is a large degree of variation in unobservable tastes. The base coefficient for Log(HPWT) is also interacted with a continuous demographic variable described below, so it should be interpreted accordingly.

Table 2.7 indicates that there exists a large degree of variation in tastes for vehicle types: cars, SUVs, and trucks. SUVs have the largest degree of heterogeneity, with an estimated standard deviation of $\sigma_{\text{SUVs}} = 9.35$ and a base coefficient of $\beta_{\text{SUVs}} = -3.5$ indicating that some household strongly prefer SUVs while other strongly dislike SUVs. Cars have the smallest degree of heterogeneity, with a standard deviation estimate of $\sigma_{\text{car}} = 0.86$ and a base coefficient of $\beta_{\text{car}} = 0.83$ indicating that the average household prefers cars with some households preferring them slightly more or less than others. The only standard deviation parameter that is not statistically significant is for vans which also has a negative base coefficient, $\beta_{\text{van}} = -0.12$.

To capture observable heterogenous tastes, several demographic variables are interacted with vehicle characteristics and are presented in Table 2.8. As expected, households with children strongly prefer vans and SUVs over other types of vehicles, presumably for their larger seating capacity. Therefore, the only instance in which a household will prefer a van is if they have a child. To capture a household's tastes for luxury vehicles, a luxury dummy variable is constructed and is interacted with household income as well as a random coefficient. As expected, wealthier households prefer luxury vehicles over non-luxury vehicles. However, the variance in unobservable tastes is large and is significant at the 1% level. This indicates that, although wealthier households prefer luxury vehicles on average, some wealthy households may be more frugal or conservative in regards to their vehicle choice and may, as a result, prefer a slightly more modest vehicle, and vice versa.

To capture households' observable and unobservable tastes for hybrid vehicles, a hybrid dummy variable is constructed and is interacted with several demographic variables in addition to a random coefficient. Table 2.9 shows that all variables are of the expected sign. Gasoline prices, income, education, and age are all positively related to hybrid vehicle choice and are significant at the 1% level. Households living in a MSA that exempts hybrid vehicles from the HOV lane minimum occupancy requirement is also found to increase the likelihood of purchasing a hybrid vehicle, however, it is only significant at the 10% level.[10] The variance in unobservable tastes for hybrid vehicles is not statistically significant indicating that all heterogeneity in tastes can be explained through observable demographics. The base coefficient for hybrid vehicles is negative, large in magnitude and is significant at the 1% level. However, it should be interpreted with caution because it is interacted with several continuous demographic variables.[11]

To capture the conspicuous environmental effect of hybrid vehicles, I construct two additional dummy variables, one for hybrid vehicles in general, and one of the Toyota Prius, and interact them with the percentage of Democrats that live in each household's county. See Table 2.9. Although the hybrid interaction coefficient is positive, it is not statistically significant. Therefore, there does not appear to be any conspicuous effects from owning hybrid vehicle in general. However, the Prius interaction coefficient is highly significant at the 1% level and is large in magnitude, thus confirming what Kahn (2007) and Sexton and Sexton (2011) found. Specifically, that one of the primary reasons consumers purchase the Prius is not necessarily to decrease fuel expenditures but rather to increase one's social status and be recognized by the community as an environmentally conscious consumer. To control for the possibility that households in green communities simply prefer more fuel efficient vehicles, I also interact the percentage of Democrats that live in each household's county with the fuel economy (mpg) of each vehicle. Though positive, there is no statistical evidence that households in greener communities prefer more fuel efficient vehicles.

Several other vehicle characteristics are interacted with the percentage of Democrats that live in each household's county. These include the vehicle characteristics: size, a truck dummy variable, Log(HPWT), and a dummy variable if the vehicle is produced by a USA manufacturer (Ford, GM, and Chrysler). Table 2.8 indicates that as the percentage of Democrats increase, households increasingly dislike larger vehicles, strongly dislike trucks, strongly prefer acceleration, and dislike vehicles produced by a USA manufacturer when compared to the average new car buyer.[12]

[10]This is somewhat expected though because California, the number one market for hybrid vehicles, ran out of Clean Pass Permits in 2007 and my model only examines the 2008 market.

[11]For example, if the household makes $50,000 per year, is 50 years old, has a college degree, lives in a county that voted 50% for the Democratic party, does not have access to HOV lanes, and the local price of gas is $3.50, then the hybrid coefficient will be $\beta_{\text{hybrid}} = -4.2$, indicating that the average household, conditional on these demographics, would dislike hybrid vehicles.

[12]There may be some multicollinearity between these variables (vehicle size, trucks, and USA manufacturers) and fuel economy (MPG), which may partially explain why the model failed to show with any statistical significance that households in greener communities prefer more fuel efficient vehicles.

Finally, I interact household age with a dummy variable for vehicles produced by a USA manufacturer. Table 2.8 shows that the coefficient is positive and is statistically significant at the 1% level indicating that older households prefer vehicles produced in the USA more than those produced by a foreign manufacturer. This result should be troubling for the Big Three Automakers because evidence suggest that brand loyalty is a key explanatory variable in vehicle choice. See Mannering and Winston (1991) and Train and Winston (2007). Therefore, over time, fewer and fewer people will prefer vehicles produced by a USA manufacturer, all else equal.

2.6.3 WTP for Hybrid Vehicles

As shown above, most households will not recover the premium of owning a hybrid vehicle through decreased fuel expenditures. On top of this, several papers have shown that fuel savings is not even the primary reason consumers purchase hybrid vehicles. Rather, consumers purchase hybrid vehicles because it is a better match to their preference type. Therefore, it is reasonable to ask how much consumers are willing-to-pay for hybrid vehicles given their observable and unobservable heterogeneous tastes. Economic theory suggests that households should only purchase a hybrid vehicle if they are willing to pay more than it costs to own.

Table 2.10 shows various WTP estimates for hybrid vehicles conditional on the household having a college degree, being 50 years old, and living in an area that does not exempt hybrid vehicles from the HOV lane minimum occupancy requirement. Furthermore, these results are reported for various gasoline prices, various household incomes, and various percentages of Democrats per household county. WTP estimates range anywhere from -$2,255 to -$13,678 indicating that the average household, regardless of their demographic type, would receive dis-utility from owning a hybrid vehicle. However, Table 2.11 suggests that if the household lives in a MSA that exempts hybrid vehicles from the HOV lane minimum occupancy requirement then they would be willing to pay an additional $1,189 to $1,625 more for them. Older households are also willing to pay more for hybrid vehicles. For example, if the head of the household is 60 rather than 50, they would be willing to pay an additional $789 to $1,095 more for a hybrid vehicle thus making them even more attractive. Finally, since preferences for hybrid vehicles vary across households for unobservable reasons, some households would be willing to pay an additional $4,331 to $5,917 more for a hybrid vehicle.[13]

[13] The variance parameter estimate for hybrid vehicles was not found to be statistically significant. Therefore,

See Table 2.12.

WTP estimates are also constructed for the Toyota Prius using the WTP estimates for hybrid vehicles in general plus the additional Prius interaction effects. Table 2.10 shows that an average household living in a county that voted 75% for the Democratic party, with local gasoline prices at $3.60 per gallon, and an annual income of $80,000 would be willing to pay $4,124 for a Toyota Prius. On the other end of the political spectrum, if the same household lived in a county that only voted 25% for the Democratic party, with gasoline prices still at $3.60 per gallon and an annual income of $80,000, the household would suffer a welfare loss from owning a Toyota Prius with a WTP estimate of -$3,939. Again, households will be willing to pay more for a Prius if they are older, have more income, and if they live in an area that exempts hybrid vehicles from the HOV lane minimum occupancy requirement, in addition to unobservable reasons.

The conspicuous environmental effect from owning a Toyota Prius is interpreted as the difference between the WTP estimates for hybrid vehicles in general and the WTP estimates for the Toyota Prius. For a household that lives in a county that voted 75% for the Democratic party, the difference in WTP estimates suggest that a household would be willing to pay $11,064 more for a Prius than any other hybrid vehicle. For the same household living in a county that voted 25% for the Democratic party, the difference suggests that they would only be willing to pay $4,618 more. These results suggest two things. First, the large difference between WTP estimates suggest that households place a high value on the unique design of the Prius. As noted earlier, other hybrid vehicles only have an inconspicuous decal on their back bumper making them virtually indistinguishable from their all gas equivalent. Second, the variation in the differences indicate that as the percentage of Democrats increase, the conspicuous Prius effect becomes increasingly more valuable. This suggests that as a household's community becomes more green, they may face more social pressure to adopt a greener lifestyle.

2.6.4 Conditional Unobservable Preferences

Using a simulation technique, I am able to construct tighter bounds on distributions of unobservable preferences by incorporating information on vehicle choice and observed

I cannot show with any statistical certainty that households are willing to pay more or less for hybrid vehicles due to unobservable reasons.

demographics into a post-estimation procedure.[14] This additional information can be used to construct more accurate WTP estimates for unobservable factors. This is done by simply dividing the distribution of unobservable preferences by the marginal utility of income parameter. Unfortunately, in this chapter the variance parameter estimate for hybrid vehicles is not statistically significant so it will not affect unobservable WTP estimates in this application. However, to demonstrate how this method could be used in future research, I provide several conditional distributions purely for demonstration. For example, Figures 1 - 3 demonstrate that Prius owners strongly dislike acceleration, strongly dislike SUVs, and dislike luxury when compared to the distribution of preferences for the population as a whole. Figure 4 shows that a Prius owner's tastes for hybrid vehicles, though not statistically significant, appears to be slightly larger than that for the average new car buyer. However, as mentioned above, all variation in tastes for hybrid vehicles is explained through observable demographics, rather than unobservable factors. Finally, Figure 5 and 6 demonstrate that Corvette owners strongly prefer acceleration and Ford F-150 owners strongly prefer trucks, respectively.

2.7 Conclusion

This chapter examined why, despite the higher cost of ownership, consumers continue to purchase hybrid vehicles. To address this question, the cost of ownership differential between hybrid vehicles and their closest all gas equivalent was examined. After factoring in vehicle depreciation, insurance premiums, interest on financing, maintenance and repair costs, fees and taxes (including tax incentives), and fuel expenditures, I find that the majority of hybrid vehicle owners will not recover the higher premium through decreased fuel expenditures after five years of ownership.

However, this chapter shows that households purchase hybrid vehicles for reasons other than just to decrease their fuel expenditures. For example, in some states, hybrid vehicles are exempt from HOV lane minimum occupancy requirements. For households that commute to work on a daily basis, this can translate into huge time savings, making owning a hybrid vehicle more valuable. For other households, owing a hybrid vehicle is more of a lifestyle choice - it fits their personality better. As this chapter points out, many hybrid vehicle owners say that they are just trying to "be more green, be more socially responsible, and reduce their carbon footprint." For other households, simply being green isn't good enough. They

[14]My procedure is based on 5 million simulations which took approximately 15 minutes to run.

want to be conspicuously green such that they are recognized by their community as an environmentally conscious consumer.

The findings of this chapter demonstrate the importance of accounting for heterogeneous preferences in consumer demand models. What one households likes, another may dislike. Future research should examine the welfare implications from the CAFE standards outlined in the 2009 Joint EPA/NHTSA National Program. The cost-benefit analysis of this program, conducted by the EPA and NHTSA, assume that the higher CAFE standards will not affect vehicle attributes including “performance, passenger- and cargo-carrying capacity, or other dimensions of utility" (see EPA/NHSTA National Program, page 177). However, this seems to be a relatively strong assumption and ignores the apparent trade-off between increased fuel efficiency and decreased horsepower or vehicle size.

Table 2.1: Summary Statistics: Vehicle Characteristics

Variable	Mean	Median	5^{th} Percentile	95^{th} Percentile
Price	$ 23,565	$ 21,610	$ 14,490	$ 41,945
Residual[b]	58%	58%	46%	68%
Score[c]	68	70	46	89
Size[d]	13.8	13.6	11.6	16.7
Horsepower	196	190	110	295
MPG Avg.[e]	24.4	23	15	38

Variable	Frequency[f]
Car	49.7%
SUV	28.9%
Van	5.7%
Truck	13.5%
Sports	2.3%
Luxury	12.2%
Hybrid	5.8%
Prius	2.5%
USA	43.5%

[a] Data obtained from the 2009 National Household Travel Survey (NHTS), Kelley Blue Book: Residual Guide, Consumer Reports, and Ward's Automotive Yearbook.
[b] The residual is the percent of a new vehicle's value that is expected to be retained 24 months into the future.
[c] Consumer Report's Road Test Score rates vehicles on a scale of 1 to 100, with 100 being the best.
[d] Size = (length x width)/10,000, measured in inches.
[e] MPG Avg. assumes 55% city driving and 45% highway driving.
[f] Frequencies are based on the NHTS sample survey (11,915 observations) rather than actual market shares.

Table 2.2: Summary Statistics: Household Demographics

Variable	Mean	Median	5^{th} Percentile	95^{th} Percentile
Gas Price[b]	\$ 3.51	\$ 3.46	\$ 3.33	\$ 3.89
%Democrat[c]	50%	51%	26%	70%
Age	57	57	34	81
Income[d]	\$ 82,472	\$ 70,560	\$ 21,993	\$ 164,846

Variable	Frequency[f]
College Educated	59.2%
Kids	28.5%
HOV[e]	14.5%

[a] Data obtained from the 2009 National Household Travel Survey (NHTS), USA Today, and the 2008 ACCRA cost-of-living database.
[b] 2008 Gasoline prices are obtained for 322 Core Based Statistical Areas (CBSAs).
[c] The 2008 Presidential election results are used to link the percentage of Democrats per county to each household.
[d] Household incomes are adjusted using the 2008 ACCRA Cost-of-Living index to account for purchasing power differences across regions.
[e] Only households that live in a MSA that exempts hybrid vehicles from the minimum occupancy requirement are included.
[f] Frequencies are based on the 2009 NHTS sample survey (11,915 observations).

Table 2.3: Five Year Cost of Ownership: (Excludes Fuel Expenditures)

		Depreciation	Financing	Insurance	Maintenance	Fees & Taxes[b]	Repairs	Total
Honda	Civic Hybrid	$16,122	$4,401	$5,160	$2,047	$1,610	$749	$30,089
Honda	Civic	$10,321	$2,857	$4,818	$1,935	$1,147	$710	$21,788
	Difference =	$5,801	$1,544	$342	$112	$463	$39	$8,301
Ford	Escape Hybrid	$16,954	$4,074	$5,150	$1,935	$1,623	$1,313	$31,049
Ford	Escape	$14,672	$3,389	$4,861	$1,935	$1,381	$1,152	$27,390
	Difference =	$2,282	$685	$289	$0	$242	$161	$3,659
Mercury	Mariner Hybrid	$20,726	$5,085	$4,708	$2,150	$320	$749	$33,738
Mercury	Mariner	$14,617	$3,344	$4,858	$1,903	$1,367	$1,152	$27,241
	Difference =	$6,109	$1,741	-$150	$247	-$1,047	-$403	$6,497
Saturn	Vue Hybrid	$14,923	$4,320	$4,876	$3,168	$59	$749	$28,095
Saturn	Vue	$14,390	$3,399	$4,491	$3,276	$1,327	$749	$27,632
	Difference =	$533	$921	$385	-$108	-$1,268	$0	$463
Nissan	Altima Hybrid	$14,552	$4,167	$6,102	$2,391	$1,560	$715	$29,487
Nissan	Altima	$12,333	$3,477	$5,292	$2,268	$1,337	$715	$25,422
	Difference =	$2,219	$690	$810	$123	$223	$0	$4,065
Toyota	Camry Hybrid	$15,740	$4,556	$5,343	$2,263	$1,682	$710	$30,294
Toyota	Camry	$13,305	$3,764	$4,709	$2,230	$1,442	$715	$26,165
	Difference =	$2,435	$792	$634	$33	$240	-$5	$4,129
Toyota	Highlander Hybrid	$24,302	$6,147	$5,293	$2,665	$2,214	$832	$41,453
Toyota	Highlander	$20,140	$5,175	$4,793	$2,702	$1,911	$832	$35,553
	Difference =	$4,162	$972	$500	-$37	$303	$0	$5,900
Toyota	Prius	$13,150	$3,978	$5,351	$2,385	$1,493	$749	$27,106
Toyota	Corolla	$9,518	$2,608	$5,298	$1,823	$1,068	$710	$21,025
	Difference =	$3,632	$1,370	$53	$562	$425	$39	$6,081
Lexus	RX 400h	$23,504	$6,823	$7,291	$2,288	$2,421	$886	$43,213
Lexus	RX 350	$22,595	$5,474	$7,264	$2,605	$2,132	$1,312	$41,382
	Difference =	$909	$1,349	$27	-$317	$289	-$426	$1,831

[a] Data obtained from Driverside.com. Costs are based on the National average.
[b] Taxes include all Federal and State hybrid vehicle tax incentives.

Table 2.4: EPA Adjusted Fuel Economy Estimates and Vehicle Prices

		EPA Estimates			Vehicle Price
Make	Model	MPG City	MPG Hwy	Combined[a]	(MSRP-Federal Tax)
Honda	Civic Hybrid	40	45	42	$21,135
Honda	Civic	25	36	29	$17,595
	Difference =	15	9	13	$3,540
Ford	Escape Hybrid	34	30	32	$24,330
Ford	Escape	20	26	22	$21,735
	Difference =	14	4	10	$2,595
Mercury	Mariner Hybrid	34	30	32	$24,020
Mercury	Mariner	20	26	22	$21,585
	Difference =	14	4	10	$2,435
Saturn	Vue Hybrid	25	32	28	$23,345
Saturn	Vue	19	26	22	$21,395
	Difference =	6	6	6	$1,950
Nissan	Altima Hybrid	35	33	34	$27,555
Nissan	Altima	19	26	22	$25,695
	Difference =	16	7	12	$1,860
Toyota	Camry Hybrid	33	34	34	$25,860
Toyota	Camry	21	31	25	$21,900
	Difference =	12	3	9	$3,960
Toyota	Highlander Hybrid	27	25	26	$34,385
Toyota	Highlander	17	23	19	$29,435
	Difference =	10	2	7	$4,950
Toyota	Prius	48	45	46	$22,985
Toyota	Corolla	26	35	29	$16,275
	Difference =	22	10	17	$6,710
Lexus	RX400h	27	24	25	$41,945
Lexus	RX350	18	23	20	$38,165
	Difference =	9	1	5	$3,780

[a] Combined fuel economy assumes 55% city driving and 45% highway driving.

Table 2.5: Fuel Expenditures: (5 years; 15000 miles/year)

		Gasoline Prices				
Make	Model	$2.00	$2.50	$3.00	$3.50	$4.00
Honda	Civic Hybrid	$3,571	$4,464	$5,357	$6,250	$7,143
Honda	Civic	$5,172	$6,466	$7,759	$9,052	$10,345
	Savings =	$1,601	$2,001	$2,401	$2,802	$3,202
Ford	Escape Hybrid	$4,688	$5,859	$7,031	$8,203	$9,375
Ford	Escape	$6,818	$8,523	$10,227	$11,932	$13,636
	Savings =	$2,131	$2,663	$3,196	$3,729	$4,261
Mercury	Mariner Hybrid	$4,688	$5,859	$7,031	$8,203	$9,375
Mercury	Mariner	$6,818	$8,523	$10,227	$11,932	$13,636
	Savings =	$2,131	$2,663	$3,196	$3,729	$4,261
Saturn	Vue Hybrid	$5,357	$6,696	$8,036	$9,375	$10,714
Saturn	Vue	$6,818	$8,523	$10,227	$11,932	$13,636
	Savings =	$1,461	$1,826	$2,192	$2,557	$2,922
Nissan	Altima Hybrid	$4,412	$5,515	$6,618	$7,721	$8,824
Nissan	Altima	$6,818	$8,523	$10,227	$11,932	$13,636
	Savings =	$2,406	$3,008	$3,610	$4,211	$4,813
Toyota	Camry Hybrid	$4,412	$5,515	$6,618	$7,721	$8,824
Toyota	Camry	$6,000	$7,500	$9,000	$10,500	$12,000
	Savings =	$1,588	$1,985	$2,382	$2,779	$3,176
Toyota	Highlander Hybrid	$5,769	$7,212	$8,654	$10,096	$11,538
Toyota	Highlander	$7,895	$9,868	$11,842	$13,816	$15,789
	Savings =	$2,126	$2,657	$3,188	$3,720	$4,251
Toyota	Prius	$3,261	$4,076	$4,891	$5,707	$6,522
Toyota	Corolla	$5,172	$6,466	$7,759	$9,052	$10,345
	Savings =	$1,912	$2,389	$2,867	$3,345	$3,823
Lexus	RX400h	$6,000	$7,500	$9,000	$10,500	$12,000
Lexus	RX350	$7,500	$9,375	$11,250	$13,125	$15,000
	Savings =	$1,500	$1,875	$2,250	$2,625	$3,000

[a] Uses EPA adjusted fuel economy estimates.

Table 2.6: Total Five Year Cost of Ownership

Make	Model	Gasoline Price $2.00	$2.50	$3.00	$3.50	$4.00
Honda	Civic Hybrid	$33,660	$34,553	$35,446	$36,339	$37,232
Honda	Civic	$26,960	$28,254	$29,547	$30,840	$32,133
	Difference =	$6,700	$6,300	$5,900	$5,499	$5,099
Ford	Escape Hybrid	$35,737	$36,908	$38,080	$39,252	$40,424
Ford	Escape	$34,208	$35,913	$37,617	$39,322	$41,026
	Difference =	$1,528	$996	$463	-$70	-$602
Mercury	Mariner Hybrid	$38,426	$39,597	$40,769	$41,941	$43,113
Mercury	Mariner	$34,059	$35,764	$37,468	$39,173	$40,877
	Difference =	$4,366	$3,834	$3,301	$2,768	$2,236
Saturn	Vue Hybrid	$33,452	$34,791	$36,131	$37,470	$38,809
Saturn	Vue	$34,450	$36,155	$37,859	$39,564	$41,268
	Difference =	-$998	-$1,363	-$1,729	-$2,094	-$2,459
Nissan	Altima Hybrid	$33,899	$35,002	$36,105	$37,208	$38,311
Nissan	Altima	$32,240	$33,945	$35,649	$37,354	$39,058
	Difference =	$1,659	$1,057	$455	-$146	-$748
Toyota	Camry Hybrid	$34,706	$35,809	$36,912	$38,015	$39,118
Toyota	Camry	$32,165	$33,665	$35,165	$36,665	$38,165
	Difference =	$2,541	$2,144	$1,747	$1,350	$953
Toyota	Highlander Hybrid	$47,222	$48,665	$50,107	$51,549	$52,991
Toyota	Highlander	$43,448	$45,421	$47,395	$49,369	$51,342
	Difference =	$3,774	$3,243	$2,712	$2,180	$1,649
Toyota	Prius	$30,367	$31,182	$31,997	$32,813	$33,628
Toyota	Corolla	$26,197	$27,491	$28,784	$30,077	$31,370
	Difference =	$4,169	$3,692	$3,214	$2,736	$2,258
Lexus	RX400h	$49,213	$50,713	$52,213	$53,713	$55,213
Lexus	RX350	$48,882	$50,757	$52,632	$54,507	$56,382
	Difference =	$331	-$44	-$419	-$794	-$1,169

[a] Factors in fuel expenditures as well as vehicle depreciation, interest on financing, maintenance and repair costs, fees and taxes (including tax incentives), and insurance premiums. All costs are based on the National average.

Table 2.7: Parameter Estimates: Vehicle Characteristics (Mixed Logit - 200 Halton Draws)

	Base Coefficient	Random Coefficient
Price Terms [b]		
Price_1	-0.1671^{***}	-
	(0.005115)	-
Price_2	-0.1572^{***}	-
	(0.004736)	-
Price_3	-0.1223^{***}	-
	(0.0043)	-
Residual Price	0.1908^{***}	-
	(0.00662)	-
Vehicle Characteristics		
Road Score	0.0194^{***}	-
	(0.001182)	-
DPM	-17.0528^{***}	0.0463
	(1.177)	(6.5197)
Size	0.4813^{***}	0.0707^{*}
	(0.0335)	(0.0435)
log(HPWT)	-1.8039^{***}	1.6244^{***}
	(0.2941)	(0.1705)
Vehicle Type		
Car	0.8599^{***}	0.8266^{**}
	(0.1468)	(0.3805)
SUV	-3.523^{**}	9.3466^{***}
	(1.414)	(2.5643)
Van	-0.1218	1.1572
	(0.7016)	(0.8385)
Truck	1.9842^{***}	2.3139^{**}
	(0.2889)	(0.9286)

[a] *** 1% significance, ** 5% significance, * 10% significance. Standard errors in parentheses.
[b] Price coefficients correspond to households in the first, second, and third income groups, respectively.

Table 2.8: Parameter Estimates: Vehicle Interactions (Mixed Logit - 200 Halton Draws)

	Base Coefficient	Random Coefficient
Luxury	-5.6818^{***}	1.1862^{***}
	(1.4173)	(0.4039)
Luxury*Log(Income)	0.4314^{***}	-
	(0.1066)	-
Kids*(SUV or Van)	1.3137^{***}	-
	(0.2331)	-
USA*Age	0.0123^{***}	-
	(0.001016)	-
%Democrat*Size	-0.2963^{***}	-
	(0.0615)	-
%Democrat*Truck	-3.5766^{***}	-
	(1.2306)	-
%Democrat*Log(HPWT)	2.3669^{***}	-
	(0.5621)	-
%Democrat*MPG	0.003322	-
	(0.014)	-
%Democrat*USA	-0.539^{***}	-
	(0.1183)	-

[a] *** 1% significance, ** 5% significance, * 10% significance. Standard errors in parentheses.

Table 2.9: Parameter Estimates: Hybrid Interactions (Mixed Logit - 200 Halton Draws)

	Base Coefficient	Random Coefficient
Hybrid	-13.1826^{***}	0.4386
	(1.3977)	(0.403)
Hybrid*Gas	1.6113^{***}	-
	(0.2594)	-
Hybrid*Log(income)	0.4211^{***}	-
	(0.0758)	-
Hybrid*Education	0.5394^{***}	-
	(0.0918)	-
Hybrid*Age	0.0122^{***}	-
	(0.00335)	-
Hybrid*%Democrat	0.5165	-
	(0.4572)	-
Hybrid*HOV	0.1987^{*}	-
	(0.1154)	-
Prius	0.2192	-
	(0.3991)	-
Prius*%Democrat	2.0269^{***}	-
	(0.681)	-

[a] *** 1% significance, ** 5% significance, * 10% significance. Standard errors in parentheses.

Table 2.10: Average WTP for Hybrid Vehicles and the Toyota Prius (Educated; Age 50; No HOV Access)

%Democrat = 75						
	All Hybrid Vehicles			Prius[b]		
	Income			Income		
Gas	$40,000	$80,000	$120,000	$40,000	$80,000	$120,000
$3.20	-$12,133	-$11,040	-$12,795	-$1,724	$24	$1,427
	($2,658)	($2,668)	($4,310)	($3,793)	($4,010)	($6,533)
$3.60	-$8,276	-$6,940	-$7,525	$2,133	$4,124	$6,697
	($1,652)	($1,601)	($2,580)	($2,879)	($3,045)	($5,030)
$4.00	-$4,419	-$2,840	-$2,255	$5,990	$8,224	$11,967
	($1,459)	($1,453)	($2,385)	($2,779)	($3,000)	($5,061)

%Democrat = 50						
	All Hybrid Vehicles			Prius[b]		
	Income			Income		
Gas	$40,000	$80,000	$120,000	$40,000	$80,000	$120,000
$3.20	-$12,906	-$11,862	-$13,851	-$5,529	-$4,020	-$3,772
	($1,966)	($1,870)	($2,969)	($2,591)	($2,617)	($4,183)
$3.60	-$9,049	-$7,762	-$8,581	-$1,672	$80	$1,498
	($1,270)	($1,154)	($1,822)	($1,927)	($1,930)	($3,124)
$4.00	-$5,192	-$3,662	-$3,311	$2,185	$4,180	$6,768
	($1,389)	($1,357)	($2,210)	($2,078)	($2,164)	($3,601)

%Democrat = 25						
	All Hybrid Vehicles			Prius[b]		
	Income			Income		
Gas	$40,000	$80,000	$120,000	$40,000	$80,000	$120,000
$3.20	-$13,678	-$12,683	-$14,906	-$9,334	-$8,065	-$8,971
	($2,209)	($2,129)	($3,374)	($3,502)	($3,623)	($5,809)
$3.60	-$9,821	-$8,583	-$9,636	-$5,477	-$3,965	-$3,701
	($1,825)	($1,764)	($2,810)	($3,090)	($3,215)	($5,196)
$4.00	-$5,964	-$4,483	-$4,366	-$1,620	$135	$1,569
	($2,255)	($2,320)	($3,782)	($3,491)	($3,727)	($6,118)

[a] Standard errors in parentheses; constructed using the Delta Method.
[b] $WTP_{\text{Prius}} = WTP_{\text{Hybrid Vehicles}} + WTP_{\text{Prius effect}}$.

Table 2.11: WTP Hybrid Vehicles: Observable Demographics

Demographic	WTP_1	WTP_2	WTP_3
College Educated	$3,228	$3,431	$4,410
	($312)	($352)	($586)
HOV Access	$1,189	$1,264	$1,625
	($481)	($544)	($900)
Age			
20	$1,460	$1,552	$1,995
	($161)	($181)	($300)
30	$2,190	$2,328	$2,993
	($361)	($408)	($674)
40	$2,920	$3,104	$3,990
	($642)	($725)	($1,198)
50	$3,651	$3,880	$4,988
	($1,003)	($1,132)	($1,873)
60	$4,381	$4,656	$5,985
	($1,445)	($1,630)	($2,696)
70	$5,111	$5,433	$6,983
	($1,966)	($2,219)	($3,670)

[a] WTP_1, WTP_2, WTP_3 correspond to households in the first, second, and third income groups, respectively.
[b] Standard errors in parentheses; constructed using the Delta Method.

Table 2.12: WTP Hybrid Vehicles: Unobservable Demographics

Percentile	WTP_1	WTP_2	WTP_3
5%	$4,331	$4,604	$5,917
	($16,783)	($18,960)	($31,325)
25%	$1,759	$1,869	$2,403
	($2,767)	($3,126)	($5,165)
75%	-$1,759	-$1,869	-$2,403
	($2,767)	($3,126)	($5,165)
95%	-$4,331	-$4,604	-$5,917
	($16,783)	($18,960)	($31,325)

[a] Unobservable preferences are assumed to follow a Normal distribution.
[b] WTP_1, WTP_2, WTP_3 correspond to households in the first, second, and third income groups, respectively.
[c] Standard errors in parentheses; constructed using the Delta Method.

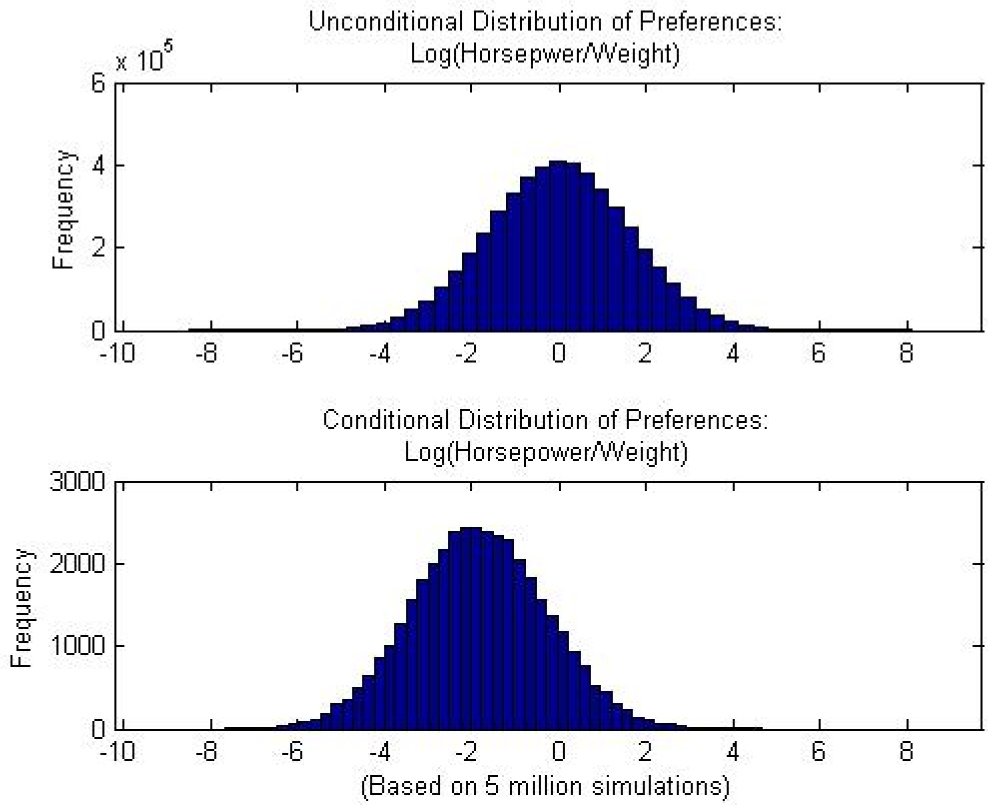

Figure 2.1: Unobservable Heterogeneous Preferences: Conditional on Purchasing a Toyota Prius

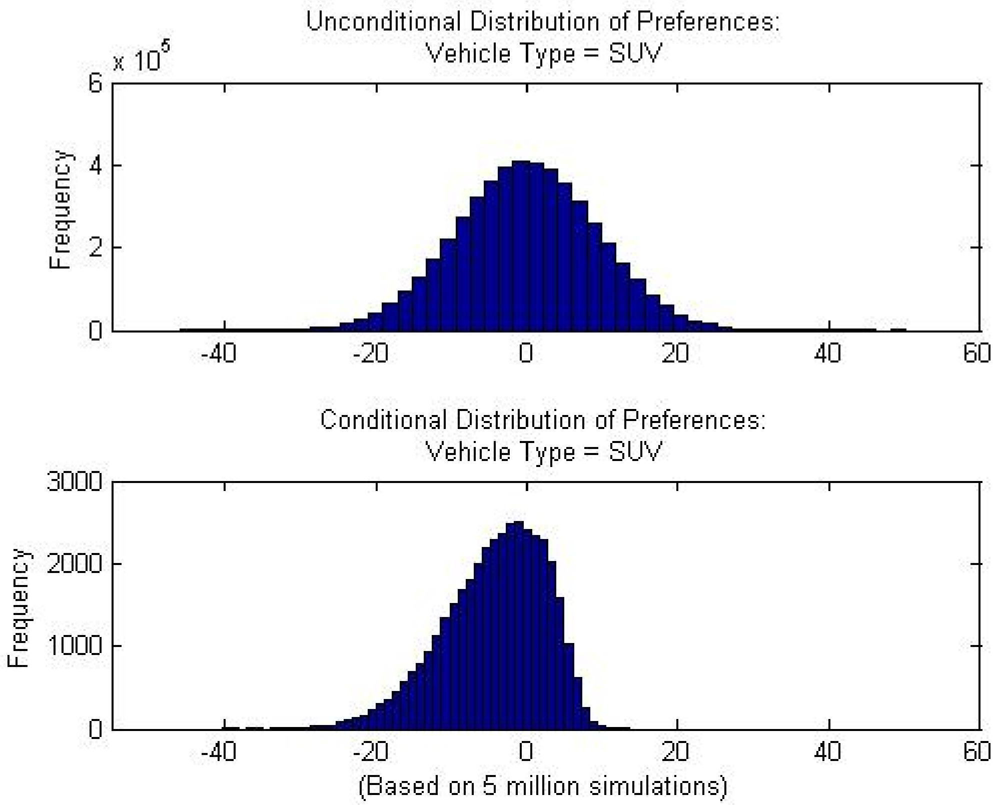

Figure 2.2: Unobservable Heterogeneous Preferences: Conditional on Purchasing a Toyota Prius

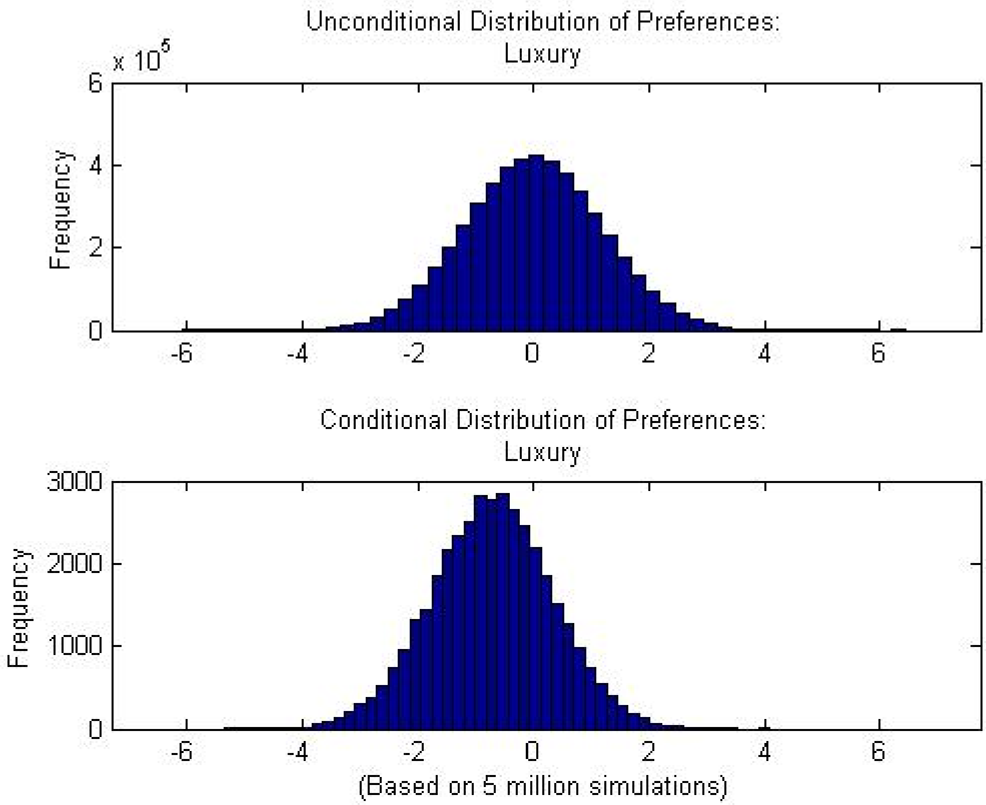

Figure 2.3: Unobservable Heterogeneous Preferences: Conditional on Purchasing a Toyota Prius

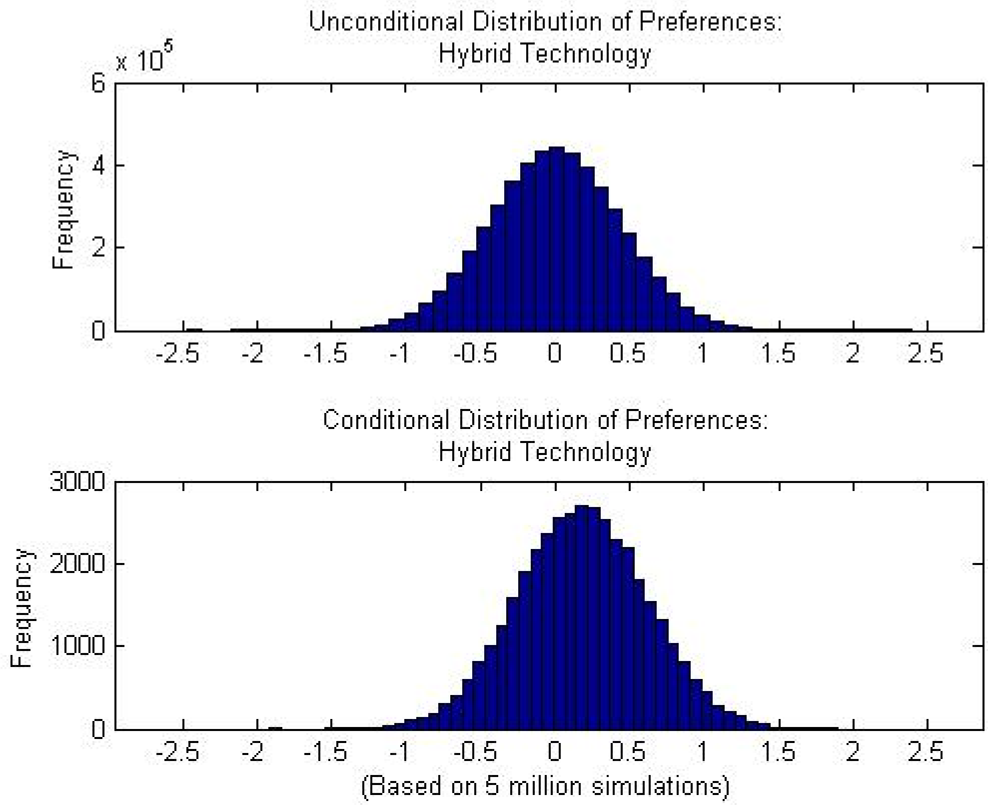

Figure 2.4: Unobservable Heterogeneous Preferences: Conditional on Purchasing a Toyota Prius

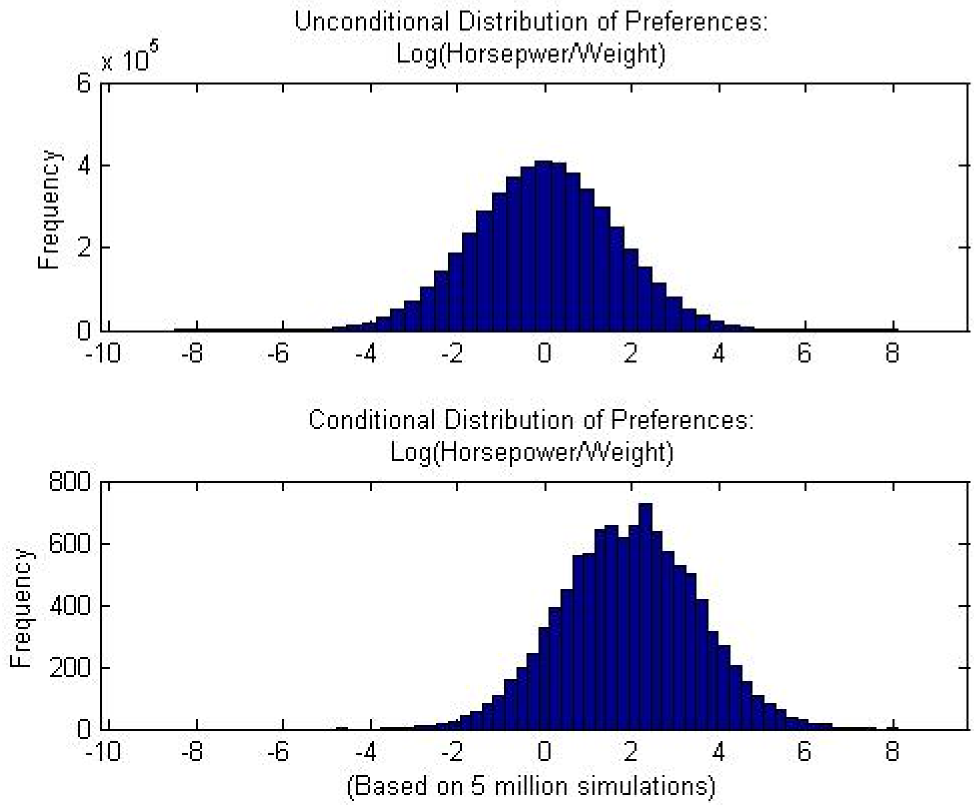

Figure 2.5: Unobservable Heterogeneous Preferences: Conditional on Purchasing a Chevrolet Corvette

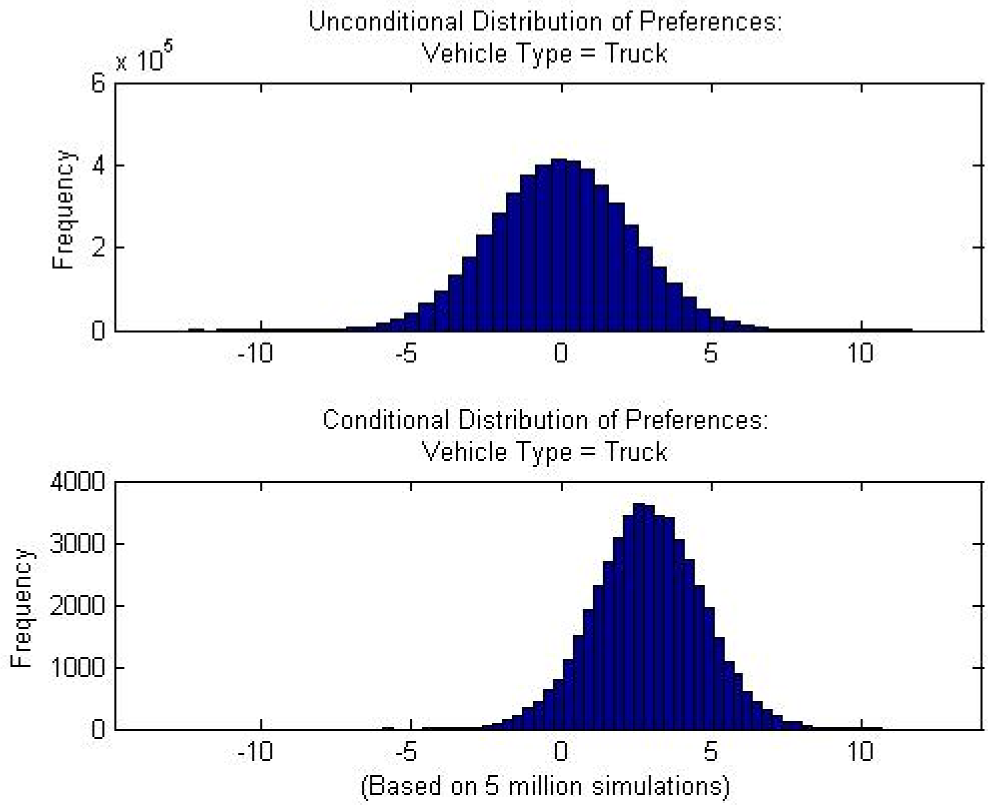

Figure 2.6: Unobservable Heterogeneous Preferences: Conditional on Purchasing a Ford F-150

Conclusion

The introduction of new goods has been a reoccurring phenomenon that has existed throughout human history. New goods are continually being created, improved, modified, and destroyed ultimately being replaced by new and better products. In monopolistically competitive markets with low barriers to entry, firms are the driving force behind this process as they seek transitory surpluses in economic profit. Firms can achieve this by either becoming the low-cost producer of a homogeneous good or differentiating their products from competitors. Consumers, as a result, benefit from this process by having new and improved products available in their choice set. However, measuring the benefits of new goods has been a long term challenge for economists primarily because the units come in dimensions other than price and quantity.

Chapter One of this dissertation examined the welfare effects from the hybrid vehicle innovation in the U.S. auto market between the years 2000 and 2008. Several structural discrete choice models are used to estimate the demand and supply curves for all vehicles. Using the parameter estimates of the models, several counterfactual analyses are conducted examining the changes in producer surplus, the extent of first mover advantages, profit cannibalization, and business stealing effects from competitors. On the consumer side, price and variety effects are estimated which help identify the benefits from a productive and allocative efficiency standpoint, respectively. Furthermore, this chapter developed a simulation technique to estimate these effects while allowing income to enter the indirect utility function nonlinearly.

Chapter Two of this dissertation took a more detailed examination at why consumers purchase hybrid vehicles. In particular, I show that in strictly financial terms, hybrid vehicles do not make sense. After factoring in vehicle depreciation, interest on financing, maintenance and repair costs, fees and taxes (including tax incentives), and fuel expenditures, I find that the majority of hybrid vehicle owners will not recover the higher premium through decreased fuel expenditures after five years of ownership. To explain why consumers continue to purchase them, a rich consumer demand model is developed using detailed household level data that explicitly links the year, make, and model of each vehicle a household owns to their specific demographic type.

The results of this dissertation suggest that, although hybrid vehicles represent a relatively small share of the overall auto market, some households are willing to pay a premium for “green" and environmentally friendly vehicles. In particular, there is strong evidence that suggests that one of the primary reasons consumers purchase hybrid vehicles is for conspicuous environmental effects; consumers not only want to be green, they want to look green as well. Toyota was the first to recognize this market niche and responded by developing the Prius. Since the introduction, Toyota has enjoyed significant first mover advantages earning substantial economic profits and sustained growth in the hybrid vehicle market segment. Toyota's success can be attributed to the strong branding and unique design of the Prius which gives consumers a vehicle with both high fuel economy and one that sends a signal to the community that they are “green" and living a low resource intensive lifestyle. Competing hybrid models merely have an inconspicuous decal on their back bumper making them virtually indistinguishable from their all gas equivalent.

The findings of this dissertation demonstrate the importance of accounting for heterogeneous preferences in consumer demand models - what one consumer likes another may dislike. However, empirically accounting for these differences is difficult for several reasons. First, economists rarely have information on all product characteristics. Second, some product characteristics may be difficult to quantify. Third, consumer demographics may only partially explain heterogeneous preferences. Fourth, empirical results depend critically on choosing an appropriate demand model and specification. This dissertation incorporated several methodologies developed in the economics literature to overcome these challenges as well as offered several contributions.

The End

REFERENCES

[1] Arizona Department of Transportation. 2011. "ADOT to release 2,500 energy-efficient vehicle plates." Last modified September 13. http://www.azdot.gov/CCPartnerships/News/NRel3041.asp .

[2] Bento, A. M., L.H. Goulder, M.R. Jacobsen, and R.H. von Haefen. 2009. "Distributional and Efficiency Impacts of Increased U.S. Gasoline Taxes." *American Economic Review* 99: 667-699.

[3] Beresteanu, A. and S. Li. 2010. "Gasoline Prices, Government Support, and the Demand for Hybrid Vehicles in the U.S." *International Economic Review* 52: 161-182.

[4] Berry, S. T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics* 25: 242-262.

[5] ——— , J. Levinsohn and A. Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63: 841-890.

[6] ——— , ——— , ——— . 2003. "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market." *Journal of Political Economy* 112: 68-105.

[7] Bresnahan, T. F. 1987. "Competition and Collusion in the American Automobile Industry: The 1995 Price War." *The Journal of Industrial Economics* 35: 457-482.

[8] ———-. 1997. "The Apple-Cinnamon Cheerios War: Valuing New Goods, Identifying Market Power, and Economic Measurement." *Stanford University Working Paper.*

[9] ———-, S. Stern, and M. Trajtenberg. 1997. "Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980's." *RAND Journal of Economics* 28: S17-S44.

[10] California Department of Motor Vehicles. 2011. "Clean Air Vehicle (CAV) Decals - High Occupancy Vehicle HOV Lane Usage." Accessed December 14. http://dmv.ca.gov/vr/decal.htm .

[11] Chamberlin, E. 1938. *The Theory of Monopolistic Competition: A re-orientation of the Theory of Value.* Cambridge: Harvard University Press.

[12] Chandra, A., S. Gulati, and M. Kandlikar. 2010. "Green Drivers or Free Riders? An Analysis of Tax Rebates for Hybrid Vehicles." *Journal of Environmental Economics and Management* 60: 78-93.

[13] Colorado Department of Transportation. 2011. "Hybrid Vehicle Use in the HOT/HOV Lanes." Last modified March 14. http://www.coloradodot.info/programs/hybrids .

[14] Diamond, D. 2009. "The Impact of Government Incentives for Hybrid-Electric Vehicles: Evidence from U.S. States." *Energy Policy* 37: 972-983.

[15] Erdem, C., I. Senturk, T. Simsek. 2010. "Identifying the Factors affecting the Willingness to Pay for Fuel-Efficient Vehicles in Turkey: A Case of Hybrids." *Energy Policy* 38: 3038-3043.

[16] Foster, V. and S. Mourato. 2000. "Valuing the Multiple Impacts of Pesticide Use in the UK: A Contingent Ranking Approach." *Journal of Agricultural Economics* 51: 1-21.

[17] Gallagher, K.S. and E. Muehlegger. 2011. "Giving Green to Get Green? Incentives and Consumer Adoption of Hybrid Vehicle Technology." *Journal of Environmental Economics and Management* 61: 1-15.

[18] Goldberg, P.K. 1995. "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry." *Econometrica* 63: 891-951.

[19] Goolsbee, A. and A. Petrin. 2004. "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV." *Econometrica* 72: 351-381.

[20] Hall, A. R. 2005. *Generalized Method of Moments*. New York: Oxford University Press.

[21] Hanemann, M.W. 1984. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses." *American Journal of Agricultural Economics* 66: 332-341.

[22] Hausman, J. A. 1997. "Valuation of New Goods under Perfect and Imperfect Competition." In *The Economics of New Goods*. Edited by T. F. Bresnahan and R. J. Gordon. Chicago: University of Chicago Press.

[23] ———. 1999. "Cellular Telephones, New Products, and the CPI." *Journal of Business and Economic Statistics* 17: 68-74.

[24] ——— and G. K. Leonard. 2002. "The Competitive Effects of a New Product Introduction: A Case Study." *The Journal of Industrial Economics* 50: 237-263.

[25] ———, A. Pakes, and G. L. Rosston. 1997. "Valuing the Effect of Regulation on New Services in Telecommunications." *Brookings Paper on Economic Activity. Microeconomics* 1997: 1-54.

[26] Hensher, D. A. and W. H. Greene. 2003. "The Mixed Logit Model: The State of Practice." *Transportation* 30: 133-176.

[27] Herriges, J.A., and C.L. Kling. 1999. “Nonlinear Income Effects in Random Utility Models." *The Review of Economics and Statistics* 81: 62-72.

[28] Hicks, J. R. 1940. “The Valuation of Social Income." *Economica* 7: 105-124.

[29] HybridCars.com. 2010. "Carpool (HOV) Lanes." Last modified July 8. http://www.hybridcars.com/local-incentives/carpool-hov-lanes.html .

[30] Imbens, G. W. and T. Lancaster. 1994. “Combining Micro and Macro Data in Microeconometric Models." *Review of Economic Studies* 61: 655-680.

[31] J.D. Power and Associates. 2008. “While Many New-Vehicle Buyers Express Concern for the Environment, Few are Willing to Pay More for An Environmentally Friendly Vehicle." Last modified March 6. http://businesscenter.jdpower.com/news/pressrelease.aspx?ID=2008029 .

[32] Kahn, M.E. 2007. “Do Greens Drive Hummers or Hybrids? Environmental Ideology as a Determinant of Consumer Choice." *Journal of Environmental Economics and Management* 54: 129-145.

[33] Knittle, C.R., and K. Metaxoglou. 2011. “Estimation of Random Coefficient Demand Models: Challenges, Difficulties and Warnings." *MIT University Working Paper.*

[34] Lancaster, K. J. 1966. “A New Approach to Consumer Theory." *Journal of Political Economy* 74: 132-157.

[35] ———. 1979. *Variety, Equity, and Efficiency.* New York: Columbia University Press.

[36] ———. 1990. “The Economics of Product Variety: A Survey." *Marketing Science* 9: 189-206.

[37] Lareau, T. J. and D. A. Rae. 1989. “Valuing WTP for Diesel Odor Reductions: An Application of Contingent Ranking Technique." *Southern Economic Journal* 55: 728-742.

[38] Mannering, F. and C. Winston. 1995. “Automobile Air Bags in the 1990s: Market Failure or Market Efficiency?" *Journal of Law and Economics* 38: 265-279.

[39] ———, ———, Z. Griliches, and R. Schmalensee. 1991. “Brand Loyalty and the Decline of American Automobile Firms." *Brookings Institute: Brookings Paper on Economic Activity. Microeconomics* 1991: 67-114.

[40] Marschak, J. 1960. “Binary Choice Constraints on Random Utility Indications." In K. Arrow *Standford Symposium on Mathematical Mothods in the Social Sciences.* Standford: Standford University Press.

[41] McFadden, D. 1980. “Econometric Models for Probabilistic Choices among Products." *Journal of Business* 53: s13- s29..

[42] ———. 2001. “Economic Choices." *The American Economic Review* 91: 351-378.

[43] Morey, E. R., V.R. Sharma, and A. Karlstrom. 2003. “A Simple Method of Incorporating Income Effects into Logit and Nested-Logit Models: Theory and Application." *American Journal of Agricultural Economics* 85: 248-253.

[44] Nahuelhual, L., M. L. Loureiro, J. Loomis. 2004. “Using Random Parameters to Account for Heterogeneous Preferences in Contingent Valuation of Public Open Space." *Journal of Agricultural and Resource Economics* 29: 537-552.

[45] Nevo, A. 2000. “A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand." *Journal of Economics and Management* 9: 513-548.

[46] ———. 2001. “Measuring Market Power In The Ready-To-Eat Cereal Industry." *Econometrica* 69: 307-342.

[47] Petrin, A. 2001. “Quantifying The Benefits of New Products: The Case of The Minivan." *NBER Working Paper Series # 8227.*

[48] ———. 2002. “Quantifying the Benefits of New Products: The Case of the Minivan. " *Journal of Political Economy* 110: 705-729.

[49] Potoglou D. and P. Kanaroglou. 2007. “Household Demand and Willingness to Pay for Clean Vehicles." *Transportation Research Part D* 12: 264-274.

[50] Rasmusen, E. 2007. “The BLP Method of Demand Curve Estimation in Industrial Organization. " *Indiana University Working Paper.*

[51] Revelt, D. and K. Train. 1998. “Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level." *The Review of Economics and Statistics* 80: 647-657.

[52] ——— and ———. 2000. “Customer-Specific Taste Parameters and Mixed Logit: Households' Choice of Electricity Supplier." *University of California at Berkley, Working Paper.*

[53] Ridley, M. 2010. *The Rational Optimist.* New York: Harper.

[54] Romer, P. 1994. “New Goods, Old Theory, and the Welfare Cost of Trade Restriction." *Journal of Development Economics* 43: 5-38.

[55] Sangkapichai, M. and J.D. Saphores. 2009. “Why are Californians Interested in Hybrid Cars?" *Journal of Environmental Planning and Management* 52: 79-96.

[56] Scarborough Research. 2007. "Hybrid Vehicle Owners are Wealthy, Active, Educated and Overwhelmingly Democratic, According to Scarborough Research" Last modified December 4. http://scarborough.com/press_releases/Scarborough-Hybrid-Vehicle-Owner-Consumer-Profile.pdf .

[57] Schumpeter, J. A. 1941. *Capitalism, Socialism, and Democracy.* New York: Harper & Brothers.

[58] Sexton, S.E. and A.L. Sexton. 2011. "Conspicuous Conservation: The Prius Effect and Willingness to Pay for Environmental Bona Fides." *University of California at Berkley, Working Paper.*

[59] Simon, H. A. 1978. "Rationality as Process and as Product of Thought." *The American Economic Review* 68: 1-16.

[60] Small, K. A. and H. S. Rosen. 1981. "Applied Welfare Economics with Discrete Choice Models." *Econometrica* 49: 105-130.

[61] Spector, M. 2007. "The Economics of Hybrids." *The Wall Street Journal,* October 29.

[62] Spence, M. 1976a. "Product Differentiation and Welfare." *The American Economic Review* 66: 407-414.

[63] ———-. 1976b. "Product Selection, Fixed Costs, and Monopolistic Competition." *The Review of Economic Studies* 43: 217-235.

[64] Stigler, G. J. and G. S. Becker. 1997. "De Gustibus Non Est Disputandum." *The American Economic Review* 67: 76-90.

[65] Sudhir, K. 2001. "Competitive Pricing Behavior in the Auto Market: A Structural Analysis." *Marketing Science* 20: 42-60.

[66] Swait, J. and J. Louviere. 1993. "The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models." *Journal of Market Research* 30: 305-314.

[67] "The Dollars and Sense of Hybrids." 2006. *Consumer Reports,* April.

[68] Thurstone, L. L. 1927. "A Law of Comparative Judgement." *Psychological Review* 34: 273-286.

[69] Train, K. E. 2003. *Discrete Choice Methods with Simulation.* New York: Cambridge University Press.

[70] ———-. and C. Winston. 2007. "Vehicle Choice Behavior and The Declining Market Share of U.S. Automakers." *International Economic Review* 48: 1469-1496.

[71] Trajtenberg, M. 1989. "The Welfare Analysis of Product Innovations, with an Application to Computer Tomography Scanners." *Journal of Political Economy* 97: 444-479.

[72] Turrentine, T.S., and K.S. Kurani. 2007. "Car Buyers and Fuel Economy." *Energy Policy* 35: 1213-1223.

[73] USA Today. 2008. "Behind the Results: 2008 Presidential Coverage." Last modified November 5. http://www.usatoday.com/news/politics/election2008/results.htm?loc=interstitialskip .

[74] Utah Department of Transportation. 2011. "C Decal Permit." Accessed December 14. http://www.udot.utah.gov/main/f?p=100:pg:0:::1:T,V:2280 .

[75] United States Environmental Protection Agency: Office of Transportation and Air Quality. 2007. "Proposed Rule for High Occupancy Vehicle (HOV) Facility Exemptions for Low Emission and Energy-Efficient Vehicles (EPA420-F-07-042)." May. Accessed December 14, 2011. http://www.epa.gov/oms/regs/ld-hwy/hov/420f07042.pdf .

[76] United State Environmental Protection Agency (EPA). 2010. "Updated Draft List of Eligible Low Emission and Energy-Efficient Vehicles Using the Hybrid-to-Gasoline Vehicle Comparison Methodology." June 2. Accessed December 14, 2011. http://www.epa.gov/oms/regs/ld-hwy/hov/420d10002.pdf .

[77] United States Department of Transportation: Federal Highway Administration. 2008. "Federal-Aid Highway Program Guidance on High Occupancy Vehicle (HOV) Lanes." August. Accessed December 14, 2011. http://www.ops.fhwa.dot.gov/freewaymgmt/hovguidance/index.htm .

[78] United States Department of Transportation: Federal Highway Administration. 2011. "SAFETEA-LU: Safe, Accountable, Flexible, Efficient Transportation Equity Act: A legacy for Users." Accessed December 14. http://www.fhwa.dot.gov/safetealu/legis.htm .

[79] Virginia Department of Transportation. 2011. "Occupancy Vehicle (HOV) Lanes - Rules and FAQs." Last modified September 30. http://www.virginiadot.org/travel/hov-rulesfaq.asp .

[80] Viton, P. A. 1985. "On the Interpretation of Income Variables in Discrete-Choice Models." *Economic Letters* 17: 203-206.

[81] Winston, C. and F. Mannering. 1984. "Consumer Demand for Automobile Safety." *The American Economic Review* 74: 316-319.

[82] Xiao, W. 2008. "The Competitive and Welfare Effects of New Product Introduction: The Case of Crystal Pepsi." *Johns Hopkins University Job Market Paper.*

APPENDICES

Appendix A

Micro-Moment Methodology

A.1 Introduction

This appendix explains how to incorporate micro-moments into the BLP (1995) methodology following the work of Petrin (2002). This idea stems from Imbens and Lancaster (1994) who were the first to combine both micro and macro data sets. In what follows below, I tie all three of the above papers together in a unifying manner to help other researchers use this method in the future. I assume the reader has read each paper.

A.1.1 Why Add Micro-Moments?

The BLP methodology relies on aggregate demographic distributions to identify heterogeneous taste parameters. No data exists between a consumer's purchase decision and their demographic type. Thus only a correlation between product choice and demographics can be inferred which leads to imprecise parameter estimates. Petrin's methodology helps correct for this by incorporating information from micro-level data which directly links a consumer's purchase to their demographic type. This requires the use of Bayes Law, properties of conditional probabilities and the conditional Logit choice probability equation.

A.2 Constructing the Micro-Moment

Additional information provided by a supplemental micro data set can be incorporated into a BLP model by constructing micro-moments in addition to the standard BLP moments. The micro-moments are based on transformations of the share equation forming probabilities (expectations) of demographic variables conditional on a household's new vehicle choice. The model's predicted probability is then matched to the "true" probability based on the micro data. For example, the probability that a consumer has an annual income of over $100,000 given that they purchased a new hybrid vehicle could be constructed from the micro data. To link this to the BLP model, the share equation must be transformed to yield the same conditional probability using Bayes Law,

$$P(A|B;\theta) = \frac{P(B|A;\theta)P(A)}{P(B)},$$

where A represents the demographical variable of interest (e.g. income, age, education, etc.) and B represents the vehicle type (e.g. hybrid, truck, van, SUV, etc.).

Next, the parameters are adjusted such that the distance between the micro level conditional probability and the model's predicted conditional probability is as small as possible,

$$\min_{\theta} ||E\left[I_{\text{micro}}\{A|B\} - \bar{P}_{\text{model}}(A|B;\theta)\right]||.$$

Note that the model's predicted conditional probability is a function of the parameters to be estimated while the micro conditional probability is not. The objective function for this expression is,

$$Q_{\text{micro}}(\theta) = \min_{\theta} \psi'_{\text{micro}}(\theta) W \psi_{\text{micro}}(\theta),$$

where W is the weight matrix and $\psi_{\text{micro}}(\theta)$ is the set of micro-moment conditions,

$$\begin{aligned}\psi_{\text{micro}}(\theta) &= E\left[I^i_{\text{micro}}\{A|B\} - \bar{P}_{\text{model}}(A|B;\theta)\right] \\ &= \frac{1}{N}\sum_{i=1}^{N}\left(I^i_{\text{micro}}\{A|B\} - \bar{P}_{\text{model}}(A|B;\theta)\right),\end{aligned}$$

where $I^i_{\text{micro}}\{A|B\}$ is an indicator variable constructed from the micro data, and N is the number of observations in the micro data set.

The weight matrix is perhaps the most difficult part of the micro-moment methodology. Before explaining how to calculate the weight matrix in detail, the three sources of randomness that enter it are,[1]

$$W = \rho * \left(\Delta_g + \Delta_h + \Delta_{ns}\right)^{-1},$$

where each element is independent and can therefore be added.

The first source of variance, Δ_g, is the variance-covariance of the micro-moment prediction errors,

$$\begin{aligned}
\Delta_g &= \psi'\psi \\
&= E\left[\left(I_{\text{micro}}\{A|B\} - \bar{P}_{\text{model}}(A|B;\theta)\right)^2\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(I^i_{\text{micro}}\{A|B\} - \bar{P}_{\text{model}}(A|B;\theta)\right)^2.
\end{aligned}$$

The next source of variance, Δ_h, is the uncertainty associated with the conditional means constructed from the micro data,

$$\begin{aligned}
\Delta_h &= E\left[\left(I_{\text{micro}}\{A|B\} - \bar{I}_{\text{micro}}\{A|B\}\right)^2\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(I^i_{\text{micro}}\{A|B\} - \bar{I}_{\text{micro}}\{A|B\}\right)^2.
\end{aligned}$$

Finally, the last source of variance, Δ_{ns}, that enters the weight matrix is the simulation variance generated from using ns simulations to reprent all U.S. households,

$$\begin{aligned}
\Delta_{ns} &= E\left[\left[E\left(\bar{I}_{\text{micro}}\{A|B\} - \bar{P}_{\text{model}}(A|B;\theta)\right)\right]^2\right] \\
&= \frac{1}{R}\sum_{r=1}^{R}\left[\frac{1}{ns}\sum_{i=1}^{ns}\left(\bar{I}_{\text{micro}}\{A|B\} - \bar{P}^r_{\text{i, model}}(A|B;\theta)\right)\right]^2 \\
&= \frac{1}{R}\sum_{r=1}^{R}(\hat{\zeta}^r)^2,
\end{aligned}$$

where R is equal to the number of sets of ns simulated individuals, $(ns^1, ns^2, ..., ns^R)$.

[1]The ρ that enters the weight matrix is used to up/down weight the variance according to how large/small the sample size of the micro data set is in relation to the main data set. Here, $\rho = J/N$, where J is the number of products and N is the sample size of the micro data.

A.3 Conditional Probability: Model's Prediction

In this section I explain how the conditional means for the model are constructed. For expositional purposes, I will refer to the following four micro-moments throughout the remainder of this appendix:

$$\psi_1(\theta) = E\left[I_{\text{micro}}\{A_1|B_1\} - \bar{P}_{\text{model}}(A_1|B_1;\theta)\right]$$
$$\psi_2(\theta) = E\left[I_{\text{micro}}\{A_2|B_1\} - \bar{P}_{\text{model}}(A_2|B_1;\theta)\right]$$
$$\psi_3(\theta) = E\left[I_{\text{micro}}\{A_1|B_2\} - \bar{P}_{\text{model}}(A_1|B_2;\theta)\right]$$
$$\psi_4(\theta) = E\left[I_{\text{micro}}\{A_2|B_2\} - \bar{P}_{\text{model}}(A_2|B_2;\theta)\right],$$

where B_1 and B_2 represent an individual purchasing a hybrid and truck, respectively, and A_1 and A_2 represent an individual with incomes ($\leq \$50,000$) and ($\$50,000 < \text{income} \leq \$100,000$), respectively.

A.3.1 Transforming The Share Equation

The probability a person of type $i \in ns$ purchases product j is,

$$\underbrace{P_{i,j}(X_j|D_i, v_i;\theta_1,\theta_2)}_{\text{J x ns}} = \frac{\exp\left(\delta_j(\theta_1) + \mu_{i,j}(\theta_2)\right)}{\sum_{r=1}^{J+1}\exp\left(\delta_r(\theta_1) + \mu_{i,r}(\theta_2)\right)}.$$

The probability a person of type $i \in ns$ purchases a particular product type (e.g. $B_1 = \text{Hybrid}$) is found by summing all probabilities with products in the subset $j \subset B_1$,

$$\underbrace{P_{i,B_1}(X_{B_1}, D_i, v_i;\theta_1,\theta_2)}_{\text{1 x ns}} = \sum_{\substack{j=1 \\ j\subset B_1}}^{J} \underbrace{P_{i,j}(X_j, D_i, v_i;\theta_1,\theta_2)}_{\text{J x ns}}.$$

Therefore the model's predicted average is,

$$\underbrace{P(B_1;\theta)}_{\text{1 x 1}} = \frac{1}{ns}\sum_{i=1}^{ns} P_{i,B_1}(X_{B_1}, D_i, v_i;\theta_1,\theta_2),$$

which is interpreted as "the probability that any person purchases vehicle type B_1."

Conditional probabilities are averaged over a subset of individuals, such that $ns_1 \leq$

ns, that fit a particular demographic type (e.g. A_1 = income $\leq$ \$50,000). Therefore, the conditional probability vector will be,

$$P(B_1|A_1;\theta) = \frac{1}{ns_1}\sum_{i=1}^{ns_1} P_i(B_1|A_1;\theta),$$

which is interpreted as the "the probability a person of type A_1 purchases vehicle type B_1."

Similarly, the conditional probability that individual i purchases vehicle type B_1 given that they are of demographic type A_2 (= \$50,000 <income $\leq$ \$100,000) is,

$$P(B_1|A_2;\theta) = \frac{1}{ns_2}\sum_{i=1}^{ns_2} P_i(B_1|A_2;\theta).$$

The same logic can be used to calculate the conditional probabilities for vehicle type B_2.

As of now, the model above predicts the probability an individual will purchase a particular vehicle type given their demographic type (e.g. $P(B_1|A_1;\theta)$ and $P(B_1|A_2;\theta)$). The micro-moments, however, require the inverse of this, namely

$$\begin{aligned}
&P(A_1|B_1;\theta)\\
&P(A_2|B_1;\theta)\\
&P(A_1|B_2;\theta)\\
&P(A_2|B_2;\theta).
\end{aligned}$$

Hence the use of Bayes Law,

$$\begin{aligned}
P_i(A_1|B_1;\theta) &= \frac{P(B_1|A_1;\theta)P(A_1)}{P(B_1)}\\
P_i(A_2|B_1;\theta) &= \frac{P(B_1|A_2;\theta)P(A_2)}{P(B_1)}\\
P_i(A_1|B_2;\theta) &= \frac{P(B_2|A_1;\theta)P(A_1)}{P(B_2)}\\
P_i(A_2|B_2;\theta) &= \frac{P(B_2|A_2;\theta)P(A_2)}{P(B_2)},
\end{aligned}$$

where $P(A_1)$ and $P(A_2)$ are directly estimated from the micro data, corresponding to the probability that any household is of demographic type A_1 and A_2, respectively. Similarly, $P(B_1)$ and $P(B_2)$ are directly observed from the aggregate level data representing the probability (market share) for vehicle types B_1 and B_2.

A.3.2 Weight Matrix: Micro-Moment Variance-Covariance

Using the micro-moments defined above for all N observations, the variance-covariance matrix can be easily calculated. For simplicity, let

$$\begin{aligned} \bar{P}_1 = &\quad P_{\text{model}}(A_1|B_1;\theta) \\ \bar{P}_2 = &\quad P_{\text{model}}(A_2|B_1;\theta) \\ \bar{P}_3 = &\quad P_{\text{model}}(A_1|B_2;\theta) \\ \bar{P}_4 = &\quad P_{\text{model}}(A_2|B_2;\theta) \end{aligned} \qquad \text{and} \qquad \begin{aligned} I_1^i = &\quad I^i_{\text{micro}}\{A_1|B_1\} \\ I_2^i = &\quad I^i_{\text{micro}}\{A_2|B_1\} \\ I_3^i = &\quad I^i_{\text{micro}}\{A_1|B_2\} \\ I_4^i = &\quad I^i_{\text{micro}}\{A_2|B_2\}. \end{aligned}$$

Then,

$$\Delta_g = \psi' * \psi$$
$$\Longrightarrow$$

$$\frac{1}{N}\begin{bmatrix} \sum_i^N (I_1^i - P_1^*)^2 & \sum_i^N (I_1^i - P_1^*) * (I_2^i - P_2^*) & \sum_i^N (I_1^i - P_1^*) * (I_3^i - P_3^*) & \sum_i^N (I_1^i - P_1^*) * (I_4^i - P_4^*) \\ \sum_i^N (I_2^i - P_2^*) * (I_1^i - P_1^*) & \sum_i^N (I_2^i - P_2^*)^2 & \sum_i^N (I_2^i - P_2^*) * (I_3^i - P_3^*) & \sum_i^N (I_2^i - P_2^*) * (I_4^i - P_4^*) \\ \sum_i^N (I_3^i - P_3^*) * (I_1^i - P_1^*) & \sum_i^N (I_3^i - P_3^*) * (I_2^i - P_2^*) & \sum_i^N (I_3^i - P_3^*)^2 & \sum_i^N (I_3^i - P_3^*) * (I_4^i - P_4^*) \\ \sum_i^N (I_4^i - P_4^*) * (I_1^i - P_1^*) & \sum_i^N (I_4^i - P_4^*) * (I_2^i - P_2^*) & \sum_i^N (I_4^i - P_4^*) * (P_3^i - P_3^*) & \sum_i^N (I_4^i - P_4^*)^2 \end{bmatrix}.$$

However, due to differences in sample sizes between the conditional micro probabilities, the indicator random variables above, I, may need to be transformed such that all elements are based on the same size sample, N. The next section explains how to do this.

A.4 Conditional Probability: Micro Data

Constructing the conditional means based on the micro data is relatively easy. The difficulty lies in constructing the variance-covariance associated with the sampling error, Δ_h, since the conditional means are most likely based on different size sub-samples. For example, the number the households in the micro data that own a vehicle of type B_1, is likely to be different than the number of households who own a vehicle of type B_2,

$N =$ size of micro data

N_1 = number of households in micro data who purchased vehicle type B_1
N_2 = number of households in micro data who purchased vehicle type B_2
$N_1 \leq N$, $N_2 \leq N$.

The conditional probabilities are the following:

$$\bar{I}_{\text{micro}}\{A_1|B_1\} = \frac{1}{N_1}\sum_{i=1}^{N_1} I_i\{A_1|B_1\}$$
$$\bar{I}_{\text{micro}}\{A_2|B_1\} = \frac{1}{N_1}\sum_{i=1}^{N_1} I_i\{A_2|B_1\}$$
$$\bar{I}_{\text{micro}}\{A_1|B_2\} = \frac{1}{N_2}\sum_{i=1}^{N_2} I_i\{A_1|B_2\}$$
$$\bar{I}_{\text{micro}}\{A_2|B_2\} = \frac{1}{N_2}\sum_{i=1}^{N_2} I_i\{A_2|B_2\},$$

where $I\{\cdot\}$ is an indicator function equal to 1 if observation i equals A_1 (or A_2) conditional on B_1 (or B_2).

A.4.1 Weight Matrix: Sample Variance

Since the micro conditional means are based on a sample of the population, there is uncertainty that needs to be accounted for. For example, we would be much more confident in our estimates if the sample size were, say, $N = 1,000,000$ versus a sample size only $N = 50$ observations. Especially since the conditional probabilities are based on a fraction of the full sample ($N_1 \leq N$). As a result, micro-moments based on conditional micro probabilities with high levels of uncertainty need to be given less weight, and vice versa.

To overcome this problem, new indicator random variables must be constructed and the Delta Method must be used to estimate the variance-covariance matrix. Consider the previous example with 4 micro-moments based on the 4 conditional probabilities,

$$\bar{I}_{\text{micro}}\{A_1|B_1\} = \frac{1}{N_1}\sum_{i=1}^{N_1} I_i\{A_1|B_1\}$$
$$\bar{I}_{\text{micro}}\{A_2|B_1\} = \frac{1}{N_1}\sum_{i=1}^{N_1} I_i\{A_2|B_1\}$$
$$\bar{I}_{\text{micro}}\{A_1|B_2\} = \frac{1}{N_2}\sum_{i=1}^{N_1} I_i\{A_1|B_2\}$$
$$\bar{I}_{\text{micro}}\{A_2|B_2\} = \frac{1}{N_2}\sum_{i=1}^{N_2} I_i\{A_2|B_2\}.$$

The Central Limit Theorem (CLT) tells us that as the sample size grows, the estimates $\bar{I}\{\cdot\}$ will converge in distribution to the true value $I^*\{\cdot\}$,

$$\sqrt{N}\begin{bmatrix} \bar{I}_{\text{micro}}\{A_1|B_1\} - I^*_{\text{micro}}\{A_1|B_1\} \\ \\ \bar{I}_{\text{micro}}\{A_2|B_1\} - I^*_{\text{micro}}\{A_2|B_1\} \\ \\ \bar{I}_{\text{micro}}\{A_1|B_2\} - I^*_{\text{micro}}\{A_1|B_2\} \\ \\ \bar{I}_{\text{micro}}\{A_2|B_2\} - I^*_{\text{micro}}\{A_2|B_2\} \end{bmatrix} \to N(0,\Delta_h),$$

where Δ_h is the variance-covariance matrix.

However, since the sample sizes used to form the conditional means are most likely different (i.e. $N_1 \neq N_2$), the variance-covariance matrix can not be calculated the usual way. To see this more clearly, consider the following two conditionals means (for simplicity),[2]

$$\underbrace{\begin{aligned} \bar{I}_1 =& \quad \bar{I}_{\text{micro}}\{A_1|B_1\} \\ \bar{I}_2 =& \quad \bar{I}_{\text{micro}}\{A_1|B_2\} \end{aligned}}_{\text{estimate}} \qquad \text{and} \qquad \underbrace{\begin{aligned} I^*_1 =& \quad I^*_{\text{micro}}\{A_1|B_1\} \\ I^*_2 =& \quad I^*_{\text{micro}}\{A_1|B_2\} \end{aligned}}_{\text{truth}}.$$

The variance would therefore take the following form,

[2] A (2 x 2) matrix is much easier to explain compared to a (4 x 4).

$$\Delta_h = E\left(\begin{bmatrix} (\bar{I}_1 - I_1^*) \\ \\ (\bar{I}_2 - I_1^*) \end{bmatrix} \begin{bmatrix} (\bar{I}_1 - I_1^*) & (\bar{I}_2 - I_2^*) \end{bmatrix} \right)$$

$$= E\left(\begin{bmatrix} (\bar{I}_1 - I_1^*)^2 & (\bar{I}_1 - I_1^*)(\bar{I}_2 - I_2^*) \\ \\ (\bar{I}_2 - I_2^*)(\bar{I}_1 - I_1^*) & (\bar{I}_2 - I_2^*)^2 \end{bmatrix} \right)$$

$$= \begin{bmatrix} \frac{1}{N_1}\sum_{i=1}^{N_1}(I_1^i - I_1^*)^2 & \overbrace{\frac{1}{?}\sum_{i=1}^{?}(I_1^i - I_1^*)(I_2^i - I_2^*)}^{\text{different sizes}} \\ \underbrace{\frac{1}{?}\sum_{i=1}^{?}(I_2^i - I_2^*)(I_1^i - I_1^*)}_{\text{different sizes}} & \frac{1}{N_2}\sum_{i=1}^{N_2}(I_2^i - I_2^*)^2 \end{bmatrix}.$$

As you can see, the variances (diagonal elements) can be calculated the normal way as the sample sizes are the same. The off-diagonal elements, on the other hand, can not be calculated because the conditional probabilities are based on different size samples. To get around this problem, new indicator functions must be constructed.

A.4.2 Indicator Random Variables

To get around the problem of different sample sizes, 4 new random variables are created all based on the same size sample, N. Recall from statistics that,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Using this, indicator random variables are constructed that are equivalent to the conditional probabilities above,

$$
\begin{aligned}
\bar{I}_{\text{micro}}\{A_1|B_1\} \equiv \bar{I}_{(A_1|B_1)} = \frac{\bar{I}_{(A_1 \cap B_1)}}{\bar{I}_{B_1}} \\
\bar{I}_{\text{micro}}\{A_1|B_2\} \equiv \bar{I}_{(A_1|B_2)} = \frac{\bar{I}_{(A_1 \cap B_2)}}{\bar{I}_{B_2}},
\end{aligned}
$$

where,

$$
\begin{aligned}
I^i_{(A_1 \cap B_1)} &= 1, && \text{if i purchases } B_1 \textit{ and } \text{is of demographic type } A_1 \\
&= 0, && \text{otherwise} \\
I^i_{(A_1 \cap B_2)} &= 1, && \text{if i purchases } B_2 \textit{ and } \text{is of demographic type } A_1 \\
&= 0, && \text{otherwise} \\
I^i_{(B_1)} &= 1, && \text{if i purchases } B_1 \\
&= 0, && \text{otherwise} \\
I^i_{(B_2)} &= 1, && \text{if i purchases } B_2 \\
&= 0, && \text{otherwise,}
\end{aligned}
$$

and

$$
\begin{aligned}
\bar{I}_{(A_1 \cap B_1)} &= \frac{1}{N}\sum_{i=1}^{N} I^i_{(A_1 \cap B_1)} \\
\bar{I}_{(A_1 \cap B_2)} &= \frac{1}{N}\sum_{i=1}^{N} I^i_{(A_1 \cap B_2)} \\
\bar{I}_{(B_1)} &= \frac{1}{N}\sum_{i=1}^{N} I^i_{(B_1)} \\
\bar{I}_{(B_2)} &= \frac{1}{N}\sum_{i=1}^{N} I^i_{(B_2)},
\end{aligned}
$$

for $i = 1, .., N$. Therefore all indicator random variables are of the same length. The Central Limit Theorem (CLT) tells us,

$$\sqrt{N}\left(\begin{bmatrix}\bar{I}_{(A_1\cap B_1)}\\ \bar{I}_{(A_1\cap B_2)}\\ \bar{I}_{(B_1)}\\ \bar{I}_{(B_2)}\end{bmatrix} - \begin{bmatrix}I^*_{(A_1\cap B_1)}\\ I^*_{(A_1\cap B_2)}\\ I^*_{(B_1)}\\ I^*_{(B_2)}\end{bmatrix}\right) \xrightarrow{d} N(0, V_I),$$

where V_I is the variance-covariance matrix (i.e. uncertainty). In interest of saving space, let,

$$\begin{aligned}I_1^i &= I^i_{(A_1\cap B_1)}\\ I_2^i &= I^i_{(A_1\cap B_2)}\\ I_3^i &= I^i_{(B_1)}\\ I_4^i &= I^i_{(B_2)}.\end{aligned}$$

Then,

$$V_I = \frac{1}{N}\begin{bmatrix}\sum_i^N (I_1^i - \bar{I}_1)^2 & \sum_i^N (I_1^i - \bar{I}_1)*(I_2^i - \bar{I}_2) & \sum_i^N (I_1^i - \bar{I}_1)*(I_3^i - \bar{I}_3) & \sum_i^N (I_1^i - \bar{I}_1)*(I_4^i - \bar{I}_4)\\ \sum_i^N (I_2^i - \bar{I}_2)*(I_1^i - \bar{I}_1) & \sum_i^N (I_2^i - \bar{I}_2)^2 & \sum_i^N (I_2^i - \bar{I}_2)*(I_3^i - \bar{I}_3) & \sum_i^N (I_2^i - \bar{I}_2)*(I_4^i - \bar{I}_4)\\ \sum_i^N (I_3^i - \bar{I}_3)*(I_1^i - \bar{I}_1) & \sum_i^N (I_3^i - \bar{I}_3)*(I_2^i - \bar{I}_2) & \sum_i^N (I_3^i - \bar{I}_3)^2 & \sum_i^N (I_3^i - \bar{I}_3)*(I_4^i - \bar{I}_4)\\ \sum_i^N (I_4^i - \bar{I}_4)*(I_1^i - \bar{I}_1) & \sum_i^N (I_4^i - \bar{I}_4)*(I_2^i - \bar{I}_2) & \sum_i^N (I_4^i - \bar{I}_4)*(I_3^i - \bar{I}_3) & \sum_i^N (I_4^i - \bar{I}_4)^2\end{bmatrix}.$$

Now all off-diagonal elements can be calculated as they are all based on the same number of observations.

A.4.3 Delta Method

If we know the variance of a random variable, the Delta Method allows us to calculate the variance of a function of the random variable. In our case, the Delta Method is defined as follows,

$$\sqrt{N}(\bar{I} - I^*) \xrightarrow{d} N(0, V_I)$$

$\Longrightarrow$

$$\sqrt{N}\big(h(\bar{I}) - h(I^*)\big) \xrightarrow{d} N\Big(0, \nabla h'(\bar{I}) V_I \nabla h(\bar{I})\Big),$$

where $h(\bar{I})$ is a function of the transformed random variables and $\nabla h(\bar{I})$ is the gradient of the transformed random variables. In this case,

$$h(\bar{I}_1, \bar{I}_2, \bar{I}_3, \bar{I}_4) = \begin{bmatrix} \bar{I}_1/\bar{I}_3 \\ \\ \bar{I}_2/\bar{I}_4 \end{bmatrix} = \begin{bmatrix} \frac{\bar{I}_{(A_1 \cap B_1)}}{\bar{I}_{B_1}} \\ \\ \frac{\bar{I}_{(A_1 \cap B_2)}}{\bar{I}_{B_2}} \end{bmatrix} = \begin{bmatrix} \bar{I}_{(A_1|B_1)} \\ \\ \bar{I}_{(A_1|B_2)} \end{bmatrix},$$

and,

$$\nabla h(\bar{I}) = < \frac{\partial h(\cdot)}{\partial \bar{I}_1}, \frac{\partial h(\cdot)}{\partial \bar{I}_2}, \frac{\partial h(\cdot)}{\partial \bar{I}_3}, \frac{\partial h(\cdot)}{\partial \bar{I}_4} >$$

$$= \underbrace{\begin{bmatrix} \frac{1}{\bar{I}_{B_1}} & 0 & \frac{-\bar{I}_{(A_1 \cap B_1)}}{(\bar{I}_{B_1})^2} & 0 \\ 0 & \frac{1}{\bar{I}_{B_2}} & 0 & \frac{-\bar{I}_{(A_1 \cap B_2)}}{(\bar{I}_{B_2})^2} \end{bmatrix}}_{2x4},$$

where V_I is a (4x4) variance-covariance matrix defined in the previous section.

A.4.4 Putting it all Together

Our original question was to calculate the variance-covariance, Δ_h, associated with the sample variance of the conditional mean estimates based on the micro data. For two of the four micro-moments, we showed the following,

$$\sqrt{N} \begin{bmatrix} \bar{I}_{\text{micro}}\{A_1|B_1\} - I^*_{\text{micro}}\{A_1|B_1\}) \\ \bar{I}_{\text{micro}}\{A_1|B_2\} - I^*_{\text{micro}}\{A_1|B_2\}) \end{bmatrix} \rightarrow N(0, \Delta_h)$$

$$\Longleftrightarrow$$

$$\sqrt{N}(h(\bar{I}) - h(I^*) \xrightarrow{d} N\Big(0, \nabla h'(\bar{I}) V_I \nabla h(\bar{I})\Big)$$

$\Longrightarrow$

$$\underbrace{\Delta_h}_{2x2} = \underbrace{\nabla h'(\bar{I})}_{2x4} \underbrace{V_I}_{4x4} \underbrace{\nabla h(\bar{I})}_{4x2},$$

as required.

A.5 Simulation Variance

The final source of variance that enters weight matrix is due to the simulation error from using $ns \in M$ simulations to represent the M total U.S. households. To estimate the simulation variance, R sets of moment conditions are averaged over using a different set of ns individuals for each $r \in R$ draw, $(ns^1, ns^2, ..., ns^R)$,

$$\begin{aligned}\Delta_{ns} &= E\left[\left[E\left(\bar{I}_{\text{micro}}\{A|B\} - \hat{P}_{\text{model}}(A|B;\theta)\right)\right]^2\right] \\ &= \frac{1}{R}\sum_{r=1}^{R}\left[\frac{1}{ns}\sum_{i=1}^{ns}\left(\bar{I}_{\text{micro}}\{A|B\} - \hat{P}^r_{\text{i, model}}(A|B;\theta)\right)\right]^2 \\ &= \frac{1}{R}\sum_{r=1}^{R}(\hat{\zeta}^r)^2,\end{aligned}$$

where R equals the number of sets of ns simulated individuals, $(ns^1, ns^2, ..., ns^R)$. Each set of simulated draws requires the mean utility parameter, δ, to be re-solved via the contraction mapping using the BLP (1995) methodology. As a result, accounting for this variance in the standard error estimates is computationally demanding.[3]

A.6 Asymptotics

The BLP moments and micro-moments are stacked and jointly estimated following the nonlinear GMM literature. The objective function is,

$$Q(\theta) = \min_{\theta} \psi'(\theta) W \psi(\theta),$$

where,

$$\psi(\theta) = \begin{bmatrix} \psi_{\text{blp}} \\ \psi_{\text{micro}} \end{bmatrix},$$

where ψ_{blp} and ψ_{micro} are the BLP and micro-moments, respectively. Since the moments are based off of different sampling processes (aggregate data vs. micro data), the weight matrix

[3]Most papers justify skipping this step by using variance reduction techniques.

is block diagonal as the covariates between the two are zero,

$$W = \begin{bmatrix} W_{\text{blp}} & 0 \\ 0 & W_{\text{micro}} \end{bmatrix}.$$

The objective function can therefore be written as,

$$\underbrace{Q(\theta)}_{1x1} = \min_{\theta} \left\{ \underbrace{\psi'_{\text{blp}}(\theta)}_{1xq} \underbrace{W_{\text{blp}}}_{qxq} \underbrace{\psi_{\text{blp}}(\theta)}_{qx1} + \underbrace{\psi'_{\text{micro}}(\theta)}_{1xk} \underbrace{W_{\text{micro}}}_{kxk} \underbrace{\psi_{\text{micro}}(\theta)}_{kx1} \right\}.$$

where q is the number of BLP moments and k is the number of micro-moments. The asymptotics for this estimator are difficult to estimate primarily due to the complexity of the weight matrix. However, the general form is the same as any other nonlinear GMM estimator,

$$\sqrt{N}(\hat{\theta}_{\text{gmm}} - \theta_0) \xrightarrow{d} N(0, V),$$

where

$$V = \left(\Gamma' W \Gamma\right)^{-1} \Gamma' W \Sigma W \Gamma \left(\Gamma' W \Gamma\right)^{-1},$$

$$\Gamma = E\left[\frac{\partial \psi(\theta)}{\partial \theta}\right],$$

and

$$\Sigma = E\left[\psi \psi'\right].$$

The optimal weight matrix is $W = \Sigma^{-1}$ which yields the most efficient estimator (smallest variance). Plugging the optimal weight matrix into the variance formula above yields,

$$V = \left(\Gamma' \Sigma^{-1} \Gamma\right)^{-1}.$$

A.6.1 Gradient of Micro-Moment

The gradient of the micro-moment is taken with respect to each parameter. For example, suppose there are two micro-moments and five parameters, $\theta = \{\theta^1, ..., \theta^5\}$,

$$\psi(\theta) = \begin{bmatrix} \psi_1(\theta) \\ \psi_2(\theta) \end{bmatrix} = \begin{bmatrix} \bar{I}_{\text{micro}}\{A_1|B_1\} - \bar{P}_{\text{model}}(A_1|B_1;\theta) \\ \bar{I}_{\text{micro}}\{A_1|B_2\} - \bar{P}_{\text{model}}(A_1|B_2;\theta) \end{bmatrix}.$$

The gradient is then,

$$\begin{aligned}
\Gamma(\theta) &= \frac{\partial\psi(\theta)}{\partial\theta} \\
&= \left\langle \frac{\partial\psi(\theta)}{\partial\theta_1}, \frac{\partial\psi(\theta)}{\partial\theta_2}, \frac{\partial\psi(\theta)}{\partial\theta_3}, \frac{\partial\psi(\theta)}{\partial\theta_4}, \frac{\partial\psi(\theta)}{\partial\theta_5} \right\rangle \\
&= \begin{bmatrix} \frac{\partial\psi_1(\theta)}{\partial\theta_1} & \frac{\partial\psi_1(\theta)}{\partial\theta_2} & \frac{\partial\psi_1(\theta)}{\partial\theta_3} & \frac{\partial\psi_1(\theta)}{\partial\theta_4} & \frac{\partial\psi_1(\theta)}{\partial\theta_5} \\ \frac{\partial\psi_2(\theta)}{\partial\theta_1} & \frac{\partial\psi_2(\theta)}{\partial\theta_2} & \frac{\partial\psi_2(\theta)}{\partial\theta_3} & \frac{\partial\psi_2(\theta)}{\partial\theta_4} & \frac{\partial\psi_2(\theta)}{\partial\theta_5} \end{bmatrix} \\
&= \begin{bmatrix} \frac{-\partial P_{\text{model}}(A_1|B_1;\theta)}{\partial\theta_1} & \frac{-\partial P_{\text{model}}(A_1|B_1;\theta)}{\partial\theta_2} & \frac{-\partial P_{\text{model}}(A_1|B_1;\theta)}{\partial\theta_3} & \frac{-\partial P_{\text{model}}(A_1|B_1;\theta)}{\partial\theta_4} & \frac{-\partial P_{\text{model}}(A_1|B_1;\theta)}{\partial\theta_5} \\ \frac{-\partial P_{\text{model}}(A_1|B_2;\theta)}{\partial\theta_1} & \frac{-\partial P_{\text{model}}(A_1|B_2;\theta)}{\partial\theta_2} & \frac{-\partial P_{\text{model}}(A_1|B_2;\theta)}{\partial\theta_3} & \frac{-\partial P_{\text{model}}(A_1|B_2;\theta)}{\partial\theta_4} & \frac{-\partial P_{\text{model}}(A_1|B_2;\theta)}{\partial\theta_5} \end{bmatrix}.
\end{aligned}$$

A.6.2 Combinging BLP and Micro-Moments

Following Imbens and Lancaster (1994), the asymptotic variance for the Stacked GMM Estimator is,

$$\sqrt{J}(\hat{\theta}_{GMM} - \theta_0) \rightarrow N(0, V),$$

where,

$$V = \left[\Gamma'_{blp} W_{blp} \Gamma_{blp} + \rho \cdot \Gamma'_{\text{micro}} \underbrace{\left(\Delta_g + \Delta_h + \Delta_{ns}\right)^{-1}}_{W_{\text{micro}}} \Gamma_{\text{micro}}\right]^{-1}.$$

The efficiency gain from the micro data is the second term, namely,

$$\text{Efficiency Gain } = \rho \cdot \Gamma'_{\text{micro}} \underbrace{\left(\Delta_g + \Delta_h + \Delta_{ns}\right)^{-1}}_{W_{\text{micro}}} \Gamma_{\text{micro}}.$$

The asymptotics of the parameter estimates, $\hat{\theta}_{GMM}$, depend on the sample size of the main data set, J, corresponding to the number of products. The constant, $\rho = J/N$, is used to up/down weight the micro-moment variance depending on how large/small the micro data set, N, is in relation to the main data set, J.

Finally, the sample analog of the variance is,

$$\text{var}(\hat{\theta}_{GMM}) = \frac{V}{J},$$

and the standard errors for the parameter estimates, $\hat{\theta}_{GMM}$, are,

$$S.E. = \sqrt{\text{var}(\hat{\theta}_{GMM})_{ii}},$$

where ii represent the diagonal elements. Finally, the t-statistics are,

$$t = \frac{\hat{\theta}_{GMM}}{S.E.}.$$

Appendix B

Isolating Price and Variety Effects with Nonlinear Marginal Utility of Income

B.1 Introduction

This appendix examines the computational challenges of conducting welfare analyses when nonlinear income effects are included in the BLP methodology. I build off the work of Herriges and Kling (1999) who provide several alternative estimation strategies to recover Compensating Variation (CV) estimates for Random Utility Models (RUMs). This appendix contributes to the literature by providing an estimation procedure to isolate price and variety effects while allowing income to enter the indirect utility function nonlinearly. These effects can be estimated conditional on both an individual's original product choice and preference type, thus giving more realistic and precise welfare measures.

Utility has long been shown to depend nonlinearly on income. High income individuals value a marginal dollar less than low income individuals. Despite this, researchers have been reluctant to include nonlinear income effects in their model specifications due to the increased computational burden it presents when conducting welfare analyses. In applications where product prices represent a small percentage of an individual's income, this seems to be a reasonable assumption. For example, income probably has very little to do with how responsive an individual is to breakfast cereal prices (see Nevo 2001). However, with high priced items, such as automobiles, income effects are expected to have a large influence on a consumer's purchase decision since the price represents a larger share of their income, thus leaving less money to spend on the numeraire good.

B.1.1 Conditional Indirect Utility

Discrete choice models in the BLP (1995) framework start by specifying an individual i's indirect utility function from consuming product $j \in J$,

$$U_{i,j} = \alpha * \log(y_i - p_j) + \delta_j + \mu_{i,j} + \epsilon_{i,j},$$

where y_i is individual i's income, p_j is product j's price, δ_j is the average utility component common to all consumers from product j, $\mu_{i,j}$ is individual i's deviation from that average, and $\epsilon_{i,j}$ is consumer i's idiosyncratic tastes. It is assumed individual i chooses product $j \in J$ that maximizes his/her utility given his/her financial constraint, $y_i = p_j + z$, with z representing the amount of the numeraire good consumed. The marginal utility of income is therefore,

$$MU_I = \frac{\partial U_{i,j}}{\partial y_i} = \frac{\alpha}{(y_i - p_j)},$$

which increases at a decreasing rate with respect to income. This results in higher income individuals being less sensitive to prices than lower income individuals.

B.2 Compensating Variation

The best metric economists have to measure the consumer welfare generated from new product innovations is Compensating Variation (CV); the dollar amount a consumer would need to be paid, in dollar terms, to adjust his/her utility level back to its original state after some change in the economic environment. To estimate this effect, several counterfactual simulations are conducted in which the new product is removed from the choice set allowing individuals to re-sort themselves to the next best alternative.

B.2.1 CV: Total Effect

Compensating variation lends itself easily to the indirect utility function,

$$\begin{aligned} &\max_{j \in 0,1,\ldots,J} \left\{ \alpha^* \log(y_i - p_j) + \delta_j + \mu_{i,j} + \epsilon_{i,j} \right\} \\ = &\max_{r \in 0,1,\ldots,J^{cf}} \left\{ \alpha^* \log(y_i - p_r^{cf} - cv_i) + \delta_r + \mu_{i,r} + \epsilon_{i,r} \right\}, \end{aligned}$$

where p_j is the original market price, p_r^{cf} is the counterfactual price, $J^{cf} \subset J$ is the counterfactual choice set, and cv_i is the amount individual i would need to be compensated to match the two utility levels. Since income enters the indirect utility function nonlinearly, the log-sum formula proposed by Small and Rosen (1981) cannot be used. As a result, a simulation technique must be used which I explain in the next section.

B.2.2 Price and Variety Effects

When quantifying the benefits of new products, two important questions to ask are 1) how much of the benefit is generated from the increased price competition and 2) how much is generated from the increased product variety. Following Hausman and Leonard (2002), the total CV can be isolated into price and variety effects. The Price Effect (PE) measures the consumer welfare generated from the increased price competition from the new product entering the market. Similarly, assuming the new product is in greater accordance to some consumer's preferences, the Variety Effect (VE) measures the allocative efficiency gain.

To estimate the PE, a household's utility is simulated using two market scenarios where the choice set (i.e. variety) remains fixed across both markets. Only prices are allowed to vary,

$$
\begin{aligned}
&\max_{j\in 0,1,\ldots,J^{cf}} \Big\{\alpha^* \log(y_i - p_j) + \delta_j + \mu_{i,j} + \epsilon_{i,j}\Big\} \\
&= \max_{r\in 0,1,\ldots,J^{cf}} \Big\{\alpha^* \log(y_i - p_r^{cf} - cv_i^{pe}) + \delta_r + \mu_{i,r} + \epsilon_{i,r}\Big\},
\end{aligned}
$$

where J^{cf} is the counterfactual choice set with all new products removed and p^{cf} is the corresponding vector of counterfactual equilibrium prices. The PE, cv_i^{pe}, is therefore the amount of money that household i would need to be compensated as a result of the price changes generated from the new products entering the market.

Similarly, the VE is estimated by simulating a household's utility, again using two market scenarios. For this effect, prices remain fixed at original market levels across both markets, while only the choice set (i.e. variety) is allowed to vary,

$$
\begin{aligned}
&\max_{j\in 0,1,\ldots,J} \Big\{\alpha^* \log(y_i - p_j) + \delta_j + \mu_{i,j} + \epsilon_{i,j}\Big\} \\
&= \max_{r\in 0,1,\ldots,J^{cf}} \Big\{\alpha^* \log(y_i - p_r - cv_i^{ve}) + \delta_r + \mu_{i,r} + \epsilon_{i,r}\Big\}.
\end{aligned}
$$

The VE, cv_i^{ve}, is therefore the amount of money that household i would need to be compensated as a result of households having additional product variety available in their choice set.

B.3 Estimation Strategy

When estimating a BLP model, there are two sets of independent simulations. The first is over $i \in ns$ simulated individual preference types approximated by a parametric distribution (e.g. normal). The second set of simulations corresponds to the vector of idiosyncratic tastes, ϵ_i^t, that each individual i has for each product $j \in J$. By specifying an individual's idiosyncratic tastes as extreme value, it can subsequently be integrated out resulting in the logistic formula,

$$P_{i,j} = \frac{\exp\left(\alpha * \log(y_i - p_j) + \delta_j + \mu_{i,j}\right)}{\sum_r^J \exp\left(\alpha * \log(y_i - p_r) + \delta_r + \mu_{i,r}\right)},$$

which is interpreted as the probability an individual of preference type $i \in ns$ purchases product $j \in J$.

Before explaining how to recover CV estimates, it is first important to point out that the same probability can be estimated via simulating, rather that integrating, over the extreme value error term,

$$\begin{aligned}\tilde{P}_{i,j} &= \frac{1}{T}\sum_{t=1}^{T} I\Big\{\max_{j\in J}\Big(\alpha * \log(y_i - p_j) + \delta_j + \mu_{i,j} + \epsilon_{i,j}^t\Big)\Big\} \\ &= \frac{1}{T}\sum_{t=1}^{T} I\Big\{\max_{j\in J} U_{i,j}^t\Big\},\end{aligned}$$

where $\epsilon^t \sim$ extreme value is a (Jx1) vector representing the t^{th} draw of individual type i's extreme tastes for each product $j \in J$ and $I\{\cdot\}$ is an indicator function equal to 1 if product j is chosen, and zero otherwise. This is repeated for each individual type, $i \in ns$, using T simulated extreme value errors to simulate utility choices.[1]

When estimating CV, however, the two sets of simulations can no longer be treated as independent and must be drawn jointly such that a simulated individual, $i \in ns$, is uniquely defined by the triplet $\{y_i, \mu_{i,j}, \epsilon_{i,j}\}$ corresponding to their income, heterogeneous preferences for product characteristics $k \in K$, and extreme idiosyncratic tastes for each product $j \in J$,

[1]The size of T should be adjusted such that $|P_{i,j} - \tilde{P}_{i,j}| <$ tolerance.

respectively.

B.3.1 Steps to Solve

For expositional purposes, assume the market consists of $j=5$ products and the new product removed from the market corresponds to $j=4$.

Step one:
Simulate an individual $i \in ns$ consisting of the triplet $\{y_i, \mu_{i,j}, \epsilon_{i,j}\}$ such that,

$$y_i \sim LogN(\mu_y, \sigma_y),$$

where y_i is a scalar representing individual i's income drawn from a log-normal distribution.[2] The heterogeneous utility component is represented by,

$$\mu_{i,j} = \sum_{k=1}^{K} \sigma_k x_j^k v_i^k,$$

where σ_k is a scale parameter estimated from the model, x_j^k is the k^{th} characteristic for product j, and v_i^k is a random draw from a standard normal distribution such that,

$$v_i^k \sim N(0,1), \qquad k = 1, 2, ..., K.$$

Lastly, draw a (Jx1) vector of extreme tastes,

$$\epsilon_i \sim -\log(-\log(t)),$$

where t is a (Jx1) vector of uniformly distributed random variables. Each element $j \in J$ corresponds to an individual's extreme taste for product j.

Step Two:
For each simulated individual $i \in ns$, simulate his/her utility maximizing decision in the

[2]The parameters, μ_y and σ_y, are estimated via a maximum likelihood procedure and a sample of U.S. household incomes obtained from the Current Population Survey (CPS). My results yield the following parameter estimates: $\mu_y = 10.6890$ and $\sigma_y = 0.9765$.

original market,

$$\begin{aligned} V_i^* &\equiv \max_{j\in J} U(y_i - p_j, x_j, v_i, \epsilon_{i,j}) \qquad \text{“Baseline Utility"} \\ &= \max_{j\in J} \Big\{ \alpha * \log(y_i - p_j) + \delta_j + \mu_{i,j} + \epsilon_{i,j} \Big\}. \end{aligned}$$

Which visually looks like,

$$\begin{bmatrix} U_{i,1} \\ U_{i,2} \\ U_{i,3} \\ U_{i,4} \\ U_{i,5} \end{bmatrix} = \begin{bmatrix} \alpha*\log(y_i - p_1) \\ \alpha*\log(y_i - p_2) \\ \alpha*\log(y_i - p_3) \\ \alpha*\log(y_i - p_4) \\ \alpha*\log(y_i - p_5) \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{bmatrix} + \begin{bmatrix} \mu_{i,1} \\ \mu_{i,2} \\ \mu_{i,3} \\ \mu_{i,4} \\ \mu_{i,5} \end{bmatrix} + \begin{bmatrix} \epsilon_{i,1} \\ \epsilon_{i,2} \\ \epsilon_{i,3} \\ \epsilon_{i,4} \\ \epsilon_{i,5} \end{bmatrix}.$$

Choose the utility with the highest value, V_i^*.

<u>Step Three:</u>

Remove the new product(s), $j = 4$, from the data set and re-solve the utility maximization problem,

$$\begin{aligned} V_i^{cf} &\equiv \max_{r\in J^{cf}} U(y_i - p_r^{cf}, x_r, v_i, \epsilon_{i,r}) \qquad \text{“Counterfactual Utility"} \\ &= \max_{r\in J^{cf}} \Big\{ \alpha * \log(y_i - p_r^{cf}) + \delta_r + \mu_{i,r} + \epsilon_{i,r} \Big\}. \end{aligned}$$

Which visually looks like,

$$\begin{bmatrix} U_{i,1} \\ U_{i,2} \\ U_{i,3} \\ U_{i,5} \end{bmatrix} = \begin{bmatrix} \alpha*\log(y_i - p_1^{cf}) \\ \alpha*\log(y_i - p_2^{cf}) \\ \alpha*\log(y_i - p_3^{cf}) \\ \alpha*\log(y_i - p_5^{cf}) \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_5 \end{bmatrix} + \begin{bmatrix} \mu_{i,1} \\ \mu_{i,2} \\ \mu_{i,3} \\ \mu_{i,5} \end{bmatrix} + \begin{bmatrix} \epsilon_{i,1} \\ \epsilon_{i,2} \\ \epsilon_{i,3} \\ \epsilon_{i,5} \end{bmatrix}.$$

Again, choose the utility with the highest value, V_i^{cf}.[3]

[3] When simulating error terms, $\epsilon_{i,j}$, each error term is unique to each product j in both the original and counterfactual market. Therefore, when the new products are removed from the market, it is critical that the simulated errors match up with the same products as they did in the original market.

Step 4:

If individual i's utility has been changed as a result of the new product being removed,

$$V_i^* \neq V_i^{cf},$$

they will need to be compensated such that to two utility levels are equal,

$$\begin{aligned} V_i^* &\stackrel{set}{=} V_i^{cf} \\ &= \alpha * \log(y_i - p_\rho^{cf} - cv_i) + \delta_\rho + \mu_{i,\rho} + \epsilon_{i,\rho}, \qquad \text{where } \rho = \arg\max_{r \in J^{cf}} U_{i,r}. \end{aligned}$$

This can easily be solved as it is one equation with one unknown,

$$\begin{aligned} &V_i^* = \alpha * \log(y_i - p_\rho^{cf} - cv_i) + \delta_\rho + \mu_{i,\rho} + \epsilon_{i,\rho} \\ &\Longrightarrow \\ &cv_i = y_i - p_\rho^{cf} - \exp\left(\frac{V_i^* - (\delta_\rho + \mu_{i,\rho} + \epsilon_{i,\rho})}{\alpha}\right). \end{aligned}$$

Step Five:

Repeat this for each $i \in ns$.

Step Six:

The average CV across all individuals is then,

$$E[cv_i] = \frac{1}{ns} \sum_i^{ns} cv_i,$$

and the total CV generated for the whole market is simply,

$$cv^{\text{market}} = M \cdot E[cv_i],$$

where M is the market size.

B.3.2 Solving PE and VE

The same logic above can be used to estimate the price and and variety effect. The only difference is the set up for steps 2 and 3.

B.4 CV Conditional on Vehicle Choice

In most applied welfare analyses, compensating variation is only reported for the average individual in the market or the market as a whole. However, when examining the benefits of new products, a more interesting question to ask is how much compensation a specific individual requires conditional on his/her first choice being removed from the choice set. This way, compensating variation is conditional on purchasing the new product, and varies from person to person due to differences in preferences.

To see this more clearly, consider a market with $ns = 10$ simulated individuals and suppose only individuals $i = 3$ and $i = 6$ purchased a new product. Visually, this takes the form,

$$\begin{bmatrix} cv_1 \\ cv_2 \\ cv_3 \\ cv_4 \\ cv_5 \\ cv_6 \\ cv_7 \\ cv_8 \\ cv_9 \\ cv_{10} \end{bmatrix}_{\text{All households}} \Longrightarrow \begin{bmatrix} 0 \\ 0 \\ cv_3 \\ 0 \\ 0 \\ cv_6 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}_{\text{Purchased new product}} .$$

Keeping only the non-zero elements in the "purchased new product" vector will yield a distribution of compensating variations conditional on purchasing the new product. As a result, the conditional CVs will trace out the demand curve (i.e. a consumer's willingness to pay) for the new product.

B.5 Conclusion

This appendix provides an estimation strategy to isolate price and variety effects while allowing income to enter the indirect utility function nonlinearly. These effects can be estimated conditional on both an individual's original product choice and preference type, thus giving more realistic and precise welfare measures. In particular, this appendix examines the welfare implications from new product innovations using the BLP (1995) methodology.

Appendix C

Marginal Utility of Income: Price and Income Specification

C.1 Introduction

The discrete choice literature has used several specifications to describe the functional relationship between household income and price, $h(y_i, p_j)$. However, each specification has caveats and so it is ultimately up to the researcher to choose the best specification in regards to their application. In what follows, I provide a detailed explanation of the various specifications used in the literature.

For simplicity, let household i's indirect utility function from consuming product j be represented by the following:

$$V(y_i, p_j, x_j) = \alpha h(y_i, p_j) + \beta f(x_j),$$

where y_i is household i's income, p_j is the the price for product j, and x_j is a vector of product characteristics. The function $h(y_i, p_j)$ represents the functional relationship between income and price, and $\beta f(x_j)$ represents all other elements of utility.

C.2 Constant Marginal Utility of Income

A common specification in the discrete choice literature is the assumption that the marginal utility of income is constant across households. This is achieved by specifying a linear relationship between price and income,

$$\alpha h(y_i, p_j) = \alpha(y_i - p_j).$$

Because only differences in utility matter in estimation, household i's income has no effect on the probability of product j being selected because it drops out of the expression,

$$\begin{aligned}
P_{i,j} &= \text{Prob}\left[\epsilon_{i,k} - \epsilon_{i,j} < V(y_i - p_j, x_j) - V(y_i - p_k, x_k)\right] \\
&= \text{Prob}\left[\epsilon_{i,k} - \epsilon_{i,j} < \left[\alpha(y_i - p_j) - \alpha(y_i - p_k)\right] + \left[\beta f(x_j) - \beta f(x_k)\right]\right] \\
&= \text{Prob}\left[\epsilon_{i,k} - \epsilon_{i,j} < \left[\alpha(p_k - p_j)\right] + \beta\left[f(x_j) - f(x_k)\right]\right].
\end{aligned}$$

The marginal utility of income is therefore

$$MU_{\text{income}} = \alpha,$$

which only depends on price and thus fails to capture any income effects.

C.3 Nonlinear Marginal Utility of Income

To overcome this problem, a common approach is to specify a nonlinear functional relationship between price and income. Most researchers use a log specification which exhibits decreasing returns,

$$\alpha h(y_i, p_j) = \alpha \log(y_i - p_j).$$

Again, only differences in utility matter in estimation. However, with this specification household i's income will not drop out of the probability expression and will, as a result, affected the probability of product j being selected,

$$\begin{aligned}
P_{i,j} &= \text{Prob}\left[\epsilon_{i,k} - \epsilon_{i,j} < V(y_i - p_j, x_j) - V(y_i - p_k, x_k)\right] \\
&= \text{Prob}\left[\epsilon_{i,k} - \epsilon_{i,j} < \left[\alpha \log(y_i - p_j) - \alpha \log(y_i - p_k)\right] + \left[\beta f(x_j) - \beta f(x_k)\right]\right] \\
&= \text{Prob}\left[\epsilon_{i,k} - \epsilon_{i,j} < \alpha\left[\log(y_i - p_j) - \log(y_i - p_k)\right] + \beta\left[f(x_j) - f(x_k)\right]\right].
\end{aligned}$$

The marginal utility of income is,

$$MU_{\text{income}} = \alpha \frac{1}{y_i - p_j},$$

which increases at a decreasing rate with respect to income, thus making high income households less sensitive to prices relative to low income households. For some examples, see Berry et al. (1995) and Petrin (2002).

Although this specification is appealing in theory, empirically it has many problems. First, when analyzing markets in which products are priced in the tens of thousands of dollars (e.g. automobiles) it is fairly common for a household to purchase a product whose price is larger than his/her reported income, $p_j > y_i$. This is problematic because the log of a negative number is undefined. This indicates that either 1) their reported income is unrepresentative of their financial constraint or 2) their reported income is inaccurate. A second problem with this specification is the increased difficulty of estimating welfare effects since the marginal utility of income varies continuously across households. Researchers are forced to use simulation techniques which significantly increases the computational burden.

C.4 Constant Marginal Utility of Income per Income Tier

A middle of the road approach is to allow the marginal utility of income to vary across income groups, but remain constant within each group. This overcomes the criticism of the first approach and greatly reduces the computational burden of conducting welfare analyses in the second. For some examples, see Morey et al. (2002) and Goolsbee and Petrin (2004).

The functional form for the price and income effect component, $h(\cdot)$, is the same as the constant marginal utility of income specification,

$$\alpha^* h(y_i, p_j) = \alpha^*(y_i - p_j),$$

however now the marginal utility of income coefficient is a piecewise function depending on household i's income,

$$\alpha^* = \begin{cases} \alpha_1 & \text{if income} \leq y_1 \\ \alpha_2 & \text{if } y_1 < \text{income} \leq y_2 \\ & \vdots \\ \alpha_m & \text{if income} > y_{m-1}. \end{cases}$$

It is expected that $\alpha_1 > \alpha_2 > \ldots > \alpha_m$ such that households in the first income bracket value a marginal dollar more than households in the second bracket, and so on. It is up to the researcher to decide how many income groups, m, to specify.

C.5 Income as a Taste Parameter

Another common approach is to treat income as a proxy variable for tastes. For example, this line of reasoning can be used to imply that wealthier households prefer luxury goods not simply because they can afford it, but rather because it fits their preference type better. With this specification two separate "income" terms must be included in the model. The first represents "taste income," I_i, and the second represents a "financial constraint," y_i. For some examples, see Winston and Mannering (1984), Viton (1984), Lareau and Rae (1989), Mannering and Winston (1995), and Train and Winston (2007).

The functional form of the price and income effect component, $h(\cdot)$, is the following:

$$\alpha h(y_i, I_i, p_j) = \alpha_p(y_i - p_j) + \alpha_{p/I}\frac{(y_i - p_j)}{I_i}.$$

Again, only differences in utility matter. Therefore, the difference in utility for the price and income effect component, $h(\cdot)$, is the following:

$$\begin{aligned} &\alpha h(y_i, I_i, p_j) - \alpha_i h(y_i, I_i, p_k) \\ &\qquad = \left[\alpha_p(y_i - p_j) + \alpha_{p/I}\frac{(y_i - p_j)}{I_i}\right] - \left[\alpha_p(y_i - p_k) + \alpha_{p/I}\frac{(y_i - p_k)}{I_i}\right] \\ &\qquad = \left[\alpha_p(y_i - p_j) - \alpha_p(y_i - p_k)\right] + \left[\alpha_{p/I}\frac{(y_i - p_j)}{I_i} - \alpha_{p/I}\frac{(y_i - p_k)}{I_i}\right] \\ &\qquad = \left[\alpha_p(p_k - p_j)\right] + \left[\alpha_{p/I}\frac{(p_k - p_j)}{I_i}\right]. \end{aligned}$$

Like the constant marginal utility of income case above, household i's "financial constraint,"

y_i, drops out, however, household i's "taste income" does not. As a result, the marginal utility of "financial income" is,

$$MU_{\text{income}} = \alpha_p + \alpha_{p/I}\frac{1}{I_i},$$

which represents an average component and one that varies with "taste income." Viton (1985) argues that "as long as one is careful not to regard the 'taste' variable as measuring income, the received specifications can be found consistent with utility-maximizing behavior." Although this specification is commonly used in the discrete choice literature, it can still be computationally burdensome when conducting welfare analysis since the marginal utility of income varies continuously across households due to differences in "taste income."

"Taste income" is also commonly interacted with other product characteristics in addition to price. For example, the vehicle demand literature indicates that wealthier households prefer hybrid vehicles more than poorer households. To capture this effect, the researcher may want to interact "taste income" with the hybrid dummy variable,

$$\text{hybrid*income effect} = \text{hybrid} * \log(\text{“taste income"}).$$

However, since this is "taste income" rather than "financial income," the marginal utility of income will be unaffected by this expression.

Appendix D

Supplemental Data: Chapter 2

D.1 Supplemental Data

Several supplemental data sources are used in Chapter Two of this dissertation regarding the following:

- Gasoline Prices
- Percentage of Democrats per County
- Median Household Income per Census Tract.

D.1.1 Gasoline Prices

Gasoline prices are obtained from the American Chamber of Commerce's ACCRA data base for 322 Core Based Statistical Areas (CBSAs). The price of gasoline is the 2008 average for one gallon of regular unleaded and includes all taxes (federal, state, and county).

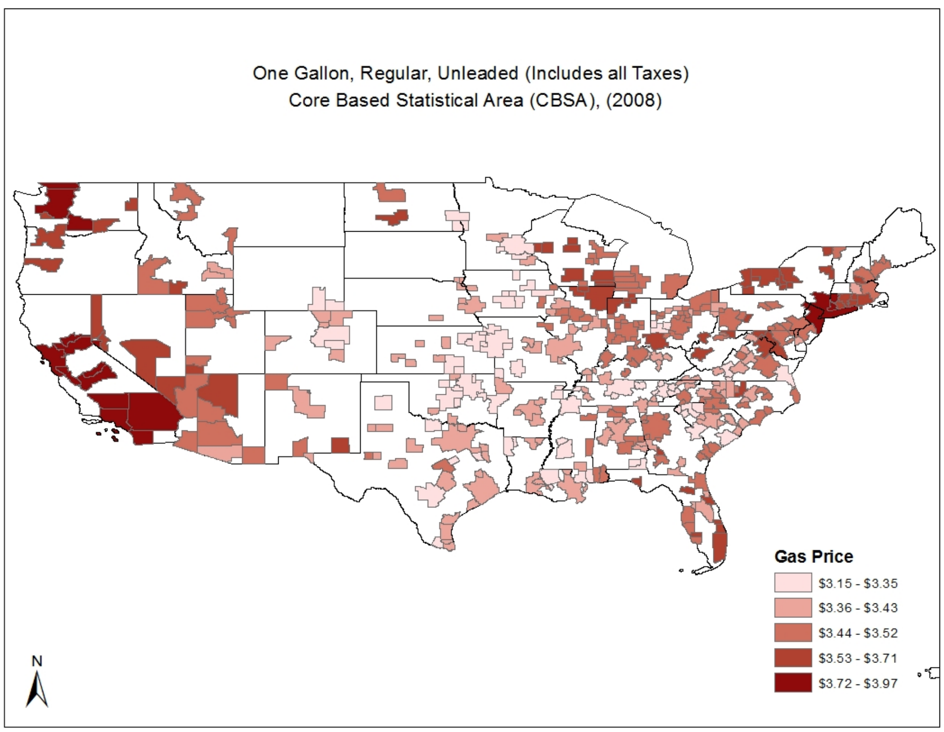

Figure D.1: Gasoline Prices per CBSA. Includes all Taxes (Federal, State, County).

Gasoline prices per CBSA are linked to each household in the NHTS. For households not living in a reported CBSA, the average gasoline price per state is used instead.

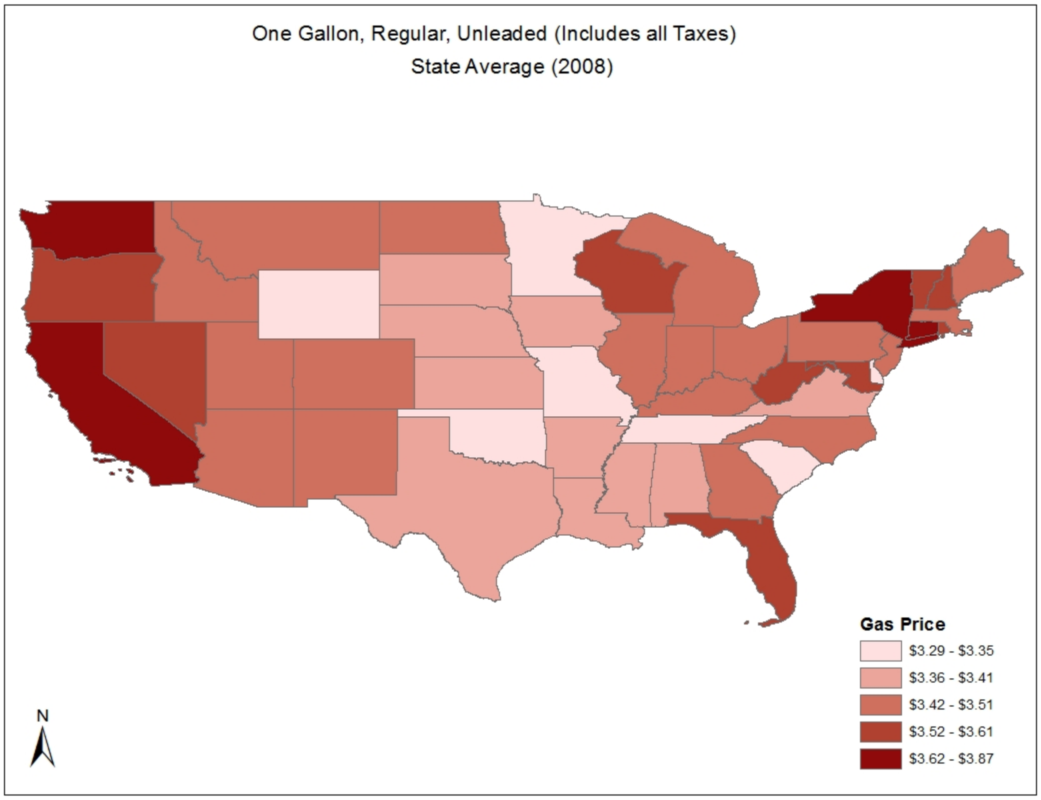

Figure D.2: State Average Gasoline Prices

D.1.2 Political Party

To capture the conspicuous environmental effect from owning a hybrid vehicle, I link the percentage of Democrats per county to each household. Although I have information on the Census Tract for each household, I could only obtain data at the county level. Specifically, I use the 2008 Presidential results per County obtained from USA Today, http://www.usatoday.com/news/politics/election2008/results.htm?loc=interstitialskip .

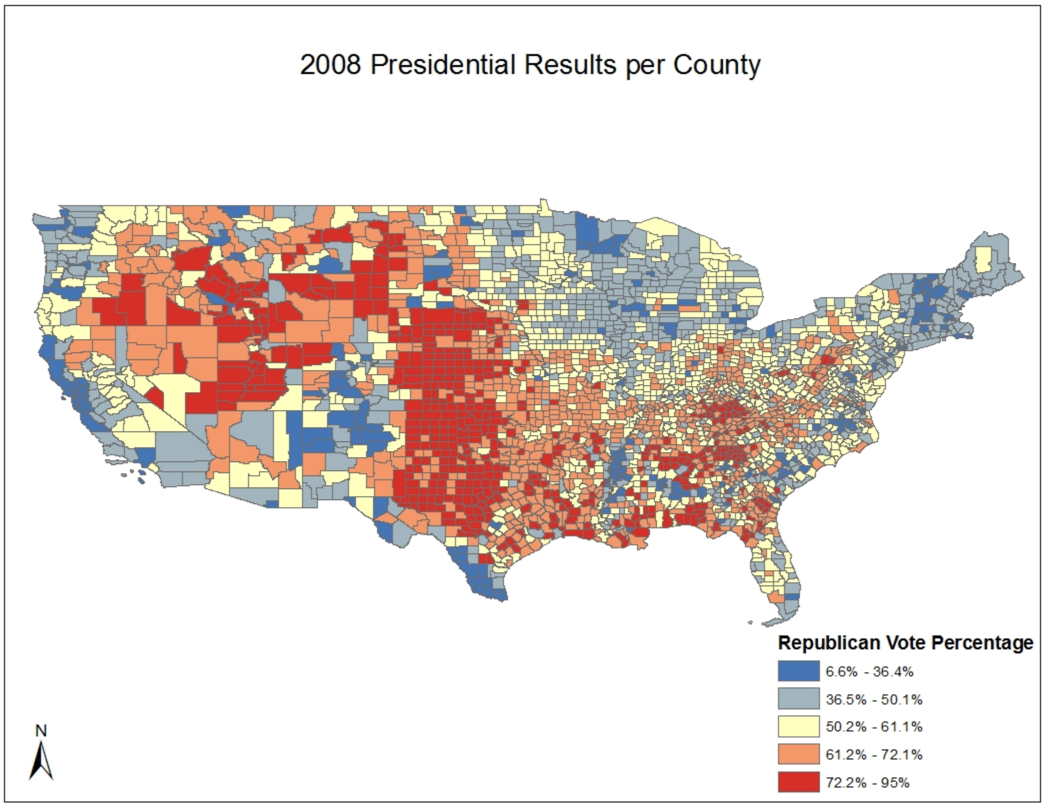

Figure D.3: Percentage of Democrats per County: 2008 Presidential Results

D.1.3 Household Income

The NHTS classifies household incomes into one of eighteen brackets, with bracket 1 representing incomes of less than $2,500 and bracket 18 representing incomes greater than $100,000. For households in the highest income bracket (18), I set their income equal to $150,000 to account the possibility that their income may be much larger than the top-coded amount of $100,000.[1]

Although the quality and richness of the NHTS data is generally high, there did appear to be a few implausible income values. To control for this, if a household's reported income differed from the Census Tract's median value by more than $50,000, I set their income equal to the median value. This information was obtained from the American Community Survey (ACS) published by the U.S. Census Bureau. Specifically, I use the "2005-2009 American Community Survey 5-Year Estimates - B19013 - Median Household Income." In total, there are 65,461 Census Tracts across the U.S. To help the reader comprehend how large a Census Tract is, the figure below maps the median income per Census Tract in North Carolina.

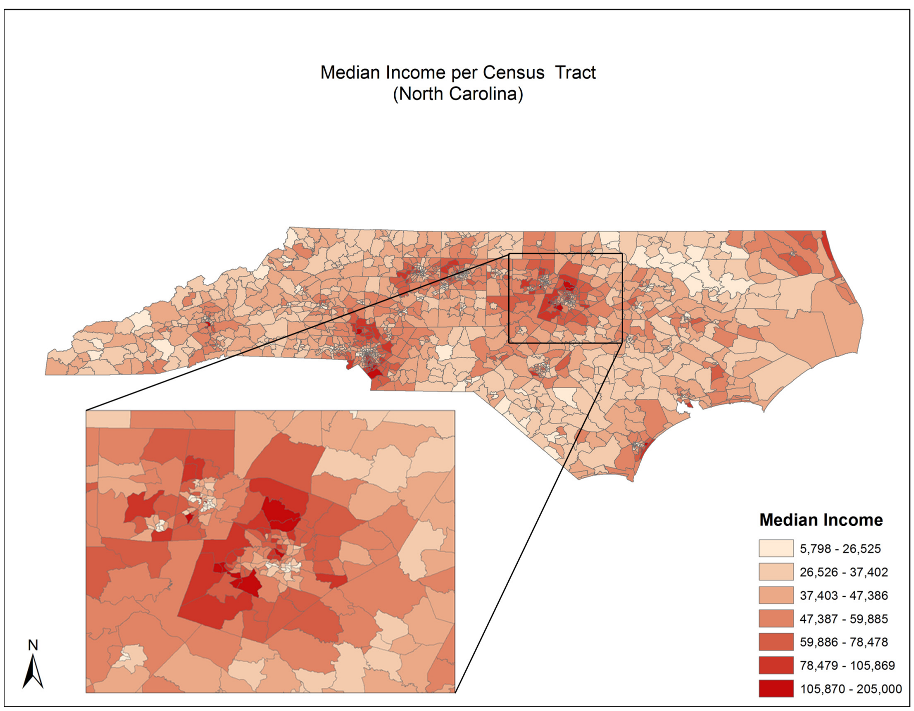

Figure D.4: Median Income per Census Tract: North Carolina

[1]This is the same assumption that Bento et al. (2008) use.

To account for cost of living differences, the ACCRA Cost of Living Index (COLI) is used to inflate/deflate household income. This index is constructed using prices for 57 commonly purchased consumer good items such as T-bone steaks, coffee, apartment rentals, house prices, phone bills, gasoline, and tennis balls. Again, if a household does not live in one of the reported CBSAs, I use the state average COLI instead.

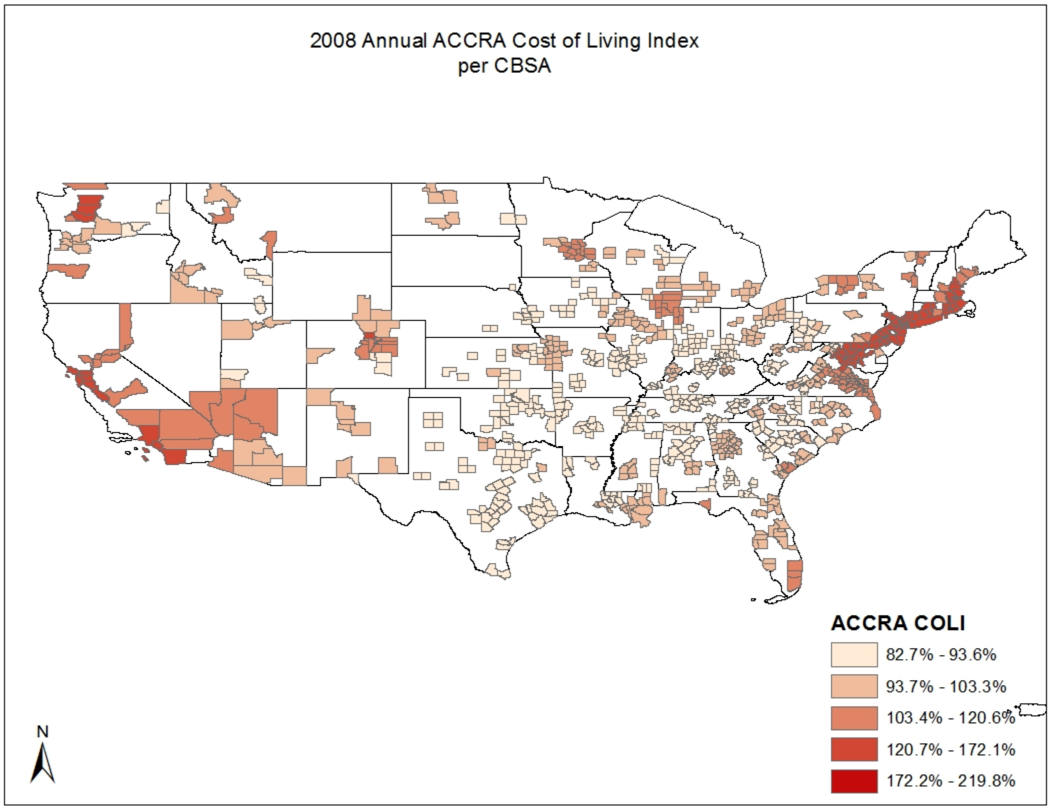

Figure D.5: Cost of Living Index: 2008 ACCRA

D.2 High Occupancy Vehicle (HOV) Lane Minimum Occupancy Requirement

D.2.1 Federal Law and HOV Lane Access

High Occupancy Vehicle (HOV) lanes were first introduced in the early 1970s primarily intended to decrease congestion on U.S. highways by encouraging car-pooling. Federal law dictates that no fewer than two occupants per vehicle may gain access to HOV lanes. However, states may, if they choose, enforce more stringent laws in order to protect the integrity and primary purpose of these lanes. For example, a state may set the minimum occupancy to 3. Today, there are over 130 highways with HOV lanes located throughout the U.S.

To help promote the adoption of low emission and fuel efficient vehicles, the Federal government has passed several laws giving states the authority to exempt certain vehicles from meeting the HOV lane minimum occupancy requirement. If a state chooses to do so, Federal law requires them to establish a program to monitor, assess, and enforce regulations such that the HOV lanes do not become degraded.[2] Therefore, states may adopt more stringent fuel economy standards than those proposed by the EPA or limit the number of special issued license plates/decals/permits. For example, several states including California only exempt the Toyota Prius, Honda Insight and Honda Civic Hybrid from the minimum occupancy requirement. States also have the responsibility to void or discontinue any special issued license plate/decal/permit if an HOV lane becomes degraded in the future.

D.2.2 State-by-State Rules

Below is a list of states that exempt hybrid vehicles from the HOV minimum occupancy requirement:

[2]An HOV facility is considered degraded if vehicles operating on it are failing to maintain a minimum average operating speed 90% of the time over a consecutive 180-day period during morning and/or evening weekday peak hours (minimum average operating speed is defined as 45mph in a 50-mph zone, or 10mph below limit when limit is less than 50mph).

Arizona

The Energy Efficient Pilot Program was started on March 9, 2007. Only the Honda Insight, Toyota Prius and Honda Civic Hybrid qualify. The state imposed a limit of 10,000 special issued license plates which was reached in May, 2008. HOV lanes are located in the Phoenix metropolitan area.

California

The Hybrid Clean Air Vehicle Decal Program was started on September 23, 2004. Only the Honda Insight, Honda Civic Hybrid, and the Toyota Prius qualify. The state imposed a limit of 85,000 decals which was reached in February, 2007. Furthermore, this program is scheduled to expire July 1, 2011 rendering all previously issued decals void. HOV lanes are located in the Los Angeles, San Diego, Sacramento, and San Francisco metropolitan areas.

Colorado

The HOV Exemption Program was started on May 15, 2008. All hybrid vehicles qualify. The state imposed an initial limit of 2,000 decals. By October 2009, 1,817 decals had been issued. HOV lanes are located in the Denver and Aspen metropolitan areas.

Florida

This program started in 2006. All hybrid vehicles qualify. Currently, there is no limit on the number of decals that will be issued. HOV lanes are located in the Ft. Lauderdale and Miami metropolitan areas.

New Jersey

The New Jersey Turnpike Authority began waving the minimum HOV lane occupancy requirement for hybrid vehicles on April 27, 2006. All hybrid vehicles qualify. Currently, there is no limit on the number of permits that will be issued. HOV lanes are located in Newark, NJ (near New York City).

New York

The Clean Pass Program was started on March 1, 2006. Only the Honda Insight, Honda Civic Hybrid and the Toyota Prius qualify. Currently, there is no limit on the number of decals that will be issued. HOV lanes are located on a 40 mile stretch of the Long-Island Expressway near the New York City metropolitan area.

Tennessee

The Smart Pass program was started on January 1, 2009. All hybrid vehicles qualify. Currently, there is no limit on the number of decals to be issued. HOV lanes are located in the Memphis and Nashville metropolitan areas.

Utah

The C “Clean" Plate program was started in 2008. All hybrids qualify. Currently, there is no limit on the number of C license plates that will be issued. HOV lanes are located in the Salt Lake City metropolitan area.

Virginia

Clean Fuel Plates were first issued in 2000 making Virginia the first state in the country to exempt hybrid vehicles from the HOV lane minimum occupancy requirement. All hybrids qualify. HOV lanes are located in the Washington, D.C. and Norfolk/Virginia beach metropolitan areas. However, due to congestion problems, clean fuel plates purchased after January 1, 2006 are no longer allowed on I-95/395 HOV lanes in the Northern Virginia area. They are, however, still are allowed on I-66. Currently, there is no limit on the number of clean fuel plates that will be issued.

D.2.3 HOV Hybrid Interaction Effect

To estimate the relationship between hybrid vehicle demand and HOV lane access, only households that live in both 1) a state that exempts hybrid vehicles from the HOV lane minimum occupancy requirement and 2) an MSA with HOV lanes, is interacted with with the hybrid dummy variable. This prevents a household that lives, for example, 200 miles from an HOV lane to be influence by a state law. Also, since I am only looking at the 2008 new vehicle market, households that live in a state with an expired clean pass program are excluded from the interactions (e.g. California). Therefore, only households that live in the following MSA's are interacted with the hybrid vehicle dummy:

- Phoenix, AZ
- Denver, CO
- Miami/Ft. Lauderdale, FL
- NY City/Long Island, NY and Newark, NJ
- Nashville, TN
- Memphis, TN
- Salt Lake City, UT
- Norfolk/VA Beach, VA
- Washington, D.C. (Northern Virginia only).

The Washington, D.C. metro area is the one exception because it encompasses: Maryland, Virginia and West Virginia, and the District of Columbia. However, only Virginia exempts hybrids from the HOV lane minimum occupancy requirement. Furthermore, hybrid vehicles are only exempt from Interstate I-66 HOV lanes. Therefore, only households that live in the following Northern Virginia counties are interacted with the hybrid dummy variable:

- Fairfax County
- Price William County
- Loudoun County
- Fauquire County
- City of Arlington
- City of Alexandria
- City of Fairfax
- City of Falls Church.

D.3 Exogenous Sample: 2009 NHTS

The 2009 NHTS is a stratified random sample consisting of a national sample plus 20 additional “add-on" areas. By design, each add-on area had a pre-specified target sample size. For example, the target sample size for California was 18,000 whereas the target sample size for Florida was 14,000. The actual number of housing units in each of these states is 13.4 million and 8.4 million, resulting in selection probabilities of 0.13% and 0.16%, respectively. As a result, households in different strata (states) had different probabilities of being selected while households within the same strata had an equal probability of being selected. Fortunately, since the strata are based on a household's location (right-hand side variable) rather than vehicle type (left-hand variable), the sample is considered exogenous in regards to my application. Therefore my model will produce consistent parameter estimates without requiring the use of sample weights.