THE CONSUMER RESPONSE TO GASOLINE PRICE CHANGES: EMPIRICAL EVIDENCE AND POLICY IMPLICATIONS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF MANAGEMENT
SCIENCE AND ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Kenneth Thomas Gillingham
June 2011

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

James Sweeney, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Lawrence Goulder

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Matthew Harding

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

John Weyant

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

When gasoline prices rise, people notice: the news is filled with reports of pinched household budgets and politicians feeling pressure to do something to ameliorate the burden. Yet, raising the gasoline tax to internalize externalities is widely considered by economists to be among the most economic efficiency-improving policies we could implement in the transportation sector. This dissertation brings new evidence to bear on quantifying the responsiveness to changing gasoline prices, both on the intensive margin (i.e., how much to drive) and the extensive margin (i.e., what vehicles to buy). I assemble a unique and extremely rich vehicle-level dataset that includes all new vehicle registrations in California 2001 to 2009, and all of the mandatory smog check program odometer readings for 2002 to 2009. The full dataset exceeds 49 million observations. Using this dataset, I quantify the responsiveness to gasoline price changes on both margins, as well as the heterogeneity in the responsiveness. I develop a novel structural model of vehicle choice and subsequent utilization, where consumer decisions are modeled in a dynamic setting that explicitly accounts for selection on unobserved driving preference at both the time of purchase and the time of driving. This utility-consistent model allows for the analysis of the welfare implications to consumers and government of a variety of different policies, including gasoline taxes and feebates.

I find that consumers are responsive to changing gasoline prices in both vehicle choice and driving decisions, with more responsiveness than in many recent studies in the literature. I estimate a medium-run (i.e., roughly two-year) elasticity of fuel economy with respect to the price of gasoline for new vehicles around 0.1 for California, a response that varies by whether the vehicle manufacturer faces a tightly binding fuel economy standard. I estimate a medium-run elasticity of driving with respect to the price of gasoline around -0.15 for new personal vehicles in the first six years. Older vehicles are driven much less, but tend to be more responsive, with an elasticity of roughly -0.3. I find that the vehicle-level responsiveness in driving to gasoline price changes varies by vehicle class, income, geographic, and demographic groups. I also find that not including controls for economic conditions and not accounting for selection into different types of new vehicles based on unobserved driving preference tend to bias the elasticity of driving away from zero – implying a greater responsiveness than the true responsiveness. This is an important methodological point, for much of the literature estimating similar elasticities ignores these two issues.

These results have significant policy implications for policies to reduce gasoline consumption and greenhouse gas emissions from transportation. The relatively inelastic estimated responsiveness on both margins suggests that a gasoline tax policy may not lead to dramatic reductions in carbon dioxide emissions, but is a relatively non-distortionary policy instrument to raise revenue. When the externalities of driving are considered, an increased gasoline tax may not only be relatively non-distortionary, but even economic efficiency-improving. However, I find that the welfare changes from an increased gasoline tax vary significantly across counties in California, an important consideration for the political feasibility of the policy. Finally, I find suggestive evidence that the "rebound effect" of a policy that works only on the extensive margin, such as a feebate or CAFE standards, may be closer to zero than the elasticity of driving with respect to the price of gasoline. This suggestive finding is particularly important for the analysis of the welfare effects of any policy that focuses entirely on the extensive margin.

Acknowledgement

I have many people to thank for to making this dissertation possible. Without question, I should start with my advisors. If it were not for Jim Sweeney, I would not have even come to Stanford. Jim taught me to think carefully about the economic logic of any contention I make in my research. I cannot thank him more for his guidance and support over the years. I also came to Stanford hoping to get a chance to work with John Weyant. I feel privileged to have had this opportunity. My perspective on energy policy modeling has been profoundly shaped by John's insights.

I have had the good fortune to have Larry Goulder as an advisor since I took a series of classes in environmental economics from him early in my time at Stanford. Larry has been a truly excellent advisor, always making sure that I see the bigger picture in any analysis that I do. In the past two years, I have greatly benefited from Matt Harding's guidance in econometrics. Matt is a world-class econometrician and I have him to thank for developing my skills as an empirical economist.

Several other economists in the economics department also greatly contributed to this dissertation. Chapters 2 and 3 of my dissertation began as a class paper for Jon Levin and Tim Bresnahan's Industrial Organization seminar. Both Jon and Tim have been extremely generous with their time and insights. Besides acting as my oral defense chair, Jon also greatly helped shaped the development of my structural model in Chapter 4 through a series of conversations we had last summer. Tim's comments helped me more deeply understand how my work is novel. In addition, another Industrial Organization professor, Liran Einav, generously agreed to meet with me several times and has very useful suggestions for improving the research.

Wes Hartmann of the Stanford Graduate School of Business taught a class in advanced empirical methods that turned out to be incredibly valuable in helping me develop the tools to estimate my structural model. Wes was also generous with his time in meeting with me to discuss the research. Conversations with Lee Schipper of the Precourt Energy Efficiency Center at Stanford were not only fun, but also very useful in keeping me honest about the details of the transportation sector.

There are also many economists outside of Stanford who provided invaluable help in bringing this dissertation to fruition. My initial interest in economic modeling began when I was a research assistant at Dartmouth College for Karen Fisher-Vanden. Karen encouraged me to major in economics and build a technical skill set. I have her to thank for leading me down the path towards this dissertation. My interest in transportation economics began while I was a research assistant at Resources for the Future (RFF), working with Winston Harrington, Ian Parry, Elena Safirova, Billy Pizer, Jim Sanchirico, and others. While at RFF, I also had the good fortune to work on projects relating to energy efficiency policy with Richard Newell and Karen Palmer. In many respects this dissertation combines these two interests, by looking at policies to improve the energy efficiency of the transportation sector. I became particularly interested in this intersection while working at the White House Council of Economic Advisers with Richard Newell. Richard and I spent a great deal of time thinking about the costs and benefits of fuel economy standards, which even more directly led me down the path to this dissertation topic.

More recently, while I was on the academic job market, I benefited greatly from discussions with a variety of economists around the country. Most notable are the excellent suggestions from Gautam Gowrisankaran at University of Arizona, Dave Rapson at UC Davis, Rob Williams at the University of Maryland, and Jeffrey Brown and Nolan Miller at the University of Illinois Urbana-Champaign. Prior to the academic job market, I also have to thank Soren Anderson of Michigan State University for pointing me to the county-level Bureau of Labor Statistics data.

Of course, this dissertation would also not have been possible without the financial support of several organizations. The US Environmental Protection Agency STAR Fellowship program funded me for several years of my doctoral career, and provided the funding to purchase the first piece of the dataset that I eventually assembled. My

EPA STAR fellowship was awarded for a project to shed light on the rebound effect of fuel economy standards and I am very pleased that my dissertation covers this topic, despite several ventures into other areas of research during my graduate career. The Stanford Global Climate and Energy Project provided the funding for my first year at Stanford, when I did a project on using hydrogen in internal combustion engine vehicles. Thus was my first foray into my own transportation research. The Precourt Energy Efficiency Center at Stanford very generously supported me during my last year at Stanford, allowing me to devote myself full-time to my dissertation research.

In assembling my dataset, I also had the generous financial support of the Stanford Institute for Economic Policy Research (SIEPR) Shultz Graduate Student Fellowship in Economic Policy, a grant administered by SIEPR from Exxon-Mobil, and a small grant from Larry Goulder's research funds. Assembling the dataset would also not have been possible without the help of Zach Richardson of the California Bureau of Automotive Repair in providing the smog check data. In addition, Ray Alvarado of R.L. Polk was very helpful in interpreting the purchased new vehicle registration data. Mark Mitchell of the Oil Price Information Service assisted in finding an academic discount that made possible the purchase of the retail gasoline price data. Finally, Hunt Allcott very generously provided the raw data on EPA fuel economy ratings.

Last, but most certainly not least, I would like to thank my friends and family for all of their support over the years and for putting up with me over the past year. My family – Steve, Carol, and Brett – have been enormously patient with me over my years in graduate school. Some of the most useful insights in my research came about through conversations with my colleagues at Stanford, such as Sébastien Houde, Anant Sudarshan, Bryan Bollinger, and Arthur van Benthem. My roommates in the past few years, Tim Kraft, Amanda Cravens, and Sarah Kate Selling, certainly played a role in helping to keep me sane and happy throughout the process. But I particularly have to thank my girlfriend Julie Fosdick for my happiness over the past several years. Julie's support has undoubtedly made this dissertation possible.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Nothing seems more noticeable to the average American than drastic increases in gasoline prices. From the long lines at gasoline stations in the 1970s, to the numerous internet searches making “high gas prices” one of the hottest search terms on Google in 2008, it is clear that when gasoline prices are high, people notice. But how do they react? As in many previous gasoline price increases in the past, the 2007-2008 spike in gasoline prices filled the news with reports of reduced sales of low fuel economy new vehicles and considerable changes in how much people drive.

With emissions from passenger vehicle and light truck gasoline use making up over 20% of the total carbon dioxide emissions in the United States in 2009 (US Environmental Protection Agency 2011) and many regions of the United States facing considerable local air pollution and congestion externalities, policymakers are extremely interested in understanding consumer behavior in personal vehicle use. In determining the cost effectiveness and economic efficiency of common policy instruments used in the transportation sector, a variety of questions arise. How do policies that change gasoline prices influence the vehicles that are purchased? How would these policies influence the amount driven? Do policies that lead to higher fuel economy vehicles in turn lead to more driving (i.e., is there a “rebound effect”)?

Policymakers may also be interested in how gasoline prices will affect the driving and the vehicle stock in order to forecast baseline emissions under different assumptions of gasoline prices. Performing a sensitivity analysis with different assumptions of gasoline prices is standard for any counterfactual policy analysis of the transportation sector. In particular, such a sensitivity analysis provides insight into how robust the conclusions of the policy analysis are.

Policymakers may be interested in more than just the aggregate response to gasoline price though. The *heterogeneity* in how different types of consumers are affected by changing gasoline prices – whether due to policy or exogenous events – is also important for political economy and equity reasons. How much consumers are affected depends on how much they drive and how much they respond to changing gasoline prices, and there may be heterogeneity in both. The result is that consumers may be affected differently if they live in a more rural area, live in an area with longer commute times, or are wealthier. The political feasibility of any policy is at least partly determined by the distributional consequences of such a policy. A deep understanding of the distributional consequences of a policy that changes gasoline prices may allow policymakers to carefully design ancillary policies, perhaps as simple as a lump-sum redistribution of the revenues, to ameliorate the equity concerns and increase the likelihood that the policy will be passed.

This dissertation explores the above questions about consumer responsiveness to gasoline price changes, delves into the heterogeneity in such responsiveness, and examines the welfare and distributional implications of the findings for policies to reduce emissions from the transportation sector. It brings together an unique and extremely rich dataset of vehicle purchases and subsequent driving in California over a period of considerable changes in gasoline prices. The dataset is sufficiently rich that summary statistics and simple regressions provide substantial insights into consumer responsiveness.

This dissertation goes further by developing a novel methodology to address an important bias in estimating the consumer responsiveness to changes in the price of gasoline. This bias, first shown in Dubin and McFadden (1984) is quite intuitive: if at the time of purchase a consumer anticipates driving a great deal, the consumer will be inclined to purchase a different vehicle, which changes the cost per mile of driving and thus the responsiveness to changes in gasoline prices. For example, a consumer who recognizes that they will drive much more than average may be inclined to purchase a vehicle with higher fuel economy, implying a lower cost per mile of driving and reduced responsiveness to gasoline price changes. Since the consumer's preference for driving is unobserved, yet correlated with the cost per mile of driving, we may misattribute the influence of this unobserved preference to the cost of driving. Assuming that consumers who know they are going to drive more purchase higher fuel economy vehicles, this endogeneity would imply that the estimated responsiveness will be over-estimated, i.e., biased away from zero. Such a bias can be considered a "selection" bias, for consumers endogenously select into a vehicle. Thus, any analysis of utilization that does not simultaneously estimate vehicle choice and utilization will have biased and inconsistent estimates for how utilization would change as the cost per mile of driving changes.

My approach to addressing this selection bias has several innovations. First, my model of vehicle choice and subsequent utilization is a dynamic model that accounts for changing gasoline prices between the time of purchase and utilization. Second, it incorporates used vehicle prices into the purchase decision. If consumers recognize that higher gasoline prices will reduce the resale price of low fuel economy used vehicles, they may be inclined to purchase a higher fuel economy new vehicle. Third, the framework developed is a utility-consistent framework that allows for the analysis of a variety of policies on personal transportation.

The focus of the policy analysis is on two key policies: a increase in the gasoline tax and a revenue-neutral "feebate" policy that uses a surcharge on low fuel economy vehicles to pay for a rebate given to high fuel economy vehicles. Contrasting these two policies is valuable for understanding how a policy focused only on one margin – the extensive margin in the case of the feebate – has a very different cost-effectiveness than a policy that affects both the intensive and extensive margins. This is particularly important given that the primary policy instrument used explicitly to reduce emissions from vehicles at both the national- and California-level is a new vehicle performance standard on either fuel economy or greenhouse gases.[1] A new vehicle performance standard is similar to a feebate, in that they both are focused only on

[1]Note a carbon dioxide emissions standard is largely equivalent to a fuel economy standard for carbon dioxide emissions are, for the most part, a linear decreasing function of the fuel economy.

the vehicle choice margin.

This dissertation is organized as follows. The remainder of this chapter discusses the meaning behind the responsiveness to gasoline price changes, describes how the work in this dissertation fits into the academic literature, and previews the methodology and results in the remaining chapters. Chapter 2 describes the unique dataset assembled for this study and presents summary statistics. Chapter 3 provides some quantitative estimates of the responsiveness on both margins, and then explores the significant heterogeneity in the responsiveness. These regression results can control for many identification concerns, but do not account for the selection bias discussed above. To do so, one must simultaneously estimate vehicle choice and subsequent utilization. Chapter 4 presents the structural model developed to simultaneously estimate vehicle choice and utilization in a utility-consistent framework that also incorporates used vehicle prices in the new vehicle purchase decision. Chapter 5 illustrates what the results of the previous chapters imply for the effects of a gasoline tax and a feebate policy. Chapter 6 summarizes the key findings and provides concluding remarks.

1.1 Quantifying Consumer Responsiveness

This section takes a big picture look at the issues involved in understanding the consumer response to gasoline price changes and other policies to reduce personal transportation emissions. Specifically, I discuss the different ways consumers may respond and what these different responses mean for quantifying the responsiveness.

What policymakers are fundamentally interested in is how total gasoline consumption changes with the implementation of a policy. Similarly, forecasters may be interested in how total gasoline consumption changes as gasoline prices change. There are several useful ways consumer responsiveness can be quantified. When it comes to policy analysis, we are interested in the absolute and percentage change in gasoline consumption (and thus change in emissions) attributable to any given policy. More generally, economists are interested in quantifying how a *marginal* change in gasoline prices will change total gasoline consumption. The metric usually used to quantify this marginal change is the elasticity of gasoline consumption with respect to the price of gasoline.[2] This gasoline price elasticity gives a sense of what the response to a small policy would be, and is useful even for larger policies if we believe that the elasticity does not change with the magnitude of the policy. Moreover, an accurate measure of the gasoline price elasticity is very useful for forecasting purposes.

The elasticity of gasoline consumption with respect to the price of gasoline is determined by several factors. The amount of gasoline consumption from vehicle use is a function of the miles driven by each vehicle in the stock and on-road fuel economy of each vehicle in the stock. So, at the most basic level, the responsiveness to changing gasoline prices is determined by how the miles driven, the attributes of the stock of vehicles, and the on-road usage of those vehicles change when gasoline prices change. Drilling down into each of these factors, it is clear there is a changing response over time. One way to break this changing response down is into the short-run, medium-run, and long-run. While there is no consensus view on the time horizon of each of these, I find it useful to think about the short-run time frame as less than a year, medium-run as greater than a year but less than three years, and the long-run as greater than three years.

In the short-run, consumers can quickly change the miles driven in response to gasoline price changes by changing routes, combining trips ("trip-chaining"), switching to public transportation, or simply not taking discretionary trips. In fact, a change in the miles driven is likely to be the most dominant factor in the short-run, for the stock of the fleet is large relative to the new vehicles entering the fleet and old vehicles being scrapped and leaving the fleet.[3] The on-road fuel economy can be greatly influenced by the average speed on roadways and individual driving behavior, but the net change in gasoline consumption from this factor depends very much on the degree of congestion at an individual location.[4]

---

[2]The elasticity of gasoline consumption with respect to the price of gasoline is defined as the percentage change in gasoline consumption when the price of gasoline is increased by one percent.

[3]Dix and Goodwin (1982) make a similar point about traffic, rather than driving in general.

[4]In most vehicles fuel economy increases to around 50 or 55 miles and subsequently decreases. If higher gasoline prices imply that people drive less in highly congested areas, then the average on-road fuel economy may actually *increase*, further reducing gasoline consumption. On the other hand, if there is a reduction in driving on only somewhat congested freeways with a high speed

In the longer-term, the stock of vehicles will evolve with changes in gasoline prices. With changing gasoline prices, consumers will purchase different new vehicles and make different decisions about when to scrap an older vehicle. If gasoline prices increase consistently over the long term then vehicle manufacturers may also choose to direct technical change towards improving vehicle efficiency to meet consumer demand. Improving vehicle efficiency may involve improving the fuel economy or providing more vehicle services, such as horsepower, while keeping fuel economy constant (Sperling and Gordon 2008). The development time for a new vehicle model is in the order of five to six years, so major changes to the engine platform can occur within this time frame (Klier and Linn 2010b). Of course, within a five to six year time frame, consumers can also make other significant adjustments in response to longer-term changes in the price of gasoline, such as moving closer to work or public transportation.

Thus, when gasoline prices change, either due to exogenous events or a policy, we would expect to see a response that increases as we move from the short-term to the long-term. Correspondingly, it is essential to define the time frame of any given elasticity estimate. Moreover, in order to fully understand the time pattern of the impacts of a policy, it is useful to break down the price elasticity of total gasoline consumption into components. The two most important components are the price elasticity of driving and the elasticity of fuel economy of new vehicles with respect to the price of gasoline. In addition, the elasticity of fuel economy of scrapped vehicles with respect to the price of gasoline can also be important to the extent that consumers scrap different types of vehicles more often when the price of gasoline changes.[5] Breaking down the elasticity is useful both to understand the temporal pattern of response, as well as for more clearly understanding the geographic pattern of response, which may have important air pollution and congestion implications.

---

limit, then the average on-road fuel economy may *decrease*. In this case, the reduction in gasoline consumption would be less. This dissertation does not examine the issue of on-road fuel economy due to lack of detailed data on congestion. However, Greene, Kahn, and Gibson (1999) and Greene and Hu (1984) provide suggestive evidence that the effects of changes in how vehicles are driven are very small.

[5] Davis and Kahn (2011) show that older vehicles with low fuel economy are actually often exported to Mexico, where in many cases they are scrapped later, leading to higher lifetime emissions.

Breaking up the gasoline consumption elasticity into components is also useful for clarifying the decision process consumers undergo that determines the total amount of gasoline consumption. The demand for gasoline is a derived demand that comes about from the production of transportation services, where we can think of the technology for producing these services as being embodied in the vehicle. Vehicles are durable goods, so the decision of which vehicle to purchase depends inherently on how that vehicle will be used. In fact, classic studies of consumer durable goods (e.g., see Dubin (1985)) assume that the demand for consumer durable goods arises entirely from the flow of services provided through ownership of such goods. Of course, consumers may also receive some utility simply from purchasing a vehicle with a particular set of attributes. If we think about demand arising from the future flow of services, it follows that the demand for fuel economy – and the responsiveness of this demand to policy – depends on consumer expectations and the trade-off between money today and money in the future (i.e., the discount rate). Chapter 4 discusses the determinants of the demand for fuel economy in new vehicle purchasing in much greater detail.

One final point of distinction is worth making at this juncture. A classic characteristic of a durable good is one in which the very short-run price elasticity of demand is actually greater than the medium- or long-run price elasticity of demand. The intuition for this characteristic is that in the short-run consumers can usually defer purchases with the hope that prices come back down, while in the long-run consumers still demand the service. For vehicles we would expect to find the same characteristic. However, this effect focuses on the decision of when to purchase a vehicle, and is distinct from the decision of which vehicle to purchase. In particular, it is distinct from the decision of the choice of fuel economy in the bundle of attributes that a vehicle comprises. So we would not expect the short-run elasticity of fuel economy with respect to the price of gasoline to be greater than the medium- or long-run elasticity.

1.2 Literature Review

There is a vast and rapidly growing literature in economics relating to how consumers respond to changes in gasoline prices and other policies to reduce emissions from transportation. For the most part, this literature is centered around quantifying the response on the margin by estimating elasticities, although the exact elasticity being estimated varies, even for studies that aim to quantify similar responsiveness. For example, some studies may estimate the elasticity of gasoline consumption with respect to the price of gasoline, while others will estimate the elasticity of gasoline consumption with respect to the cost per mile of driving.

This literature review does not attempt to cover the entire literature (an appropriate subject for a book), but focuses on papers that relate more closely to this dissertation. For comparability purposes, I focus on studies using data from the United States. There are many studies using international data or European data that I do not review. Similarly, I focus the review on more recent studies and do not include many studies from before 1990. I divide up the literature review into subsections on total gasoline consumption responsiveness, new purchase vehicle purchase responsiveness, utilization responsiveness, heterogeneity in responsiveness, the rebound effect, and policy analysis.

1.2.1 Estimates of Responsiveness in Gasoline Consumption

There is an enormous literature going back many decades attempting to estimate the responsiveness of gasoline consumption to changes in the price of gasoline. Most of this literature uses aggregate data at the national or state level. Most of these studies primarily use time series variation to identify the responsiveness, although several studies use cross-sectional variation.

A starting point in any review of gasoline demand elasticities is a list of the many extensive survey articles. These articles survey the literature, and usually include a meta-analysis bringing together the many different estimates found in previous work. Some of the more well-known survey articles include Blum, Foos, and Gaudry (1988), Dahl and Sterner (1991), Sterner and Dahl (1992), Dahl (1995), Espey (1998), de Jong and Gunn (2001), Graham and Glaister (2002), Goodwin, Dargay, and Hanly (2004), and Basso and Oum (2007). All of these surveys list the results of many studies and some go further to include details and comparisons of specific studies. Most of these surveys provide surprisingly little critical analysis of studies being cited. Basso and Oum (2007) is one exception, which goes into great depth in providing a critical analysis, yet neglects to discuss whether the reviewed studies can properly identify the coefficients of interest, as is the recent focus in applied econometrics. I will not critically review each and every study here, but will provide a brief overview of the general findings.

To give a sense of the range of values in the literature, I put together a fairly complete list of the recent empirical work estimating the price elasticity of gasoline demand in the US. Table 1.1 lists these studies, along with relevant details about the studies, and the estimated short-run or long-run elasticity values.

There are a few clarifying points worth mentioning about Table 1.1. First, there is no consistent definition of short-run and long-run in these studies. As discussed above, the concept behind separating out the short-run from the long-run elasticities is simply that in the long-run, consumers will have more time to adjust. One way to think about interpreting the time frame of the elasticities is to consider that the time frame depends at least partly on the identifying variation used in the estimation. For example, if monthly fixed effects are used with time series data, as in Hughes, Knittel, and Sperling (2008), then within-month variation is being used, so the elasticity estimate is perhaps best viewed as a monthly elasticity. With no fixed effects, then the length of the time period of the data may be a useful guideline, although it raises the question of whether time-varying unobservables may be confounding the estimate. If cross-sectional data are used, then it is common to interpret the elasticity estimate as a long-run elasticity. The rationale for this interpretation is that the price differences being used to identify the elasticity are based on long-run or "steady state" differences between regions. Of course, it may be extremely difficult to identify the price elasticity with purely cross-sectional data for it is impossible to account for unobserved heterogeneity across regions that may be correlated with the coefficient.[6]

A second way to think about interpreting the time frame of the elasticities is to use a dynamic model. If we are using the time frame of the identifying variation,

[6] For example, (Pickrell and Schimek 1999) find that state gasoline prices are correlated with population density, so that unobserved differences between urban and rural areas may be a major issue in studies using cross-sectional survey data

Table 1.1: Literature on the Price Elasticity of Gasoline Demand

Paper	data type	source	time frame	method	short-run	long-run
Gately (1992)	national time series	FHWA	1966-1989	OLS; LDV	-0.1	–
Hausman and Newey (1995)	repeated cross-section	RECS	1979-1981	Semip	–	-0.81
Haughton and Sarkar (1996)	panel, US states	FHWA	1970-1991	OLS; LDV	-0.12 to -0.17	-0.23 to -0.36
Schimek (1996)	national time series	FHWA	1950-1994	OLS; LDV	-0.12	-0.7
Puller and Greening (1999)	repeated cross-section	CEX	1980-1990	2SLS	-0.49	–
Agras and Chapman (1999)	meta-analysis		1982-1995		-0.25	-0.92
Schmalensee and Stoker (1999)	cross-section	RTECS	1988-1991	OLS; Semip		-0.29 to -1.1
Kayser (2000)	panel, households	PSID	1981	Heckman	-0.23	–
Graham and Glaister (2002)	meta-analysis				-0.2 to -0.5	-0.24 to -0.8
Goodwin, Dargay, and Hanly (2004)	meta-analysis				-0.25	-0.64
West and Williams (2004)	cross-section	CEX	1996-1998	AIDS		-0.18 to -0.74
Small and Van Dender (2007)	repeated cross-section	FHWA	1966-2001	3SLS; LDV	-0.045 to -0.089	–
Small and Van Dender (2007)	repeated cross-section	FHWA	1997-2001	3SLS; LDV	-0.022 to -0.067	-0.43
Hughes, Knittel, and Sperling (2008)	national time series	EIA	1975-1980	3SLS	-0.21 to -0.34	–
Hughes, Knittel, and Sperling (2008)	national time series	EIA	2001-2006	3SLS	-0.034 to -0.077	–
Wadud, Graham, and Noland (2009)	repeated cross-section	CEX	1984-2003	SUR	-0.351 to -0.203	–
Bento et al. (2009)	cross-section	NHTS	2001	Bayes	–	-0.35
Lin and Prince (2009)	CA time series	CA	1970-2007	OLS; LDV; IV	–	-0.22
Wadud, Graham, and Noland (2010a)	repeated cross-section	CEX	1997-2002	OLS; RE	–	-0.5
Wadud, Graham, and Noland (2010b)	repeated cross-section	CEX	1997-2002	Semip	–	-0.21 to -0.63
Hymel, Small, and Van Dender (2010)	repeated cross-section	FHWA	1966-2004	3SLS, MI	-0.075	-0.36
Hymel, Small, and Van Dender (2010)	repeated cross-section	FHWA	1984-2004	3SLS	-0.054	-0.18
Davis and Kilian (2011b)	panel, US states	EIA	1989-2008	OLS; 2SLS	-0.10 to -0.46	–
Manzan and Zerom (2011)	repeated cross-section	RTECS	1991-1994	Semip	–	-0.2 to -0.5

Notes: FHWA = Federal Highway Administration; RECS = Residential Energy Consumption Survey; CEX = Consumer Expenditure Survey; RTECS = Residential Transportation and Energy Consumption Survey; PSID = Panel Study of Income Dynamics; EIA = Energy Information Administration; NHTS = National Household Travel Survey; CA = California Energy Commission and California Equalization Board; OLS = Ordinary Least Squares; LDV = Lagged Dependent Variable; 2SLS = Two Stage Least Squares; Semip = Semiparametric estimation; Heckman = Heckman selection model; AIDS = Almost Ideal Demand System; 3SLS = Three Stage Least Squares; RE = Random Effects; Bayes = Bayesian Markov Chain Monte Carlo; MI = Multiple Imputation.

then the estimation can yield only a single elasticity estimate. However, for decades there have been studies that have attempted to identify both the long-run and short-run elasticity by incorporating dynamics through the inclusion of a lagged dependent variable. This approach is often called a "partial adjustment process" because it roughly captures the inertia in gasoline consumption due to the time it takes for the stock of vehicles to turn over. We can formalize this common approach as follows. Denote Y_t as the dependent variable at time t (i.e., gasoline consumption), P_t is the price of gasoline, and $\boldsymbol{X}_t$ is a vector of other covariates (e.g., income). Without loss of generality, assume $\boldsymbol{X}_t$ is one dimensional for ease of notation. Thus we can denote $\boldsymbol{X}_t$ as X_t. The most common specification in most studies with dynamic models is simply

$$\log(Y_t) = \beta_0 + \beta_1 \log(P_t) + \beta_2 \log(Y_{t-1}) + \beta_3 \log(X_t) + \varepsilon_t \tag{1.1}$$

where ε is a mean-zero stochastic error term. The common approach runs ordinary least squares (OLS) on this specification and then interprets β_1 as the short-run price elasticity of gasoline demand. To get the long-run elasticity, studies assume that we have a dynamic system that converges in the long-run to a steady state. In steady state $\log(Y_t) = \log(Y_{t-1}) \equiv \log(Y)$, so the equation can be rewritten as

$$\log(Y) = \beta_0 + \frac{\beta_1}{1-\beta_2} \log(P_t) + \frac{\beta_3}{1-\beta_2} \log(X_t) + \frac{1}{1-\beta_2} \varepsilon_t.$$

With this specification in mind, $\frac{\beta_1}{1-\beta_2}$ is commonly interpreted as the long-run price elasticity of gasoline demand. Papers such as Haughton and Sarkar (1996) and Schimek (1996) are classic examples of this approach.

Despite being relatively commonly accepted in the energy economics literature, there are a few reasons why we may be concerned about using this approach. The first is simply whether we believe the model of the gasoline markets approaching a theoretical steady state. From historical observation, it appears that the gasoline market does not settle down to a steady state, but is in a constant state of dynamic disequilibrium. This observation does not necessarily invalidate the approach, but it does raise the question of what the long-run estimates from this approach really mean.

A second concern is that the OLS estimator of β_2 is likely to be biased in a finite sample. Moreover, if there is serial correlation of the error terms, then the OLS estimator is not consistent. To see this, consider the simplified setting where the true model is

$$Y_t = \rho Y_{t-1} + \varepsilon_t,$$

where ε_t is an i.i.d stochastic error term.

For unbiasedness we require that $\mathbb{E}[\hat{\rho}|Y_{t-1}] = \rho$ where $\hat{\rho}$ is the OLS estimate of ρ. This condition is equivalent to showing the more common zero conditional mean condition: $\mathbb{E}[\varepsilon_t|Y_{t-1}] = 0$. But by rolling back Y_t to the starting point, we can recognize that $Y_0 = \varepsilon_0$ and $Y_1 = \rho Y_0 + \varepsilon_1$. So it is easy to see that $\mathbb{E}[\varepsilon_1|Y_0] = \mathbb{E}[\varepsilon_1|\varepsilon_0] = \varepsilon_1 \neq 0$. Since $\mathbb{E}[\varepsilon_1|Y_0] \neq 0$, we can follow the logic up the chain to see that $\mathbb{E}[\varepsilon_t|Y_{t-1}] \neq 0 \; \forall t$. Thus, the conditional moment condition necessary for unbiasedness in the classical OLS model does not hold.

However, OLS may still be consistent, for if the more restrictive conditional moment condition $\mathbb{E}[\varepsilon|Y_{t-1}, ..., Y_0] = 0$ holds, then we can still say that $\frac{1}{T}\sum_t Y_{t-1}\varepsilon_t \xrightarrow{p} 0$. But this is not the case if we have serial correlation of the errors. Suppose we have serial correlation, so that $\varepsilon_t = \theta\eta_{t-1} + \eta_t$, where $\eta_t \sim$ i.i.d. $\mathcal{N}(0, \sigma^2)$. We can then see that $\mathbb{E}[\varepsilon_t\varepsilon_{t-1}] = \theta\mathbb{E}[\eta_{t-1}\eta_{t-1}] = \theta\sigma^2$. Using this, we can see that $\forall t$

$$\begin{aligned}
\mathbb{E}[\varepsilon_t Y_{t-1}] &= \mathbb{E}[\varepsilon_t(\rho Y_{t-2} + \varepsilon_{t-1})] \\
&= 0 + \mathbb{E}[\varepsilon_t\varepsilon_{t-1}] \\
&= \theta\sigma^2 \\
&\neq 0.
\end{aligned}$$

This result shows that the stronger conditional moment condition $\mathbb{E}[\varepsilon|Y_{t-1}, ..., Y_0] = 0$ cannot hold and that $\frac{1}{T}\sum_t Y_{t-1}\varepsilon_t \xrightarrow{p} \theta\sigma^2 \neq 0$. Going back to the specification in (1.1), this result indicates that we certainly should be concerned about the estimate of β_2 when there is evidence of serial correlation of the error terms. To the extent that the lagged dependent variable is correlated with price, we may also be concerned about the estimate of β_1, which gave us the short-run elasticity.

What does this mean for the validity of the estimates in the literature? It suggests that if we have a small sample (e.g., a yearly national time series over only a few decades), estimates using a specification with a lagged dependent variable are very likely to be biased. Just as importantly, it suggests that if we have serial correlation of the error terms, as is very common in time series studies, then OLS is very likely to yield inconsistent estimates. I view this as suggesting that we should be somewhat careful about the interpretation of coefficients from specifications that claim to provide both short-run and long-run estimates. If there is no evidence of serial correlation (e.g., the Durbin-Watson test fails to reject the null of no serial correlation) and the sample is very large, then we may not have a issue. From my review, it appears that much of the older literature includes a lagged dependent variable and suffers from this issue. The importance of this issue likely varies by study.

A second clarifying point about the studies reviewed in Table 1.1 relates to endogeneity from the simultaneous determination of prices and quantities in the gasoline market. The specification given in (1.1) appears to be a demand equation, with the quantity demanded for gasoline as the dependent variable and price included as a regressor. However, since prices and quantities are determined in the market by both supply and demand, without estimating the system of equations simultaneously, the coefficient on the price variable will inherently be biased and inconsistent. Many of the early studies only specify and estimate a demand equation using market-level data and thus suffer from this bias. Several of the more recent studies, such as Davis and Kilian (2011b), attempt to address this bias using instrumental variables or other techniques. The difficulty is in finding convincing instrumental variables, for even well-considered instruments (e.g., state tax rates as in Davis and Kilian (2011b)) can easily be criticized as being invalid.

Despite these caveats about the estimation approaches in much of the previous literature, Table 1.1 still provides a sense of the magnitude of the price elasticity of gasoline demand and the trends over time. To generalize, it appears that the short-run price elasticity of gasoline demand has been in the range of -0.1 to -0.3 and more recently has been closer to -0.05 to -0.1. The long-run elasticity appears to have been all over the map, with a range of estimates as wide as -0.2 to -0.8. More recent estimates appear to be in the -0.18 to -0.4 range, with more of the estimates in the upper half of that range.

Small and Van Dender (2007) and Hughes, Knittel, and Sperling (2008) provided some of the first discussion of this apparent decline in responsiveness. Small and Van Dender suggest that this decline in responsiveness may be due to increased incomes and congestion influencing the consumer decisions about how much to drive. Small and Van Dender hypothesize that with greater wealth, consumers will have a greater value of time, and thus the fuel cost of driving will be less important than the time cost of driving – so that consumers will be less responsive to changes in the fuel price. Similarly, Small and Van Dender suggest that a similar process is at work for congestion: as congestion increases and consumers spent more time in traffic, they become more concerned about the time cost of driving and less concerned about the fuel cost of driving.

These stories for why the responsiveness to gasoline prices may be decreasing are quite plausible and correspond well with the empirical evidence. Another possibility is simply that there was not much variation in gasoline prices over many of the more recent years in the time frame of both Small and Van Dender (2007) and Hughes, Knittel, and Sperling (2008). If consumers do not respond to small changes in gasoline prices, such as those within the range of the seasonality of gasoline prices, then recent studies would not be likely to pick up much responsiveness. An important contribution of this dissertation is to examine more recent data that covers a time with very substantial gasoline price changes. I will discuss this issue more and its implications for policy in several chapters in this dissertation.

A final point to address in Table 1.1 is the occasional use of semi-parametric techniques to estimate the price elasticity of gasoline demand. For example, in Table 1.1, Hausman and Newey (1995), Schmalensee and Stoker (1999), Wadud, Graham, and Noland (2010b), and Manzan and Zerom (2011) all use semi-parametric estimation approaches. Yatchew (2001) use a similar semiparametric approach on data from Canada. These approaches allow for extremely flexible functional specifications of the relationship between gasoline demand and other covariates, including the gasoline price. In each of these models, some of the covariates enter in a parametric fashion, and hence the approaches are semiparametric, rather than purely nonparametric. Hausman and Newey (1995) find that large differences between the semiparametric and parametric approaches in the estimated elasticities. More recently, Wadud, Graham, and Noland (2010b) find much smaller differences and suggest that there are not major gains to be had by using the semiparametric approach. Given that the estimates of elasticities from the semiparametric approaches are well within the range of the rest of the literature, it may be reasonable to consider a semiparametric approach as a useful robustness check, rather than as the primary modeling approach. The concern about finding clean identification of the coefficients may well be far more important than the choice of a parametric or semiparametric approach.

1.2.2 Estimates of Responsiveness in Vehicle Utilization

Much of the responsiveness in gasoline demand to higher gasoline prices can be expected to occur by consumers choosing to drive less. One would expect this vehicle utilization response to be just slightly less responsive than the full gasoline consumption response, with the difference being made up by the purchase of higher fuel economy vehicles and other adjustments (e.g., pressing on the gas pedal more lightly at green lights). Thus, many of the same estimation approaches and concerns apply just as much to estimating vehicle utilization elasticities as to estimating gasoline demand elasticities. In fact, many of the papers in the literature use a similar specification to (1.1), only with vehicle miles traveled (VMT) as the dependent variable.

There are a surprising number of papers that estimate a utilization elasticity for vehicles. Many of these papers estimate this elasticity in order to weigh in on the magnitude of the "rebound effect" of increased driving due to lowering the cost per mile of driving with higher fuel economy. Starting with Greene (1992), many analysts have equated the rebound effect with the elasticity of VMT with respect to the cost per mile of driving. The elasticity of VMT with respect to the cost per mile of driving is a pure utilization elasticity (i.e., it captures how utilization changes when the price of utilization changes). Both a change in the price of gasoline and a change in the fuel economy of the vehicle would affect the cost per mile of driving. One might expect that changing either of these would lead to a similar response by consumers, for both change the cost per mile similarly. One story for why the response may differ is consumers respond differently to changing gasoline prices at the pump than if they are induced to purchase a vehicle with higher fuel economy. I discuss this concept in greater depth, along with a review of the rebound effect in Section 1.2.5.

Instead of the elasticity of VMT with respect to the cost per mile of driving, many studies are more interested simply in how consumers change the amount driven when the price of gasoline changes, i.e., the elasticity of VMT with respect to the price of gasoline. These studies are estimating a similar elasticity to the price elasticity of gasoline demand, and often use the same approaches. Table 1.2 lists the most widely cited recent empirical work on the vehicle utilization elasticity, separating out studies that estimate a VMT elasticity with respect to the cost per mile of driving from the VMT elasticity with respect to the price of gasoline.

Just as with the price elasticity of gasoline demand, we can learn a few things simply by looking at the range of estimates in the literature. For both the VMT elasticity with respect to the price of gasoline and the VMT elasticity with respect to the cost per mile of driving, the range of estimates appears to be around -0.05 to -0.2 in the short-run and around -0.15 to -0.8 in the long-run. Of course, the same caveats mentioned above about the difficulty of ascertaining the time frame of the elasticity apply equally here.

Not surprisingly, just as there was evidence that the responsiveness in gasoline consumption has been declining over time, there is also some evidence that the responsiveness in driving has been declining over time (e.g., note the low estimates of Small and Van Dender (2007), Hymel, Small, and Van Dender (2010), and Greene (2011)). Similarly, the low estimates of responsiveness may be due to rising incomes and congestion. They may also be partly due to the limited changes in gasoline prices in more recent years.

Table 1.2: Literature on the Elasticity of VMT in the United States

VMT elasticity with respect to the price of gasoline						
Paper	data type	source	time frame	method	short-run	long-run
Sweeney (1979)	national time series	FHWA	1957-1974	OLS		-0.12 to -0.23
Gately (1992)	national time series	FHWA	1966-1989	OLS	-0.11	-0.11
Haughton and Sarkar (1996)	panel, US states	FHWA	1970-1991	OLS; LDV	-0.09 to -0.16	-0.22
Agras and Chapman (1999)	meta-analysis		1982-1995		-0.15	-0.32
Pickrell and Schimek (1999)	repeated cross-section	NPTS	1969-1995	OLS	–	-0.04 to -0.34
Kayser (2000)	panel, households	PSID	1981	Heckman	-0.23	–
de Jong and Gunn (2001)	meta-analysis				-0.16	-0.26
Goodwin, Dargay, and Hanly (2004)	meta-analysis				-0.1	-0.3
Austin (2008)	meta-analysis				-0.10 to -0.16	-0.26 to -0.31
Lin and Prince (2009)	CA time series	CA	1970-2007	OLS; LDV; IV	–	-0.07
Hymel, Small, and Van Dender (2010)	repeated cross-section	FHWA	1966-2004	3SLS; LDV; MI	-0.047	-0.241
Hymel, Small, and Van Dender (2010)	repeated cross-section	FHWA	1984-2004	3SLS; LDV	-0.048	-0.159
Greene (2011)	national time series	FHWA	1967-2007	OLS; LDV	-0.05	-0.3
VMT elasticity with respect to the cost per mile of driving						
Paper	data type	source	time frame	method	short-run	long-run
Mannering and Winston (1985)	repeated cross-section	EIA	1977-1979	OLS; LDV	-0.228	-0.279
Mayo and Mathis (1988)	national time series	FHWA	1958-1984	GLS	-0.22	-0.26
Greene (1992)	national time series	FHWA	1957-1989	OLS; LDV	-0.05 to -0.19	–
Greene (1992)	national time series	FHWA	1966-1989	OLS; LDV	-0.09	–
Jones (1993)	national time series	FHWA	1966-1990	OLS; LDV	-0.11 to -0.13	-0.3
Schimek (1996)	repeated cross-section	FHWA	1950-1994	OLS; LDV	-0.05 to -0.17	-0.13 to -0.42
Goldberg (1998)	repeated cross-section	CEX	1984-1990	IV	0.0 to -0.2	–
Greene, Kahn, and Gibson (1999)	repeated cross-section	RTECS	1979-1994	3SLS	–	-0.23
Greening, Greene, and Difiglio (2000)	meta-analysis		1990-2000		-0.1	-0.20 to -0.30
West (2004)	cross-section	CEX	1997	CF	–	-0.87
Small and Van Dender (2007)	repeated cross-section	FHWA	1966-2001	3SLS; LDV	-0.045	-0.22
Bento et al. (2009)	cross-section	NHTS	2001	Bayes	–	-0.74

Notes: FHWA = Federal Highway Administration; NPTS = Nationwide Personal Transportation Survey; PSID = Panel Study of Income Dynamics; CA = California Energy Commission and California Air Resource Board; EIA = Energy Information Administration; CEX = Consumer Expenditure Survey; RTECS = Residential Transportation and Energy Consumption Survey; NHTS = National Household Travel Survey; OLS = Ordinary Least Squares; LDV = Lagged Dependent Variable; Heckman = Heckman selection model; 3SLS = Three Stage Least Squares; MI = Multiple Imputation; GLS = Generalized Least Squares; IV = Instrumental Variables; CF = Control Function Approach; Bayes = Bayesian Markov Chain Monte Carlo.

A few studies here are worth further elaborating on. Kayser (2000) uses a Heckman selection model to address a selection issue in the particular dataset being used in the study. In the PSID data, many households do not own a vehicle, and thus the demand for driving is censored at zero. The approach taken by Kayser aims to address this issue. West and Williams (2004) use a similar approach in the estimation of the price elasticity of gasoline demand. However, the selection issue dealt with in Kayser (2000) and West and Williams (2004) applies only in datasets that include consumers that do not own a vehicle. The Dubin and McFadden (1984) selection issue (i.e., consumers with different expected driving select into different fuel economy vehicles) is more broadly an issue, and certainly affects any study using microdata. The Dubin and McFadden (1984) selection issue is dealt with in varying ways in Goldberg (1998), West (2004), Bento et al. (2009), and arguably Small and Van Dender (2007).[7]

The methods used to address the Dubin and McFadden (1984) selection issue all are conceptually similar in that the estimated choice probabilities from a vehicle choice model are used to adjust the utilization estimation. Dubin and McFadden (1984) suggests three approaches to address the issue. Goldberg (1998) takes the "instrumental variables" approach and instruments for the price of utilization with the fitted choice probabilities from a separate nested logit vehicle choice estimation. Mannering and Winston (1985) takes the second approach, the "reduced form approach," which includes an indicator for vehicle of type i being chosen in the utilization equation and then replaces this indicator with the estimated choice probability of vehicle i. West (2004) takes the third approach by including a control function (i.e., the inverse Mills ratio) in a Heckman-style selection model, where the exclusion restrictions are the previously estimated choice probabilities from a vehicle choice estimation. Dubin and McFadden (1984) describes this approach as the "conditional expectation correction method." Bento et al. (2009) (as well as Feng, Fullerton, and Gan (2005) and Jacobsen (2010)) use the Dubin and McFadden (1984) framework, but simultaneously estimate vehicle choice and utilization in a static setting. Chapter 4 discusses

[7] I use "arguably" in reference to Small and Van Dender, because Small and Van Dender estimate a system of equations using aggregate data, with one of the three equations being a fuel intensity equation, so that the choice of fuel economy of the fleet is effectively chosen simultaneously with the choice of how much the fleet is driven.

the selection bias and how these approaches address it in much greater detail.

1.2.3 Estimates of Responsiveness in New Vehicle Purchases

In the long-run, how the fuel economy of new vehicles entering the light duty fleet adjusts with changing gasoline prices is a critical determinant of the price elasticity of gasoline demand. There is a large and important empirical literature on estimating the demand for new vehicles, with key papers including Bresnahan (1981), Berry, Levinsohn, and Pakes (1995), Petrin (2002), and Berry, Levinsohn, and Pakes (2004). However, there is a more limited literature on how the demand for new vehicles adjusts with gasoline prices. Moreover, not all of the studies estimate an elasticity value, for some simply calculate the shifts in sales in each of the fuel economy quartiles. Busse, Knittel, and Zettelmeyer (2010) and Li, Timmons, and von Haefen (2009) fall into this category, and both find evidence of a shift towards higher fuel economy vehicles when gasoline prices rise. Table 1.3 lists several of the studies that estimate a new vehicle fuel economy elasticity for vehicles in the United States.

The studies in Table 1.3 suggest that consumers do respond to gasoline price changes by purchasing higher fuel economy vehicles, but the responsiveness is relatively inelastic. The short-run elasticity of fuel economy with respect to the price of gasoline appears to be in the range of 0.05 to 0.2. The long-run response may be on the higher end of this range if we take the long-run estimates in the literature seriously. However, the same critiques discussed above about short-run and long-run estimates also apply here, and perhaps even more strongly. Some studies, such as Austin and Dinan (2005) are not clear whether the estimated elasticity is long-run or short-run (I place it as long-run because the dataset is cross-sectional). But even more worrisome is that in the long-run, there will be both supply-side and demand-side responses, while nearly all of the models only include a demand-side (Austin and Dinan (2005) is an exception). Klier and Linn (2010b) uses monthly fixed effects and vehicle model-year (i.e., the vehicle model interacted with year) fixed effects to attempt to nonparametrically control for endogeneity of the average retail price. The concern Klier and Linn (2010b) are attempting to address is the possibility that the average

Table 1.3: Literature on the Elasticity of New Vehicle Fuel Economy

Fuel economy elasticity with respect to the price of gasoline						
Paper	data type	source	time frame	method	short-run	long-run
Mayo and Mathis (1988)	national time series	FHWA	1958-1984	OLS	0.21	–
Gately (1992)	national time series	FHWA	1966-1989	OLS	0.001	–
Schimek (1996)	national time series	FHWA	1950-1994	OLS; LDV	0.05	0.23
Agras and Chapman (1999)	meta-analysis		1982-1995		0.12	0.60
Austin and Dinan (2005)	cross-section	Sev	2001	Simul	–	0.22
Klier and Linn (2010b)	national time series	Wards	1970-2007	OLS; FE	0.12	–
Gallons per mile elasticity with respect to the price of gasoline						
Paper	data type	source	time frame	method	short-run	long-run
Haughton and Sarkar (1996)	panel, US states	FHWA	1970-1991	OLS; LDV	-0.09 to -0.17	-0.51 to -0.66

Notes: FHWA = Federal Highway Administration; Sev = Several sources (Edmonds, EPA); OLS = Ordinary Least Squares; LDV = Lagged Dependent Variable; Simul = Simulation model; FE = fixed effects.

retail price may be correlated with unobserved vehicle characteristics, as discussed in many papers, such as Berry, Levinsohn, and Pakes (1995) and Nevo (2000).

The Klier and Linn (2010b) approach may perhaps be useful for the short-run, but estimation of the long-run responsiveness inherently involves modeling the supply-side responses. Recent evidence suggests that firms certainly take into account incentives, such as gasoline prices and fuel economy standards in product design decisions (Whitefoot, Fowlie, and Skerlos 2011; Knittel 2011), which determine the long-run fuel economy of the fleet. Given this, I view long-run estimates that do not include a supply-side as at least somewhat suspect.

How exactly might we expect the response in the new vehicle fleet fuel economy to occur? In the short-run, Busse, Knittel, and Zettelmeyer (2010) suggest that there is more of an adjustment in the quantity of new vehicles sold in different fuel economy quartiles than in the prices of those vehicles. For example, when the price of gasoline goes up, fewer vehicles in the lowest quartile are sold. Following this logic, the manufacturer response to higher gasoline prices must be to cut production of low fuel economy vehicles rather than reducing prices of those vehicles to keep sales up.[8] Of course, there may also be selective marketing techniques such as low-interest financing or rebates to dealers when gasoline prices increase. To the best of my knowledge, no articles in the literature have been able to address the link between gasoline prices and these other manufacturer responses.

In the medium-run, Klier and Linn (2010a) suggest that vehicle manufacturers can make some adjustments to vehicle weight, power, and fuel economy, but are unlikely to redesign the engine of the vehicle. In this context, Klier and Linn (2010a) suggest that such major adjustments may occur every four years or so, while the design cycle for a new engine technology is on the order of ten years. To the extent that this is true, we may expect to see full supply-side responses to changes in the price of gasoline in the four to ten year range. The full dynamics of this process of manufacturer response has yet to be explored, although the model in Whitefoot, Fowlie, and Skerlos (2011) may be a promising framework for such a research endeavor.

[8]In contrast, Busse, Knittel, and Zettelmeyer (2010) show that in the used vehicle fleet, the prices of vehicles tend to adjust more than the quantity of vehicles sold - which is an intuitive result when one considers that the supply of used vehicles is nearly fixed.

1.2.4 Evidence of Heterogeneity in Responsiveness

The vast majority of the empirical work on the responsiveness to gasoline price changes has been focused on pinning down a particular elasticity value. Only a handful of papers have discussed the heterogeneity in responsiveness. One might expect that different types of consumers would respond very differently to changes in the price of gasoline depending on their driving needs and budget. One motivation for understanding the heterogeneity in responsiveness is that it plays a role in any policy analysis of the distributional consequences of a policy that raises gasoline prices. Indeed, several of the studies that examine the heterogeneity in the responsiveness of gasoline demand to changing gasoline prices also perform an analysis of the distributional consequences of a gasoline tax.

Schmalensee and Stoker (1999) was one of the first papers to examine heterogeneity in the elasticity of gasoline demand, yet provided only limited guidance on the price elasticity and focused much more on heterogeneity in the income elasticity of gasoline demand. West (2004) and West and Williams (2004) examine heterogeneity in responsiveness by estimating the response by expenditure decile and quintile respectively. West (2004) looks at the heterogeneity in the price elasticity of VMT demand and finds that the lowest household expenditure decile has over a 50% greater responsiveness to gasoline price than the highest household expenditure decile. The estimates in West (2004) suggest that there is an U-shaped pattern of responsiveness, where the least responsive decile (-0.75 elasticity) is the eighth decile, so that the first (i.e., lowest income) and last (i.e., highest income) deciles are more responsive. West (2004) does not provide an explanation for this particular pattern, although it may be related to a tighter budget constraint at the low end and lower marginal utility of additional driving at the higher end. West and Williams (2004) also find that the lowest expenditure quintile (elasticity of -0.7) has over three times the responsiveness of the highest expenditure quintile (elasticity of -0.2). There is no evidence of a U-shaped pattern in the estimates of West and Williams (2004).

Bento et al. (2009) use the 2001 National Highway Transportation Survey to examine the heterogeneity in the price elasticity of gasoline demand by household type, class of vehicle, and age of vehicle. Bento et al. find that families with children and owners of trucks and sport utility vehicles are more responsive to changes in gasoline prices. The magnitudes of the differences are surprisingly small though. For example, households with children have a long-run price elasticity of -0.39, while the estimated elasticity for all other households is -0.32. Similarly, the elasticities by type of vehicle range from -0.27 for compact cars to -0.34 for small trucks. Nearly all other classes are in the -0.28 to -0.32 range. Bento et al. (2009) also find that the age of the vehicle seems to make very little difference in the elasticity. Bento et al. (2009) do not include standard errors on the different elasticity estimates, so it is difficult to know whether these differences are statistically significant.

Wadud, Graham, and Noland (2009) and Wadud, Graham, and Noland (2010a) use household expenditure data from the Consumer Expenditure Survey to examine heterogeneity in responsiveness. Wadud, Graham, and Noland (2009) perform separate estimations for each household expenditure quintile and find a U-shaped pattern of price elasticities of gasoline demand, similar to the finding in West (2004). In Wadud, Graham, and Noland (2009), the least responsive quintile is the third quintile, so that the first and last quintiles are more responsive. The first (i.e., lowest income) quintile is the most responsive, with an estimated price elasticity of gasoline demand of -0.35.

Wadud, Graham, and Noland (2010a) take a different approach and follow studies in Canada and the United Kingdom (Yatchew 2001; Santos and Catchesides 2005) by interacting the price of gasoline with household expenditure. Wadud, Graham, and Noland (2010a) also interact the price of gasoline with an indicator variable for a rural location. Wadud, Graham, and Noland (2010a) then examine the price elasticity of gasoline demand for households with different characteristics by simply plugging in these characteristics into the estimated model. They find that the responsiveness to gasoline prices declines monotonically with income. The monotonicity may not be surprising in this analysis, for it was effectively imposed by the specification. Wadud, Graham, and Noland also find that the responsiveness is higher for consumers in urban areas than rural areas and and generally somewhat higher for households with multiple vehicles. Since much of the variation in Wadud, Graham, and Noland (2010a) is cross-sectional, and gasoline prices are generally higher in urban areas than rural areas (largely due to higher gasoline sales taxes), it is possible that some of the urban versus rural result stems from the differing price levels. This would be a concern if consumers respond more when gasoline prices are already higher. One argument for this is that higher gasoline prices are a larger fraction of the consumer budget and thus are more salient (Greene 2011).

1.2.5 The Rebound Effect

Background

The "rebound" or "take-back" effect is a critical issue for policies to promote higher fuel economy of the vehicle fleet, and accordingly has received a great deal of attention in the energy economics and policy literature. It is defined in various ways in different studies, making comparability across studies difficult at times. The basic idea is simple: as fuel economy is improved, the cost per mile of driving decreases, leading to a "rebound" of energy use. There are several pathways by which this effect may occur and interpretations of the attempts to quantify this effect in the literature. The rebound effect is important for energy efficiency policy analysis simply because it determines the reduction in energy and emissions possible from the policy.

To begin, I find it constructive to present a definition of the rebound effect that gets at the heart of the issue and can be applied to any policy to improve energy efficiency.

Definition 1 ***The Rebound Effect*** - *The rebound effect of a policy leading to an improvement in the technical energy efficiency of a good is the additional energy use due to the decrease in the cost of utilization of the good.*

In the context of vehicles, this can be translated to say "the rebound effect of a policy to improve fuel economy is the increase in energy use due to the decrease in the cost per mile of driving." Implicit in this definition is that the non-driving energy services provided (i.e., the other attributes of the vehicle) do not change along with the energy efficiency improvement, although for a policy analysis we may also be interested in the additional energy use including any changes in energy services.

The definition here differs slightly from the previous literature by linking the rebound effect to a particular policy and stating it in a more precise and useful way. In contrast, much of the literature is either quite vague about what is meant by the rebound effect or not careful about the relating the rebound effect to a particular policy.

To clarify further, we can begin with the genesis of the concept of the rebound effect. In 1865, William Stanley Jevons noted that the improvement of engine technology not only increased the efficiency at which coal is used, but it also made coal economical as a fuel for many other uses, corresponding with an increase in overall coal use (Jevons 1865). This effect is sometimes called "Jevons' Paradox." Brookes (1979) and Khazzoom (1980) applied a similar concept to suggest that the use of policies to promote energy efficiency would lead to an overall *increase* in energy use, above the no-policy counterfactual. Saunders (1992) coined the term the "Khazzoom-Brookes Postulate" to refer to this possibility and provided some theory to suggest that it could occur in a neo-classic growth model by spurring additional economic growth. Others have referred to the possibility that energy efficiency policy could increase energy use above the no-policy counterfactual as "backfire" (Saunders 1992). "Backfire" can be considered an extreme case, where the rebound effect is so large that it overtakes all of the energy savings from improved energy efficiency. Even in this extreme case, the policy may be economic-efficiency improving, for more energy services are being provided.

There are several pathways that could lead to energy use increasing when improved energy efficiency reduces the cost of utilization. Greene (1992) lays out what is possibly the most useful typology with the following three categories: the direct rebound effect, indirect rebound effect, and macroeconomic or general equilibrium rebound effect. The *direct rebound effect* is the most obvious. An energy efficiency policy that reduces the cost per mile of driving makes it less expensive to drive, so that consumers drive more. Thus, some of the energy savings from the energy efficiency policy are lost from the energy use of the additional driving. The *indirect rebound effect* is slightly more subtle. An energy efficiency policy that reduces the cost per mile of driving makes it less expensive to drive, so that consumers effectively have more income to spend on other goods. The purchase and use of these other goods may require energy, eroding the energy savings from the energy efficiency policy. Finally, the *macroeconomic rebound effect* refers to broader price and economic growth effects. An energy efficiency policy that reduces the cost per mile of driving for all vehicles in a country as large as the United States may actually shift the global gasoline demand curve inwards enough to reduce the price of gasoline. Assuming a downward sloping demand curve for the rest of the world, consumers in other countries will then consume more gasoline, and the market price will re-equilibrate. The same effect would occur for other uses of gasoline in the United States. The additional gasoline use would again cut into the energy reductions from the policy. A reduction in the cost per mile of driving has been argued to induce economic growth, possibly leading to additional energy use (Saunders 1992). Whether this is possible or likely depends on the model of economic growth that is assumed.

Others have added further elements to the typology, although these additional elements are not as straightforward. For example, Greening, Greene, and Difiglio (2000) adds a fourth category of transformational effects that refer to the possibility that changes in technology have the potential to alter preferences, social institutions, and the organization of production. These transformational effects are of such a vague nature, that most more recent authors (e.g., Sorrell and Dimitropoulos (2008)) do not include them in the typology. At the very least, they would be extremely hard to quantify in a policy analysis setting. Similarly, Sorrell (2007) suggests that any additional energy used in creating a more energy-efficient good should be included in the rebound effect. This perhaps misses the meaning of the rebound effect. Of course, we would like to include any change, positive or negative, in the amount of energy used in the production of the good due to a policy, but calling this effect as part of the rebound effect seems misguided.

To formalize the rebound effect, I introduce some notation. Consider a policy that improves the energy efficiency of a durable good. Furthermore, consider a particular time when we wish to measure the rebound effect (e.g., one year after implementation). Absent the policy, there would have been some amount of fuel use of this good – the baseline fuel use F_0. F_0 is defined as

$$F_0 = \frac{U_0}{E_0},$$

where U_0 is the baseline use of the good (e.g., miles driven) and E_0 is the baseline technical efficiency (e.g., miles per gallon). The policy improves the energy efficiency of the good, so that the new fuel use, *absent a rebound effect*, can be calculated as

$$F_{nr} = \frac{U_0}{E_1},$$

where E_1 is the new technical efficiency due to the policy. The difference $F_0 - F_{nr}$ is the technical fuel savings from the policy before accounting for the rebound effect. In other words, this difference captures what the fuel savings would have been if there was no rebound effect at all. This is historically what engineering analyses of energy efficiency policies have calculated.

The final, observed, energy use after the policy F_1 (including the rebound effect) can be defined as

$$F_1 = \frac{U_1}{E_1} + I + M,$$

where U_1 is the observed use of the good with the policy, I is the fuel use due to the indirect rebound effect from additional income being used to purchase other energy-using goods or services, and M is the fuel use due to the macroeconomic rebound effect from price effects or economic growth.[9] The energy use from the direct rebound effect D can easily be seen to be

$$D = F_1 - I - M - F_{nr} = \frac{U_1 - U_0}{E_1}.$$

It is instructive to think about the overall rebound effect in percentage terms, rather than in terms of fuel use. Indeed, much of the literature writes about the rebound effect as a percentage, where 100 percent refers to a case of "backfire," whereby all of the fuel savings from improving the energy efficiency are overtaken by the additional fuel use induced by the lower utilization cost. In percentage terms, the

[9] Here I abstract from the influence of the macroeconomic effect back to U_1.

rebound effect from a policy that changes fuel efficiency from E_0 to E_1 can be defined as

$$R = \frac{D + I + M}{F_0 - F_{nr}} \cdot 100,$$

where R is the magnitude of the rebound effect (in percent).

In many cases, it is useful to consider a policy that improves fuel economy on the margin. In this case, the direct rebound effect can be thought of in terms of the elasticity of utilization with respect to fuel efficiency

$$\beta_{U,E} = \frac{\partial U}{\partial E}\frac{E}{U}. \tag{1.2}$$

This elasticity allows for an easy calculation of the change in utilization due to a marginal policy $(U_1 - U_0)$, and thus of D. This elasticity alone does not tell us the full rebound effect, and it technically does not even tell us the magnitude of the direct rebound effect $D/(F_0 - F_{nr})$ (in percentage terms), yet it does provide useful information.

Correspondingly, if we are interested in $F_1 - F_0$ from a policy that influences fuel economy on the margin, we can use the elasticity of gasoline demand with respect to fuel efficiency:

$$\beta_{F,E} = \frac{\partial F}{\partial E}\frac{E}{E}.$$

This elasticity does not quite give us $F_0 - F_{nr}$, for it should be thought of as the elasticity of gasoline demand with respect to fuel efficiency *net of the rebound effect*. However, both of these two elasticities play a key role in efforts to quantify the rebound effect.

Quantifying the Rebound Effect

Many of the studies estimating the gasoline demand and VMT responsiveness to gasoline price changes have been attempting to quantify the rebound effect, motivated by (1.2). More recently a handful of studies have attempted to address the magnitude of the macroeconomic rebound effect. While each of these literatures provides useful insights, there may be reason to be cautious about interpreting the estimates as a rebound effect in both.

Let's begin by considering the ideal experiment to identify the rebound effect. We would have two identical regions, both representative of the United States. In one region, we implement a policy requiring that consumers purchasing a more efficient version of a particular good than they would have otherwise. Then we could identify the direct rebound effect by examining how much more the consumers use the good. We could look at how much more of other energy-using goods they purchase to understand the indirect rebound effect. This ideal experiment would still not allow us to identify the macroeconomic rebound effect, which we could only learn about by understanding supply and demand in the market fuel.

Since this ideal experiment is impossible, previous studies have attempted to estimate the rebound effect with non-experimental data. The vast majority of the studies attempting to identify the rebound effect from policies to raise fuel economy use time series or cross-sectional data to estimate the VMT elasticity with respect to the cost per mile of driving. Such an estimation is intended to provide estimates that can be used to determine the direct rebound effect from a policy that increases fuel economy on the margin. The usefulness of the VMT elasticity with respect to the cost per mile of driving, rather than the fuel economy as in (1.2), stems from the mathematical relationship between the VMT elasticity with respect to the cost per mile of driving and the VMT elasticity with respect to fuel economy. Denote P_U as the cost of utilization, i.e., the cost per mile of driving. Denote P_F as the price of fuel, i.e., the price per gallon of gasoline. By construction, $P_F = P_U E$. Then we have:

$$
\begin{aligned}
\beta_{U,E} &= \frac{\partial U}{\partial E}\frac{E}{U} \\
&= \frac{\partial U}{\partial P_U}\frac{\partial P_U}{\partial E}\frac{E}{U} \\
&= \frac{\partial U}{\partial P_U}\frac{\partial (P_F E^{-1})}{\partial E}\frac{E}{U} \\
&= -\frac{\partial U}{\partial P_U}\frac{P_F}{E^2}\frac{E}{U} \\
&= -\frac{\partial U}{\partial P_U}\frac{P_U}{U} \\
&= -\beta_{U,P_U}.
\end{aligned}
$$

One can also easily derive a similar relationship replacing fuel economy (in miles per gallon) with gasoline consumption (in gallons per mile) to show that $\beta_{U,G} = \beta_{U,P_U}$ where G is gasoline consumption (Greene 2011). What this static relationship suggests is that we can estimate how driving changes when fuel economy changes by estimating how driving changes when the price per mile of driving changes.

Most of the studies in the bottom half of Table 1.2 had this relationship in mind and tended to call the VMT elasticity with respect to the cost per mile of driving as the "rebound effect." In order to estimate the VMT elasticity with respect to the cost per mile of driving, variation in gasoline prices is used. Thus, the estimated elasticities in the bottom half of Table 1.2 turn out to be remarkably similar to those in the top half of the same table, which are estimates of the VMT elasticity with respect to the gasoline price.

In fact, a static relationship can also be derived to show that the VMT elasticity with respect to the gasoline price should be identical to the VMT elasticity with respect to the cost of utilization. We can see this as follows:

$$
\begin{aligned}
\beta_{U,P_F} &= \frac{\partial U}{\partial P_F}\frac{P_F}{U} \\
&= \frac{\partial U}{\partial P_U}\frac{\partial P_U}{\partial P_F}\frac{P_U E}{U} \\
&= \frac{\partial U}{\partial P_U}\frac{P_U}{U} \\
&= \beta_{U,P_U}.
\end{aligned}
$$

Thus, the top half of Table 1.2 should be entirely equivalent to the bottom half of the same table. Both should provide equal insight into the direct rebound effect. Moreover, a similar static analysis shows the relationship between the estimated values of the elasticity of gasoline demand with respect to the gasoline price and the VMT elasticity with respect to the cost per mile of driving. Yet in this case, we can see that there is an additional term:

$$
\begin{aligned}
\beta_{F,P_F} &= \frac{\partial F}{\partial P_F}\frac{P_F}{F} \\
&= \frac{\partial (UE^{-1})}{\partial P_U}\frac{\partial P_U}{\partial P_F}\frac{P_U E}{VE^{-1}} \\
&= \left[E^{-1}\frac{\partial U}{\partial P_U} + U\frac{\partial E^{-1}}{\partial P_U}\right] E^{-1}\frac{P_U E^2}{VE^{-1}} \\
&= \frac{\partial U}{\partial P_U}\frac{P_U}{U} + \frac{\partial E^{-1}}{\partial P_U}\frac{P_U}{E^{-1}} \\
&= \beta_{U,P_U} + \beta_{E^{-1},P_U}.
\end{aligned}
$$

If we assume that fuel consumption (E^{-1}) is exogenously determined, as is the case in the very short run, then this result suggests that $\beta_{F,P_F} = \beta_{U,P_U}$. The intuition for this static result is clear: if we ignore how the price of fuel influences new vehicle fuel economy, then the price elasticity of gasoline demand is simply equal to the VMT elasticity with respect to the price per mile of driving. Some of the papers listed in Table 1.1 estimate the price elasticity of gasoline demand and claim to be shedding light on the rebound effect. To the extent that we can ignore how the price of fuel influences new vehicle fuel economy, there may be some validity to this claim. Of course, in a dynamic setting, new vehicle fuel economy may also influence scrappage decisions, which would add another term to the above analysis.

Note that this relationship between the price elasticity of gasoline demand and the VMT elasticity with respect to the cost per mile of driving was derived assuming that $\frac{\partial P_U}{\partial P_F} = 1$, and thus imposes that the cost of driving does not change when the price of gasoline changes. This is a reasonable assumption for any given vintage in the vehicle fleet, and is also a reasonable assumption in the very short term. It would cetainly not be true for the entire vehicle fleet in the longer term, for consumers may purchase higher fuel economy vehicles if the price of gasoline increases. Chapter 5 includes a further discussion of how the price elasticity of gasoline demand can be broken down into components. Allowing the cost of driving to change with the price of gasoline would add a third "rebound term" to account for how consumers who are induced to purchase a higher fuel economy vehicle end up driving more.

Other authors, such as Wadud, Graham, and Noland (2009), estimate the elasticity of gasoline demand with respect to the fuel economy and claim to provide guidance on the rebound effect. The argument that would have to be made in order to justify this assertion is based on the relationship between the elasticity of gasoline demand with respect to the fuel economy and the VMT elasticity with respect to the price per mile of driving:

$$\begin{aligned}
\beta_{F,E} &= \frac{\partial F}{\partial F}\frac{E}{F} \\
&= \frac{\partial (UE^{-1})}{\partial E}\frac{E}{UE^{-1}} \\
&= -1 + \frac{\partial U}{\partial E}\frac{E}{U} \\
&= -1 + \beta_{U,E}.
\end{aligned}$$

By the result shown above that $\beta_{U,E} = -\beta_{U,P_U}$, we also have that $\beta_{F,E} = -1 - \beta_{U,P_U}$.

All of these static relationships provide the intellectual backing for the many approaches to estimating parameters that are useful for improving our understanding of the rebound effect. However, there may be some reasons to believe that some of these results may not be as useful as we might wish them to be. Greene (2011) suggests that if a policy mandates higher fuel economy, it would raise the capital cost of the vehicles and thus reduce the magnitude of the rebound effect. How might this occur? Greene (2011) claims that the higher amortized capital cost of the vehicle would make the fuel cost of driving a smaller percentage of the total cost of driving, thus reducing how much consumers respond to a changing fuel cost of driving. Greene (2011) then shows empirical results that suggest this may be true, although further work on this issue is clearly warranted.

I posit another possibility for why we may not be able to simply assume that the consumer VMT response to gasoline price changes is equivalent to the consumer response to higher fuel economy, as is the case if $\beta_{U,P_F} = -\beta_{U,P_U}$. Suppose that consumers respond asymmetrically to changes in the cost per mile of driving, so that decreases in the cost per mile of driving (e.g., from increased fuel economy) lead to little responsiveness, but increases in the cost per mile of driving (e.g., from increased gasoline prices) lead to much more responsiveness. This hypothesis corresponds to the evidence that elasticities tend to be higher for periods of rising prices than for periods of falling prices (Gately 1992, 1993). Furthermore, it may also be that consumers respond more to dramatic price changes than to subtle, smaller price changes. A change in the fuel economy of the vehicle due to a policy may lead to a more subtle price change than a major change in gasoline price. There is some evidence of such an asymmetric response in the literature, for we can see that more recent studies that are identified off small changes in the price of gasoline tend to have much lower estimates of responsiveness.

I also posit a third possibility. When a policy requires higher fuel economy for new vehicles, it serves to incentivize consumers to purchase different vehicles. These vehicles are not the first choice vehicles to drive and may be a bundle of attributes that consumers do not wish to utilize as much. In this case, the responsiveness to a change in the cost per mile of driving from a gasoline price change *conditional on the* *vehicle*, may be different than the responsiveness to a change in the cost per mile of driving that *influences the vehicle choice*. For example, suppose a consumer would have purchased an exciting sports car with low fuel economy, but under the policy instead purchased a more sedate sedan. The consumer may be inclined to drive the sedan less than they would have driven the sports car. In addition, given the heterogeneity in responsiveness, it is possible that the consumers who are induced to change the vehicle they purchase under the policy are systematically different than the rest of the consumers. For example, they may be less responsive to changing gasoline prices – which could also bring down the rebound effect relative to the elasticity of driving with respect to the price of gasoline. I am not aware of these concept in the literature, but I discuss it further in Chapter 4.

All this said, the prevailing evidence in the literature on the direct rebound effect suggests that most recently it has been in the range of -0.5 to -0.15 in the short-run, and slightly higher than that in the long-run (Small and Van Dender 2007; Hughes, Knittel, and Sperling 2008; Hymel, Small, and Van Dender 2010). This recent range of estimates tend to be closer to zero than the range of estimates over the past three decades, a point emphasized in several papers that suggest that the direct rebound effect is declining over time.

In contrast to the direct rebound effect, there is very little evidence on the indirect rebound effect and macroeconomic rebound effect. One might expect that both of these are small. There is no evidence that I am aware of on the importance of the indirect rebound effect, although Sorrell (2007) speculates that it is not large. There are a handful of papers on the macroeconomic rebound effect, often using Computable General Equilibrium (CGE) models of the United Kingdom economy. Some of the first cursory work attempting to look at the macroeconomic rebound effect indicated it may be an important effect (Kydes 1997). Dimitropoulos (2007) surveys these models and comes to the conclusion that the macroeconomic rebound effect may be quite large. A working paper, Saunders (2011), and a well-known report by the Breakthrough Institute (Jenkins, Nordhaus, and Shellenberger 2011) come to this same conclusion. However, there may be reasons to view this conclusion as suspect.

When looking at changes in energy efficiency and energy in the long-run, it is extremely easy to view the long-run trends of greater energy efficiency and greater energy use as a correlation that has some causal predictive power. This is a classic case of confounding correlation with causation, for in the long-run technology also improves. The energy services provided from the new technology are vastly superior to what was previously available, so it seems odd to attribute the additional energy use from the considerably different energy services to the rebound effect. Moreover, in the long-run we have seen an incredible growth in income around the world. This begs the question: is the greater use of energy due to technical change and income, or to improved energy efficiency? Even more importantly, if we implement a policy to improve energy efficiency, can we expect energy use to increase based on the previous correlations?

The recent work in CGE modeling to look at the macroeconomic rebound effect is promising, but tends to rely upon many structural assumptions about how technology improvement interacts with energy efficiency. For example, Barker, Ekins, and Foxon (2007) find a relatively large (19 percent) macroeconomic rebound effect in a CGE model of the United Kingdom economy. I view exploring the macroeconomic rebound effect as a promising area for future research for CGE modelers, with much still to learn about the implications of different assumptions.

1.2.6 Policy Analysis

There is a rich literature of economic policy analysis on policies to reduce emissions from the transportation sector. The literature most related to this dissertation can be considered to fall into two categories. The first contains papers on gasoline taxes, and in particular, the incidence and distributional effects of gasoline taxes. The second contains papers on fuel economy standards and how fuel economy standards compare to other policies along a variety of dimensions.

Poterba (1989, 1991) provided some of the first in-depth analysis of the distributional consequences of gasoline taxes. Hausman and Newey (1995) use household-level data from the Residential Energy Consumption Survey to estimate the price elasticity of gasoline demand and the deadweight loss from gasoline taxes. West and Williams

(2004) use household-level data from the Consumer Expenditure Survey to estimate the price elasticity of gasoline demand and use the results to assess the optimal gasoline tax. West and Williams (2004) also examine the distributional consequences of gasoline taxes, and note that gasoline taxes are considerably less regressive (or even progressive) if the revenues are used to reduce labor taxes or to fund lump-sum transfers.

West (2004) examines the distributional consequences of gasoline taxes and other policies and finds that gasoline taxes are actually less regressive than taxes on engine size or subsidies for new vehicles. West and Williams (2007) also use the Consumer Expenditure Survey data, only this time to estimate the cross-price elasticity between gasoline and leisure. This cross-price elasticity turns out to be an important factor in the calculation of the optimal gasoline tax. Along the same lines, Parry and Small (2005) are also interested in the optimal gasoline tax, and do a careful job of quantifying the externalities that would justify a Pigouvian gasoline tax.

Bento et al. (2009) use the 2001 National Highway Transportation Survey to perform perhaps the most comprehensive study of the welfare and efficiency impacts of increased gasoline taxes in the US. Bento et al. (2009) point out that the way the gasoline tax revenues are recycled is extremely important for the distributional impacts of the policy. When revenue recycling is taken into account, Bento et al. find that a 25-cent gasoline tax increase impacts the average household by about $30 per year (in 2001 dollars).

The literature on fuel economy standards is equally as rich. Mayo and Mathis (1988) perform a demand estimation to see whether Corporate Average Fuel Economy (CAFE) standards over the 1977 to 1983 period seemed to have any effect. The empirical results in Mayo and Mathis suggest that CAFE standards did not appear to have an independent, statistically significant effect on driving or vehicle fuel efficiency. Krupnick, Walls, and Collins (1993) examine the cost effectiveness of increasing CAFE standards versus substituting methanol, compressed natural gas (CNG), and reformulated gasoline for gasoline if the goal is to reduce greenhouse gas emissions. CNG is found to be the most cost-effective, followed by CAFE standards, and then methanol and reformulated gasoline.

Goldberg (1998) is interested in the cost of CAFE standards and how they compare to the costs of a gasoline tax. Goldberg estimates a nested logit discrete choice model of vehicle choice using the Consumer Expenditure Survey data and links this demand side with firm-level Bertrand differentiated products oligopoly model for the supply of vehicles. Goldberg finds that a doubling of the gasoline price would be necessary to achieve the reductions in fuel consumption that CAFE (up to that point) had achieved. This result stems in part from the fact that Goldberg finds that consumers are only barely responsive to changes in gasoline prices, as indicated in Table 1.2.

West and Williams (2005) examine the cost of reducing gasoline consumption using either CAFE standards or gasoline taxes. West and Williams find that because gasoline taxes increase the cost per mile of driving, they tend to discourage leisure (encourage labor). West and Williams find that CAFE standards tend to have the opposite effect. This set of findings leads to the result that the marginal cost of increasing the gasoline tax is less than half of the marginal cost of increasing CAFE standards. Without accounting for this interaction with the labor market, West and Williams find that the marginal cost of the two policies would be similar.

Greene et al. (2005) provide possibly the only published economic analysis of feebates in the literature. As I will discuss in Chapter 5, a feebate policy is similar in many respects to a fuel economy standard, for the shadow cost of the fuel economy standard constraint may be thought of as an implicit tax on low fuel economy vehicles. Greene et al. (2005) use a nested logit demand side with a highly stylized supply side to quantify how much feebates would increase the fuel economy of the fleet under different assumptions. Greene et al. suggest that the fuel economy increase that one could expect from fuel economy standards may come predominantly from the adoption of new fuel economy technologies, rather than from shifts or decreases in the sales of new vehicles.[10] However, it is not clear whether this result stems from the data or structural assumptions in the model.

More recently, there has been a set of papers using different methodologies to quantify the costs of CAFE standards. Klier and Linn (2010a) attempt to use a

[10]This result is claimed to hold only for a national feebate program and may differ depending on the geographic coverage of the program.

clever identification strategy based on the claimed ability of automakers to change many of the attributes of a vehicle before overhauling the engine design. The goal of the study is to examine at the medium-run costs of CAFE standards. Their findings suggest that consumers value an increase in power more than an increase in fuel economy. Klier and Linn (2010a) simulate the medium-run effects of an increase in CAFE and find that the regulatory costs are much lower in the medium-run than in the short-run, for automakers have time to adjust product designs, rather than being forced to re-optimize only in pricing and production quantity decisions.

Anderson and Sallee (2011) use loopholes in the CAFE regulation that allow automakers to earn credits for producing flex-fuel vehicles that can use both gasoline and ethanol in order to quantify the marginal cost of the regulation. The idea behind the identification strategy in the paper is that automakers presumably choose to use the loopholes up to the point where the marginal cost of producing flex-fuel vehicles is equal to the marginal cost of compliance of CAFE. Anderson and Sallee (2011) find that the marginal cost of compliance of CAFE is surprisingly low: $9 to $27 per vehicle in recent years.

Jacobsen (2010) use the approach in Bento et al. (2009) to analyze the cost and impacts of CAFE standards. Jacobsen finds that the profit impacts of CAFE fall almost entirely on domestic producers. Jacobsen performs a welfare analysis to compare CAFE standards to gasoline taxes. He finds that increasing gasoline taxes to save a gallon of gasoline would have one-sixth the welfare cost of increasing CAFE standards. More than half of the costs of the CAFE standard are caused by distortions in used car markets. These welfare effects are estimated to fall disproportionately on low-income households.

Each of these above studies about the welfare implications of CAFE standards does not address what is perhaps the most controversial topic relating to CAFE standards: whether consumers "undervalue" fuel economy by seeming to require higher implicit discount rates for improvements in fuel economy than for other investments. If consumers truly undervalue fuel economy, some authors have argued that a fuel economy standard may make consumers ex post happier by influencing them to purchase higher fuel economy vehicles than they might have otherwise (Allcott and Wozny

2010). Yet, the current empirical evidence on whether there is truly an undervaluation of fuel economy is mixed. Several recent studies use adjustments in the relative prices of higher and lower fuel economy vehicles to examine whether consumers appear to undervalue fuel economy and come to differing conclusions, from full adjustment (Busse, Knittel, and Zettelmeyer 2010; Sallee, West, and Fan 2009), to partial under-adjustment (Allcott and Wozny 2010), to significant under-adjustment (Kilian and Sims 2006).

This literature review reveals that although the issues in this dissertation have been of great interest and widely studied, there remain many open questions where new data and techniques may shed significant insights for policy development.

1.3 Preview of Methodology and Results

This section provides a more in-depth summary of the methodology and results in each of the following chapters in my dissertation.

In Chapter 2, I begin by describing the unique dataset used in this study in greater detail. Summary statistics from such a rich dataset paint a fascinating picture of driving and purchase behavior and heterogeneity in such behavior. The difference in miles driven and the change in vehicles purchased in years with different gasoline prices provide the first descriptive evidence of consumer response to changing gasoline prices. Simply by looking at the data, it appears that there has been a consumer response to changing gasoline prices in terms of both driving and vehicle purchase decisions. When gasoline prices increase, consumers appear to drive less. Similarly, when gasoline prices increase, consumers appear to switch to higher fuel economy vehicles across vehicle classes, within-vehicle classes, and then even within vehicle models. These findings provide a sense of the variation in the data that is underpinning the results in the following chapters.

Chapter 3 provides the first regression evidence of the responsiveness on both margins to the changing gasoline prices. Ordinary least squares (OLS) and fixed effects regressions provide a first quantification of the responsiveness. The rich dataset allows me to control for a wide variety of possible confounds, and also facilitates a careful analysis of the heterogeneity in how different people respond. I examine vehicle-level heterogeneity in the responsiveness to gasoline prices by geography, income, and demographics to an extent that had not been possible in the previous literature.

The regression results show that consumers clearly responded to changing gasoline prices. I find a medium-run (i.e., roughly two-year) elasticity of driving with respect to the price of gasoline in the order of -0.17 to -0.25 for new personal vehicles in the first six years of life. I find much less responsiveness in driving to changing fuel economy, with an elasticity of driving with respect to fuel economy in the range of 0.05. Similarly, the responsiveness of vehicle purchasers to changes in gasoline prices is noticeable, yet still quite inelastic: the elasticity of the fleet-wide average fuel economy with respect to the price of gasoline is around 0.1. During the period of my study, CAFE standards remained largely constant and were not binding for several manufacturers, thus allowing for the shorter-term response in the amount of higher versus lower fuel economy vehicles being manufactured Jacobsen (2010). For those manufacturers that faced a binding CAFE standard constraint, I find slightly less responsiveness. All of these elasticity estimates are the result of estimations controlling for economic conditions and zip code-level income and demographics, while the driving responsiveness estimates are also conditional on vehicle characteristics. To see how representative new personal vehicles are, I can examine the responsiveness of the rest of the vehicles in the light duty vehicle fleet (i.e., the older vehicles and non-personal vehicles) by looking at all biennial smog test results in 2002 to 2009. I find that older vehicles are more price responsive, with a one to two-year elasticity of driving with respect to the price of gasoline in the range of -0.3 to -0.5.

I also find considerable evidence of heterogeneity in responsiveness. Quantile estimations immediately show the considerable differences between the percentiles of vehicles by responsiveness. I also find differences between vehicles owned by different groups of drivers that are both statistically and economically significant. I find that higher fuel economy vehicles tend to be less responsive. Leased vehicles tend to be driven more and are more responsive.

The responsiveness tends to vary with income with a largely U-shaped pattern, so that the more responsive vehicles are those driven by wealthy and low-income households. The very highest and very lowest income categories are the only exception to this, with the very highest income drivers of new personal vehicles being some of the least responsive. I also find some evidence that vehicles in different counties in California show differences in responsiveness.

The use of OLS and fixed effects estimators in Chapter 3 provides for useful and transparent results in which it is clear what variation in the data is driving the results. However, these estimation results do not account for the possible selection bias first described in Dubin and McFadden (1984). Exploration of this selection bias is saved for Chapter 4.

Chapter 4 focuses on improving our understanding of the importance of the selection bias in the utilization elasticity caused by different types of consumers "selecting" into vehicles of different fuel economy. I begin by explaining the nature of the selection bias and why we may be concerned about it. Then I develop a novel structural model of vehicle choice and subsequent utilization that explicitly accounts for unobserved heterogeneity in expected driving. The model presents a utility-consistent framework whereby consumers weigh the benefits against the costs of both vehicle choice and subsequent utilization. This model allows me to simultaneously estimate the consumer responsiveness to gasoline price changes on both margins, while taking into account the sequential nature of the decisions. In addition, the framework is well-suited to analyze a pure form of the direct rebound effect: the responsiveness of driving when a policy induces consumers to purchase different fuel economy vehicles. The model is also designed to allow for a clear analysis of the importance of the selection bias.

The structural model estimation restricts the sample to consumers who purchased a new personal vehicle within the past six years, rather than the entire stock. This subsample choice is made in part for data reasons and in part because most new vehicles are held by an individual owner for an average of about six years based on national-level data (Todorova 2007; NHTS 2009; Polk 2010). The subsample choice also eases the computational burden of estimating the model and allows me to avoid explicitly modeling the used vehicle market.

The primary result of the structural model estimation is a medium-run elasticity of driving with respect to the price of gasoline of 0.15. Without accounting for the bias, the result of estimating the structural model is 0.21. This finding indicates that the selection bias is economically important and without accounting for such a bias, we would *over-estimate* the elasticity of driving. Moreover, there is an intuition for the sign of this bias. It captures the idea that those who anticipate driving a great deal will purchase a higher fuel economy vehicle (rather than a more comfortable vehicle), which lowers the cost per mile of driving and implies a lower responsiveness. As a result, the marginal cost of driving is negatively correlated with miles driven, so the resulting responsiveness is biased. We might expect this to be the result, but for vehicles it need not be the case, for those who anticipate driving more may choose to purchase a more comfortable, low fuel economy vehicle. The results of my estimation indicate that the unobserved heterogeneity works in the direction we might initially suppose, so that there is less responsiveness when we address the selection bias.

Another result of the structural model estimation is a medium-run elasticity of the fuel economy of new vehicles with respect to the price of gasoline of 0.10. This result matches extremely closely with the result from the reduced form estimation. Both this result and the utilization responsiveness imply that consumers do respond to changes in the prices of gasoline, but that the response remains quite inelastic in the medium-run. Thus, policies to reduce emissions from the transportation sector are likely to require large changes in the price of gasoline to achieve substantial emissions reductions.

The third result of the structural model estimation is an estimate of the direct rebound effect. By providing an incentive for consumers to purchase more efficient vehicles through a feebate, I find that the elasticity of driving with respect to the fuel economy of the vehicle (for those consumers that changed the vehicle purchased) is 0.06. This more pure estimate of the rebound effect from a feebate policy can again be interpreted as a medium-run response. It differs from the negative of the VMT elasticity because it captures the responsiveness of people who changed vehicle purchases based on where they live and what different new vehicle they purchased. It has important policy implications for CAFE standards, feebates, or any other policy instrument to encourage fuel economy on the extensive margin. Specifically, a rebound effect that is small in magnitude, such as 0.06, implies that the reduced emissions from the policy will not be overtaken by the increased emissions from the rebound effect.

Chapter 5 delves more fully into the policy implications of my previous results. I examine the effects on consumers of two policies to reduce greenhouse gases from the transportation sector: a tax that raises the price of gasoline by one dollar per gallon and a revenue-neutral feebate policy. The contrast between the two policies is useful, for the gasoline tax works on both the driving and vehicle choice margins, while the feebate policy works only on the vehicle choice margin – just as fuel economy standards do. I use the structural model from Chapter 4 to examine the welfare effects of gasoline taxes and feebate policies for new personal vehicles. To understand the broader effects of a gasoline tax policy, I develop a vintage model of the entire vehicle fleet in California.

The first result stems directly from the estimated elasticities of the previous chapters. I find that the price elasticity of gasoline demand is predominantly determined by the responsiveness in driving to gasoline prices in both the short-run and medium-run. This result differs from some of the previous literature that suggests the fuel economy responsiveness is the more important factor in determining the price elasticity of gasoline demand. Of course, in the longer run, manufacturer design decisions may also respond to gasoline prices, so the fuel economy responsiveness (absent tightly binding fuel economy standards) may perhaps be a more important factor.

The relatively inelastic estimates of driving responsiveness also imply that the gasoline tax policy leads to relatively small fuel savings and carbon dioxide emissions reductions. However, the policy brings in a substantial amount of revenue. If this revenue is redistributed lump-sum to consumers, the resulting consumer deadweight loss from the policy, absent externalities, is quite small. If we only consider the global warming externality, the implied cost of carbon from the consumer deadweight loss (including a distortion from the pre-existing gasoline tax) is in the range of \$100 to \$120 per tonne of CO_2. This estimate is above many estimates of the social cost of carbon. However, there are other important externalities from driving, such as energy security, accidents, congestion, and local air pollution. If the values used for the externalities in the literature are correct, such that the other externalities make up just over half of the total external cost, then the implied cost of carbon from the one dollar gasoline tax is closer to the estimates of the social cost of carbon in the literature – but still above them. Given this, a somewhat smaller increase in the gasoline tax may be more likely to be economic efficiency-improving. The political feasibility of any increase in the gasoline tax may be suspect though. Importantly, I find that there is considerable and systematic heterogeneity in the household-level consumer surplus loss across counties without a careful redistribution of the tax revenues.

I examine the effect of a $50,000 per gallon per mile revenue-neutral feebate policy on the cohort of vehicles that were purchased in 2002 using the structural model. This policy incentivizes some consumers to purchase higher fuel economy vehicles. The results suggest an increase in the average fuel economy of the fleet of 15 percent. This increase in fuel economy leads to an increase in driving of roughly one percent for the entire fleet, implying a modest direct rebound effect of about 0.07. The increase in driving is greater if we examine only those consumers who are induced to purchase a different vehicle.

Using my estimated structural model, I find a one-time welfare loss to consumers from purchasing a less desirable but higher fuel economy vehicle in the range of $5.6 per vehicle. However, the feebate results assume that the future nationwide fuel economy standards are not tightly binding on all manufacturers – otherwise the feebate policy may just make it easier for manufacturers to meet the fuel economy standards, but not change the fleet fuel economy. This assumption is not likely to hold given the large planned increase in CAFE standards over the next decade.

Chapter 6 provides concluding remarks that bring together the findings of the dissertation. In addition, this chapter includes a discussion of the many promising areas of future research that could profitably build upon the work accomplished here.

Chapter 2

Data

The dataset assembled for this study is novel in its breadth and detail of vehicle choices and driving behavior. The focus of this dissertation is on the state of California, which is the most populated state in the US and has considerable variety in demographics and levels of urbanization. California's stringent air quality regulations have also served to nearly eliminate the number of diesel vehicles in the light-duty fleet, simplifying the analysis. The time frame for the study is 2001 to 2009, a period containing the striking gasoline price changes in 2006 to 2008. These gasoline price changes, along with gasoline price differences across counties, provides useful identifying variation in gasoline prices. Just as in the classic papers in the literature on estimating vehicle demand (e.g., Bresnahan (1981), Berry, Levinsohn, and Pakes (1995), Petrin (2002), Berry, Levinsohn, and Pakes (2004)), this paper focuses on new vehicles. The richness of my dataset allows for a quite different methodology than in the classic papers.

I use two distinct versions of the dataset, a choice necessitated by the nature of the data and the different questions being answered. The first version of the dataset is used in all of the remaining chapters of this dissertation. It focuses on all new personal vehicles purchased in 2001-2004 and on the first six years of driving. This dataset includes detailed information about the characteristics of the vehicle purchasers and thus is very useful for an exploration of the heterogeneity in the responsiveness. The restriction to vehicles during the first six years of life implies that the vehicle owners tend to be wealthier than the overall California population and that the responsiveness should be viewed as representative for this (still large) population of drivers.

The second version of the dataset is used only in Chapter 3. It contains nearly the entire stock of vehicles (including older vehicles, company cars, and rental cars) in California in 2002 to 2009 and allows for the use of a different source of variation to examine the responsiveness to gasoline price changes. By looking at all vehicles, I can compute a useful estimate of the what the overall VMT responsiveness is to gasoline price changes. The downside of this approach is that much of the detailed information about the consumers (e.g., income, zip code of residence) is lost. However, exploring this source of variation in concert with the more detailed information provides a broader picture of how consumers in general respond to gasoline price changes.

For each dataset, I give a summary of the dataset and provide summary statistics to paint a picture of the broader factors at work in the dataset that will be influencing the estimation results in the subsequent chapters. For personal vehicles, I also provide initial descriptive evidence showing a responsiveness to gasoline price changes in both new vehicle purchasing and driving behavior.

2.1 Personal Vehicles in First Six Years

The primary foundation of this first dataset contains all 12.3 million new personal vehicle registrations in California from January 2001 to May 2009. These data were acquired from R.L. Polk and primarily originate from the CA Department of Motor Vehicles (DMV). Each vehicle is identified in the data by the 17-digit Vehicle Identification Number (VIN), and contains information on the make, model, model year, trim, body type, engine cylinders, engine size, weight, drive type (Four/All Wheel Drive or Two Wheel Drive), existence of a turbocharger or supercharger, the Manufacturer Suggested Retail Price (MSRP), whether the vehicle is leased or purchased, and the zip code that the vehicle was registered in. The buyer type (i.e., personal, firm, or government) is also observed, allowing me to restrict this dataset to only personal vehicles (just over 80 percent of all new vehicles). In addition, R.L. Polk acquired data from dealer financing forms on the household income of the purchaser for a large sub-sample of personal purchases (over 70 percent of the categorical income variable is observed and in most years over 85 percent is observed).

The registration data are first matched with US Environmental Protection Agency fuel economy ratings. The ratings were adjusted in 2008 to more accurately reflect the fuel economy that is realized in common on-road driving conditions. In this dissertation, I use the 2008 ratings throughout the entire time frame for a consistent measure that more closely reflects on-road fuel economy. The data are matched with vehicle safety ratings from the National Highway Transportation Safety Administration (NHTSA) Safercar.gov website. These safety ratings are based on a 5 star rating scheme that is qualitatively similar to the ratings from Insurance Institute for Highway Safety and Consumer Reports.

An important differentiator of this study from previous studies is the source of the vehicle utilization data. Rather than than using largely self-reported survey data on the distance of trips, I am able to use actual odometer readings taken by the mechanic and electronically reported to the California Bureau of Automotive Repair during the mandatory smog check. Since 1984, every vehicle in California that is covered by the biennial smog check program must be in compliance in order to be registered with the DMV.[1] To be in compliance, vehicles must meet the California criteria air pollution standards for several local air pollutants. Since 1998, vehicles covered by the program are required to receive a smog test at the seventh registration renewal (usually at the end of six years of vehicle life), and then biennially thereafter.[2] A smog test is also required at the time of a title transfer outside of the family for any non-exempt vehicle in California that is more than four model years old. Earlier

[1] The following vehicles have been exempt since 1998: hybrids, gasoline powered vehicles 1975, diesel powered vehicles manufactured prior to 1998 or with a gross vehicle weight rating of more than 14,000 lbs, electric vehicles, natural gas vehicles over 14,000 lbs, and motorcycles. Some hybrids are in the dataset, perhaps because when a vehicle is traded-in to some dealers, the dealers have the vehicle smog-checked, regardless of whether it is exempt. Appendix A lists further details on the coverage of the smog check program.

[2] Technically, "owners of vehicles six or less model years old will pay an annual smog abatement fee for the first six registration years instead of being required to provide a biennial smog certification." This means that some vehicles that were a model year early or late relative to the year of selling (e.g., a model year 2000 vehicle sold in 2001) might have the mandatory smog check at either the fifth or seventh registration. This detail only applies to a small portion of the vehicles.

incarnations of this smog check dataset have been used in previous studies, such as Kahn (1996), who looked at emissions rates by different vehicle types, and Hubbard (1998) who investigated fraud by smog check testing stations in allowing non-passing vehicles to pass the test by under-reporting the pollutant readings. Fortunately for this study, there is no obvious incentive for mechanics to falsely report the odometer readings. These readings are perhaps the best revealed preference measure of how much the vehicles actually have been driven. The only other work in economics using these smog check data that this author is aware of is Knittel and Sandler (2010), which provides cursory evidence of the responsiveness on the intensive margin using the full smog check dataset without the detail of the R.L. Polk data.

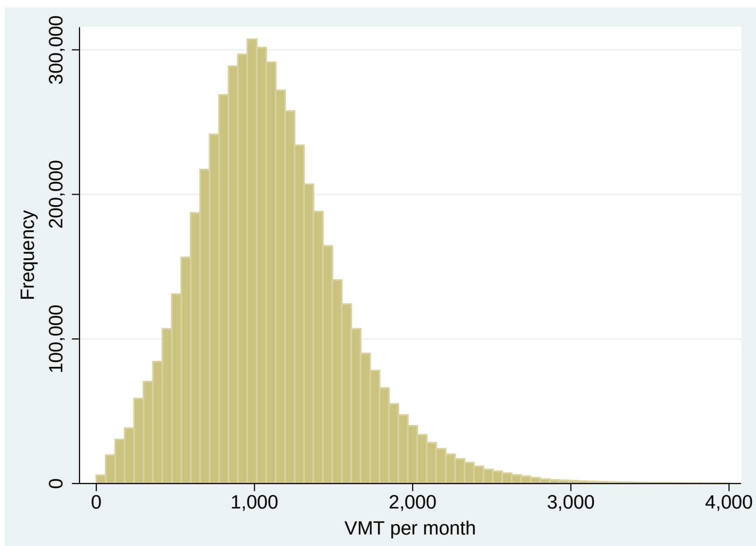

Figure 2.1: Driving per month by vehicles during their first six years of use in California has been remarkably unimodal.

For the new personal vehicles dataset, I use smog check data from 2005 to mid-2009. Besides odometer and pollutant readings, the smog check dataset also includes the make of the vehicle, the transmission type, the zip code of the test site, and the zip code of the vehicle registration (for smog checks after 2007). Vehicles are identified in the smog data by VIN and thus can be matched exactly to vehicles in the R.L. Polk registration dataset. Hence, I observe whether the owner of the vehicle moved by whether the registration location changed between the initial registration and the test. Figure 2.1 shows the distribution of vehicle miles traveled (VMT) per month for personal vehicles. The mean of VMT per month is 1,089, with a surprisingly high variance of 465. This high variance provides the first evidence that there is substantial heterogeneity in how vehicles are driven.

An important factor that could influence vehicle choice is the expected depreciation of the vehicle. For example, certain makes are known to depreciate more than others (e.g., Hondas are known to hold value well), and how well a particular vehicle model holds its value may depend on the gasoline price at the time (e.g., the resale price of light trucks might drop if gasoline prices are high). I use data from the National Automobile Dealers Association (NADA) on average used car prices in California by make-model-model year-trim. For vehicles where the model was not available six years prior, I use a similar model of the same make and vehicle class. The NADA data also include an adjustment factor to account for higher or lower odometer readings.

The monthly average retail gasoline price (tax-inclusive) at the county level in California is acquired from the Oil Price Information Service (OPIS). There is some limited cross-county variation in gasoline prices, particularly at the beginning of the time frame of the study, but most of the variation is time series (Figure 2.2).

The seasonality in Figure 2.2 is very clear and makes it a bit more difficult to see the trend. There are several ways to de-season the gasoline price data. One way to very simply see the trend without the seasonality is to plot the California average gasoline price over time for each month. Figure 2.3 shows the upward trend to 2008 and the subsequent dip without seasonality.[3]

To address economic conditions that may affect driving, I bring in the monthly county-level unemployment rate in California from the Bureau of Labor Statistics (BLS) and the monthly national-level Consumer Confidence Index (CCI) from the Conference Board. Figure 2.4 shows that the gasoline price spike in 2007-2008 preceded when the recession began to have a major impact on employment in 2008.

In addition to the unemployment rate and CCI, rapidly declining housing prices may also influence consumer decisions about large durable goods purchases, such as

[3]The author thanks Tom Wenzel of Lawrence Berkeley National Laboratory for this suggestion.

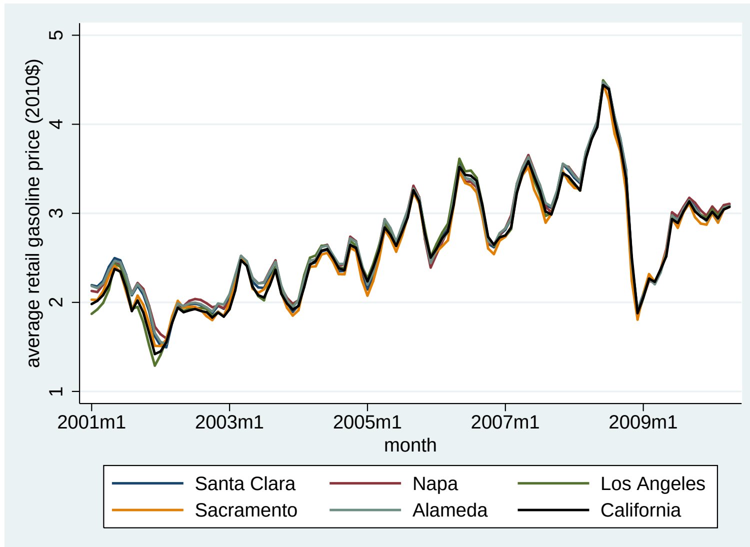

Figure 2.2: Retail gasoline prices in California were relatively flat and then rose substantially until 2008, providing substantial time series variation in addition to some cross-county variation. Four representative counties are shown here. Sources: Oil Price Information Service for the county time series and US Energy Information Administration (EIA) for the California average.

the purchase of a new vehicle. Accordingly, I bring in monthly county-level average housing prices from the California Association of Realtors. Figure 2.5 shows that housing prices were high during much of the time that gasoline prices were increasing, and then when the housing bubble began bursting in late 2007 and early 2008, there was a precipitous decline in housing prices.

The story that these graphs tells is that the increase in gasoline prices largely occurred while the economy was doing well – unemployment was low and housing prices were high. The 2008 spike in gasoline prices came after the housing bubble burst and consumer confidence dropped, but before unemployment reached high levels.

In the final addition to the dataset, I merge in zip code-level demographics and county-level commute times from the US Census Bureau. Appendix B provides a much more detailed description of the data merging and cleaning process.

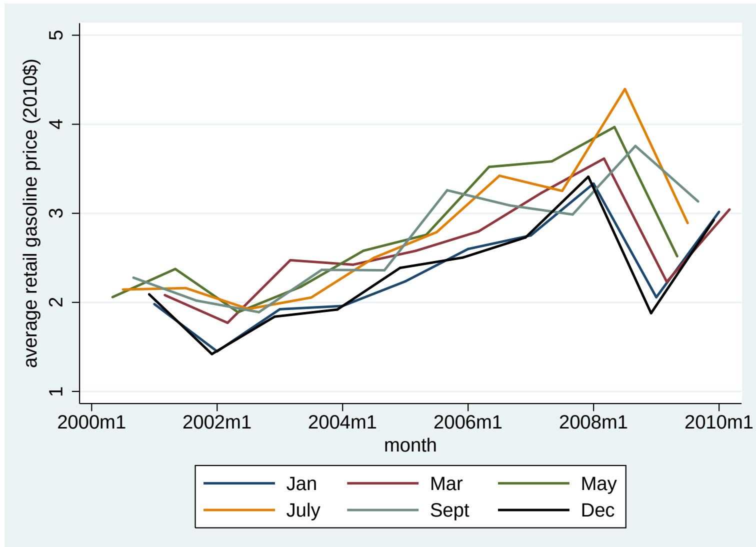

Figure 2.3: The upward trend in retail gasoline prices in California is even more noticeable by following months over time to address seasonality. Source: US Energy Information Administration (EIA).

2.1.1 Summary Statistics

Table 2.1 contains summary statistics of the entire merged dataset, where an observation is a personal vehicle registration. All dollar values are adjusted to 2010 dollars by the BLS Consumer Price Index. There are a few notable points to make from a quick glance at the summary statistics. The first group of summary statistics show characteristics of new vehicles entering the California personal vehicle fleet. The cylinders and liters variables showing the engine size and displacement are of typical magnitudes for vehicles available in the United States. The automatic transmission variable is most useful during the years that the smog check covers (2001 to 2004), for the transmission type of the vehicle is from the smog check data. During the years 2001 to 2004, about 93 percent of new vehicles have automatic transmissions. After 2004, all of the vehicles in the dataset are coded as having an automatic transmission.[4] Roughly 2 percent of the new vehicles in the time period covered by the dataset are hybrid-electric vehicles. 66 percent of the vehicles are from foreign firms

[4] I do not include this variable in some of the later analysis.

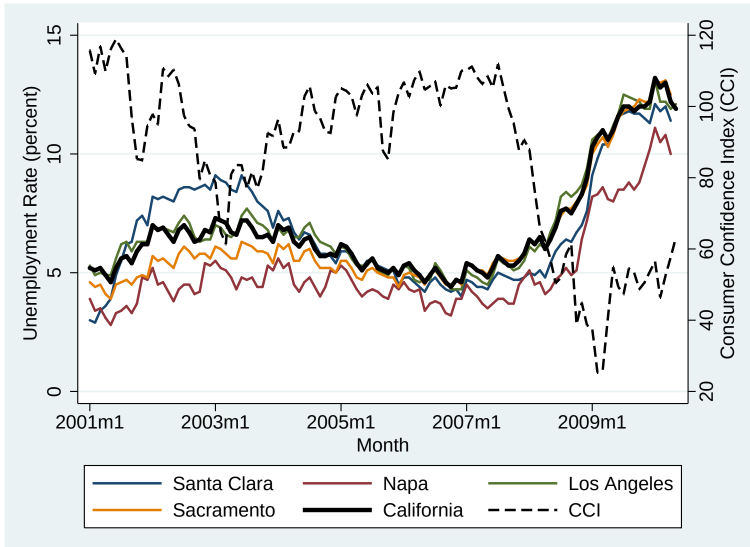

Figure 2.4: Unemployment and the consumer confidence index show that the economy was doing quite well until 2008, when the CCI plummeted and unemployment began increasing. Sources: Bureau of Labor Statistics for unemployment data and The Conference Board for CCI data.

(coded as imported vehicles). 18 percent of the vehicles have all-wheel or four-wheel drive. The fleet-wide harmonic average fuel economy is 19.22 miles per gallon.[5] The arithmetic average fuel economy for the fleet is similar: 20.53 miles per gallon. The mean manufacturer suggested retail price (MSRP) is \$29,709, although this value is skewed upwards by a few extremely expensive sports cars purchased in California. The median MSRP is \$27,473.

[5]The harmonic mean of positive real numbers $x_1, x_2, ..., x_n$ is defined as $H = n/(\frac{1}{x_1} + \frac{1}{x_2} + ... + \frac{1}{x_n})$. The harmonic mean captures the average fuel economy from driving each vehicle in the fleet the same number of miles.

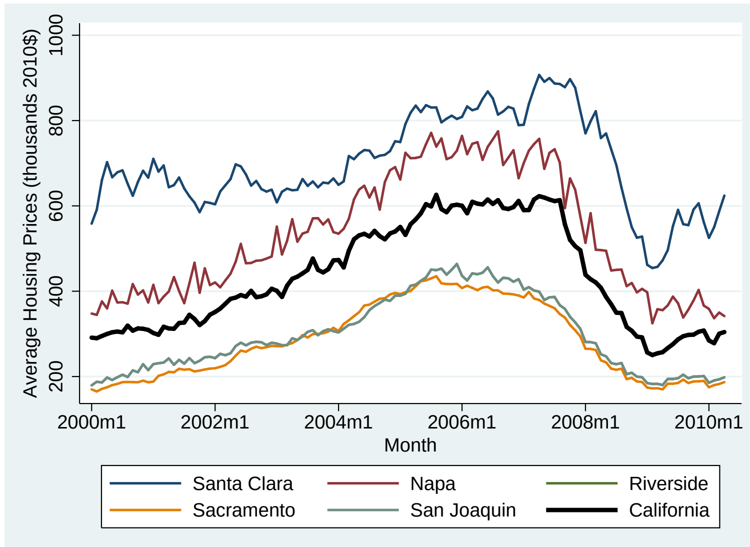

Figure 2.5: Housing prices in California were high during much of the time of the gasoline price increase, but dropped precipitously in late 2007 and early 2008.

In the second group of variables, we have smog check details, income, prices, and economic conditions. I show the variables for gasoline price and economic conditions both at the time of the vehicle purchase and the average over the interval between smog checks. The variable at the time of purchase is used in the estimation of the responsiveness on the extensive margin, while the average over the interval between smog checks is used for the estimation on the intensive margin. The months to first smog check variable has a mean of nearly 70 months until a smog check, which is just under six years. This reflects the feature of the smog check program that most consumers have to perform a smog check within a few months of six model years from the purchase of the vehicle, with exceptions being consumers who transfer the vehicle title outside of the family and hybrid owners.[6] To understand this “months to test” variable better, we can look at the histogram of the variable (Figure 2.6).

The histogram shows that most consumers have a smog check performed within a few months of 72 months (six years). There is another small mass of consumers who have smog checks within a few months of 84 months. These consumers nearly all have

[6]It is worth mentioning that some hybrid vehicles are in the smog check data, presumably because fleets and dealers have all of their vehicles smog checked regardless.

Table 2.1: Personal New Vehicle Dataset Summary Statistics

Variable	Mean	Std Dev	Min	Max	Observations
cylinders	5.81	1.56	2	16	12,334,583
liters	3.33	1.3	0.4	8.4	12,334,583
automatic transmission	0.96	0.21	0	1	12,334,583
gross veh weight rating (000s)	5.32	1.26	0.44	14.05	12,334,583
hybrid	0.02	0.15	0	1	12,334,583
import	0.66	0.47	0	1	12,334,583
safety rating	4.31	0.43	1	5	12,334,583
convertible	0.03	0.16	0	1	12,334,583
turbo	0.03	0.17	0	1	12,334,583
all-wheel drive	0.18	0.39	0	1	12,334,583
fuel economy 2008 ratings	19.22	5.72	8	50	12,334,583
vehicle MSRP (2010$)	29,709	12,492	9,034	1,500,000	12,334,583
months to smog test	69.31	10.87	13	107	4,652,064
VMT	1,088.68	464.95	0	4,993	4,652,064
income category	5.87	2.29	1	9	8,723,983
income (thousands $)	72.74	42.19	7.5	150	8,723,983
resale price of same model 6 yrs old	12,371.14	5948.68	1522	734,325	12,334,583
gas price at purchase (2010$)	2.6	0.63	1.25	5	12,334,583
avg gas price (2010$)	2.92	0.23	2.19	3.69	12,334,508
county unemployment rate	5.97	2	2.8	27.1	12,334,583
county house prices (000s $)	490.27	194.75	94.44	1195.37	12,334,583
consumer confidence index	93.51	18.53	25.3	118.9	12,334,583
avg unemployment rate	7.38	2.37	3.58	29.03	12,334,583
avg housing prices	478.24	173.22	106.95	1072.55	12,334,583
avg consumer conf index	78.84	15.35	47.04	104.36	12,334,508
zip density (000/mi^2)	5.08	5.5	0	52.18	12,334,583
commute time 2000(min)	27.1	4.28	13.4	43.1	12,334,583
zip businesses 2000	1,510.74	957.96	1	6,521	12,334,583
zip businesses/capita	0.05	0.59	0	104.82	12,334,583
zip population 2007	41,404.93	20,470.53	1	109,549	12,334,583
zip pop growth rate 00-07	1.77	3.1	-32.5	199.2	12,334,583
zip median hh income 2007	70,633.08	27,370.53	0	375,000	12,334,583
zip % pop age 65+	11.14	5.29	0	100	12,334,583
zip % pop under 18	25.73	6.06	0	41.3	12,334,583
zip % pop white 2007	59.66	18.55	4.4	100	12,334,583
zip % pop black 2007	5.15	7.45	0	86.60	12,334,583
zip % pop hispanic 2007	31.93	21.46	0	97.8	12,334,583
zip lawn & garden SPI	118.48	55.89	0	486	12,334,583

Note: harmonic mean given for fuel economy (arithmetic mean = 20.5), all-wheel drive includes four-wheel drive

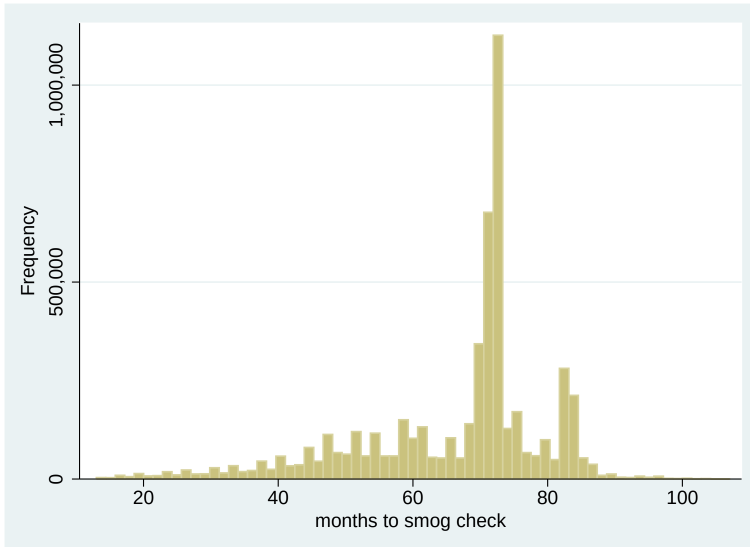

Figure 2.6: The histogram of the months until a smog test in California shows the largest peak right around six years.

a vehicle with a model year greater than the year of purchase (e.g., a model year 2008 vehicle during the year 2007). This makes sense since the law mandates a smog check after six *model years*, rather than six calendar years. Curiously, nearly half of the consumers who bought a vehicle with a model year greater than the year of purchase seem to have done a smog check within a few months of six years anyway. I cannot find a systematic reason for this, and it appears to be a randomization in the administration of the smog check program. Similarly, some of the consumers who purchase a vehicle where the model year is less than the year of purchase appear to have done a smog check right around 60 months (five years). Again, most of the consumers who fall into this category have a standard six year smog check. Attempts to discuss these features of the data and to verify that there truly was random assignment with California BAR have thus far been unsuccessful, but this may be a promising avenue to explore in the future as a source of plausible exogenous variation.

Figure 2.6 also shows that the number of vehicles that have an early smog check is relatively small and bunched at a few points, primarily corresponding to when leases expire. The law does not require a smog check before 48 months (four years), but I do observe a surprising number of smog checks before this time. I attribute these smog check to vehicles being traded-in to dealerships that automatically smog check all incoming vehicles.

Back to Table 2.1, we can note that both the "months to smog check" and VMT variables have roughly 4.6 million observations, which make up 37 percent of the entire 12.3 million observation dataset. This is again due to the fact that only vehicles purchased in 2001 through 2004 have any data from the smog check program, for it is at least four years before a smog check for any vehicle is required and my smog check data are only through 2010. However, for the years 2001 through 2004, the smog check dataset is quite complete. For 2001, 96 percent of the personal vehicles in the R.L. Polk dataset have a VMT and months to test reading. For years 2002, 2003, and 2004 the percentages are 95 percent, 82 percent, and 19 percent respectively. The declining percentages reflect the fact that my smog data only runs through June 2009, so any consumers who were required to do a smog check after June 2009 would not be matched.

The income variable from R.L. Polk gives the household income of the vehicle purchaser at the time of purchase. There are 8.7 million observations with an income reported out of the 12.3 million total observations (70 percent of the dataset). The income variable is quite incomplete (40 percent in 2001) during the first few years of the dataset, but becomes increasingly complete by the end of the dataset (85 percent in each of the years after 2007). The variable is a categorical variable with nine income brackets.[7]

I believe that these income data are quite representative of the actual household income of households that purchased new vehicles. One way to assess this is to examine the distribution of the income in my dataset and compare this to the distribution of household income in California from the Census. Figure 2.7 shows that the distribution in the dataset is sensible in light of the Census household income. The data contains only new vehicle purchasers, who tend to be wealthier than the average

[7]For the analysis, I can also create an second income variable by taking the mean of the endpoints of each income bracket. For the highest income bracket I use $150,000. This interpolated variable is useful in the analysis for both scaling reasons and for ease of interpretation. I recognize that this introduces some measurement error, possibly implying a degree of attenuation bias in the empirical results. Thus, the primary results do not use this interpolated variable.

population. The bins used in Figure 2.7 are the brackets R.L. Polk provides for the income variable.

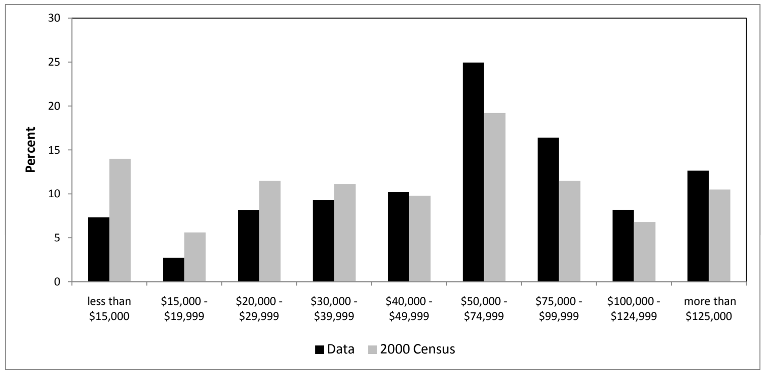

Figure 2.7: The income distribution in the dataset appears sensible relative to the Census household income.

Figure 2.7 shows the extent to which new vehicle purchasers are wealthier. This can be seen by the somewhat heavier weight in the distribution in the higher income brackets in my data than the Census data. However, the overall distribution appears quite plausible, with no very strange masses. The only mass that may be surprising is at the bottom of the income distribution - approximately 7 percent of new vehicles in California are purchased by consumers in the lowest income bracket. There are two explanations for this. One is simply that even low income households may save up for a new vehicle if they have strong preferences for driving a new vehicle. The second, perhaps more likely, explanation is that some low income households may have low earned income, but still have access to wealth. For example, a college student would be considered a separate household making under $15,000, but may have wealthy parents who will buy a new vehicle for them. Table 2.2 provides the cross-tabulation of income and vehicle class in order to give a clearer picture about what consumers in different income brackets are purchasing. The final two columns of the table indicate the observations that are missing income and provide a total for each vehicle class (see Appendix C for a definition of each vehicle class).

An implication from the "Fraction of Vehicles" cross-tabulations in Table 2.2 is that the unconditional probability that a low-income consumer purchases a small car

Table 2.2: Personal New Vehicle Registrations in California

	Counts of Vehicles (thousands)										
	income brackets (000s $)										
	<15	15-20	20-30	30-40	40-50	50-75	75-100	100-125	>125	Missing	Total
Small Car	183	65	137	145	157	409	302	167	174	656	2,393
Large Car	115	44	98	108	122	327	239	133	128	523	1,836
Sporty Car	18	7	17	19	23	66	52	31	34	140	407
Prestige Sporty	3	2	4	4	5	19	20	18	30	63	168
Luxury	33	16	36	41	51	176	174	131	182	399	1,240
Prestige Luxury	4	2	5	5	6	24	28	29	56	101	261
Pickup	49	17	37	40	45	116	78	39	35	189	644
Full Pickup	96	34	78	88	101	275	189	96	90	383	1,430
Sport Utility	123	49	108	120	144	427	350	222	268	725	2,535
Full Utility	38	15	34	38	46	140	115	75	107	245	853
Minivan	27	11	24	27	32	91	74	47	49	188	568
Total	689	261	577	634	731	2,069	1,620	989	1,153	3,611	12,335
	Fraction of Vehicles										
Small Car	0.27	0.25	0.24	0.23	0.21	0.20	0.19	0.17	0.15	0.18	0.19
Large Car	0.17	0.17	0.17	0.17	0.17	0.16	0.15	0.13	0.11	0.14	0.15
Sporty Car	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.03
Prestige Sporty	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.03	0.02	0.01
Luxury	0.05	0.06	0.06	0.06	0.07	0.09	0.11	0.13	0.16	0.11	0.10
Prestige Luxury	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.03	0.05	0.03	0.02
Pickup	0.07	0.06	0.06	0.06	0.06	0.06	0.05	0.04	0.03	0.05	0.05
Full Pickup	0.14	0.13	0.14	0.14	0.14	0.13	0.12	0.10	0.08	0.11	0.12
Sport Utility	0.18	0.19	0.19	0.19	0.20	0.21	0.22	0.22	0.23	0.20	0.21
Full Utility	0.05	0.06	0.06	0.06	0.06	0.07	0.07	0.08	0.09	0.07	0.07
Minivan	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.05	0.04	0.05	0.05
Total	1	1	1	1	1	1	1	1	1	1	1

Vehicle Classes are defined in Appendix C is much higher than the probability of a purchase of a vehicle in any other vehicle class. The unconditional probability that a high-income consumer purchases a sport utility vehicle is much higher than the probability of a purchase of a vehicle in any other vehicle class. Not surprisingly, high income households are also more likely to purchase a luxury or prestige sporty vehicle. Other vehicle classes, such as minivans and sporty cars are consistently purchased across income classes.

Back to the summary statistics in Table 2.1, we can see that the average resale price of the same vehicle model six years old is about $12,370. Comparing this to the average MSRP shows just how substantial the depreciation is for new vehicles during the first several years of life. Interestingly, the depreciation for different types of vehicles changed as gasoline prices changed. I will present this evidence in the next subsection that shows initial evidence of responsiveness to gasoline prices.

The third group of variables in Table 2.1 contains zip code-level demographics and the county-level commute time. Average county-level commutes range from 13 minutes to over 40 minutes. The zip code-averaged median household income in 2007 is just over $70,600, with the maximum zip code median household income of $375,000. The lawn and garden spending-power-index (SPI) roughly captures the amount that consumers spend on lawns and gardens, which may be a reasonable indicator for a suburban area.

2.1.2 Initial Evidence of Responsiveness

Responsiveness in driving

Even from a quick look at the data we can see some descriptive evidence of responsiveness to gasoline price changes. To begin, Figure 2.8 shows the longer-term trend in driving by Californians based on traffic count data on state highways. From the figure, we can see that the amount Californians drive has been on an increasing trend for the past 30 years. However, there are a few noticeable dips in the increase in driving, which appear to correspond with high gasoline prices. For example, there is a clear drop-off in driving in 2007 and 2008.

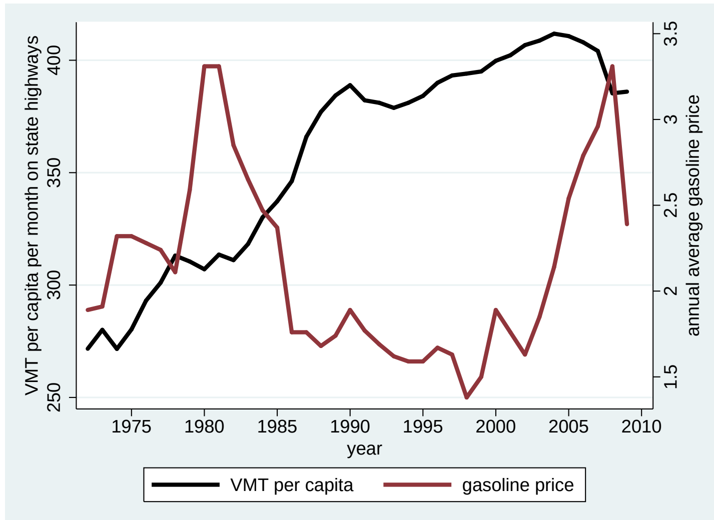

Figure 2.8: Vehicle miles traveled per capita has been increasing steadily, with the only deviations from this trend occurring during times of high gasoline prices (in real 2010$). Sources: California Department of Transportation state highway traffic counts and US Energy Information Administration.

This same suggestive relationship between gasoline prices and driving is also evident in the smog check data. Figure 2.9 illustrates that the average VMT over the first six years of vehicle life has decreased along with the average gasoline price over that same time frame. The x-axis in Figure 2.9 indicates the month of registration for a vehicle that is then smog tested within a few months of six years later. The average VMT per month covers the full six year period. Of course, the evidence in both Figure 2.8 and Figure 2.9 is only suggestive of the inverse relationship between gasoline prices and driving, for economic conditions may also play an important role, and historically have been highly correlated with gasoline prices. Fortunately in my dataset, the high gasoline prices began well before economic conditions deteriorated, so one can view at least the initial decline in driving (and change in purchasing) as influenced by gasoline prices rather than economic conditions.

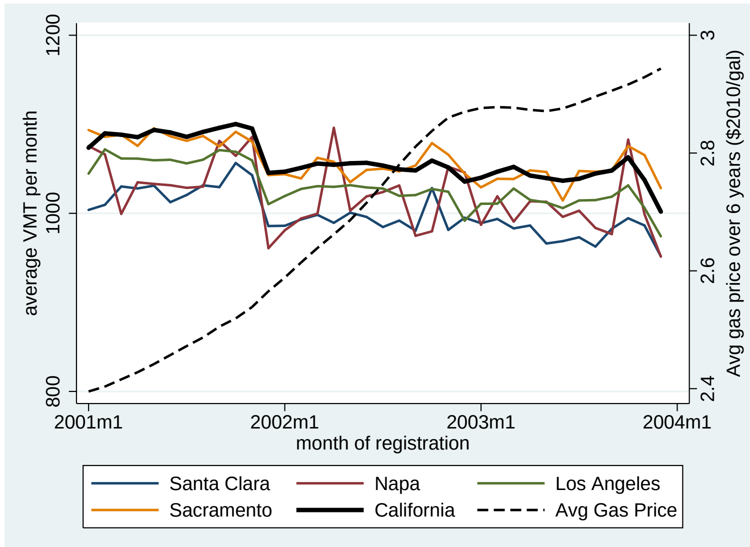

Figure 2.9: Average vehicle miles traveled per month has dipped slightly over the time frame of the study. The average in this graph is taken for all personal vehicles that received a smog check within two months of six years after registration (over 70% of the data), and is the average over the six years between the registration and smog check. The average gasoline price shown is the average over those same six years.

Responsiveness in vehicle purchasing

In new vehicle purchasing, we can also see evidence of responsiveness to gasoline price changes. I show descriptive evidence of responsiveness across vehicle classes, within vehicle classes, and even within models. In driving, I show descriptive evidence of responsiveness in the lower odometer readings of vehicles that faced higher gasoline prices.

Table 2.3 shows the number of personal vehicles registered by year broken down by vehicle class. There are a few striking features of the data that we can see from this table. The first is the decline in the number of new vehicles purchased after 2006. The drop in sales is particularly dramatic between 2007 and 2008. There is then a flattening out from 2008 to 2009 if the numbers from 2009 are extrapolated to 12 months. But what is equally noticeable is the *shift* in sales between vehicle classes with different fuel economy.

Table 2.3: Personal New Vehicle Registrations in California Over Time

	2001	2002	2003	2004	2005	2006	2007	2008	2009*
	Counts of Vehicles (thousands)								
Small Car	287	263	268	279	302	314	316	288	76
Large Car	262	237	222	217	233	232	219	169	46
Sporty Car	80	67	52	47	48	50	36	21	6
Prestige Sporty	19	20	20	20	24	26	21	14	4
Luxury	146	148	153	154	165	161	150	124	38
Prestige Luxury	33	30	34	33	38	36	30	22	6
Pickup	107	92	86	78	87	80	64	40	10
Full Pickup	195	191	198	211	220	179	142	75	18
Sport Utility	297	307	323	334	343	330	308	222	71
Full Utility	125	131	135	123	109	102	77	41	10
Minivan	92	82	74	77	80	70	50	33	10
Total	1,644	1,567	1,566	1,572	1,648	1,581	1,412	1,052	294
	Fraction of Vehicles								
Small Car	0.17	0.17	0.17	0.18	0.18	0.20	0.22	0.27	0.26
Large Car	0.16	0.15	0.14	0.14	0.14	0.15	0.16	0.16	0.16
Sporty Car	0.05	0.04	0.03	0.03	0.03	0.03	0.03	0.02	0.02
Prestige Sporty	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.01
Luxury	0.09	0.09	0.10	0.10	0.10	0.10	0.11	0.12	0.13
Prestige Luxury	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Pickup	0.07	0.06	0.05	0.05	0.05	0.05	0.05	0.04	0.04
Full Pickup	0.12	0.12	0.13	0.13	0.13	0.11	0.10	0.07	0.06
Sport Utility	0.18	0.20	0.21	0.21	0.21	0.21	0.22	0.21	0.24
Full Utility	0.08	0.08	0.09	0.08	0.07	0.06	0.05	0.04	0.03
Minivan	0.06	0.05	0.05	0.05	0.05	0.04	0.04	0.03	0.03
Total	1	1	1	1	1	1	1	1	1

*Only January through May are available for 2009

As we can see in Table 2.3, the change in the fleet began just as gasoline prices began increasing in 2005. In 2004, the fraction of new small cars began creeping up from a steady 0.17 to 0.18. In 2006, 2007, and 2008 the increase continued, corresponding closely with the increase in gasoline prices (and the beginning of the economic downturn). When gasoline prices dropped again in 2009, the fraction declined slightly to 0.26. On the flip side, sales of full utility vehicles began at 0.08 and began dropping in 2006, just as small car sales started increasing. The drop continues for each year until the end of the study time frame, such that by the first several months of 2009, the fraction is only 0.03 – less than half of what it was before. A very similar pattern occurs for the full pickup vehicle segment.

Table 2.3 indicates that there was a clear shift in vehicles purchased across vehicle classes. However, there is also some evidence suggesting a shift in vehicle purchases *within* vehicle classes towards more fuel efficient vehicles. Table 2.4 provides this descriptive evidence by dividing each new vehicle class into quartiles based on the fuel economy of the vehicle. For any vehicle class, the first quartile contains the quarter of the models with the lowest fuel economy. The fourth quartile contains the quarter of models with the highest fuel economy. For example, the first vehicle class in the table is the "small car" vehicle class. The lowest quartile in the small car vehicle class has a harmonic mean fuel economy of 21.9 miles per gallon, while the highest quartile has a harmonic mean fuel economy of 31.7 miles per gallon. The three vehicle classes presented in Table 2.4 account for much of the new vehicle fleet and were chosen for they are the three large vehicle classes with sufficient variance in the mean fuel economy across the quartiles to make an interesting comparison.

The fractions shown in Table 2.4 show a pattern suggestive of within-class shifts in vehicle purchasing behavior over time. For example, in the small car vehicle class, the lowest quartile begins with 11 percent market share in 2001, but has that market share decrease to 7 percent by 2006 – with the major decrease occurring when the price of gasoline began increasing in 2005. On the other hand, the highest quartile begins with an 50 percent market share in 2001 and by 2008 the market share has increased to 61 percent. Pickups show a similar, but slightly more nuanced, story. The lowest quartile decreases over the time frame. The highest quartile exhibits considerable

Table 2.4: Fuel Economy Quartiles of Vehicle Classes Over Time

Quartile	FE	2001	2002	2003	2004	2005	2006	2007	2008	2009*
	Counts of Vehicles (thousands)									
	Small Car									
1	21.9	32	29	28	28	25	23	23	20	21
2	24.1	48	44	49	45	42	36	36	37	32
3	25.8	61	58	66	73	88	79	79	66	57
4	31.7	143	131	125	132	146	175	175	191	175
Total		283	262	268	278	301	313	313	314	284
	Pickup									
1	15.5	11	8	6	6	6	5	5	3	1
2	16.7	29	24	22	14	14	14	14	10	5
3	17.4	43	39	39	36	48	47	47	40	26
4	18.9	24	20	18	21	20	15	15	11	8
Total		106	91	86	78	87	80	80	63	40
	Sport Utility									
1	14.8	104	97	88	82	75	54	54	46	29
2	16.7	76	84	93	94	98	94	94	69	41
3	18.6	13	14	13	26	30	29	29	52	35
4	21.1	103	113	129	131	140	152	152	141	116
Total		296	307	323	333	342	329	329	308	221
Quartile	**Fraction of Vehicles**									
	Small Car									
1	21.9	0.11	0.11	0.11	0.10	0.08	0.07	0.07	0.06	0.07
2	24.1	0.17	0.17	0.18	0.16	0.14	0.12	0.12	0.12	0.11
3	25.8	0.22	0.22	0.25	0.26	0.29	0.25	0.25	0.21	0.20
4	31.7	0.50	0.50	0.47	0.48	0.49	0.56	0.56	0.61	0.61
	Pickup									
1	15.5	0.10	0.09	0.07	0.08	0.07	0.06	0.06	0.04	0.02
2	16.7	0.27	0.26	0.25	0.18	0.16	0.17	0.17	0.16	0.13
3	17.4	0.40	0.43	0.46	0.47	0.55	0.59	0.59	0.63	0.64
4	18.9	0.22	0.22	0.21	0.27	0.23	0.18	0.18	0.17	0.21
	Sport Utility									
1	14.8	0.35	0.31	0.27	0.25	0.22	0.17	0.17	0.15	0.13
2	16.7	0.26	0.27	0.29	0.28	0.29	0.28	0.28	0.23	0.18
3	18.6	0.04	0.04	0.04	0.08	0.09	0.09	0.09	0.17	0.16
4	21.1	0.35	0.37	0.40	0.39	0.41	0.46	0.46	0.46	0.53

*Only January through May are available for 2009

FE refers to the quartile harmonic mean fuel economy in miles per gallon variability, but remains roughly flat. However, the second-highest quartile shows an increase in market share from around 40 percent to over 60 percent. Far more of the popular models are in the second-highest quartile, so it is understandable that the switch moves into this quartile, rather than the highest fuel economy quartile. Sport utility vehicles display a very similar pattern to small cars. Other vehicle classes also show a similar pattern, but in many cases with more nuances, as in the pickup class.

So far, the descriptive evidence is suggestive of across-vehicle class switching and within-vehicle class switching. In addition, there is even some suggestive evidence of within-model switching. This could be due to consumers choosing a model first and then choosing a different engine size (e.g., four cylinder rather than six cylinder) in order to gain higher fuel economy. To see this, we must look at individual models. For example, the Honda Accord comes in a four cylinder and six cylinder version for all of the model years in my dataset. In 2001, the market share of the six cylinder engine version (21 miles per gallon versus 24 miles per gallon for the four cylinder) was 31 percent. In 2009, the difference in fuel economy between the two versions remains roughly the same, but the market share of the six cylinder engine version decreases to 25 percent. Of course, other aspects of the vehicle model changed as well, so this evidence is highly suggestive. However, I observe this feature for many vehicles in the dataset. Some others may show the opposite behavior, but this can be attributed to the introduction of new models that emphasize the larger engine, which was especially common in SUVs and pickups before the highest prices in 2007 and 2008.[8]

With evidence of across-vehicle class switching, within-vehicle class switching, and within-vehicle model switching towards higher fuel economy vehicles, it is no surprise that the fuel economy of the fleet increased in the years with higher gasoline prices. This is shown clearly in Figure 2.10. One way to interpret this figure is as follows. Until 2006, the harmonic mean fuel economy of the new vehicle fleet in California had not changed very much in nearly a decade. For many years, automakers chose to improve other desirable attributes of vehicles rather than the fuel economy, with this

[8]The length of time it takes for a new model to go from drawing board to showroom is considered to be about five to six years, although some tweaks can be made along the way (Klier and Linn 2010b).

choice at least partly due to low gasoline prices (Knittel 2010). Yet in 2006 as the gasoline price started increasing, the harmonic mean fuel economy of new vehicles began inching upwards. It then peaked in 2008 along with the gasoline price peak, before dropping again as gasoline prices returned to lower levels. The dip in the mean fuel economy in the summer of 2005 may be attributable to the "Employee Pricing Sale" of General Motors, Ford, and Chrysler, which succeeded in greatly bolstering the sales of domestically-made sport utility vehicles and light trucks. This underscores the need for the econometrics in the following chapters to account for such price changes.

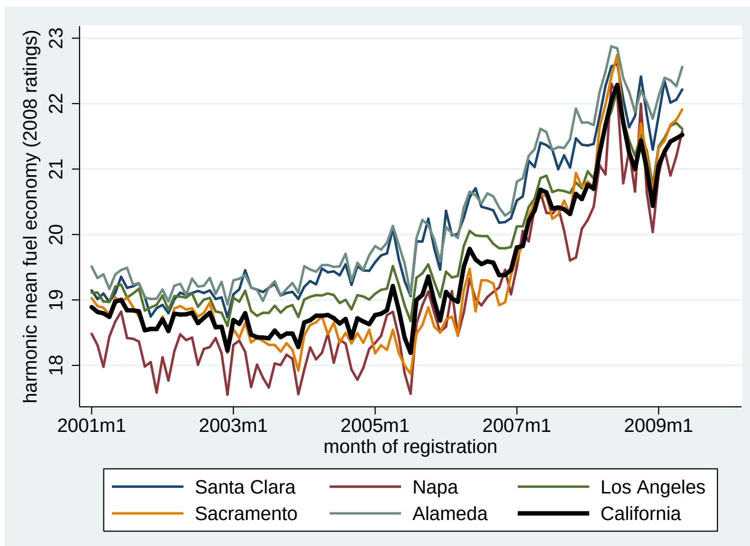

Figure 2.10: The mean fuel economy of the new fleet in California was flat and then peaked at the same time as the gasoline price peaked.

Responsiveness in used vehicle prices

There is also some suggestive evidence of the prices of different classes of used vehicles have changed along with gasoline prices. Table 2.5 shows the harmonic mean fuel economy for each vehicle class and then the mean ratio of the resale price to the price of the vehicle in each year. The ratios over time can be viewed in two ways. On one hand, there is a clear shift in depreciation *within* broader vehicle categories.

For example, in 2001 the ratio for sport utility vehicles is 0.45 and for full utility vehicles is 0.47, while in 2008 the ratio for sport utility vehicles is 0.44 and for full utility vehicles is 0.37. This suggests that when gasoline prices are low, as in 2001, the lower fuel economy vehicles within the "utility vehicle" class depreciate less, but when gasoline price are high, as in 2008, the higher fuel economy vehicles within the same class depreciate less. The same holds within several other vehicle classes: sporty cars, luxury cars, and pickups. Small cars and large cars appear to depreciate similarly over time.

Table 2.5: Ratio of Resale Price After Six Years to MSRP

	FE	2001	2002	2003	2004	2005	2006	2007	2008	2009*
Small Car	27.6	0.42	0.40	0.37	0.38	0.42	0.43	0.42	0.42	0.41
Large Car	22.4	0.37	0.35	0.33	0.33	0.38	0.41	0.38	0.40	0.38
Sporty Car	20.2	0.43	0.43	0.42	0.42	0.41	0.42	0.41	0.42	0.39
Prestige Sporty	18.3	0.50	0.48	0.46	0.44	0.45	0.44	0.41	0.40	0.36
Luxury	20.3	0.43	0.43	0.43	0.42	0.42	0.42	0.42	0.41	0.39
Prestige Luxury	17.5	0.40	0.43	0.42	0.35	0.38	0.37	0.35	0.33	0.33
Pickup	17.3	0.46	0.43	0.41	0.40	0.46	0.52	0.53	0.52	0.49
Full Pickup	14.6	0.48	0.46	0.45	0.45	0.48	0.47	0.42	0.40	0.32
Sport Utility	18.0	0.45	0.44	0.44	0.44	0.45	0.45	0.44	0.44	0.39
Full Utility	13.9	0.47	0.44	0.40	0.38	0.38	0.39	0.39	0.37	0.34
Minivan	18.6	0.37	0.38	0.36	0.32	0.36	0.35	0.33	0.31	0.28

Note: Fuel economy is the harmonic mean fuel economy over all observations in units of miles per gallon

On the other hand, when looking *across* broader vehicle categories, there is more noise and somewhat weaker suggestive evidence of a relationship. Figure 2.11 plots the ratio of the 2008 depreciation to the 2001 depreciation in Table 2.5. The linear trendline plotted on the graph indicates the positive relationship. However, the wide scatter in the plot indicates that the relationship is quite noisy. More detailed inspection of individual points suggests that the highest points on the graph are the higher fuel economy vehicle classes within broader vehicle categories. For example, the highest point in the graph is the "pickup" vehicle class, which actually displayed less depreciation than in 2001 than 2008 – despite a relatively low fuel economy. Of course, this may be simply because when gasoline prices changes, consumers who were looking for a used pickup truck were willing to pay more for smaller, higher fuel economy pickup trucks than lower fuel economy full pickups.

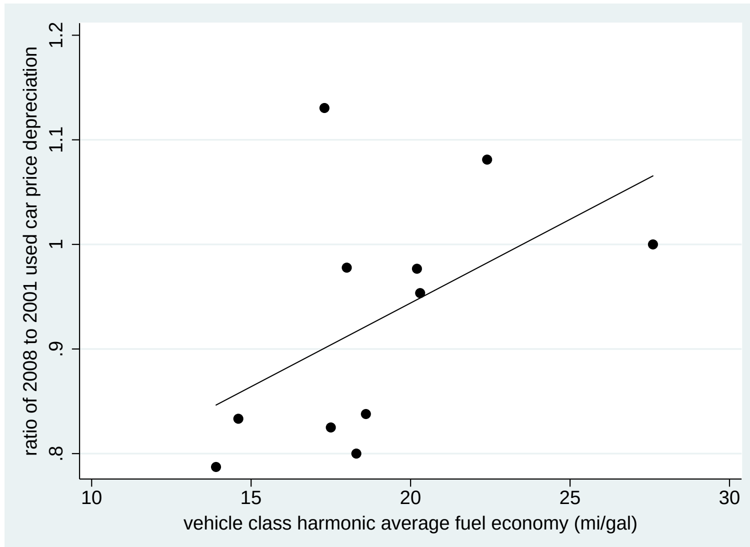

Figure 2.11: The mean fuel economy of the new fleet in California was flat and then peaked at the same time as the gasoline price peaked.

The evidence shown above is of course only indicative that the demand for lower fuel economy used vehicles is decreasing with higher gasoline prices. This view of the data is consistent with the results in Busse, Knittel, and Zettelmeyer (2010), but used car prices are the result of supply and demand, and I have no data on the supply side of the used car market. The changes in used car prices do show that the *equilibrium* price has been adjusting. Under the assumption that supply reacts slowly in the used vehicle market due to the relatively small inflow and outflow into the stock of vehicles each year, the equilibrium price changes in the short run must be from changes in demand.

Of course, this evidence, and the evidence given above on changing purchasing patterns and driving behavior should be taken as descriptive evidence that is only suggestive of a relationship. However, it helps to clarify what the dataset is telling us and helps to elucidate the variation that will be identifying the parameters in the remaining chapters of this dissertation.

2.2 All Vehicles

The second dataset I use in this dissertation allows me to examine the driving responsiveness to gasoline price changes in the broader California light-duty vehicle fleet, rather than only personal vehicles during the first six years of life. The greater breadth of the dataset comes at a cost: I know much less about the vehicles and the vehicle owners in this dataset. The primary foundation for this dataset is the smog check data, rather than the R.L. Polk data, although the R.L. Polk data are included as well. Instead of restricting the dataset to vehicles purchased in 2001 to 2004, I now include vehicles going back all the way to model year 1976.[9] In this sense, the dataset includes nearly all of the California light duty vehicle fleet over several years. For this dataset, I bring in smog check data from the beginning of 2002 to the end of 2009.

A major advantage of this dataset is that I can use the biennial smog check readings for all vehicles registered before 2002 in addition to the readings at the end of six years. Using a two year interval rather than a six year interval is very useful for there is a shorter period for the variation in gasoline prices to be averaged out in the calculation of the average gasoline price. An unfortunate feature of this dataset is that very recently purchased vehicles (e.g., purchased after 2004) are not included, for they have not yet received the first smog check.

The number of observations in the raw smog check data ranges from 9.6 million tests in 2002 to 11.6 million tests in 2009. Of course, many of these tests are multiple tests for the same vehicle. For each test, I observe the VIN of the tested vehicle, the date of the test, the test station, the odometer reading, the test result, the body type of the vehicle, the engine size and displacement of the vehicle, and the transmission type of the vehicle (e.g., automatic or manual). These data are quite messy, so I perform several rounds of data cleaning, which are documented in detail in Appendix B. Then, I convert the smog check dataset to one where each observation is a "vehicle-period between two tests." In other words, if vehicle A was tested in 2002, 2004, and 2006, there would be an observation for 2002-2004 and 2004-2006.

[9]Vehicles first registered prior to 1976 are not required to have a smog check.

This approach allows me to examine the relationship between the average gasoline price and economic conditions over a period of time and the actual number of miles driven over that same time frame.

Besides gasoline prices and economic conditions, I also merge in the R.L. Polk new vehicle registration data from 2001 to 2004. Bringing in the R.L. Polk data allows me to capture new vehicles that may have had a smog check within the time frame of my smog check data. In this sense, the "personal vehicle in first six years" dataset described above is subsumed within my "all vehicles" dataset. As in the new personal vehicles dataset, I assume an odometer reading of zero for the new vehicle registrations.[10]

2.2.1 Summary Statistics

The full dataset contains 49.7 million observations. Among these there are 19.3 million unique vehicles that had a period between two tests in 2002 to 2009. We can get a sense of how much of the total stock of light duty vehicles in California in any given year that this sample covers by examining vehicles for which the period overlaps that year. Table 2.6 compares the full stock of vehicles in the Department of Motor Vehicles (DMV) registration data compiled by the California Air Resources Board for the EMFAC Air Quality Model.[11] The years that include the most observations are shown in the table.

Table 2.6 indicates that for several of the key years of the sample, as much as 81 percent of the total registrations reported in the DMV are included in the smog check data. The remaining 19 percent may be exempt vehicles or miscoded VINs. The drop-off in the counts of vehicles in the dataset in 2003 and after 2006 are due to the fact that the R.L. Polk data began in January 2001 and are only relevant through 2004. After 2004, the length of time to the first smog test has not passed yet, so vehicles

[10]Of course, there may be some miles on all vehicles from test drives, but by law these cannot be more than 3,000 miles. This issue would influence all new registrations equally so it is not likely to be more than measurement error in my results. Future work to bring in the actual DMV data would give me the actual odometer reading at the time of the new vehicle registration.

[11]These numbers are publicly available at http://www.arb.ca.gov/app/emsinv/trends/ems_trends_results.php. The EMFAC Model itself can be downloaded at http://www.dot.ca.gov/hq/env/air/pages/emfac.htm.

Table 2.6: Comparison of the Entire Light Duty Vehicle Stock to Full Dataset

	2003	2004	2005	2006	2007	2008
DMV Total Registrations (000s)	22,460	23,304	23,501	23,280	23,255	23,646
DMV Passenger Vehicles (000s)	12,745	13,050	12,923	12,810	12,799	13,021
DMV SUVs/Light Trucks/Vans (000s)	9,715	10,254	10,578	10,469	10,456	10,625
Dataset Total Vehicles(000s)	15,533	18,916	19,018	18,287	17,289	15,610
Passenger Vehicles (000s)	8,884	10,778	10,742	10,237	9,599	8,614
SUVs/Light Trucks/Vans (000s)	6,649	8,134	8,276	8,050	7,690	6,996
Fraction of Total Fleet in Data	0.69	0.81	0.81	0.79	0.74	0.66
Passenger Vehicles	0.70	0.83	0.83	0.80	0.75	0.66
SUVs/Light Trucks/Vans	0.68	0.79	0.78	0.77	0.74	0.66

purchased after 2004 cannot be included (with perhaps some exceptions at dealerships and government offices). Interestingly, the difference between the count of vehicles in the 2007 and 2008 datasets is very similar to the number of new registrations in those years (1.6 million). The upshot of this feature of the data is that the results that include the greatest changes in the gasoline price will be missing some of the newest vintages of vehicles.

The last two rows of Table 2.6 indicate that the missing registrations in my dataset do not differ much by the type of vehicle. Slightly more of the non-passenger vehicles are missing, but the percentage difference is small – and is constant enough over time that it may simply be due to a difference in vehicle definitions. This is reassuring and suggests that there is not a major selection issue in what registered vehicles are missing.

We can also examine the counts of vehicles by body type, both in terms of observations and in terms of vehicles. Table 2.7 shows that the fraction of observations with each body type is very similar to the fraction of vehicles with each body type. This is important because any analysis run on the observations will effectively be weighting different types of vehicles more heavily if they are observed more often. Any differences are due to more observed smog checks for any given body type, which may relate to the liquidity in the used car market of vehicles with that body type. Fortunately, the differences appear extremely minor.

I find that looking at tabulations of the body type by year yields a similar result. The same can be said for other observed attributes of the vehicles.

Table 2.7: Counts of Observations and Vehicles by Body Type

	Obs (000s)	Fraction
Sedan/Coupe	27,655	0.56
Pickups	8,294	0.17
Sport Utility	7,878	0.16
Passenger Vans	3,640	0.07
Station Wagon	1,012	0.02
Van	1,220	0.02
Total	49,698	1.00
	Vehicles (000s)	Fraction
Sedan/Coupe	10,197	0.53
Pickups	3,449	0.18
Sport Utility	3,255	0.17
Passenger Vans	1,470	0.08
Station Wagon	361	0.02
Van	534	0.03
Total	19,266	1.00

Similar to the new personal vehicle dataset, I can examine the months between two tests. Figure 2.12 shows that the vast majority of smog tests are done with a two year interval. There is also some mass of the density throughout the first two years, corresponding to title transfers to consumers outside of the family of the original owner. There is less mass after two years, which may be due to vehicles temporarily taken out of service, violators, or perhaps in some instances vehicles that had another smog check in between that had a miscoded VIN. There is a smaller peak at six years, which is entirely from the 2001 to 2004 R.L. Polk data being merged into the full smog dataset. If the R.L. Polk data are dropped from the dataset, the mass at six years disappears.

The basic summary statistics for the dataset are given in Table 2.8. Comparing Table 2.8 to Table 2.1 shows how many fewer variables I have to work with in the "all vehicles" dataset. It also illustrates the differences between the full stock of vehicles and vehicles that are in the first six years of life. The most important difference is that vehicles in the first six years of life are driven more (a mean of 1,089 miles per month) than in the all vehicles dataset (a mean of 833 miles per month). Of course, the dataset includes some vehicles in the first six years of life from the R.L. Polk data.

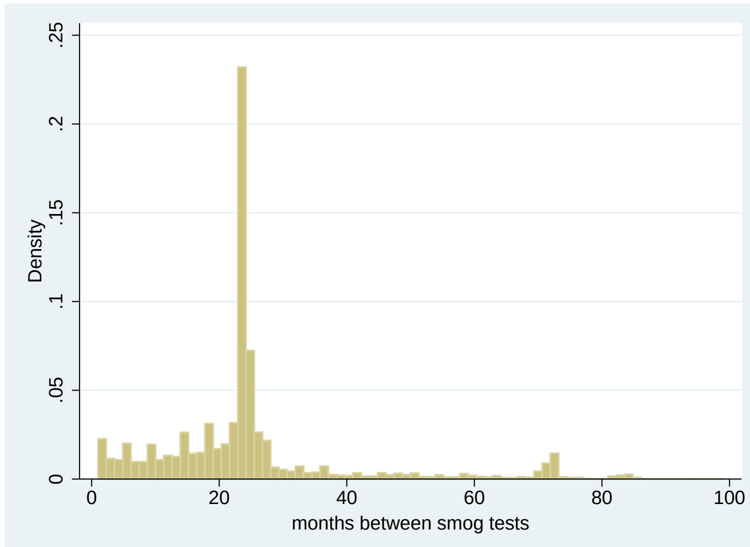

Figure 2.12: The histogram of the months until a smog test for *all* vehicles in California shows a primary peak around two years, with a small peak around six years corresponding to the R.L. Polk data from 2001 to 2004.

If the R.L. Polk data are excluded, the mean drops to 816 miles per month.

Table 2.8: All Vehicles Dataset Summary Statistics

Variable	Mean	Std Dev	Min	Max	N
miles driven per month	833.61	592.66	0	6083.34	49,698,340
average gas price between tests	2.80	0.44	1.25	4.91	49,698,340
average unemployment rate between tests	6.14	1.88	3.1	31.3	49,698,340
average CCI between tests	90.55	15.06	25.3	111.9	49,698,340
average house price between tests	514.11	186.4	99.2	1195.37	49,698,340
cylinders	5.61	1.54	1	16	49,667,869
liters	3.23	1.34	0.5	10	49,698,340
automatic transmission	0.84	0.37	0	1	49,698,340
commute times (min)	27.63	2.89	13.4	34.5	49,698,340

I can further explore how driving changes with the age of vehicles by examining how vehicles of different model year vintages are driven differently. Table 2.9 shows the number of observations in the dataset in each model year vintage and the mean amount of driving per vehicle per month over the periods in 2002 to 2009 for which smog check odometer readings were taken. While there is plenty of survey evidence that utilization declines with the age of a vehicle (e.g., Pickrell and Schimek (1999)), this may be some of the very first large-scale revealed preference evidence of how driving declines with the age of a vehicle.

I perform a variety of checks to confirm that the nature of the data is not influencing the decline in driving with the age of the vehicle. For example, I was concerned that the timing of the checks may make a difference and influence the results. I find that none of the observables such as test dates, average gasoline prices, and economic conditions have any pattern along with these data, as I expected. To show that changing gasoline prices are not the reason for the decline in driving seen in Table 2.9, I include the real average gasoline price over the time between the tests in an additional column. This column shows that the only pattern works in the opposite direction: more recent model years face slightly higher gasoline prices for they are more likely to be covering six years that include the peak in gasoline prices in 2007 and 2008.

Table 2.9 helps fill out the story for how the mean amount driven varies with the model year of a vehicle. Old vehicles are driven less than a quarter of the amount that newer vehicles are driven. Newer vehicles make up a very large share of the fleet, with the oldest model years having only a tiny fraction of more recent model years on the road. This corresponds with the results in Davis and Kahn (2011), which indicate that many of the oldest, most emitting vehicles are exported to Mexico or other countries in Latin America. It is also possible that some of the oldest vehicles are only driven in very rural areas and are not smog checked (or registered). Of course, the miles driven by such vehicles would be small and thus are not likely to be a major concern.

An issue with Table 2.9 is that it covers all years of data in the dataset and thus is partly confounding the model year of the vehicle with the year of calculation of the means of VMT and average gasoline prices. For an even more clear presentation of how driving of different model year vehicles changed over time as new vehicles enter the fleet and gasoline prices change, Table 2.10 shows the average VMT by model year and the midpoint year between tests (i.e., the year that is the average of the years between the two observed tests). Table 2.10 indicates that older vehicles are

Table 2.9: Driving per Vehicle by Model Year

Model Year	Obs(000s)	VMT(mi/month)	avg gas price(2010$)
2002	2,763	1,099.9	2.73
2001	3,866	1,030.1	2.82
2000	4,128	963.1	2.94
1999	4,417	929.4	2.86
1998	4,175	920.3	2.79
1997	3,796	863.5	2.83
1996	3,215	831.6	2.77
1995	3,292	769.5	2.81
1994	2,763	746.0	2.77
1993	2,365	715.1	2.77
1992	1,960	684.3	2.76
1991	2,005	646.9	2.76
1990	1,775	611.5	2.74
1989	1,567	575.7	2.72
1988	1,199	549.8	2.70
1987	1,029	518.8	2.69
1986	832	487.3	2.69
1985	592	452.6	2.68
1984	421	422.3	2.68
1983	221	393.8	2.68
1982	162	372.4	2.68
1981	119	350.2	2.69
1980	9	333.7	2.71
1979	14	289.8	2.72
1978	12	274.9	2.73
1977	10	267.9	2.73
1976	7	249.0	2.74

driven substantially less than newer vehicles.

Table 2.10: Driving per Vehicle by Model Year Over Time

	average VMT in mi/month by the midpoint year between tests							
Model Year	2002	2003	2004	2005	2006	2007	2008	2009
2003					1,058	1,006	1,019	949
2002				1,061	1,037	1,031	941	911
2001				1,045	1,054	950	898	890
2000			1,042	1,020	963	903	883	843
1999		1,077	1,027	991	901	875	802	807
1998	1,021	1,053	1,024	924	900	813	811	783
1997	1,020	1,041	940	913	814	803	730	760
1996	966	955	922	824	803	726	736	725
1995	944	929	832	800	714	707	650	696
1994	897	867	815	744	704	655	651	675
1993	869	830	769	715	668	640	622	668
1992	836	789	735	681	639	613	599	644
1991	789	741	689	640	602	582	572	623
1990	749	697	650	602	566	549	541	588
1989	713	659	612	562	527	511	501	547
1988	682	626	582	534	501	485	479	524
1987	653	592	545	504	469	457	447	489
1986	624	556	515	469	441	428	422	456
1985	592	520	478	435	402	390	383	422
1984	562	491	445	404	369	364	355	386
1983	530	457	418	374	343	339	332	355
1982	514	434	393	354	323	321	314	331
1981	483	410	369	328	307	309	301	308
1980	419	398	346	333	287	303	277	260
1979	469	334	319	283	247	260	226	231
1978	418	323	286	265	232	237	269	264
1977	369	305	271	239	243	285	254	232
1976	384	289	248	230	221	275	195	221

2.2.2 Initial Evidence of Responsiveness

The all vehicles dataset is best suited to look at the responsiveness in driving when gasoline prices change. Perhaps the most straightforward way to examine the driving responsiveness is simply to look at how the average amount of driving between two tests changes over time as the gasoline price and economic conditions change. In looking at the average driving, it is important to be careful about how new vehicles are dealt with. Specifically, the new vehicles merged in from the R.L. Polk data were only given a smog check after roughly six years, and thus only had a smog check in 2006, 2007, 2008, or 2009. As shown above, these vehicles are likely to be driven more than the larger stock of vehicles, so one would expect to find more driving in the years for which these are included in the data.

Figure 2.13 shows the average amount of driving between smog tests and the average gasoline price over that interval. The x-axis plots the date of the mid-point between the two observed tests. For consistency, I drop all of the R.L. Polk new vehicle data. So the VMT numbers are only for vehicles older than six years and the period between tests to keep in mind is roughly two years. However, the graph looks nearly the same when the R.L. Polk new vehicle data are included.

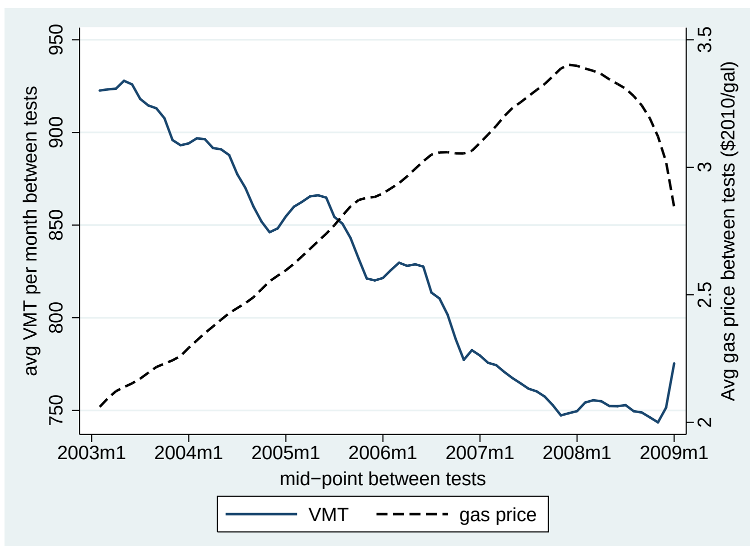

Figure 2.13: The plot of average VMT between tests and average gasoline price over the same time frame shows a clear negative relationship. This plot includes only vehicles over six years old. The x-axis denotes the mid-point between the two observed test dates.

The relationship between VMT and the gasoline price in Figure 2.13 is very clearly negative, providing further suggestive evidence that as gasoline prices rise, consumers cut back on the amount they drive. Just as before, we may be concerned that other factors may also be influencing driving, such as economic conditions. I find that the unemployment rate, house prices, and Consumer Confidence Index all do not appear to be as highly correlated with VMT as the gasoline price. In fact, the correlation coefficient between VMT and the gasoline price is roughly -0.9, while the correlation coefficient between all of the economic indicators and VMT are on the order of -0.1. We can see why graphically in Figure 2.14.

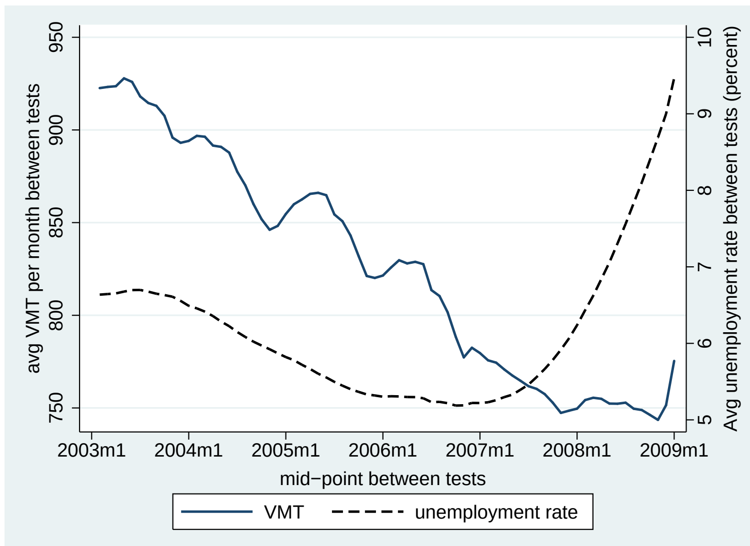

Figure 2.14: The plot of average VMT between tests and average unemployment rate over the same time frame indicates that while economic conditions may play a role in influencing driving, it is less than we might expect. Note the averaging here is over the observations and the time between tests, not the population, as in Figure 2.4

Figure 2.14 shows that economic conditions were actually improving during part of the time when driving was decreasing. In 2009, the economic conditions had greatly deteriorated, but the price of gasoline had dropped back down – two effects that may have canceled each other out and led to roughly constant driving.

Chapter 3

Regression Evidence and Heterogeneity

This chapter presents a set of ordinary least squares (OLS) and fixed effects regression results to examine both the consumer responsiveness to gasoline price changes and the heterogeneity in this responsiveness.[1] OLS and fixed effects regressions facilitate exploration of the different sources of variation in the data to deepen our understanding of what the data can and cannot identify. The richness of the dataset is essential for exploring the heterogeneity in the responsiveness across geography, income groups, and demographic groups.

The point estimates in this chapter of the responsiveness to gasoline price changes fall squarely within the literature, with only relatively minor differences based on the model specification and variation being used. Perhaps the most striking finding in this chapter is how heterogeneous the responsiveness is. In addition to showing evidence of heterogeneity by geography, income, and demographics, this chapter also uses the "new personal vehicles" dataset and the "all vehicles dataset" to show how the responsiveness differs by the vintage of the vehicle.

[1] An early version of this analysis is in the paper "Identifying the Elasticity of Driving: Evidence from a Gasoline Price Shock in California," which was awarded the Dennis O'Brian US Association for Energy Economics Best Student Paper Award.

These results enrich our perspective about how and where the response to a gasoline price changes actually occurs. In a policy analysis, we may be interested in where the response occurs in order to quantify the co-benefits of the policy from reducing local air pollutants and congestion. We may be interested in who makes the adjustments when gasoline prices change in order to better understand the distributional consequences of policies to reduce emissions from the vehicle fleet. Chapter 5 uses the insights from this chapter to shed light on policy.

The analysis in this chapter addresses a variety of possible confounds. However, it does not address the selection bias that may occur when consumers who have different unobserved preferences for driving select into different vehicles. Chapter 4 tackles this issue, develops a new framework for addressing it, and quantifies its importance.

This chapter is structured as follows. I begin by discussing the estimation of the price elasticity of driving and performing several regressions to give a sense of what this elasticity might be for both new personal vehicles and all vehicles. Then I move to examining the heterogeneity in the elasticity of driving in several different ways. Finally, I examine the elasticity of fuel economy with respect to the price of gasoline and the heterogeneity in the responsiveness of new vehicle purchases to changes in gasoline prices.

3.1 Responsiveness in Driving

Most authors have thought about quantifying how consumers change how much they drive when gasoline prices change by modeling driving demand as a function of the cost of driving, consumer characteristics, and vehicle characteristics. In modeling this relationship, there are several important details to keep in mind. The first is the choice of specification. The second is the horizon or time frame that is being modeled. This has implications for the validity of different specifications as well as for whether the estimates can be viewed as long-run or short-run. The third is the choice of regressor: cost per mile of driving or gasoline price. The last is the level of aggregation at which the decision of how much to drive is being modeled.

In this chapter, I examine some of the most common specifications used by previous authors. My detailed vehicle-level dataset allows me to model driving choice at the vehicle level. To formalize such a vehicle-level model, consider the demand for driving by a vehicle owner i at time $t \in \{1, ..., T\}$. We can model this demand, VMT_{it}, as a function of the cost per mile of driving $C_{it} \in \mathbb{R}^1$, a vector of characteristics of the vehicle being driven $\boldsymbol{V}_i \in \mathbb{R}^{\mathcal{V}}$, location- or driver-specific demographics $\boldsymbol{D}_i \in \mathbb{R}^{\mathcal{D}}$, and economic conditions $\boldsymbol{E}_{it} \in \mathbb{R}^{\mathcal{E}}$:

$$VMT_{it} = f(C_{it}, \boldsymbol{V}_i, \boldsymbol{D}_i, \boldsymbol{E}_{it}). \tag{3.1}$$

From this basic setup, many variants are possible. The exact specification of this functional relationship varies by study. A log-log specification is perhaps the most common, but a linear specification is common as well, and some papers model the relationship semi-parametrically. In addition, studies differ in how exactly the cost of driving is treated. Studies also differ in whether VMT is treated as a function of the lagged VMT to model hysteresis. As discussed in detail in Chapter 1, this may lead to biased and inconsistent estimates, so I will not examine such specifications in this chapter.

Choice of specification

To begin, consider the common log-log specification.[2] A pure form of the log-log specification assumes a model where VMT is a multiplicative function of all of the covariates and an error term. Then taking logs, we have a linear-in-logs form for the relationship between VMT and the covariates:

$$\begin{aligned} \log(VMT_{it}) =& \beta_0 + \beta_C \log(C_{it}) + \beta_{V_1} \log(V_i^1) + ... + \beta_{V_{\mathcal{V}}} \log(V_i^{\mathcal{V}}) + \beta_{D_1} \log(D_i^1) + ... \\ &+ \beta_{D_{\mathcal{D}}} \log(D_i^{\mathcal{D}}) + \beta_{E_1} \log(E_{it}^1) + ... + \beta_{E_{\mathcal{E}}} \log(E_{it}^{\mathcal{E}}) + \varepsilon_{it}, \end{aligned} \tag{3.2}$$

[2] All logs are natural logarithms.

where V_i^k is the k-th vehicle characteristic, D_i^k is the k-th individual characteristic, E_{it}^k is the k-th economic condition, and ε_{it} is a mean-zero stochastic error term. Including VMT and C_{it} in logarithm form in the specification is extremely convenient, for it allows us to interpret the coefficient β_C as the elasticity of driving with respect to the cost per mile of driving.

An even more common variant of a log-log specification is one where all the covariates except $\log(C_{it})$ enter linearly in levels:

$$\log(VMT_{it}) = \beta_0' + \beta_C' \log(C_{it}) + \boldsymbol{\beta}_V' \boldsymbol{V}_i + \boldsymbol{\beta}_D' \boldsymbol{D}_i + \boldsymbol{\beta}_E' \boldsymbol{E}_{it} + \varepsilon_{it}', \tag{3.3}$$

where ε_{it}' is again a mean-zero stochastic error term. The interpretation of β_C' here is again the elasticity of driving with respect to the cost per mile of driving.

The two specifications are very similar and should provide similar estimates of the coefficient on $\log(C_{it})$. In fact, the only reason that β_C' may differ from β_C is if there is a different relationship between $\log(C_{it})$ and a non-linear form of the covariates than a linear form of the covariates. This would raise questions about the robustness of the results to the specification of the model.

The exact specifications in the literature of course vary in minor ways from the two given above. Including the lagged VMT is one way that many studies differ. Another way is that many studies do not include economic conditions in the model. To the extent that gasoline prices and economic conditions are correlated, this presents a case of a clear case of omitted variables bias.

A more significant departure from (3.2) and (3.3) is to use a linear specification where all covariates enter in levels. For example, consider the model

$$VMT_{it} = \gamma_0 + \gamma_C C_{it} + \boldsymbol{\gamma}_V \boldsymbol{V}_i + \boldsymbol{\gamma}_D \boldsymbol{D}_i + \boldsymbol{\gamma}_E \boldsymbol{E}_{it} + u_{it}, \tag{3.4}$$

where u_{it} is a mean-zero stochastic error term. For convenience, I henceforth refer to such a linear-in-levels specification as a "linear specification."

The coefficients in such a linear specification do not have the convenient elasticity interpretation as those in the log-log model, which likely explains why fewer studies use a linear specification. In order to obtain an elasticity estimate from (3.4), we must choose a value of VMT and C at which to evaluate the elasticity. The most common assumption is at the means: $\overline{VMT}$ and $\overline{C}$. Since β_C is equal to $\frac{\partial VMT}{\partial C}$, we can calculate the estimated elasticity at the mean as $\hat{\beta}_{VMT,C} = \hat{\beta}_C \frac{\overline{C}}{\overline{VMT}}$. Note that different prices of gasoline imply a different elasticity. Gately (1992) suggests that this feature makes the linear specification preferable because we might expect the elasticity to vary with the price of gasoline and the amount consumer drive.

The linear specification may have some additional advantages. Most obviously, a linear specification would be preferable to a log-log specification if the true relationship is linear. Thus, one could conceivably try both models and see which model fits the data better.[3] Preferably, we would have an economic theory that would suggest a particular specification. Less obviously, a linear specification may also be preferred because of Jensen's Inequality and the time horizon of the data. I discuss this in more detail below.

Time horizon

Up to this point, we have not specified what a time period t is exactly. In some datasets, the data are monthly, so an observation includes the monthly average gasoline price and monthly amount driven. In other datasets, the data are yearly, so we have the averages over the year. This averaging is not an issue in a linear specification, but may be an issue in a log specification.

With some simple algebra, it is easy to see why taking the average is a problem in the log specification. The issue is simply that if consumers actually make decisions over a shorter time frame than the time frame available in the data, then it is incorrect to use the variable in the data (averaged over the shorter time frames) in a log specification. The intuition comes directly from Jensen's inequality.

Without loss of generality, we can abstract from all other covariates and drop the i subscript. Suppose that consumers make short-run decisions about how much to drive on a weekly basis and that the relationship between the cost of driving and the

[3] The R-squared, Akaike Information Criterion, or Bayesian Information Criterion are all metrics of model fit that could be calculated.

amount driven is log-log. Then we can write the true model of consumer decision-making as

$$\log(VMT_t) = \beta_0 + \beta_1 \log(C_t) + \varepsilon_t, \tag{3.5}$$

where t refers to a week. Suppose we have monthly, or even yearly, data. The average VMT over a year that we observe would be $VMT_y = \frac{1}{|\mathcal{T}_y|}\sum_{t\in\mathcal{T}_y} VMT_t$, where $\mathcal{T}_y$ is the set of all weeks in year y and $|\mathcal{T}_y|$ is the cardinality of $\mathcal{T}_y$. C_y is defined analogously from C_t.

Thus, if we wish to use yearly data with (3.5), we can take the average of both sides to get

$$\frac{1}{|\mathcal{T}_y|}\sum_{t\in\mathcal{T}_y} \log(VMT_t) = \beta_0 + \beta_1 \frac{1}{|\mathcal{T}_y|}\sum_{t\in\mathcal{T}_y} \log(C_t) + \varepsilon_t.$$

By Jensen's inequality we know that unless VMT_t is a linear function

$$\frac{1}{|\mathcal{T}_y|}\sum_{t\in\mathcal{T}_y} \log(VMT_t) \neq \log\left(\frac{1}{|\mathcal{T}_y|}\sum_{t\in\mathcal{T}_y} VMT_t\right).$$

The right-hand side of the equation is what is available in the data, while the left-hand side is what the model suggests we should use. The same issue holds for C_t. Thus, if we simply plug in VMT_y and C_y and estimate the model

$$\log(VMT_y) = \beta_0' + \beta_1' \log(C_y) + \varepsilon_t',$$

then the estimated coefficients, $\hat{\beta}_0'$ and $\hat{\beta}_1'$, will inherently be biased and inconsistent estimates of β_0 and β_1. Mannering and Winston (1985) was perhaps the only paper to recognize this point, although other authors who chose to estimate linear models may have had this issue in mind.[4]

How important might this issue be? If we have monthly data, as in Hughes,

[4] Incidentally, Mannering and Winston (1985) take this point very seriously and actually set up a model where the *cumulative* VMT in a time period is a function of the cumulative cost of operating it in that time period and vehicle/household characteristics. Note that this specification is completely equivalent to (3.4), only with VMT_{it} and C_{it} rescaled.

Knittel, and Sperling (2008), and we assume that consumers make short-run decisions at the weekly or monthly level, this may not be much of a concern for estimating a short-run elasticity. If VMT_t and C_t are a linear functions, then Jensen's inequality does not apply and we certainly do not have an issue. However, if the data contain averages over longer time frames, such as two years or six years as in my dataset, we may have a concern. The short-run consumer decision-making time horizon is most certainly shorter than two years, so we certainly have an issue if we wish to interpret the coefficient as a short-run elasticity. If we consider the consumer decision-making time horizon to be longer, such as a year or two years, so that we see t as equal to a year or two years, then we again may not have as much of a problem with data over two years in my dataset. We may still have a problem with data over six years. Given this, it seems prudent in my analysis to focus on the results of linear specifications.

This still leaves the question of how to interpret the time horizon of estimates based on the six year average data. As discussed in Chapter 1, most authors either use a lagged dependent variable to find estimates of long-run and short-run responsiveness, do not describe whether the estimate is long-run or short-run, or (more recently) use the time horizon of the variation to determine the elasticity. For example, when monthly fixed effects are included, as in Hughes, Knittel, and Sperling (2008), only within-month variation is used, so the estimates can be interpreted as a monthly (i.e., short-run) elasticity. Following this logic, the estimated elasticity using the two year average data can best be interpreted as a two-year elasticity. Similarly, the estimated elasticity using the six year average data is also probably best interpreted as based on the time horizon of the variation in the price variable in the data. As was shown in Chapter 2, there was very little variation in the price of gasoline in the first few years. In fact, the primary variation was almost entirely over a two year time horizon. Thus, the most appropriate interpretation of the estimated elasticity may again be a roughly two-year elasticity.

Cost per mile or price of gasoline?

As was discussed in Chapter 1, authors aiming to estimate the same thing – consumer price responsiveness in driving – differ in the key regressor used. In (3.1), the utilization of a vehicle is modeled as a function of the utilization cost C_{it}, which can be thought of in units of dollars per mile driven. This is in many respects the most natural "price" of driving to consider. However, many other studies estimate the responsiveness to the retail price of gasoline.

Of course, the cost per mile of driving and the retail price of gasoline are directly related, for the fuel cost of driving can be written as P_{it}/MPG_i, where P_{it} is the retail price of gasoline the owner of vehicle i faces (in dollars per gallon) and MPG_i is the fuel economy of the vehicle (in miles per gallon).[5] If we assume that there are no interactions between the fuel cost of driving and the other variable costs of driving (e.g., time cost, maintenance, insurance), then we can let $C_{it} = P_{it}/MPG_i$ without loss. Nearly all of the literature follows this assumption and does not include variables for the other costs of driving. As long as consumers respond to a dollar per mile increase in the cost of driving the same regardless of its source, this assumption should hold. This is what we would expect for rational consumers.

To see the difference between estimating a model with $C_{it} = P_{it}/MPG_i$ and a model with P_{it}, consider the simple linear specification in (3.4). In this specification, since MPG_i enters C_{it}, we can assume that it does not enter the vector of vehicle characteristics $\boldsymbol{V}_i$. One implication of this specification is that if the cost per mile of driving declines – regardless of the source whether it is due to changing fuel economy or gasoline prices – the consumer response in terms of the amount driven will remain the same.

We could imagine relaxing this restriction by estimating a coefficient for both the price of gasoline and the fuel economy of the vehicle, as in the following model:

$$VMT_{it} = \gamma_0' + \gamma_P' P_{it} + \gamma_{MPG}' MPG_i + \boldsymbol{\gamma}_V' \boldsymbol{V}_i + \boldsymbol{\gamma}_D' \boldsymbol{D}_i + \boldsymbol{\gamma}_E' \boldsymbol{E}_{it} + u_{it}', \tag{3.6}$$

where u_{it}' is again a mean-zero stochastic error term.

In order for γ_P' and γ_{MPG}' to be separately identified in (3.6), we must have sufficient variation in both retail gasoline prices and fuel economy. Herein lies a considerable weakness of this approach. CAFE standards have been binding or close to

[5] Technically, MPG_i should be MPG_{it}, for fuel economy degrades very slightly over time. As long as we believe that the time period is sufficiently short, then MPG_i is a reasonable assumption.

binding for most of the past two decades, effectively limiting the amount of variation in the nationwide fuel economy. Thus, any analysis based entirely on times series variation (e.g., Greene (2011)) is extremely unlikely to have sufficient variation to identify these coefficients separately. Studies based on cross-sectional variation in fuel economy may hold somewhat more promise if other possibly correlated unobservables can be controlled for – although ideally we would like to have experimental or quasi-experimental variation.

Assuming they are well-identified in the data, both (3.4) and (3.6) can be used to obtain elasticity estimates. We can calculate an elasticity from (3.6) just as described for (3.4). Since $\gamma_P = \frac{\partial VMT}{\partial P}$, the estimated elasticity of driving with respect to the retail price of gasoline at the mean of VMT and P is simply $\hat{\gamma}_{VMT,P} = \hat{\gamma}_P \frac{\overline{P}}{\overline{VMT}}$.

Recall that in Section 1.2.5 we saw that in a static setting the VMT elasticity with respect to the gasoline price should be identical to the VMT elasticity with respect to the cost of utilization. So we would expect to see $\hat{\gamma}_{VMT,C} = \hat{\gamma}_{VMT,P}$. Similarly, we would expect consumers to respond to changes in the cost per mile of driving due to changing fuel economy in exactly the same way as changes in the cost per mile of driving due to changes in the prices of gasoline. It follows then that we would expect to see $\hat{\gamma}_{VMT,C} = \hat{\gamma}_{VMT,P} = \hat{\gamma}_{VMT,MPG}$. Moreover, in the log-log formulation, the relationship between the specification with only $\log(C_{it})$ and with both $\log(P_{it})$ and $\log(MPG_i)$ is even more clear, for $\beta_C \log(C_{it}) = \beta_C \log(P_{it}) - \beta_C \log(MPG_i)$. Thus, in the log-log specification, including only $\log(C_{it})$ effectively imposes the restriction that the elasticity of driving with respect to the price of gasoline is the negative of the elasticity of driving with respect to fuel economy.

When might such an equality not hold? The only idea I am aware of in the previous literature is that policies that raise fuel economy tend to also increase the upfront capital cost of a new vehicle, so that the fuel cost of driving is a smaller percentage of the total cost of driving, and thus treated as less important by consumers (Greene 2011). In Section 1.2.5, I suggest other possibilities. Suppose the response to changes in the cost per mile of driving is asymmetric, so that that decreases in the cost per mile of driving (e.g., from increased fuel economy) lead to little responsiveness, but increases in the cost per mile of driving (e.g., from increased gasoline prices) lead to much more responsiveness. Along these lines, suppose that a small subtle change in the cost per mile of driving (e.g., from increased fuel economy) is not noticed by most consumers and leads to little responsiveness, while larger changes (e.g., from changing gasoline prices) may lead to much more responsiveness relative to the magnitude of the change in the cost per mile of driving. Alternatively, when policy mandates change fuel economy, some vehicle options (i.e., combinations of price and attributes) are no longer as desirable to some consumers, so consumers may choose to purchase vehicles that they would not like to drive as much. Of course, there also remains the final possibility that there is not enough variation in fuel economy in the data to identify the coefficient β_{MPG}. In my empirical results, I estimate both (3.4) and (3.6) to gain a sense of whether my data can speak to this issue.

Aggregation and simultaneity

A final important detail about using regression evidence to quantify the price responsiveness in driving is in the degree of aggregation in the data. Many of the studies in the literature on estimating the responsiveness to gasoline price changes rely upon highly aggregate data, usually at the national-level or state-level. Simply looking at the tables of estimates in Chapter 1 does not appear to yield a clear pattern about whether aggregate-level studies present a bias.

Yet we might be concerned about studies using market-level data for the standard simultaneity endogeneity concern. With market-level data, it is common to consider the gasoline price P_{it} as endogenous in (3.6), for the market price is determined by both supply and demand. This simultaneity concern can also be thought of as unobserved supply-side factors that are correlated with the price of gasoline. Ideally, we would need to instrument for the price with a supply-shifter – a variable that influences the supply of gasoline, but does not influence the demand (except through price). The same concern will also hold for the price per mile of driving C_{it} in (3.6), for C_{it} is a function of the price of gasoline. Only a few of the studies in the literature attempt to address this concern. Moreover, those that do often use instrumental variables that are hard to believe are actually supply-shifters.

Do studies that use individual-level data face the same issue? Quite possibly.

On one hand, some studies have either implicitly or explicitly argued that individual consumers are price-takers when it comes to the price of gasoline, and do not individually influence the price of gasoline. In a sense, this is assuming that there are no supply-side shocks that influence the price of gasoline and driving. This would suggest that the simultaneity concern is much less of an issue. On the other hand, it is possible that there are local shocks that influence both how much consumers drive and the local gasoline price. If such shocks are truly local, then controlling for local economic conditions by including $\boldsymbol{E}_{it}$ may be sufficient to address the concern. Of course, it may also be possible that there are other local shocks that influence both gasoline prices and how much consumers drive. I find it difficult to think of a story for such shocks that does not relate to local economic conditions. The data also suggest that there may be few local shocks affecting gasoline prices: Figure 2.2 in Chapter 2 indicates that the retail gasoline prices in most counties in California follow the California average retail gasoline price quite closely.

Some of the studies using individual-level data discuss this issue. For example, West and Williams (2007) use the Consumer Expenditure Survey data to estimate the price elasticity of gasoline demand (among other things). West and Williams (2007) discuss the possibility of endogeneity from local economic booms, but find no suitable options for instruments. Other studies come to similar conclusions, or instrument with what could be argued to be dubious instruments.

Besides the simultaneity concern, might we expect there to be another aggregation bias? If we are estimating a log-log specification, then a similar Jensen's inequality concern to the one discussed above may occur. However, in a linear specification, we should not expect other issues. Finally, while it is not an econometric issue, it is true that only individual-level data allows for a full examination of the heterogeneity in responsiveness, which is quite interesting for local air pollution and distributional consequences.

3.1.1 Personal Vehicles in First Six Years

I begin my empirical exploration with my dataset of new personal vehicles during the first six years of use. I present evidence using a linear specification and a log-log specification. Then I present some initial evidence that the responsiveness may be different when gasoline prices are high than when they are low.

Linear specification

My estimations with a linear-in-levels specification are all based on (3.4). I repeat this equation here for convenience:

$$VMT_{it} = \gamma_0 + \gamma_C C_{it} + \boldsymbol{\gamma}_V \boldsymbol{V}_i + \boldsymbol{\gamma}_D \boldsymbol{D}_i + \boldsymbol{\gamma}_E \boldsymbol{E}_{it} + u_{it},$$

In the estimation u_{it} may be split into an error components structure, e.g., $u_{it} = \mu_t + \xi_m + \nu_{it}$, where μ_t are time fixed effects, ξ_m are model fixed effects, and ν_{it} is a mean-zero stochastic error term. Time fixed effects may be month-of-the-year fixed effects or year fixed effects. I can examine specifications with more than one of these fixed effects at once by including one by way of the within transformation and others by including indicator variables. I can also examine other fixed effects, or combinations of fixed effects as well, such as county fixed effects, time of the smog check fixed effects, or vehicle class fixed effects.

In the estimation, $\boldsymbol{V}_i$ includes such characteristics as engine liters, engine cylinders, an indicator for having a turbocharger, an indicator for having an automatic transmission, the gross vehicle weight rating, an indicator for whether the vehicle has four wheel drive, the safety rating of the vehicle, and an indicator for whether the vehicle is imported. In some specifications, fuel economy is included in $\boldsymbol{V}_i$, while in others it is subsumed in C_{it}. When fuel economy is included separately in $\boldsymbol{V}_i$, P_{it} is used in place of C_{it}.

$\boldsymbol{D}_i$ includes such individual and zip-code demographics as whether the driver leased the vehicle, the log of the zip code population, the zip code population growth from 2000 to 2007, the log of the number of businesses in the zip code, the log of the zip code household income, the percentage of the zip code population over 65 or under 18, the percentage of the population that is different races, and the county-level commute times. $\boldsymbol{E}_i$ includes economic conditions such as the county-level unemployment rate, the consumer confidence index, and county-level housing prices.

There may be several identification concerns in this analysis. Both gasoline prices and VMT display seasonality, with both higher gasoline prices and more driving in the summer. One way to address this concern is to include a variable indicating what fraction of the time between registration and the test occurs over the summer months.[6] Second, if the patterns of where consumers moved over the time between registration and the smog check are somehow correlated gasoline prices, then the coefficients would be biased and inconsistent. Since I observe whether the registration is in the same county as the test and subsequent registration, I can perform the same estimation only using the average of the gasoline prices in the two locations or even restrict the sample to only those who did not move.[7]

Next, selection may confound the estimates in four different ways. First, there is the selection issue discussed by Dubin and McFadden (1984), in which consumers who plan to drive more purchase different vehicles. This selection issue is the motivation of the structural analysis in Chapter 4 and is not addressed here.

A related issue is that consumers who anticipate high gasoline prices may choose to purchase a more efficient vehicle, thus lowering the cost per mile of driving and perhaps reducing the subsequent responsiveness to gasoline price changes. This could be a significant concern in using time series variation to identify the relationship between gasoline prices and driving. However, my study time frame presents a unique circumstance where gasoline prices were low and relatively stable during 2001 through mid-2004, when the consumers for whom I observe VMT purchased the new vehicles. This is several years before the gasoline price spike of 2007-2008, so under the reasonable assumption of imperfect foresight of future gasoline prices, this selection issue is not likely to be an important concern.

A third selection issue may confound the estimation if consumers of different unobserved driving preferences selected into an early or late smog check. Figure 2.6

[6]The summer months are defined here as June, July, and August.

[7]I find that restricting the sample to those who do not move leads to another type of sample selection, for the population of those who do not move is inherently different than the full population.

in Chapter 2 showed that roughly 35% of the sample has either an early or late smog check, either because the title was transferred, the model year of the vehicle allowed a test earlier or later, or the consumer was negligent in getting the registration renewed. This may be an issue if vehicles that had an early smog check because the vehicle title was transferred were driven more – and the interval of the smog check is correlated with gasoline prices. Similarly, vehicles may have had a late smog check because the vehicle was unused for a period of time. Those with early and late smog checks would face a different average price of gasoline, and possibly leading to a spurious correlation between the gasoline price and driving. To address the possibility of this selection issue, I can either control for the months until the smog test or focus entirely on consumers who had a smog check within a few months of the normal six years.

A fourth selection issue may be that economic conditions could be influencing the decrease in driving. Fortunately, in the time frame of my study, the highest gasoline prices occurred before the economic downturn really hit, so that gasoline prices and economic conditions are not as highly correlated as in most previous gasoline price shocks.[8] By conditioning on the average county-level unemployment rate, CCI, and housing prices, I can control for changes in economic factors that could influence driving.

Finally, we may be concerned that different types of consumers purchase vehicles at different times of the year. Copeland, Dunn, and Hall (2011) show that dealers drop the price of a particular model year vehicle over the year until the introduction of the next model year in early summer. So, it is plausible that different types of consumers may buy new cars at different times of year if consumers time new vehicle purchases. A pattern of this sort is not obvious from the summary statistics, yet it is possible. Including month-of-the-year fixed effects should help address this selection issue. Along these lines, some of the consumers may have more of the summer, when there are higher gasoline prices and different driving habits, within the interval of time between the two smog checks. I can include a control for the percentage of time between the two tests in the summer.

[8]In my dataset, the Pearson correlation coefficient between the gasoline price and unemployment is -0.14 and between the gasoline price and the CCI is -0.12.

The results from estimating (3.6) by ordinary least squares and fixed effects estimation are given in Table 3.1. All of the estimations are run on the entire 4.65 million observation dataset. The columns increasingly add additional controls. Column (1) begins with a simple regression of VMT on the average retail gasoline price without any controls to give a sense of the correlation between the two variables. The coefficient on the average gasoline price indicates that without any controls, a one dollar increase in the average gasoline price is associated with a decrease in driving by -118 miles per month (a VMT elasticity with respect to the price of gasoline of roughly -0.29 at the means). As discussed above, the responsiveness in these results is probably best considered a medium-run, or two-year responsiveness since the primary identifying variation is is time series variation over roughly two years. Column (2) adds controls for the months between the registration and test (M) to address the possibility of a selection by consumers of different driving needs or preferences into these different intervals. The controls are all relative to "normal" smog checks right around six years (53 percent of the observations). The resulting coefficient on the gasoline price suggests that these additional controls brings down the responsiveness slightly.

Column (3) adds the control variables for economic conditions. These make a much larger difference to the estimated responsiveness to gasoline prices, bringing the coefficient on the gasoline price down to -63. This suggests that if gasoline prices are increased by one dollar, driving would decrease by 63 miles per month. This corresponds to an elasticity of VMT with respect to the price of gasoline of -0.15 at the means. The signs of the coefficients on the controls generally make sense as well: higher unemployment corresponds to less driving and higher consumer confidence corresponds to more driving. Higher home prices are associated with less driving, a result that perhaps can be attributed to the fact that higher home prices are in regions with more congestion, and thus, less driving. Higher home prices also may be correlated with more vehicles per household, so the per-vehicle driving may be less. Column (4) adds variables to deal with a possible endogeneity relating to the time of the year of the purchase and test. Specifically, I add month-of-the-year indicators and the variable for the percentage of the time between registration and the smog

Table 3.1: Intensive Margin Regressions: Linear Model

Dependent variable: vehicle-miles-traveled per month (mean = 1,089 miles per month)

	(1) base	(2) months-to-test	(3) econ cond	(4) mon-of-yr summer	(5) demog	(6) vehicle chars	(7) model FE
avg gasoline price	-117.7***	-110.7***	-62.8***	-62.8***	-78.8***	-69.6***	-69.3***
	(1.1)	(1.1)	(1.5)	(1.5)	(1.5)	(1.5)	(8.0)
M < 58		112.2***	109.5***	109.6***	106.7***	115.6***	114.7***
		(0.7)	(0.7)	(0.8)	(0.7)	(0.7)	(4.0)
57 > M > 63		68.8***	67.0***	69.1***	67.9***	78.6***	78.4***
		(1.0)	(1.0)	(1.0)	(0.9)	(0.9)	(2.8)
62 > M > 70		54.8***	52.7***	54.2***	54.3***	60.1***	60.1***
		(0.8)	(0.8)	(0.8)	(0.8)	(0.8)	(3.4)
73 > M > 82		67.7***	65.5***	65.2***	66.8***	66.8***	67.0***
		(0.7)	(0.7)	(0.7)	(0.7)	(0.7)	(3.2)
81 > M > 86		-36.2***	-36.2***	-32.4***	-27.7***	-27.6***	-25.3***
		(0.7)	(0.7)	(0.7)	(0.7)	(0.7)	(2.4)
M > 86		42.0***	38.3***	40.7***	44.7***	44.8***	44.1***
		(1.7)	(1.7)	(1.8)	(1.7)	(1.7)	(3.7)
avg unempl rate			-3.7***	-3.7***	-3.2***	-2.4***	-2.7***
			(0.2)	(0.2)	(0.2)	(0.2)	(0.5)
avg CCI			1.3***	1.4***	1.2***	1.2***	1.3***
			(0.1)	(0.1)	(0.1)	(0.1)	(0.2)
avg housing prices			-0.3***	-0.3***	-0.1***	-0.1***	-0.1***
			(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
% summer months				126.6***	128.3***	107.7***	113.4**
				(23.6)	(23.3)	(23.0)	(36.9)
zip density					-7.2***	-6.6***	-6.4***
					(0.1)	(0.0)	(0.2)
zip bus/capita					-40.6***	-29.5***	-26.7***
					(11.2)	(8.0)	(7.4)
log(zip population)					-15.6***	-16.8***	-17.6***
					(0.5)	(0.4)	(0.8)
zip pop growth					3.6***	3.4***	3.4***
					(0.1)	(0.1)	(0.2)
log(zip income)					-67.5***	-39.4***	-32.7***
					(1.0)	(1.0)	(3.4)
commute time					5.4***	5.4***	5.4***
					(0.1)	(0.1)	(0.2)
liters						-5.8***	-26.2***
						(0.6)	(5.3)
cylinders						-11.1***	-7.6*
						(0.4)	(3.6)
turbo						-29.3***	-23.5**
						(1.2)	(7.7)
auto transmission						-31.7***	-40.1***
						(0.8)	(8.1)
gross veh weight						-6.5***	-5.6**
						(0.3)	(1.9)
all-wheel drive						2.5***	21.7***
						(0.7)	(3.5)
safety rating						-15.0***	3.6
						(0.5)	(12.3)
import						16.7***	
						(0.6)	
fuel economy						2.6***	-1.3
						(0.1)	(2.3)
constant	1,405.6***	1,357.8***	1,272.2***	1,214.7***	1,871.2***	1,558.8***	1,613.0***
	(3.0)	(3.0)	(9.8)	(11.4)	(17.6)	(17.5)	(68.2)
month-of-year FE	N	N	N	Y	Y	Y	Y
lease, race & age	N	N	N	N	Y	Y	Y
veh body & class	N	N	N	N	N	Y	Y
model FE	N	N	N	N	N	N	Y
R-squared	0.002	0.012	0.019	0.020	0.039	0.065	0.035
Observations	4.65m	4.65m	4.65m	4.65m	4.65m	4.65m	4.65m

Heteroskedasticity-robust standard errors in parentheses, clustered on model in col (7)
*** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level
“M” refers to the months between registration and the first smog test

test occurs over the summer. Adding these makes no noticeable difference to the coefficient on the gasoline price.

Column (5) adds a variety of zip code demographic and location-specific variables to address the possibility of endogeneity due to a correlation between gasoline prices and median household income, population density, businesses per capita, or average commute times. I also include an indicator for whether the vehicle is leased in this specification. The addition of these controls increases the responsiveness to imply an elasticity of about -0.19 at the means. The signs of the coefficients again make sense: areas with higher density, more businesses, and greater wealth tend to have less driving per vehicle, while faster-growing areas and areas with longer commute times have more driving per vehicle.

Column (6) adds a variety of vehicle characteristics to account for the possibility that drivers of different types of vehicles not only choose to drive differently but may also face systematically different gasoline prices. Adding these additional controls brings down the responsiveness slightly to suggest that if gasoline prices increase by one dollar, driving will decrease by 70 miles per month. The corresponding elasticity is -0.17 at the means. The signs of several of the vehicle characteristics coefficients are interesting, although for the most part I have no priors on what they should be. The only exception is the coefficient on the fuel economy of the vehicle, which implies that higher fuel economy vehicles tend to be driven more. This makes sense, as higher fuel economy vehicles cost less per mile to drive. The responsiveness is quite small though – a one mile per gallon increase in the fuel economy of a vehicle is associated with 2.6 more miles driven per month. The corresponding elasticity of driving with respect to fuel economy is roughly 0.05 at the means.

The coefficient on fuel economy is identified partly by the time series variation in the fuel economy and partly by the cross-sectional variation in fuel economy. The time series variation in fuel economy is somewhat limited in 2001-2004, as indicated by Figure 2.10 in Chapter 2. The use of the cross-sectional variation in fuel economy may be somewhat limited as well due to the relatively limited cross-sectional variation in gasoline prices. Moreover, there are nationwide CAFE standards that were binding nationwide on many of the manufacturers over this time period. Thus, the coefficient on fuel economy must be interpreted recognizing that there may have been interactions with the California market in the manufacturer decision process for meeting the nationwide standard. To the extent that there is enough variation in fuel economy and that the coefficient is well-identified (i.e., no major interactions with nationwide CAFE standards), this coefficient on fuel economy is perhaps a better estimate of the "direct rebound effect" than most estimates in the literature. Later in this section I will examine a specification that includes the fuel cost per mile of driving rather than the gasoline price and the fuel economy, and in Chapter 4, I estimate what I consider to be an even better estimate of the direct rebound effect.

Column (7) adds vehicle model fixed effects to account for any model-specific unobserved heterogeneity that could be correlated with gasoline prices. Not surprisingly, these fixed effects render many of the vehicle characteristics insignificant, including the fuel economy. Including the model fixed effects appears to make nearly no difference to the coefficient on the price of gasoline.

Recall that one of the possible selection issues in these specifications is that there is unobserved heterogeneity relating to the time between tests. All but the first specification estimated above include controls for the number of months between registration and the smog check, and these controls did not seem to make a substantial difference. However, we may still be concerned that these controls do not sufficiently address the issue. For a cleaner, but perhaps less representative estimation, we can examine only those vehicles that had a smog check within two months of the normal six years. Table 3.2 presents the same estimation results as in Table 3.1, only restricted to this "normal" subsample and no longer including months-to-test controls. The idea for this restricted subsample is that the variation in gasoline prices for drivers who receive the notice in the mail right around six months is very likely to be random, for the exact date when the drivers get the notice in the mail and then get a chance to take the vehicle into the smog check is likely to be random.[9]

The results in Table 3.2 are similar to the results in Table 3.2, with the primary difference being a slightly larger responsiveness. Across all specifications, the coefficient on the average gasoline price is in the range of -95 to -120 miles per month.

[9] I thank Liran Einav for this idea.

Table 3.2: Intensive Margin Regressions: Six Year Subsample

Dependent variable: vehicle-miles-traveled per month (mean = 1,060 miles per month)

	(1) base	(2) econ cond	(3) mon-of-yr summer	(4) demog	(5) vehicle chars	(6) model FE
avg gasoline price	-116.7***	-121.9***	-118.4***	-102.9***	-99.7***	-94.9***
	(1.6)	(3.7)	(3.7)	(3.7)	(3.6)	(15.2)
avg unempl rate		-4.3***	-4.3***	-3.2***	-2.4***	-2.7***
		(0.3)	(0.3)	(0.3)	(0.3)	(0.6)
avg CCI		-1.9***	-1.6***	-0.5**	-0.7***	-0.3
		(0.2)	(0.2)	(0.2)	(0.2)	(0.5)
avg housing prices		-0.3***	-0.3***	-0.1***	-0.1***	-0.1***
		(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
% summer months			1,316.9***	1266.6***	1225.0***	1144.3***
			(59.9)	(59.3)	(58.5)	(170.9)
zip density				-7.1***	-6.7***	-6.6***
				(0.1)	(0.1)	(0.2)
zip bus/capita				-68.2***	-51.3***	-48.0***
				(5.7)	(4.8)	(5.6)
log(zip population)				-20.0***	-20.5***	-21.1***
				(0.6)	(0.6)	(0.9)
zip pop growth rate				3.8***	3.7***	3.6***
				(0.1)	(0.1)	(0.2)
log(zip income)				-61.0***	-35.6***	-31.3***
				(1.4)	(1.4)	(3.8)
commute time(min)				6.0***	5.9***	5.9***
				(0.1)	(0.1)	(0.3)
liters					-3.0***	-24.3***
					(0.8)	(6.3)
cylinders					-12.8***	-9.9*
					(0.5)	(4.3)
turbo					-19.4***	-19.0**
					(1.9)	(7.3)
auto transmission					-44.0***	-52.7***
					(1.1)	(8.6)
gross veh weight					-5.8***	-5.1**
					(0.4)	(1.7)
all-wheel drive					10.8***	25.6***
					(1.0)	(3.6)
safety rating					-11.7***	-0.8
					(0.8)	(13.2)
import					30.0***	
					(0.9)	
fuel economy					2.8***	-0.8
					(0.2)	(2.6)
constant	2,028.3***	1,723.2***	1,332.8***	1,770.8***	1,500.1***	1,599.8***
	(24.9)	(25.0)	(29.4)	(34.6)	(35.1)	(129.8)
month-of-year FE	N	N	Y	Y	Y	Y
lease, race & age	N	N	N	Y	Y	Y
veh body & class	N	N	N	N	Y	Y
model FE	N	N	N	N	N	Y
R-squared	0.002	0.010	0.010	0.030	0.056	0.028
Observations	2.19m	2.19m	2.19m	2.19m	2.19m	2.19m

Heteroskedasticity-robust standard errors in parentheses, clustered on model in col (7)
*** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level

Columns (5) and (6) are the preferred specifications and corresponds to an elasticity of VMT with respect to the price of gasoline in the range of -0.25 at the means. The difference between the results in Table 3.2 and 3.2 may stem partly from the slightly better identification, but also from the heterogeneity in responsiveness – it is possible that the population of drivers who have a smog check after six years is slightly more responsive than the larger population of drivers in my new personal vehicles dataset.

Interestingly, the coefficient on fuel economy in Column (5) is nearly the same as the coefficient on fuel economy in the same specification (Column (6)) in Table 3.2. This provides some reassurance that the estimate of this coefficient is robustly estimated.

Log-log specification

Since much of the literature runs a specification similar to the log-log specification in (3.3), I find it useful to follow suit for the sake of comparison. As in the linear specification, I will focus on specifications where the gasoline price and fuel economy enter separately. These specifications can be written as follows

$$\log(VMT_{it}) = \beta_0'' + \beta_P'' \log(P_{it}) + \beta_{MPG}'' \log(MPG_i) + \boldsymbol{\beta}_V'' \boldsymbol{V}_i + \boldsymbol{\beta}_D'' \boldsymbol{D}_i + \boldsymbol{\beta}_E'' \boldsymbol{E}_{it} + \varepsilon_{it}'',$$

where ε_{it}'' is again a mean-zero stochastic error term that may be divided into error components including fixed effects. I also include a specification that restricts $\beta_P'' = -\beta_{MPG}''$ by including the cost per mile rather than the gasoline price and the fuel economy.

The results are given in Table 3.3. Each column in Table 3.3 refers to a similar specification as the equivalent column in Table 3.1, only with the variables for the gasoline price and fuel economy subsumed with a single variable for the cost per mile. The coefficients on the log of the gasoline price and the log of the fuel economy in Table 3.3 can be interpreted as driving elasticities with respect to the price of gasoline and fuel economy respectively. The values of the driving elasticity with respect to gasoline price are in fact quite similar to those calculated at the means from the estimated coefficients in the linear specification.

The results in Columns (6) and (7) of Table 3.3 suggest an elasticity of -0.19, which I take as my preferred estimate of the driving responsiveness to gasoline prices for the log-log specification. The coefficient on the log of the fuel economy in Columns (6) and (7) is *negative*, which is less reassuring. It is very small and in many respects economically insignificant. In addition, I also run a specification where fuel economy enters linearly and find a small, but positive coefficient (not shown in the Table). The difference in the sign of the coefficient may relate to a misspecification of the relationship between fuel economy and VMT in one of the two specifications. Given that the results are more sensible in the linear specification, and the Jensen's inequality concern does not apply to the linear specification, I view the linear estimation results as my preferred results.

Gallons per mile regressions

To explore the implications of the non-linearity of fuel economy in miles per gallon (GPM), I can examine specifications where fuel consumption, in gallons per mile, enters the model instead of the miles per gallon. Table 3.4 presents linear specifications equivalent to Columns (6) and (7) in Table 3.1 and log-log specifications equivalent to Columns (6) and (7) in Table 3.3. I find entirely analogous results to those in the previous tables. Just as was the case for the coefficient on fuel economy, the coefficient on the gallons per mile fuel consumption is statistically insignificant when model fixed effects are included. The coefficient on GPM in Column (1) indicates that an increase in the fuel consumption of the vehicle (i.e., lower fuel economy) by one gallon per mile decreases driving by 333 miles per month. The mean of GPM is 0.05 gallons per mile. Thus, the elasticity of driving with respect to GPM is about -0.02 at the means.

The result in Column (3) shows that in the log-log specification, we again have a coefficient with the unexpected sign. The coefficient suggests that vehicles with lower fuel economy are driven more, the opposite of the relationship we see in the linear specifications (with an elasticity of 0.01). In case there is a non-linearity in the relationship that the logarithm is capturing and the linear relationship is not

Table 3.3: Intensive Margin Regressions: Log-log Model

Dependent variable: log vehicle-miles-traveled per month

	(1) base	(2) months-to-test	(3) econ cond	(4) mon-of-yr summer	(5) demog	(6) vehicle chars	(7) model FE
log(gasoline price)	-0.33***	-0.30***	-0.19***	-0.19***	-0.21***	-0.19***	-0.19***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)
M < 58		0.11***	0.11***	0.11***	0.10***	0.11***	0.11***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
57 > M > 63		0.07***	0.07***	0.07***	0.07***	0.08***	0.08***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
62 > M > 70		0.05***	0.05***	0.05***	0.05***	0.06***	0.06***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
73 > M > 82		0.07***	0.07***	0.07***	0.07***	0.07***	0.07***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
81 > M > 86		-0.03***	-0.03***	-0.03***	-0.02***	-0.02***	-0.02***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
M > 86		0.05***	0.05***	0.05***	0.05***	0.06***	0.05***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
avg unempl rate			-0.01***	-0.01***	-0.01***	-0.00***	-0.01***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
avg CCI			0.00***	0.00***	0.00***	0.00***	0.00***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
avg housing prices			-0.00***	-0.00***	-0.00***	-0.00***	-0.00***
			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
% summer months				0.13***	0.13***	0.11***	0.12**
				(0.02)	(0.02)	(0.02)	(0.04)
zip density					-0.01***	-0.01***	-0.01***
					(0.00)	(0.00)	(0.00)
zip bus/capita					-0.04***	-0.03***	-0.03***
					(0.01)	(0.01)	(0.01)
log(zip population)					-0.01***	-0.01***	-0.02***
					(0.00)	(0.00)	(0.00)
zip pop growth rate					0.00***	0.00***	0.00***
					(0.00)	(0.00)	(0.00)
log(zip income)					-0.06***	-0.03***	-0.03***
					(0.00)	(0.00)	(0.00)
commute time					0.00***	0.00***	0.00***
					(0.00)	(0.00)	(0.00)
liters						-0.02***	-0.03***
						(0.00)	(0.01)
cylinders						-0.01***	-0.01*
						(0.00)	(0.00)
turbo						-0.02***	-0.03*
						(0.00)	(0.01)
auto transmission						-0.03***	-0.04***
						(0.00)	(0.01)
gross veh weight						-0.01***	-0.01***
						(0.00)	(0.00)
all-wheel drive						0.00	0.02***
						(0.00)	(0.00)
safety rating						-0.01***	0.00
						(0.00)	(0.01)
import						0.03***	
						(0.00)	
log(fuel economy)						-0.01***	-0.06
						(0.00)	(0.04)
constant	7.21***	7.16***	7.10***	7.05***	7.68***	7.50***	7.56***
	(0.00)	(0.00)	(0.01)	(0.01)	(0.02)	(0.02)	(0.14)
month-of-year FE	N	N	N	Y	Y	Y	Y
lease, race & age	N	N	N	N	Y	Y	Y
veh body & class	N	N	N	N	N	Y	Y
model FE	N	N	N	N	N	N	Y
R-squared	0.002	0.010	0.015	0.015	0.031	0.064	0.030
Observations	4.65m	4.65m	4.65m	4.65m	4.65m	4.65m	4.65m

Heteroskedasticity-robust standard errors in parentheses, clustered on model in col (7)
*** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level
"M" refers to the months between registration and the first smog test

Table 3.4: Intensive Margin Regressions: Gallons per Mile

	(1) Linear	(2) Linear	(3) Log-log	(4) Log-log
avg gasoline price	-69.34***	-69.06***		
	(1.48)	(8.25)		
GPM	-332.94***	974.19		
	(53.89)	(640.07)		
ln(avg gasoline price)			-0.19***	-0.19***
			(0.00)	(0.02)
log(GPM)			0.01***	0.06
			(0.00)	(0.04)
constant	1,638.24***	1,544.35***	7.49***	7.55***
	(17.09)	(64.73)	(0.02)	(0.14)
summer months control	Y	Y	Y	Y
months-to-test controls	Y	Y	Y	Y
month-of-year FE	Y	Y	Y	Y
demographics	Y	Y	Y	Y
veh characteristics	Y	Y	Y	Y
model FE	N	Y	N	Y
R-squared	0.065	0.035	0.064	0.029
Observations	4.65m	4.65m	4.65m	4.65m

Heteroskedasticity-robust s.e. in parentheses, clustered on model in (2) and (4)
*** indicates significant at 1% level

capturing, I also examine a specification several other similar models to Column (1) (not shown). For example, I examine a model with a quadratic term for GPM. I also examine models with higher order polynomials in GPM. These should more flexibly allow for a non-linear relationship between GPM and VMT. I find that with the inclusion of the higher order terms, all of the GPM variables are statistically insignificant. However, the GPM coefficients are relatively large and the signs of the coefficients suggest a negative and convex relationship between GPM and VMT. Since the logarithm is a concave function, this result suggests that a specification that includes the logarithm of the GPM is misspecified. I find a similar issue with the log-log specification with the fuel economy instead of GPM.

Of course, the relationship between VMT and fuel economy may not be well identified by my data. There is not a great deal of time series variation in the fuel economy of the vehicles purchased in 2001 to 2003 (Figure 2.10 in Chapter 2). Only after 2005 did the fuel economy of the fleet begin increasing more substantially. Thus, cross-sectional variation in fuel economy is the primary variation identifying the coefficients. Given the rich set of covariates, it is very possible that correlated unobservables are entirely controlled for, but if not, we might be concerned about identification of the coefficient. Given this, I view the coefficients on fuel economy and GPM with at least some caution. Fortunately, removing the fuel economy variable from the model does not seem to influence the coefficient on the price of gasoline very much at all – and it is clear from Figure 2.9 in Chapter 2 that there is sufficient variation in the price of gasoline.

Cost per mile specification

Just as there is an argument to be made to include GPM rather than fuel economy (even if it seems to make little difference), there is also an argument to be made to include the fuel cost per mile of driving rather than the gasoline price and fuel economy. The primary argument is simply that if we are interested the demand for utilization, the marginal cost of utilization is the cost per mile of driving. This is much of the reason why many of the studies in the literature use the cost per mile as a regressor, rather than separating out the price of gasoline from the fuel economy.

Table 3.5 has the same columns as Table 3.4, only with the fuel cost per mile or log of the fuel cost per mile included rather than the gasoline price or fuel economy/gasoline consumption. Columns (1) and (2) indicate that if the fuel cost per mile (mean = 0.145 dollars per mile) is increased by ten cents, then driving per month would decline by -61 or -85 miles per month respectively. These correspond to an elasticity of VMT with respect to the price per mile of driving of -0.08 and -0.11 at the means respectively. Columns (3) and (4) show similar results, with elasticities of -0.06 and -0.13 at the means respectively. Intuitively, these responsiveness estimates are in between the VMT responsiveness to gasoline prices and to fuel economy, for they are driven by both of these factors. In a sense, these responsiveness estimates are seemingly "brought down" towards zero from the responsiveness to gasoline prices by the responsiveness to fuel economy. I find similar results if I restrict the sample to only "normal" six year intervals between the registration and the smog check.

Table 3.5: Intensive Margin Regressions: Cost per Mile

	(1) Linear	(2) Linear	(3) Log-log	(4) Log-log
fuel cost per mi	-605.36***	-846.88***		
	(15.88)	(87.21)		
log(fuel cost per mi)			-0.06***	-0.12***
			(0.00)	(0.02)
constant	1,382.34***	1,406.88***	6.97***	6.82***
	(15.21)	(64.48)	(0.02)	(0.06)
summer months control	Y	Y	Y	Y
months-to-test controls	Y	Y	Y	Y
month-of-year FE	Y	Y	Y	Y
demographics	Y	Y	Y	Y
veh characteristics	Y	Y	Y	Y
model FE	N	Y	N	Y
R-squared	0.065	0.035	0.064	0.030
Observations	4.65m	4.65m	4.65m	4.65m

Heteroskedasticity-robust s.e. in parentheses, clustered on model in (2) and (4)
*** indicates significant at 1% level

Greater responsiveness to high prices than low prices?

There is a literature starting with Gately (1992) that suggests an asymmetric response to gasoline price changes, with more response as gasoline prices increase and less as gasoline prices decrease. My time period does not contain sufficient downwards variation in gasoline prices to examine this question, but I can examine a related question. Even as long ago as Gately (1992), it has been argued that the consumer responsiveness to changes in the price of gasoline varies with the *level* of the price of gasoline. More recently Manzan and Zerom (2011) make a similar argument using data from 1991 to 1994 and a semi-parametric approach to suggest that the price elasticity ranges between -0.2 at low prices to -0.5 at high prices. While the authors do not state whether these are short-run or long-run elasticities, the use of primarily cross-sectional variation suggests that they can best be interpreted as long-run.

To get a rough sense of whether there appears to be more of a response when prices are high than when they are low, I estimate (3.6) using only data including registrations in 2001 to 2002 and then again using data including registrations from 2003 to 2004. I again restrict the subsample to vehicles that had a smog check within a few months of six years. The subsample covering 2001 and 2002 consists of 1.4 million vehicles and has a mean of the average gasoline price in the interval between the smog tests of \$1.99 per gallon. The subsample covering 2003 and 2004 consists of 0.75 million vehicles (96% in 2003) and has a mean average gasoline price of \$2.22 per gallon. This is not a dramatic difference in average gasoline price, but the prices rose to a much higher level over the six years after registration for the vehicles registered in 2003 to 2004.

Table 3.6 runs the baseline linear specification (3.6) for each of the subsamples. The results in Table 3.6 indicate that the responsiveness appears to be greater in 2003 to 2004 than in 2001 to 2002. The corresponding elasticity values at the means for the two periods are roughly -0.13 in the data for vehicles registered in 2001 to 2002 and -0.19 for vehicles registered in 2003 and 2004. Since I control for changing economic conditions, the differences in the elasticities should not be confounded by the different economic conditions over those two time periods. Table 3.6 only shows a selected few of the coefficients for brevity, but interestingly enough the coefficients on nearly all of the control variables did not change between the two time frames.

Table 3.6: Intensive Margin Regressions: Higher and Lower Prices

Dependent variable: vehicle-miles-traveled per month

	(1) Pre-2003	(2) Pre-2003	(3) Post-2002	(4) Post-2002
avg gasoline price	-68.72***	-67.05***	-105.93***	-91.43***
	(2.58)	(13.69)	(4.38)	(9.16)
fuel economy	1.61***	-3.72	4.56***	0.68
	(0.16)	(2.80)	(0.24)	(2.38)
constant	1,522.82***	1,658.00***	1,702.36***	1,758.88***
	(25.56)	(107.78)	(31.94)	(96.51)
summer months control	Y	Y	Y	Y
months-to-test controls	Y	Y	Y	Y
month-of-year FE	Y	Y	Y	Y
demographics	Y	Y	Y	Y
veh characteristics	Y	Y	Y	Y
model FE	N	Y	N	Y
R-squared	0.065	0.034	0.067	0.038
Observations	4.65m	4.65m	4.65m	4.65m

Heteroskedasticity-robust s.e. in parentheses, clustered on model in (2) and (4)
*** indicates significant at 1% level

These cursory results deserve more exploration with a longer time frame and additional low and high periods, but are quite interesting in the context of the findings of Small and Van Dender (2007), Hughes, Knittel, and Sperling (2008), and Hymel, Small, and Van Dender (2010). All of these studies use variation prior to the higher gasoline prices in 2007 and 2008 and find a short-run responsiveness either in the gasoline demand or driving demand elasticity closer to zero than -0.10. Perhaps these results indicate that policy analysts should take into account the level of the gasoline price when using estimates of the price elasticity of gasoline or driving demand. On the other hand, it is also possible that the greater responsiveness at times of higher gasoline prices may also be due to the very noticeable speed of the increases in the gasoline prices in 2007 and 2008, rather than the level of the prices. The time frame of my dataset does not contain times of high and level gasoline prices, so I can not separately identify these two effects.

Robustness

I examine a variety of additional specifications to check the robustness of my results. Throughout all of the specifications, I find that the results were quite insensitive to taking out any one variable. If I remove the economic conditions variables, the estimated responsiveness to gasoline price changes is greater, as would be expected. If I remove the fuel economy variable, I find that there is nearly no change in the coefficient on the price of gasoline, which is reassuring. If I include a variable for whether the vehicle is leased, the estimated responsiveness appears to barely change at all.

Interestingly, if I include an indicator for whether the vehicle is a hybrid (not in my baseline specifications above), I find a negative coefficient, suggesting that conditional on having the higher fuel economy, hybrids are driven less. This may partly be because they are most effective for city driving, when the regenerative breaking is more useful. The reason I did not include this variable for whether a vehicle is a hybrid is that hybrids are only observed in the dataset when there is a title change, and are not required to have the standard smog check after six years. Thus, I am concerned that such a variable is partly confounded with selection based on title changes. This is a topic ripe for future investigation in this dataset. Fortunately for this analysis, I find that the coefficients of interest, such as on the price of gasoline, do not change at all if I add the hybrid vehicle indicator variable.

I find that the fact that some consumers moved makes nearly no difference to the results. To examine this, I run the specifications where vehicles that had the initial registration county differ from the test county or registration county at the time of the test facing either the gasoline price in the initial county or in the subsequent county. Moreover, I can use the average of the two gasoline prices. None of these tests make any noticeable difference to the results, most likely due to the fact that most vehicles did not move and that time series variation in gasoline prices is the primary identifying variation.

What does change the results more noticeably is if substantially different variation is used. The results are fairly unstable with the inclusion of county-level fixed effects: the coefficients show more responsiveness in some specifications and not others. County fixed effects not only remove all cross-sectional variation in the gasoline price, but along with all of the other controls lead identification to be based on possibly more limited time series variation. At the same time, an argument could be made that the rich set of demographic covariates in this study captures any heterogeneity that could be cross-sectionally correlated with the gasoline price. Including fixed effects for the quarter or month of the registration appears to have a similar effect. It again may not make sense to do this when I already have variables to account for economic conditions, for it limits the variation being used in the study. In both cases I am including so many fixed effects that I run the risk of over-fitting, so that there is not much variation left to identify the coefficients and it is difficult to decipher what this remaining variation indicates. Thus, I do not include these results.

3.1.2 All Vehicles

The all vehicles dataset contains nearly all of the vehicles in the fleet during a several year period in the early 2000s. It contains all of the biennial test results for all of the non-exempt operated vehicles in California older than six model years. In addition, it contains all of the vehicles in the R.L. Polk data (including non-personal vehicles such as rental cars and company cars) registered in 2001 to 2004 that matched with a smog check result. Thus, the only portion of the fleet missing includes some vehicles less than six model years old that I do not have R.L. Polk data for (e.g., registered first in 2000). As was demonstrated in Chapter 2, the vast majority of the fleet is included, with as much as 80 percent of all registered vehicles in California included in some years. Chapter 2 also shows that this dataset contains many fewer controls than the personal vehicle dataset.

Linear specification

Table 3.7 contains the first set of results using the 49.7 million observation dataset. I again use a linear-in-levels specification. The dependent variable is VMT per month, which has a mean of 833 miles per month in the larger dataset. Column (1) only includes the average gasoline price as a regressor, and it indicates that without any controls, an increase in the gasoline price by one dollar is associated with a decrease in driving by 147 miles per month. This corresponds to a VMT elasticity with respect to the price of gasoline around -0.49 at the means. We can loosely consider this elasticity to be a one to two year elasticity, for most of the smog checks in the dataset (i.e., all but those from the R.L. Polk data) were within a two years interval. When compared to the elasticity results using the new personal vehicle dataset that uses a six year interval between the registration and smog test, this elasticity should be viewed as having a shorter time horizon.

Column (2) adds economic conditions, demonstrating that if we do not account for economic conditions, we will likely overestimate the responsiveness. Adding a variable for the percentage of the interval between smog tests that is summer and a variable for county-level commute times does not appear to make any noticeable difference to the coefficient on the average gasoline price. The results in Columns (2) through (4) all correspond to an elasticity of VMT with respect to the gasoline price in the range of -0.44 at the means. Interestingly, this estimate corresponds very closely with the findings in Knittel and Sandler (2010). Knittel and Sandler (2010) use the roughly same smog check data, only they have several more years in the sample, use only a log-log specification, and do not have access to the R.L. Polk data.

In Column (5) we see that adding vehicle characteristics, including fixed effects for the vintage of the vehicle (i.e., model year) serves to increase the estimated responsiveness. The result in Column (5) corresponds to an elasticity of -0.57 at the

Table 3.7: Intensive Margin Regressions: All Vehicles

Dependent variable: vehicle-miles-traveled per month (mean = 833 miles per month)

	(1) base	(2) econ cond	(3) summer	(4) commute	(5) vehicle chars	(6) VIN FE	(7) VIN & year FE
avg gasoline price	-147.52***	-131.83***	-133.24***	-132.29***	-170.99***	-151.23***	-86.40***
	(0.20)	(0.21)	(0.21)	(0.21)	(0.20)	(0.68)	(1.64)
avg unemployment		-7.25***	-7.52***	-6.61***	-2.91***	-12.70***	-9.28***
		(0.07)	(0.07)	(0.07)	(0.07)	(0.40)	(0.41)
avg CCI		0.78***	0.75***	0.81***	1.78***	1.03***	0.04
		(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.04)
avg house price		-0.17***	-0.17***	-0.17***	-0.16***	-0.20***	-0.18***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
% summer months			66.98***	66.25***	74.69***	65.42***	31.61***
			(0.93)	(0.93)	(0.92)	(2.86)	(2.99)
commute				2.47***	1.98***	1.69***	1.87***
				(0.03)	(0.03)	(0.23)	(0.23)
liters					-36.76***	-24.62	-20.85
					(0.17)	(13.00)	(13.00)
cylinders					1.75***	1.80	-1.69
					(0.13)	(5.93)	(5.92)
auto transmission					1.88***	-48.81***	-48.91***
					(0.23)	(3.24)	(3.24)
constant	1,246.09***	1,263.72***	1,254.98***	1,173.31***	787.37***	1,331.10***	1,297.01***
	(0.58)	(1.24)	(1.25)	(1.56)	(6.69)	(68.30)	(68.67)
model year FE	N	N	N	N	Y	N	N
vehicle body FE	N	N	N	N	Y	Y	Y
VIN FE	N	N	N	N	N	Y	Y
test year FE	N	N	N	N	N	N	Y
R-squared	0.012	0.013	0.014	0.014	0.126	0.028	0.030
Observations	49.7m	49.7m	49.7m	49.7m	49.7m	4.97m	4.97m

Heteroskedasticity-robust standard errors in parentheses, clustered on VIN in cols (6)-(7)

*** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level

means. Column (6) includes VIN fixed effects, which can be included here because I observe several tests of the same vehicle in this dataset. VIN fixed effects have an advantage in that they nonparametrically control for any time invariant vehicle-level unobservables that may be correlated with the gasoline price. To the extent that each vehicle is held by the same household over time, this approach controls for household-level unobservables. Of course, many vehicles are sold or change hands within a family. Time invariant VIN fixed effects are not useful if these changes in vehicle driver are very common throughout the dataset. I suspect that they are, but without household-level information (e.g., from actual DMV data), it is impossible to know. It is certainly worth examining the results with VIN fixed effects. Unfortunately, with 49.7 million observations, I found it impossible to run VIN fixed effects on my computer with the full dataset. Thus, I took a randomly drawn 10 percent subsample of VINs, and performed the VIN fixed effects estimation on this subsample. I find a coefficient on the average gasoline price of -151, which corresponds to a one-year elasticity of driving with respect to the price of gasoline of -0.50 at the means.

The final column of Table 3.7 includes year fixed effects in addition to VIN fixed effects using the same 10 percent subsample. There are enough different years covered in this larger dataset (2001 to 2009) that year fixed effects may make sense. I view the results with VIN fixed effects as an exploration of a different source of variation: within-year variation. Of course, they also serve to nonparametrically control for any unobserved heterogeneity that changes over time and may be correlated with gasoline prices. Given that I control for changing economic conditions, it is not entirely clear what additional possible correlated heterogeneity needs to be controlled for. But, I do find examining a specification with year fixed effects as useful for getting a sense of what the *one-year* elasticity may be. I find that an increase in the gasoline price by one dollar is associated with a decrease in driving by 86 miles per month in that first year since the change in the gasoline price. This corresponds to an elasticity of -0.29 at the means.

Log-log specification

Just as for the new personal vehicle dataset, we can look at a log-log specification for an alternative functional form and more convenient interpretation of the coefficient. Table 3.8 contains the same columns as Table 3.7 only with VMT and the gasoline price entering in logarithms. The results are in general quite similar to those in the linear specification. One noticeable difference is that the responsiveness appears to be greater across the board in the log-log specification. This is particularly true in Column (7), where adding fixed effects for the test year (i.e., year of the test that begins the interval) appears to make entirely no difference to the coefficient on the gasoline price.

Table 3.8: Intensive Margin Regressions: All Vehicles (Log-log)

Dependent variable: log vehicle-miles-traveled per month

	(1) base	(2) econ cond	(3) summer	(4) commute	(5) vehicle chars	(6) VIN FE	(7) VIN & year FE
log(gasoline price)	-0.54***	-0.50***	-0.50***	-0.50***	-0.73***	-0.68***	-0.67***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
avg unemployment		-0.01***	-0.01***	-0.01***	-0.00***	-0.02***	-0.02***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
avg CCI		0.00***	0.00***	0.00***	0.00***	0.00***	0.00***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
avg house price		-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
% summer months				0.07***	0.09***	0.08***	0.11***
				(0.00)	(0.00)	(0.01)	(0.01)
commute				0.00***	-0.00***	-0.00**	-0.00**
				(0.00)	(0.00)	(0.00)	(0.00)
liters					-0.12***	-0.02	-0.02
					(0.00)	(0.03)	(0.03)
cylinders					0.03***	-0.03*	-0.03*
					(0.00)	(0.01)	(0.01)
auto transmission					0.05***	-0.08***	-0.08***
					(0.00)	(0.01)	(0.01)
constant	6.93***	6.91***	6.91***	6.87***	5.32***	7.28***	7.12***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.15)	(0.15)
model year FE	N	N	N	N	Y	N	N
vehicle body FE	N	N	N	N	Y	Y	Y
VIN FE	N	N	N	N	N	Y	Y
test year FE	N	N	N	N	N	N	Y
R-squared	0.007	0.007	0.007	0.007	0.165	0.025	0.028
Observations	49.7m	49.7m	49.7m	49.7m	49.7m	4.97m	4.97m

Heteroskedasticity-robust standard errors in parentheses, clustered on VIN in cols (6)-(7)
*** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level

There are two caveats worth noting about these results. First, the issues with the log-log specification may apply here as well, even though the time interval between the smog tests is smaller. Second, the fit of the model, based simply on the R-squared criterion, appears to be better with the linear specification than the log-log specification. The R-squared is certainly far from a perfect model selection criterion, but it provides some guidance. For nearly all of the specifications, the R-squared is higher for the linear specifications than the log-log specifications. I also examine the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) and, not surprisingly, find similar results. Thus, I consider the linear specification model results to be the preferred results.

Comparison of subsamples

There may be an entirely different concern though about the results in Table 3.7. The sample used to estimate the results above in Table 3.7 and Table 3.8 can be thought of as weighted towards older vehicles, for vehicles older than six years make up more than half of the fleet and I do not have data for all of the vehicles under six years (e.g., vehicles registered in 2005 and have not had the first smog check yet). To understand whether this sample selection is substantially biasing my results, we can first quickly compare the results in Table 3.7 to the results using only the sample of new personal vehicles. We found an elasticity of in the range of -0.17 to -0.25 at the means from the new personal vehicle dataset, while with the larger dataset we are finding an elasticity in the range of -0.3 to -0.57 at the means. This greater estimated responsiveness in the larger sample with a shorter time interval immediately suggests that new vehicles may be driven somewhat differently.

We can examine this possibility in another way by looking at the subsample of vehicles that had a five year smog check or greater. Nearly all (over 99 percent) of these vehicles are vehicles from the R.L. Polk dataset. Running the same estimation as in Table 3.7 on this subsample are mixed. The results given in Columns (1) through (4) of Table 3.9 suggest that the responsiveness is greater relative to the estimates from the new personal vehicles dataset as well as the estimates from the "all vehicles dataset" in Table 3.7. However, the results in Columns (5), where vehicle characteristics are controlled for, suggests a much lesser response, an elasticity in the range of -0.13 at the means. The model year fixed effects are the most important addition that leads to a changed response. Not surprisingly, when VIN fixed effects are added, nearly all of the coefficients are insignificant. This is simply because very few vehicles that had an interval between smog tests more than five years also had another smog test. If there is only one smog test for each VIN, VIN fixed effects are not identified. Thus, Column (5) is the preferred specification in Table 3.9.

Table 3.9: Intensive Margin Regressions: All Vehicles with Smog Check > 5 Years

Dependent variable: vehicle-miles-traveled per month (mean = 1,012 miles per month)

	(1) base	(2) econ cond	(3) summer	(4) commute	(5) vehicle chars	(6) VIN FE	(7) VIN & year FE
avg gasoline price	-204.81***	-194.10***	-194.17***	-192.02***	-48.08***	-36.12	11.60
	(5.28)	(2.88)	(2.88)	(2.88)	(3.61)	(45.86)	(51.87)
avg unemployment		-1.62***	-1.63***	-1.01***	0.15	-4.73	-5.15
		(0.27)	(0.27)	(0.27)	(0.26)	(3.16)	(3.16)
avg CCI		-0.17	-0.17	-0.07	6.00***	1.17	-0.86
		(0.12)	(0.12)	(0.12)	(0.13)	(1.58)	(2.27)
avg house price		-0.15***	-0.15***	-0.15***	-0.22***	-0.20***	-0.21***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)
% summer months			60.93	60.46	459.54***	709.62	584.52
			(34.54)	(34.54)	(31.81)	(372.67)	(380.57)
commute				1.83***	2.68***	0.24	0.26
				(0.11)	(0.10)	(1.13)	(1.13)
liters					7.94***	18.87	19.44
					(0.66)	(30.62)	(30.63)
cylinders					-22.35***	-14.71	-16.17
					(0.49)	(12.33)	(12.34)
auto transmission					-4.96***	-58.05***	-57.88***
					(1.06)	(13.85)	(13.84)
constant	1,570.77***	1,646.15***	1,631.26***	1,561.81***	-108.66*	1,142.24***	1,267.01***
	(14.37)	(18.08)	(19.94)	(20.31)	(46.00)	(290.23)	(339.73)
model year FE	N	N	N	N	Y	N	N
vehicle body FE	N	N	N	N	Y	Y	Y
VIN FE	N	N	N	N	N	Y	Y
test year FE	N	N	N	N	N	N	Y
R-squared	0.005	0.007	0.007	0.007	0.116	0.006	0.006
Observations	3.1m	3.1m	3.1m	3.1m	3.1m	0.3m	0.3m

Heteroskedasticity-robust standard errors in parentheses, clustered on VIN in cols (6)-(7)

*** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level

How can we interpret these results? First, we can consider the difference between these results and the results for the entire "all vehicles" dataset in Table 3.7. With Column (5) as the preferred specification, the result here suggests that drivers of newer vehicles in the fleet are less responsive than drivers of older vehicles. This is consistent wealthier drivers driving newer vehicles. It is also consistent with drivers who value the luxury and comfort of newer vehicles receiving more enjoyment per dollar of fuel cost, so that they are less responsive to fuel cost changes. It is also consistent with households with multiple vehicles switching to newer, more fuel efficient vehicles when the price of gasoline increases. Of course, the variation identifying these results is based on the six-year intervals between registrations and smog tests, rather than the roughly two-year intervals. We would generally expect that the responsiveness would be greater for the vehicles with a six-year interval between tests, for there is more time for adjustment. Since we find the opposite, this perhaps can be viewed as additional evidence of heterogeneity in responsiveness between the drivers of new versus older vehicles.

In interpreting these results, we can also compare these results from the larger dataset restricted to newer vehicles (Table 3.9) to the results from the new personal vehicles dataset (Table 3.7). The responsiveness estimated in Column (5) is less than the responsiveness in the new personal vehicles dataset. This suggests that drivers of new *personal* vehicles in the first six years are more responsive than *all* drivers of new vehicles in the first six years. One might imagine that drivers of company vehicles and government vehicles have an almost entirely inelastic response to changing gasoline prices if they are not paying for the cost of the gasoline, as may often be the case. The other difference between the two estimations is simply all of the additional control variables in the new personal dataset. If I drop many of these and run an identical specification on the new vehicle dataset, I find roughly the same result, indicating that these additional covariates are not a major reason for the difference.

To understand better how older vehicles differ from newer vehicles, I run the same estimations as before, only restricted to the older vehicles that were under the biennial smog check program. Table 3.10 presents these results. Without controls for vehicle characteristics, it appears that the older vehicles are less responsive than the newer vehicles. However, once vehicle characteristics are added in Column (5), the relative responsiveness flips. Comparing Column (5) in Table 3.9 and Table 3.10, we can see that the older vehicles are more responsive to gasoline prices. The coefficient on gasoline price in Column (5) suggests than an increase in gasoline prices by one dollar is associated with a decrease in driving by 171 miles per month, or an elasticity of -0.57 at the means. Including VIN fixed effects is a more useful approach here for there are many vehicles in the dataset where I observe multiple tests. Of course, the issues of using VIN fixed effects when there may be different owners of the vehicles still certainly applies. But the results follow those in the new personal vehicles dataset: adding VIN fixed effects brings down the responsiveness. The corresponding elasticity in Column (6) of -0.49 still indicates more of a response than for newer vehicles. Column (7) adds test year fixed effects to explore within-test year variation. As before, this brings the responsiveness down and perhaps gives a sense of what the one-year responsiveness is. This one-year elasticity estimate is roughly -0.30 at the means.

Table 3.10: Intensive Margin Regressions: All Vehicles with Smog Check <= 5 Years

Dependent variable: vehicle-miles-traveled per month (mean = 821)

	(1) base	(2) econ cond	(3) summer	(4) commute	(5) vehicle chars	(6) VIN FE	(7) VIN & year FE
avg gasoline price	-143.89***	-127.26***	-128.69***	-127.78***	-171.52***	-148.15***	-90.38***
	(0.64)	(0.21)	(0.21)	(0.21)	(0.20)	(0.70)	(1.83)
avg unemployment		-7.40***	-7.68***	-6.80***	-2.98***	-12.49***	-9.73***
		(0.07)	(0.07)	(0.07)	(0.07)	(0.41)	(0.43)
avg CCI		0.71***	0.68***	0.74***	1.82***	0.95***	0.20***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.04)
avg house price		-0.17***	-0.17***	-0.17***	-0.15***	-0.18***	-0.18***
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
% summer months			67.32***	66.61***	73.69***	60.39***	32.89***
			(0.93)	(0.93)	(0.93)	(2.90)	(3.07)
commute				2.39***	1.97***	1.65***	1.83***
				(0.03)	(0.03)	(0.24)	(0.24)
liters					-38.39***	-3.10	-2.29
					(0.17)	(17.14)	(17.14)
cylinders					2.31***	-12.51	-12.21
					(0.13)	(9.02)	(9.00)
auto transmission					2.25***	-49.99***	-49.85***
					(0.24)	(3.35)	(3.35)
constant	1,224.81***	1,248.47***	1,239.90***	1,160.95***	787.95***	1,311.02***	1,230.73***
	(1.85)	(1.26)	(1.27)	(1.60)	(6.76)	(94.18)	(94.11)
model year FE	N	N	N	N	Y	N	N
vehicle body FE	N	N	N	N	Y	Y	Y
VIN FE	N	N	N	N	N	Y	Y
test year FE	N	N	N	N	N	N	Y
R-squared	0.012	0.013	0.013	0.014	0.122	0.027	0.028
Observations	46.6m	46.6m	46.6m	46.6m	46.6m	4.6m	4.6m

Heteroskedasticity-robust standard errors in parentheses, clustered on VIN in cols (6)-(7)
*** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level

Six-year versus two-year data

In the new personal vehicles dataset, only six year data were available. One concern in using this dataset is the length of the time period between smog tests. For example, we might be concerned that the averaging over such a long period of time may mask changes in gasoline prices, economic conditions, and VMT. Fortunately, the all vehicles dataset contains smog test odometer readings at largely two-year intervals, rather than six-year intervals. Combining three of the two-year intervals together yields a six-year interval, so I can compare the two-year results to the six-year results. The analysis in this subsection is based on this idea.

The data in the all vehicles dataset are not always in two-year intervals, due to the required smog test whenever the vehicle title is changed. Rather than drop these data, I use the odometer readings from the first and last tests in my dataset for any particular VIN, thus combining all of the intervals between tests for each VIN. Vehicles that only have two tests are dropped from the analysis. In the all vehicle dataset, a several observations may have the same VIN, but in the converted dataset there is one observation for each VIN.

In combining the intervals between tests, I have to drop a significant amount of data (i.e., all of the middle intervals), which leads to a re-weighting of the results. Thus, to illustrate the effect this might have, Table 3.11 shows how the results in Table 3.7 change with incremental changes in dataset. Columns (1) and (2) replicate Columns (2) and (5) in Table 3.7. Columns (3) and (4) contain the same specifications, but restricted to all VINs that have more than two tests (i.e., all VINs that are a single observation are dropped). This effectively drops nearly all of the R.L. Polk data. Columns (5) and (6) keep only the first and last intervals between tests for each VIN in preparation for combining these into one observation that contains the first and last test for each VIN. All of these columns still use the two-year (or shorter) time intervals between tests and show the results as the dataset progressively evolves towards the fully converted dataset.

The results in Table 3.11 show that the coefficient on the average gasoline price over the time interval remains fairly consistent when the observations are dropped in the process of converting the dataset. This suggests that the changing of the dataset

Table 3.11: Intensive Margin Regressions: Effect of Converting Dataset

Dependent variable: vehicle-miles-traveled per month

	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline		Drop Single Intervals		No Middle Intervals	
	econ cond	vehicle chars	econ cond	vehicle chars	econ cond	vehicle chars
avg gasoline price	-131.83***	-170.99***	-134.90***	-159.95***	-139.94***	-163.65***
	(0.21)	(0.20)	(0.22)	(0.21)	(0.26)	(0.25)
avg unemployment	-7.25***	-2.91***	-7.44***	-3.06***	-8.00***	-3.84***
	(0.07)	(0.07)	(0.08)	(0.07)	(0.09)	(0.09)
avg CCI	0.78***	1.78***	0.75***	1.73***	1.46***	2.14***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
avg house price	-0.17***	-0.16***	-0.17***	-0.15***	-0.19***	-0.17***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
% summer months		74.69***		69.31***		80.44***
		(0.92)		(0.95)		(1.32)
commute		1.98***		1.86***		2.27***
		(0.03)		(0.03)		(0.04)
liters		-36.86***		-38.29***		-37.26***
		(0.16)		(0.17)		(0.22)
cylinders		1.82***		1.26***		1.82***
		(0.13)		(0.13)		(0.18)
auto transmission		1.89***		0.74**		-1.57***
		(0.23)		(0.24)		(0.32)
constant	1,263.72***	787.35***	1,265.41***	767.19***	1,242.37***	752.80***
	(1.24)	(6.69)	(1.30)	(6.91)	(1.58)	(8.76)
model year FE	N	Y	N	Y	N	Y
vehicle body FE	N	Y	N	Y	N	Y
R-squared	0.013	0.126	0.015	0.121	0.020	0.135
Observations	49.7m	49.7m	44.3m	44.3m	27.6m	27.6m

Heteroskedasticity-robust standard errors in parentheses

*** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level

is not likely to importantly influence the coefficients.

Switching to the longer interval dataset by looking at the time between the first test and the last test seems to make some difference. Table 3.12 shows the results using the converted dataset where an observation is the time between the first and last test for each VIN (in the 2002 to 2009 time period). The mean length of the interval between tests is 64 months, rather than 25 months in the complete all vehicles dataset. The mean VMT is 563 miles per month, which is less than the mean of 833 in the all vehicles dataset due to the fact that both the R.L. Polk data and many of the shorter time intervals between tests are not included. Columns (1) to (3) contain the same specifications as Columns (1), (2), and (5) in Table 3.7. Columns (4) to (6) contain the same specifications, but include controls for the number of months between smog tests. The same controls that are included in the specification in Table 3.1 are included here. I include these controls for the same reason that they were included in the six-year data analysis: vehicles that have had title changes (or have very late tests) will have a different time between tests and may be driven differently. In the all vehicle dataset, the time between tests is short enough that there is less variation in gasoline prices based simply on the time between tests, so these controls are unnecessary. However, in the converted dataset used for this analysis, there is considerable variation in the time between tests (standard deviation of 17 month), and thus very significant variation in the gasoline price faced by drivers who have different time intervals between tests.

A comparison of the results in Tables 3.12 and 3.11 indicates that there is some difference when the longer intervals between smog tests are used. In the first three columns, the coefficients on the average gasoline price indicate greater responsiveness than the equivalent specification in the earlier tables (Table 3.7 or 3.11). The corresponding elasticity of driving with respect to the average price of gasoline in Columns (1) and (2) is around -0.9 at the means. When the months to test controls are included, the coefficient on the gasoline price is somewhat closer to zero. The coefficient in Column (5) corresponds to a decrease in driving of -84 miles per month with a one dollar increase in the price of gasoline. The corresponding elasticity of driving with respect to the price of gasoline is around -0.6 at the means. The responsiveness is

Table 3.12: Intensive Margin Regressions: Longer Interval Between Tests

Dependent variable: vehicle-miles-traveled per month (mean = 563)

	(1) base	(2) econ cond	(3) vehicle chars	(4) base	(5) econ cond	(6) vehicle chars
avg gasoline price	-181.56***	-178.56***	-220.66***	-107.63***	-84.09***	-188.58***
	(0.55)	(0.59)	(0.59)	(0.55)	(0.60)	(0.60)
avg unemployment		-0.77***	3.77***		0.37***	4.65***
		(0.10)	(0.09)		(0.09)	(0.09)
avg CCI		-4.79***	-3.14***		1.46***	0.39***
		(0.02)	(0.02)		(0.02)	(0.02)
avg house price		-0.09***	-0.06***		-0.07***	-0.04***
		(0.00)	(0.00)		(0.00)	(0.00)
% summer months			230.83***			139.03***
			(7.07)			(6.86)
commute			1.33***			1.08***
			(0.04)			(0.03)
liters			-32.55***			-30.08***
			(0.20)			(0.20)
cylinders			5.63***			4.28***
			(0.17)			(0.17)
auto transmission			-6.18***			-7.17***
			(0.31)			(0.30)
constant	1,072.98***	1,553.52***	1,101.79***	828.50***	664.32***	692.00***
	(1.55)	(2.85)	(6.59)	(1.57)	(3.21)	(6.71)
model year FE	N	N	Y	N	N	Y
vehicle body FE	N	N	Y	N	N	Y
time-to-test controls	N	N	N	Y	Y	Y
R-squared	0.010	0.016	0.122	0.059	0.060	0.139
Observations	13.6m	13.6m	13.6m	13.6m	13.6m	13.6m

Heteroskedasticity-robust standard errors in parentheses

*** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level

even greater in Column (6) when vehicle characteristics and model fixed effects are added.

In some respects, the results are not dramatically different between the longer and shorter time intervals. Comparing Column (6) in Table 3.12 and Table 3.11 shows this point. In Table 3.11, the coefficient on the average gasoline price is -163, while in Table 3.12, the coefficient is -188. While these are certainly different, they are in the same general range of responsiveness.

When comparing the coefficients, the results suggest that the responsiveness in driving to gasoline price changes is even *larger* when a longer time interval between tests is used. This may be due a variety of factors: a slightly different sample being used (e.g., more of 2007 and 2008 being covered) and perhaps simply because of the averaging over the longer time interval. To the extent that the latter is the case, this finding reinforces the result that the responsiveness for newer vehicles is less than the responsiveness for older vehicles in the fleet. Without more data at shorter intervals over a different time path of gasoline prices, it is difficult to know whether the difference in results is simply due to a different subsample and the heterogeneity in responsiveness, or due to an issue with the averaging over the longer period of time.

3.1.3 Summary Discussion

The findings in this section suggest an elasticity of VMT with respect to the price of gasoline on the order of -0.17 to -0.25 (taken at the means) for new personal vehicles being driven during the first six years. The -0.17 elasticity was estimated using all new personal vehicles for which I observe a smog check in 2001 to 2004, while the -0.25 elasticity was estimated on a subsample of drivers who had a smog check within two months of six years. Furthermore, I show that these results are relatively insensitive to the choice of specification: linear-in-levels or log-log. I also show that a specification that includes fuel consumption in gallons per mile, rather than fuel economy in miles per gallon, yields very similar results.

I find very little responsiveness in driving to fuel economy, with an estimated elasticity of VMT with respect to fuel economy in the range of zero to -0.05. One reason for this low responsiveness may relate to asymmetric response by consumers. Another may simply be the somewhat limited variation in fuel economy in 2001 to 2004. In fact, during that period, fleet-wide average fuel economy was largely flat, so much of the identification of the fuel economy coefficient is based on cross-sectional variation.

Interestingly, a specification with the fuel cost per mile of driving entering the model instead of the price of gasoline and fuel economy tends to lead to an estimated driving responsiveness in between the responsiveness with respect to the gasoline price and the responsiveness with respect to fuel economy. To the extent that this is generalizable, this may have implications for the many studies that estimate a responsiveness based on the cost per mile of driving.

I separately estimate the linear model for vehicles registered in the years 2001 to 2002 and 2003 to 2004 to see whether there is any evidence that a lower gasoline price level (and variance of price) tends to lead to less responsiveness. Indeed, this is the suggestive result of running these separate estimations: the elasticity in the earlier two years is -0.13 at the means, while the elasticity in the later two years is -0.19 at the means. Of course, there may be other differences between these years, so this result is best interpreted as a suggestive result.

My estimations for new personal vehicles include an extremely rich set of covariates, but are a particular subsample of the vehicles in the light duty fleet, all of which may be affected by changes in gasoline prices. In fact, we might imagine that drivers of older (i.e., usually lower fuel economy) vehicles are more responsive. This is exactly what I find when I estimate a similar linear-in-levels specification using my larger "all vehicles" dataset. I find a roughly two-year elasticity of driving with respect to the price of gasoline around -0.5 at the means for the full dataset including the newer vehicles that I observe. I find the same result when I restrict my dataset to only the older vehicles for which I observe multiple the biennial smog test results. If I include year of test fixed effects, I find an elasticity value of -0.3 at the means, which perhaps is best interpreted as a one-year elasticity.

If I restrict my sample to only vehicles that receive a smog check after five years or more, I find that this subsample of newer vehicles tends to be less responsive, as one might expect. The elasticity for this subsample, which includes personal vehicles, company cars, rental cars, and government cars, is estimated to be -0.13 in my preferred specification. The variation from the five-plus year intervals between registration and test is again identifying the coefficients, so I interpret the elasticity as a roughly two-year elasticity, just as for the personal vehicle dataset. The logic behind this is again that the variation in gasoline prices was largely over the two years and not over the entire interval.

These results provide perhaps the most complete explorations of the elasticity of driving with respect to the gasoline price in the literature. The bottom-line is simply that while the responsiveness to gasoline prices is still quite inelastic, there is more responsiveness than some previous studies, such as Hymel, Small, and Van Dender (2010), Small and Van Dender (2007), and Hughes, Knittel, and Sperling (2008), may have suggested. However, there is also some initial evidence that consumers respond differently to gasoline price increases than to changes in fuel economy, perhaps because the responsiveness is a function of both the sign of the change, the speed of the change, and the level of the change in the cost per mile of driving. Finally, these results also begin to point to some important heterogeneity in how different types of drivers respond to gasoline prices. Drivers of older vehicles appear to be more responsive than drivers of newer vehicles. Drivers of personal vehicles appear to be more responsive than drivers of other types of vehicles. The next section of this chapter explores heterogeneity in driving responsiveness in much more detail.

3.2 Heterogeneity in the Driving Responsiveness

It is intuitive that there would be heterogeneity in the responsiveness of driving to gasoline price changes. The previous section already provides suggestive evidence that there is heterogeneity in responsiveness based on the age of the vehicle, where drivers of older vehicles tend to be more responsive than those of newer vehicles. However, consumers also differ in where they live, how much money they have, and their driving needs. All of these factors influence how much consumers are willing to alter driving plans when gasoline prices change. This heterogeneity has important implications for the effects of gasoline price changes on local air pollution and congestion. It also has implications for the distributional consequences of policies that raise the price of gasoline.

In this section, I examine heterogeneity in the driving responsiveness using several different approaches. Quantile regression is relatively new and increasingly popular technique to give us a sense of what the distribution of the responsiveness is for different consumers by allowing us to estimate the response at different quantiles of the population (Koenker and Hallocks 2001). The more standard approach to examine heterogeneity would simply be to interact the gasoline price with different variables for the group in the population we are interested in. Following this approach, I first interact fuel economy and different vehicle classes with the gasoline price to see whether drivers of lower fuel economy vehicles are more responsive to gasoline prices, as we might expect. I then interact different demographics with the gasoline price to paint a picture of what groups of consumers are most able to adjust to changing gasoline prices. Along these lines, I can also use the vehicle-specific household income variable that is present in my dataset. This variable is only available for a subsample of the dataset, so I do not include it in the analysis above. But, for examining heterogeneity in the responsiveness based on income, this vehicle-specific variable is more useful than the zip code median household income variable.

For understanding the implications of a policy that raises gasoline prices for local air pollution and congestion, we are most interested in geographic heterogeneity. I examine interactions with location-specific variables, such as the density of the zip code. Even more interestingly, I interact indicator variables for the counties in California with the gasoline price to quantify which counties are more or less responsive.

For the analysis of heterogeneity, I use the dataset of new personal vehicles over the first six years for nearly all of the heterogeneity analysis, due to the much richer set of variables to work with. Thus, there are two important caveats to make about this analysis. First, the analysis only applies to new personal vehicles, rather than all vehicles. Second, the dataset applies to *vehicles*, rather than *households*. Using vehicles, rather than households, is certainly useful for examining the effects of gasoline price changes on local air pollution and congestion. On the other hand, if we are interested in the implications of heterogeneity in driving responsiveness for the distributional consequences of policies, using households as the unit of observation has a more natural interpretation. Since some households own several vehicles, while others own one or none, it is important to carefully interpret the responsiveness at the vehicle-level. For example, to the extent that wealthier households own more vehicles, by using vehicle-level data, we may underestimate the heterogeneity in the responsiveness by household-level income. The recognition that these are vehicle-level results should be considered in the interpretation of the results.

3.2.1 Quantile Regression

Quantile regression allows us to estimate the quantiles of responsiveness in a population. For example, the median is the 0.5 quantile. A median quantile regression is known as the "least absolute deviation" estimator, for it is minimizing the sum of the absolute value of the error terms (or deviations), rather than the sum of the squares of the error terms as in OLS. A median quantile regression is more robust to outliers than OLS, and thus is becoming a more popular robustness check when an undue influence of outliers is suspected. A median quantile regression is only one possibility; we can examine a quantile for any number between zero and one. For example, the 0.2 quantile regression yields coefficients that indicate the relationship between the regressor and the dependent variable for the 0.2 quantile of the population.

Background

To illustrate what quantile regression is doing, I will give a brief formal description, loosely following the treatment in Koenker and Hallocks (2001). For an in-depth treatment of quantile regression, see Koenker (2005). To begin, consider any $\tau \in [0, 1]$. The τ-quantile of any random variable $X \sim F_X$ is given by

$$Q_X(\tau) = F_X^{-1}(\tau) = \inf\{x : F_X(x) \geq \tau\}.$$

For quantile regression we are interested in the conditional quantile function $Q_{Y|X}(\tau) = X\beta_\tau$, where $Y \sim F_Y$. To understand how quantile regression is estimated, first consider the loss function

$$\rho_\tau(x) = x(\tau - \mathbb{1}_{\{x<0\}}),$$

where $\mathbb{1}_{\{x<0\}}$ is an indicator function for $x < 0$. The optimization problem for quantile estimation is then

$$\min_\beta \mathbb{E}\left[\rho_\tau(Y - X\beta)\right].$$

Note here we are minimizing the loss function ρ_τ evaluated at the residuals, just as the minimization problem for OLS minimizes the loss function x^2 evaluated at the residuals. The sample analogue, which is actually used for estimation, simply replaces the expectation with the average, so that the estimator of β_τ is given by

$$\hat{\beta}_\tau = \arg\min_\beta \frac{1}{N}\sum_{i=1}^{N}(\rho_\tau(Y_i - X\beta)),$$

where N is the number of observations. This minimization problem can be solved as a linear programming problem using the simplex method.

Results

I perform a 0.25, 0.5, and 0.75 quantile regression using the new personal vehicles dataset. To speed the computation of the LP problem, I run the quantile regression on a 10 percent subsample of 0.46 million observations. I use the linear specification given in Column (6) of Table 3.1, which had a nearly identical result to the result in Column (7) that included model fixed effects.[10]

The results of the quantile regression estimation are given in Table 3.13. For reference, when I run the fixed effects regression given by the specification in Column

[10] I could include vehicle model fixed effects by manually de-meaning all of the variables in the dataset by model first (i.e., use the within transformation) and then running the quantile regression on the transformed dataset, but I did not see this as worth the extra effort, since the OLS results in Columns (5) and (6) of Table 3.1 were so close.

(6) of Table 3.1 on this 10 percent subsample, I find a coefficient on the gasoline price of -72 miles per month (recall the coefficient in Table 3.1 was -70). The median (0.5) quantile regression result gives a similar result, with an estimated coefficient on the gasoline price of -73. The fact that the OLS (conditional mean) result and the quantile (conditional median) result are so close is encouraging, and suggests that outliers are not driving the results. This may not be surprising considering how large the sample is. I posit that if the 0.5 quantile regression was run on the entire sample of 4.6 million vehicles, then we would find that the median would be nearly identical to mean, for the influence of outliers would be mitigated by the sample size.

The 0.2, 0.25, 0.75, and 0.9 quantile regression results in Table 3.13 map out the heterogeneity in responsiveness. For example, we can interpret the 0.1-quantile regression coefficient on the gasoline price as suggesting that if the gasoline price increases by one dollar, then the 0.1 quantile of the population of drivers of new vehicles in the first six years will decrease driving by 95 miles per month. The 0.9-quantile regression result suggests that if the gasoline price is increased by one dollar, then the 0.9 quantile of this population will decrease driving by 48 miles per month. The corresponding elasticities at the means range from -0.23 for the 0.1 quantile to -0.11 for the 0.9 quantile. This range provides a useful starting point into quantifying the heterogeneity, but it does not help to answer the question of *who* is more or less responsive.

3.2.2 Heterogeneity by Vehicle Type

To begin examining which drivers are more or less responsive, we can begin by looking at heterogeneity in responsiveness based on the type of vehicle and the fuel economy of the vehicle. Are consumers who drive gas guzzlers and other lower fuel economy vehicles more or less responsive? On one hand, we might think that they would be more responsive because the fuel cost of driving is a greater fraction of the total cost of driving, perhaps making changes in gasoline prices more salient to these drivers. On the other hand, consumers who drive low fuel economy vehicles may be wealthier and perhaps have more more inelastic driving needs with such vehicles. The net of

Table 3.13: Intensive Margin Regressions: Quantile Regression

Dependent variable: vehicle-miles-traveled per month (mean = 1,089 miles per month)

	(1) 0.10 quantile	(2) 0.25 quantile	(3) 0.50 quantile	(4) 0.75 quantile	(5) 0.90 quantile
avg gasoline price	-95.71***	-84.94***	-73.00***	-60.05***	-48.24***
	(6.32)	(5.29)	(5.03)	(6.21)	(9.15)
M < 58	97.53***	93.54***	94.13***	111.35***	150.94***
	(3.15)	(2.63)	(2.50)	(3.10)	(4.58)
57 > M > 63	73.74***	73.79***	66.99***	75.91***	94.82***
	(4.04)	(3.38)	(3.21)	(3.97)	(5.86)
62 > M > 70	51.94***	51.76***	53.78***	57.84***	67.92***
	(3.43)	(2.87)	(2.73)	(3.38)	(4.98)
73 > M > 82	46.68***	63.01***	75.29***	79.65***	82.51***
	(2.97)	(2.50)	(2.38)	(2.95)	(4.36)
81 > M > 86	-7.17*	-8.35**	-16.28***	-32.02***	-45.82***
	(3.41)	(2.86)	(2.73)	(3.38)	(5.01)
M > 86	72.40***	76.53***	65.83***	34.85***	-4.14
	(8.24)	(6.88)	(6.54)	(8.08)	(11.94)
avg unempl rate	-5.95***	-6.10***	-5.35***	-1.23	3.83**
	(0.80)	(0.68)	(0.66)	(0.83)	(1.21)
avg CCI	0.15	0.77**	1.11***	1.03***	1.71***
	(0.30)	(0.25)	(0.24)	(0.30)	(0.44)
avg housing prices	-0.07***	-0.08***	-0.09***	-0.07***	-0.07***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
% summer months	90.19	30.24	74.56	152.86	374.79**
	(96.98)	(80.58)	(76.88)	(96.04)	(142.71)
zip density	-3.61***	-5.10***	-6.42***	-7.65***	-8.29***
	(0.21)	(0.18)	(0.18)	(0.23)	(0.35)
zip businesses/cap	-3.81	-31.63***	-30.33***	-24.19**	-4.65*
	(2.17)	(2.77)	(4.33)	(8.63)	(2.17)
log(zip population)	-8.45***	-13.20***	-19.28***	-21.97***	-24.52***
	(1.62)	(1.36)	(1.31)	(1.64)	(2.44)
zip pop growth rate	2.16***	2.80***	3.02***	4.52***	6.22***
	(0.39)	(0.33)	(0.30)	(0.36)	(0.50)
log(zip income)	-2.73	-16.64***	-36.25***	-63.72***	-83.18***
	(4.32)	(3.59)	(3.38)	(4.13)	(6.03)
commute time	1.27***	2.78***	4.83***	7.54***	10.03***
	(0.24)	(0.20)	(0.19)	(0.24)	(0.35)
liters	-18.47***	-14.21***	-10.20***	-3.22	2.70
	(2.54)	(2.07)	(1.91)	(2.30)	(3.36)
cylinders	-9.11***	-11.00***	-11.55***	-13.10***	-13.91***
	(1.63)	(1.35)	(1.27)	(1.54)	(2.26)
turbo	9.09	-6.39	-25.58***	-43.20***	-69.04***
	(6.19)	(5.16)	(4.87)	(5.98)	(8.76)
auto transmission	-24.09***	-29.14***	-30.39***	-32.03***	-31.86***
	(3.37)	(2.81)	(2.68)	(3.31)	(4.90)
gross veh weight	-6.98***	-6.38***	-7.00***	-8.99***	-7.02***
	(1.29)	(1.06)	(0.98)	(1.20)	(1.77)
all-wheel drive	-6.03*	-1.82	3.74	7.04*	4.79
	(2.98)	(2.50)	(2.37)	(2.92)	(4.30)
safety rating	-10.03***	-14.12***	-18.43***	-23.61***	-24.81***
	(2.31)	(1.93)	(1.84)	(2.27)	(3.35)
import	29.87***	21.24***	15.14***	6.77**	-6.34
	(2.76)	(2.28)	(2.13)	(2.61)	(3.83)
fuel economy	-3.14***	-1.01*	1.36**	3.92***	8.22***
	(0.58)	(0.48)	(0.45)	(0.55)	(0.82)
constant	1023.51***	1282.49***	1628.90***	2064.38***	2211.90***
	(73.58)	(61.36)	(58.17)	(71.73)	(105.68)
month-of-year FE	Y	Y	Y	Y	Y
lease, race & age	Y	Y	Y	Y	Y
veh body & class	Y	Y	Y	Y	Y
Observations	0.46m	0.46m	0.46m	0.46m	0.46m

Heteroskedasticity-robust standard errors in parentheses

*** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level

"M" refers to the months between registration and the first smog test these is an empirical question.

To examine heterogeneity by fuel economy and vehicle type, I again use Column (6) in Table 3.1 as my baseline specification. I show the results using both the full subsample of new personal vehicles, and the subsample of those who had "normal" six year interval between registration and smog test. The results are clear: drivers of lower fuel economy vehicles tend to be more responsive to changes in gasoline prices. Table 3.14 presents these results.

Columns (1) and (4) in Table 3.14 present the same results as Column (6) in Table 3.1 and Column (5) in Table 3.2 for reference. Columns (2) and (5) interact the gasoline price with fuel economy. The results indicate that the responsiveness is a decreasing function of (i.e., less negative) of the fuel economy. In other words, drivers of lower fuel economy vehicles tend to be more responsive to changing gasoline prices. Interestingly, once this interaction with the gasoline price is controlled for, we see that those who drive higher fuel economy vehicles tend to drive less. One reason for this might simply be that drivers of smaller, higher fuel economy vehicles tend to live in more densely populated areas that require less driving. Looking at the data, this is clearly the case: the average population density in the dataset for vehicles with fuel economy greater than the subsample harmonic mean of 18.6 miles per gallon is 5.4 thousand people per mile squared, while the average density for vehicles with fuel economy less than the mean is 4.5 thousand people per mile squared.[11]

Columns (1),(2),(4), and (5) also indicate how much vehicles of different vehicle classes are driven on average, relative to small cars, which is the omitted vehicle class. The results imply that large vehicles are driven slightly more, sporty vehicles are driven much less, luxury vehicles are driven less, pickups are driven about the same, full pickups are driven more, sport utility vehicles are driven less, and both full utility vehicles and minivans are driven more. These results are useful to keep in mind in interpreting the results in Columns (3) and (6).

Columns (3) and (6) in Table 3.14 include interactions between the gasoline price and vehicle classes to examine the heterogeneity in responsiveness across vehicle classes. Table 3.15 summarizes the responsiveness by vehicle class and provides

[11] The subsample being referred to here is the 4.65 million subsample for which I observe VMT.

Table 3.14: Heterogeneity in Driving Responsiveness by Vehicle Class

Dependent variable: vehicle-miles-traveled per month (mean = 1,089 miles per month)

	Full Personal Vehicle Sample			Six Year Subsample		
	(1)	(2)	(3)	(4)	(5)	(6)
	base	interact fuel econ	interact veh class	base	interact fuel econ	interact veh class
avg gasoline price	-69.6***	-212.5***	-37.9***	-99.7***	-232.3***	-79.6***
	(1.5)	(5.0)	(2.9)	(3.6)	(7.9)	(5.2)
fuel economy	2.6***	-17.0***	2.2***	2.8***	-14.9***	2.4***
	(0.1)	(0.7)	(0.1)	(0.2)	(1.0)	(0.2)
gaspr*fuel economy		7.3***			6.6***	
		(0.2)			(0.4)	
Large Cars	20.3***	19.6***	-88.4***	5.8***	4.9***	-140.9***
	(0.9)	(0.9)	(10.6)	(1.4)	(1.4)	(15.1)
Sporty Cars	-38.1***	-37.6***	269.8***	-29.4***	-29.0***	250.9***
	(1.3)	(1.3)	(15.3)	(2.0)	(2.0)	(24.1)
Prestige Sporty	-279.0***	-279.2***	-324.2***	-292.4***	-292.7***	-358.6***
	(2.3)	(2.3)	(22.1)	(3.5)	(3.5)	(35.3)
Luxury	-37.1***	-38.2***	20.3	-49.2***	-50.2***	-22.6
	(1.1)	(1.1)	(11.1)	(1.7)	(1.7)	(16.7)
Prestige Luxury	-76.2***	-76.5***	-33.0	-86.7***	-86.8***	-16.6
	(1.7)	(1.7)	(17.9)	(2.6)	(2.6)	(27.6)
Pickup	1.5	-3.1	209.9***	29.1*	22.9	261.8***
	(9.6)	(9.6)	(17.8)	(14.7)	(14.7)	(26.1)
Full Pickup	92.3***	88.1***	398.9***	119.1***	113.3***	412.4***
	(9.5)	(9.5)	(15.7)	(14.7)	(14.7)	(23.3)
Sport Utility	-36.8***	-37.0***	4.1	-46.1***	-46.6***	-47.8**
	(2.7)	(2.7)	(10.0)	(4.0)	(4.0)	(14.8)
Full Utility	62.7***	63.4***	322.2***	55.5***	55.9***	281.3***
	(2.8)	(2.8)	(12.4)	(4.3)	(4.3)	(18.1)
Minivan	194.9***	194.1***	279.0***	191.6***	190.8***	256.3***
	(2.0)	(2.0)	(14.7)	(3.0)	(3.0)	(20.7)
gaspr*Large Cars			40.1***			54.2***
			(3.9)			(5.6)
gaspr*Sporty Cars			-115.4***			-105.3***
			(5.7)			(9.0)
gaspr*Prestige Sporty			16.5*			24.3
			(8.2)			(13.0)
gaspr*Luxury			-22.0***			-10.6
			(4.1)			(6.2)
gaspr*Prestige Luxury			-16.6*			-26.5**
			(6.6)			(10.2)
gaspr*Pickup			-80.2***			-90.3***
			(5.6)			(8.0)
gaspr*Full Pickup			-116.2***			-112.3***
			(4.6)			(6.7)
gaspr*Sport Utility			-15.3***			0.4
			(3.6)			(5.3)
gaspr*Full Utility			-96.0***			-83.8***
			(4.5)			(6.5)
gaspr*Minivan			-31.6***			-24.5**
			(5.4)			(7.6)
constant	1,558.8***	1,943.4***	1,476.7***	1,500.1***	1,867.3***	1,459.0***
	(17.5)	(21.7)	(18.8)	(35.1)	(40.1)	(36.4)
veh characteristics	Y	Y	Y	Y	Y	Y
veh body	Y	Y	Y	Y	Y	Y
demographics	Y	Y	Y	Y	Y	Y
month-of-year FE	Y	Y	Y	Y	Y	Y
econ conditions	Y	Y	Y	Y	Y	Y
summer control	Y	Y	Y	Y	Y	Y
R-squared	0.065	0.065	0.066	0.056	0.056	0.057
Observations	4.7m	4.7m	4.7m	2.2m	2.2m	2.2m

Heteroskedasticity-robust standard errors in parentheses

*** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level

the corresponding elasticities. For reference, it also includes the harmonic mean fuel economy by vehicle class in the subsample for which I observe VMT. These estimates are based on Column (3) in Table 3.14 and I use the delta method to calculate the standard errors and statistical significance.

Table 3.15: Summary of Responsiveness in Driving by Vehicle Class

Dependent variable: vehicle-miles-traveled per month (mean = 1,089 miles per month)

	Vehicles (000s)	fuel economy (mi/gal)	coefficient	s.e.	elasticity at means	s.e.
All Vehicle Classes	4,652	18.6	-69.6***	(1.5)	-0.172***	(0.004)
Small Cars	771	26.4	-37.9***	(2.9)	-0.094***	(0.007)
Large Cars	723	21.9	2.2	(3.0)	0.005	(0.007)
Sporty Cars	198	20.7	-153.3***	(5.1)	-0.379***	(0.013)
Prestige Sporty	60	18.5	-21.4***	(7.7)	-0.053***	(0.019)
Luxury	444	19.9	-59.9***	(3.2)	-0.148***	(0.008)
Prestige Luxury	99	17.7	-54.5***	(6.1)	-0.135***	(0.015)
Pickup	287	17.4	-118.1***	(5.0)	-0.292***	(0.012)
Full Pickup	523	14.7	-154.1***	(3.9)	-0.381***	(0.010)
Sport Utility	914	17.3	-53.2***	(2.5)	-0.132***	(0.006)
Full Utility	395	13.6	-134.0***	(3.7)	-0.331***	(0.009)
Minivan	238	18.3	-69.5***	(4.8)	-0.172***	(0.012)

Heteroskedasticity-robust standard errors in parentheses
*** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level

The coefficients on the interactions in both Columns (3) and (6) indicate that there is considerable heterogeneity across vehicle classes. Relative to small cars, most vehicle classes are more responsive. Full utility vehicles and full pickups are the lowest fuel economy vehicle classes, and are two of the most responsive vehicle classes. The responsiveness for each of these vehicle classes (-134 miles per month for full utility and -154 for full pickup) is over three times greater than the responsiveness for small cars (-38 miles per month). Interestingly, drivers of sporty cars are also relatively more responsive. For full utilities, full pickups, and sporty vehicles, we might expect the drivers to also own another vehicle that has higher fuel economy, so some of the difference in responsiveness may be due to a switching between vehicles. It is also possible that the large SUVs and pickups are often used for recreation trips and when gasoline prices rise dramatically, consumers cut back on such trips.

Interestingly, the responsiveness of prestige sporty vehicles tends not to be statically significantly different than small cars, which may relate to the much greater wealth of drivers of prestige sporty vehicles. Large cars tend to be relatively less responsive than small cars. One explanation for this is that many households may only have two vehicles, where one is a large car and the other an SUV or pickup. For these households the within-household switching will likely be to the large car. In addition, drivers of large cars are slightly more likely to be in rural areas (the zip code density is slightly greater for small car drivers), so it is plausible that drivers of large cars do not have as many options to reduce driving as drivers of small cars. Later in this section, I explore the heterogeneity in responsiveness by geography and show that consumers who live in different places in California respond differently to changes in gasoline prices.

Table 3.15 indicates that the estimated responsiveness for each of the vehicle classes is statistically significant. In addition, we can examine whether the *differences* in the responsiveness are statistically significant. It is clear from the t-statistics of the interaction terms that small cars have a statistically significantly different responsiveness than nearly all other vehicle classes at the one percent confidence level. The only exceptions are prestige sporty and prestige luxury.

We may also be interested in how the responsiveness varies by different characteristics of the vehicle. Table 3.16 presents the estimation results when all of the other vehicle characteristics are interacted with the gasoline price. I include two specifications, both again based on Column (6) of Table 3.1. Column (1) replicates Column (6) of Table 3.1 for reference. Column (2) includes the interactions between the vehicle characteristics and the gasoline price, along with the vehicle class and vehicle body fixed effects that are included in Column (1). Column (3) removes these vehicle class and vehicle body fixed effects in case these are confounding the interactions of interest.

Perhaps the most interesting of these interactions is the coefficient on the gross vehicle weight rating, for heavier vehicles on the road impose an externality on smaller vehicles by increasing the probability of a fatality of an occupant in the smaller vehicle in the event of an accident (White 2004; Anderson 2008; Jacobsen 2010). The coefficient on this interaction is statistically significant to the one percent confidence level and appears to be at least somewhat economically significant as well.[12] The sign of the coefficient is negative, suggesting that heavier vehicles are more responsive to

[12]The units on the gross vehicle weight rating are in thousands of pounds, with a mean of 5.4.

Table 3.16: Heterogeneity in Driving Responsiveness by Vehicle Characteristics

Dependent variable: VMT per month (mean = 1,089 mi per mon)

	(1) base	(2) interact veh chars	(3) no veh class or body FE
avg gasoline price	-69.6***	-18.1	42.3***
	(1.5)	(12.1)	(12.0)
fuel economy	2.6***	-2.5***	2.3***
	(0.1)	(0.1)	(0.1)
liters	-5.8***	189.5***	95.4***
	(0.6)	(6.9)	(6.7)
cylinders	-11.1***	-99.2***	-27.7***
	(0.4)	(4.6)	(4.6)
turbo	-29.3***	51.2**	106.6***
	(1.2)	(15.7)	(15.4)
auto transmission	-31.7***	-134.9***	-186.1***
	(0.8)	(10.4)	(10.1)
gross veh weight	-6.5***	20.1***	3.5
	(0.3)	(3.9)	(3.8)
all-wheel drive	2.5***	-8.3	-17.1*
	(0.7)	(8.3)	(8.2)
safety rating	-15.0***	-16.0*	21.2**
	(0.5)	(6.8)	(6.7)
import	16.7***	36.9***	3.5
	(0.6)	(8.1)	(8.0)
gaspr*liters		-64.2***	-37.7***
		(2.5)	(2.5)
gaspr*cylinders		29.3***	6.0***
		(1.7)	(1.7)
gaspr*turbo		-40.2***	-50.8***
		(5.8)	(5.7)
gaspr*auto transmission		51.0***	57.6***
		(3.9)	(3.7)
gaspr*gross veh weight		-7.1***	-3.8**
		(1.4)	(1.4)
gaspr*all-wheel drive		13.4***	7.2*
		(3.1)	(3.0)
gaspr*safety		-3.1	-13.4***
		(2.5)	(2.5)
gaspr*import		-11.6***	5.0
		(3.0)	(3.0)
avg gas price	-69.6***	-18.1	42.3***
	(1.5)	(12.1)	(12.0)
fuel economy	2.6***	-2.5***	2.3***
	(0.1)	(0.1)	(0.1)
constant	1,558.8***	1,866.5***	1,262.9***
	(17.5)	(36.9)	(36.5)
veh body & class	Y	Y	N
demographics	Y	Y	Y
month-of-year FE	Y	Y	Y
econ conditions	Y	Y	Y
summer control	Y	Y	Y
R-squared	0.065	0.042	0.065
Observations	4.7m	4.7m	4.7m

Heteroskedasticity-robust standard errors in parentheses

*** indicates significant at 1% level, ** significant at 5% level

changes in gasoline prices. We may have suspected this result already from the earlier results showing that the heavier vehicle classes (e.g., full utility vehicles and full pickups) are more responsive. This result is encouraging from a policy standpoint, for it suggests that policies that raise the price of gasoline may have additional safety co-benefits from leading to less driving of some of the heaviest vehicles.

The interaction with engine size (cylinders) indicates that vehicles with larger engines are less responsive, a result perhaps relating to the relative non-responsiveness of the prestige luxury and prestige sporty vehicle classes. The interaction with engine displacement (liters) suggests that vehicles with greater engine displacement are more responsive, which likely relates to the larger, more powerful SUVs and trucks being more responsive. Similarly, the interaction with turbo suggests that vehicles that have a turbocharger are relatively more responsive, a result likely driven by the high responsiveness of sporty cars.

3.2.3 Heterogeneity by Zip Code Demographics

If we are interested in the distributional consequences of a policy that increases gasoline prices, there are two factors to consider: the amount that vehicles are driven and the responsiveness to changes in gasoline prices. If the response is fairly inelastic, then the amount driven should be the dominant factor, as will be discussed in more detail in Chapter 5. However, the responsiveness may vary by different groups, which may influence the burden of a policy. Thus, if we are concerned about the impact of increases in gasoline prices on the poor, we should first quantify how much the poor drive, and then quantify how the poor respond to changing gasoline prices. This subsection focuses on heterogeneity in responsiveness by zip code demographics, from the census. This includes zip code median household income, but does not include the vehicle-level income variable included in the R.L. Polk data.

To analyze the heterogeneity in a variety of zip code demographics, I interact each demographic with the gasoline price. Table 3.17 contains these results. Column (1) replicates Column (5) in Table 3.1 for reference, while Column (2) adds interactions between the demographic variables and the gasoline price. I do not include interactions for geographic metrics, such as the density of the zip code, zip code population, zip code population growth rate, zip code businesses per capita, and commute times.

The results in Table 3.17 suggest that there is some heterogeneity across the demographics presented here. To begin, from Column (1), we can see that vehicles registered in higher income zip codes are driven less. This does not mean that households in higher income zip codes drive less, it just means that each individual vehicle owned by vehicles in higher income zip codes is driven less, perhaps because wealthier households own more vehicles. From Column (2), we can see that vehicles registered in higher income zip codes are more responsive to gasoline price changes. This may seem surprising, but there are several reasonable explanations. First, since wealthier households are more likely to own several vehicles, they may have more opportunities to switch to a higher fuel economy vehicle. Wealthier households may also be able to afford, and be more inclined to switch to, air travel for some trips. Finally, wealthier households may take some trips with a low marginal utility, such as pleasure-rides with a new vehicle; when gasoline prices rise significantly, it may be low cost not to partake in these trips. In the next subsection, I explore the heterogeneity by income in much more detail.

The leased vehicles coefficients indicates that leased vehicles are driven more and are slightly more responsive. Consumers who lease a vehicle tend to be wealthier (mean zip code household income of $75,000 for leased vehicles versus $70,000 for non-leased vehicles). The difference in household income is statistically significant to a one percent level. Of course, unobserved characteristics about those who lease vehicles undoubtedly play a role as well.

The population over 65 coefficients suggest that vehicles registered in zip codes with a higher percentage of the population greater than 65 years old tend to have a less responsiveness than those in zip codes with a lower percentage. However, this coefficient is only statistically significantly to a five percent level, suggesting that perhaps vehicles driven by these households are not very different in responsiveness than the rest of the vehicles. On the other side of the coin, vehicles registered in zip codes with a higher percentage of the population under 18 years old tend to be more

Table 3.17: Heterogeneity in Driving Responsiveness by Demographics

Dependent variable: VMT/month (mean = 1,089 mi/mon)

	(1) base	(2) interact with demographics
avg gasoline price	-69.6***	161.4***
	(1.5)	(47.4)
lease	7.2***	106.5***
	(0.6)	(7.2)
log(zip income)	-39.4***	-5.2
	(1.0)	(11.1)
zip % pop age 65+	-3.7***	-5.2***
	(0.1)	(0.7)
zip % pop under 18	4.1***	11.5***
	(0.1)	(0.7)
zip % pop white	0.4***	1.3***
	(0.0)	(0.2)
zip % pop black	-0.2***	1.1*
	(0.0)	(0.5)
zip % pop hispanic	0.1**	-0.3
	(0.0)	(0.3)
gaspr*lease		-37.4***
		(2.7)
gaspr*log(zip inc)		-12.7**
		(4.1)
gaspr*pop age 65		0.6*
		(0.3)
gaspr*pop age 18		-2.7***
		(0.3)
gaspr*white pop		-0.3***
		(0.1)
gaspr*black pop		-0.5**
		(0.2)
gaspr*hispanic pop		0.1
		(0.1)
constant	1,558.8***	928.5***
	(17.5)	(128.4)
geographic demogs	Y	Y
veh body & class	Y	Y
vehicle char	Y	Y
month-of-year FE	Y	Y
econ conditions	Y	Y
summer control	Y	Y
R-squared	0.065	0.042
Observations	4.7m	4.7m

Heteroskedasticity-robust standard errors in parentheses

*** indicates significant at 1% level, ** significant at 5% level

responsive. Finally, there is no strong pattern of a difference in elasticities by the percentage of the population being of different races. If I test the difference between the responsiveness of the different zip code race variables, I cannot reject the null that they are the same. These age and race zip code-level demographic results are not as useful as household-level demographic data may be, for inherently there is measurement error due to any within-zip code heterogeneity. Future work to acquire individual household-level demographics and match these to vehicles could greatly enhance this analysis.[13]

3.2.4 Heterogeneity by Income

Household income is perhaps the most interesting vehicle-level characteristics that I observe. Unfortunately, the income variable is not complete, and so I must consider the income subsample separately. Despite this, we may be particularly interested in understanding the heterogeneity in responsiveness by income for policy development purposes, so it is worthwhile to examine this subsample. I observe 2.3 million vehicles that include the household income categorical variable and an odometer reading. Of these, 48 percent were registered in 2003, 23 percent in 2002, 17 percent in 2001, and the remainder in 2004. Thus, this subsample is by its very nature over-representing 2003, so we should be at least somewhat careful in interpreting the results from this subsample.

To begin examining the heterogeneity by income, I interact the household income categorical variable with the gasoline price to see how vehicles driven by households of different incomes respond. The results are shown in Table 3.18. Column (1) again presents the results from Column (6) in Table 3.1 for reference. Column (2) presents exactly the same specification, only restricted to the subsample where I observe income. Column (3) adds indicator variables for the different income groups. Recall that in Figure 2.7 in Chapter 2, we can see that the distribution of the categorical income variable in the data seems to make sense relative to the distribution of income in the population in California. Column (4) includes interactions between the income

[13] The full DMV data may be very useful for such an endeavor.

category indicators and the price of gasoline.

The results in Table 3.18 first indicate that the subsample where I observe income is different than the 4.7 million vehicle sample of new personal vehicles. Specifically, we see that the responsiveness to gasoline prices is greater than in the full new personal vehicle sample. This may not be surprising, since the income subsample is more heavily weighted towards 2003, when there were higher gasoline prices and more significant changes in gasoline prices. More importantly, I find that when I include the income controls to the previous specification, there is very little change to the coefficient on the gasoline price. This suggests that the zip code demographics do a reasonable job at capturing income-specific unobservables that could be correlated with gasoline prices.

The coefficients on the income indicator variables in Column (3) indicate how much new vehicles are driven by households in each of the income groups. The coefficients are all relative to the lowest income category, "$<$ 15k" per year. The results suggest that driving per vehicle increases with income until the we reach the highest income category, when it drops again slightly. The decrease in per vehicle driving for the highest income category may due at least in part to the fact that the wealthiest tend to own more vehicles.

The coefficients on the interactions between the gasoline price and the income indicator variables in Column (4) provide a sense for the differences in how drivers of vehicles owned by households with different income respond. To further clarify, Table 3.19 summarizes the responsiveness by income category and provides the corresponding elasticities. These estimates are calculated from Column (4) in Table 3.18 and I use the delta method to calculate the standard errors and statistical significance.

The first row in Table 3.19 shows the estimated responsiveness from Column (3) of Table 3.18 where the estimation was performed on the income subsample without including interactions between income and the gasoline price. Interestingly, when the interactions are added, the estimated responsiveness for all income groups is considerably lower.[14] It is the least for vehicles driven by new vehicle purchasers from the lowest income group, which have nearly an entirely inelastic response. This

[14] I do not currently have an explanation for this, although it is a curious result.

Table 3.18: Heterogeneity in Driving Responsiveness by Income

Dependent variable: VMT per month (mean = 1,089 mi per mon)

	(1) base	(2) income subsample	(3) income controls	(4) income inter
avg gasoline price	-69.6***	-93.2***	-94.9***	-9.0
	(1.5)	(2.4)	(2.4)	(7.5)
fuel economy	2.6***	3.0***	2.9***	2.9***
	(0.1)	(0.2)	(0.2)	(0.2)
$15k - $20k			8.7***	107.1**
			(2.2)	(39.5)
$20k - $30k			4.4**	161.1***
			(1.6)	(28.5)
$30k - $40k			4.4**	219.2***
			(1.6)	(27.7)
$40k - $50k			9.5***	253.1***
			(1.5)	(27.0)
$50k - $75k			17.4***	255.6***
			(1.4)	(23.3)
$75k - $100k			26.7***	334.8***
			(1.5)	(24.3)
$100k - $125k			26.1***	372.4***
			(1.6)	(26.9)
>125k			19.1***	317.8***
			(1.6)	(24.8)
gaspr*($15k - $20k)				-17.8*
				(7.1)
gaspr*($20k - $30k)				-18.7***
				(3.4)
gaspr*($30k - $40k)				-19.3***
				(2.5)
gaspr*($40k - $50k)				-17.5***
				(1.9)
gaspr*($50k - $75k)				-14.3***
				(1.4)
gaspr*($75k - $100k)				-15.8***
				(1.2)
gaspr*($100k - $125k)				-15.5***
				(1.2)
gaspr*(>125k)				-11.9***
				(1.0)
constant	1,558.8***	1,652.2***	1,736.8***	1,496.0***
	(17.5)	(23.9)	(24.4)	(31.6)
demographics	Y	Y	Y	Y
veh body & class	Y	Y	Y	Y
vehicle char	Y	Y	Y	Y
month-of-year FE	Y	Y	Y	Y
econ conditions	Y	Y	Y	Y
summer control	Y	Y	Y	Y
R-squared	0.065	0.065	0.066	0.066
Observations	4.7m	2.3m	2.3m	2.3m

Heteroskedasticity-robust standard errors in parentheses

*** indicates significant at 1% level, ** significant at 5% level

Table 3.19: Summary of Responsiveness in Driving by Income

Dependent variable: vehicle-miles-traveled per month (mean = 1,089 miles per month)

	Vehicles (000s)	coefficient	s.e.	elasticity at means	s.e.
All Income	2,332	-94.9***	(2.4)	-0.244***	(0.006)
<15k	165	-9.0	(7.5)	-0.023	(0.019)
$15k - $20k	67	-26.8***	(7.2)	-0.069***	(0.018)
$20k - $30k	172	-27.7***	(5.6)	-0.071***	(0.014)
$30k - $40k	196	-28.3***	(7.7)	-0.072***	(0.015)
$40k - $50k	223	-26.5***	(6.2)	-0.068***	(0.016)
$50k - $75k	568	-23.3***	(6.4)	-0.060***	(0.016)
$75k - $100k	391	-24.8***	(5.0)	-0.064***	(0.017)
$100k - $125k	206	-24.6***	(6.7)	-0.063***	(0.017)
>125k	344	-20.9***	(6.7)	-0.054***	(0.017)

Heteroskedasticity-robust standard errors in parentheses
*** indicates significant at 1% level

response may be related to the fact that consumers who make less than $15,000 per year and yet still purchase a new vehicle are likely to be unusual. They may have the vehicle purchased for them (e.g., from parents), and perhaps even have the gasoline purchased for them. They may only make $15,000 per year, but rely on previously earned wealth to purchase the vehicle. Alternatively, although perhaps less likely, they may have a very strong need to purchase a particular type of new vehicle and to drive, and thus would quite inelastic in driving responsiveness. Without further information, I can not disentangle these effects. Accordingly, I view the results for the lowest income category with caution.

What can be made of the results? As income increases, the responsiveness appears to slightly increase, level off, and finally drop slightly with the highest income bracket. In one sense this runs counter to the “U-shaped” pattern found in West (2004),Wadud, Graham, and Noland (2009), given the ups and downs in the responsiveness as income increases. However, in another sense they correspond to this U-shaped pattern. If we ignore the lowest income category and consider all vehicles purchased by households with income between $15k and $40k as the most budget constrained, we see that the vehicles purchased by these households are the most responsive. Then responsiveness declines somewhat to the $50k to $75k income range. Using a standard t-test, the difference between the elasticity in the $30k to $40k range is statistically significantly different than the elasticity in the $50k to $75k income range at the one percent confidence level.

The elasticity appears to to increase again for households in the \$75k to \$125k range, but the difference between the elasticity of the \$50k to \$75k income range and the elasticity of the \$75k to \$125k income range is only statistically significantly different at the ten percent confidence level using a two sided t-test. Using a one-sided t-test, the \$75k to \$125k income category is statistically significantly greater than the \$50k to \$75k income category at the five percent confidence level. Household in the \$75k to \$125k are more likely to have multiple vehicles, and thus can switch between vehicles. In addition, households in these groups may have less need to respond to higher gasoline prices, yet it may also be easier for them to do so, as some of their driving may have low marginal utility. Households in the \$75k to \$125k range are also larger households on average, and accordingly have more options at their disposal for arranging their travel more efficiently (e.g., through ride sharing). In addition, there is the possibility of substitution by air travel for long distance driving for holiday trips.

The very highest income group has the least responsiveness, and is statistically significantly less than all of the others at a one percent level, using a one-sided t-test. This result may simply be because the cost of gasoline is such a small fraction of the household budget for the wealthiest households that changes in gasoline price are less noticeable and worrisome.

Of course, from a bigger picture, these results are all conditional on a new vehicle being purchased. To gain a more full picture of how the responsiveness in driving to gasoline price changes varies by income, we should also be interested in the amount used vehicles are driven when gasoline prices change. Unfortunately, this dataset does not allow me to examine the responsiveness by income for this larger population of drivers. That said, these results are novel to the literature as being perhaps the first non-survey evidence of the heterogeneity in driving by income.

3.2.5 Heterogeneity by Geography

To understand the implications of a policy that changes gasoline prices for local air pollution and congestion, we can examine the heterogeneity in driving responsiveness by geography. I am using the term geography here broadly in order to include characteristics of the location, as well as simply the physical location itself.

To begin, I interact several of the zip code demographic variables that are more geographic in nature with the gasoline price. Table 3.20 presents these results. Again, Column (1) replicates Column (6) in Table 3.1 for reference. Column (2) includes the interactions.

The interaction coefficients in Table 3.20 sheds some light on how the responsiveness differs for consumers who live in different places. Both Columns (1) and (2) indicate that higher population density is associated with less driving. This is not surprising, for more densely populated areas are both more congested and have better access to public transportation. Interestingly, the coefficient in Column (2) of the interaction between the population density and the gasoline price suggests that consumers in more densely populated areas are *less* responsive.[15] One might have expected the opposite result, where rural areas are the least responsive and urban areas are the most responsive. Given this, I perform a variety of specification checks – yet do not find a change in this result. One way to view this result is that in most cities in California (with the exception of San Francisco), the public transportation options are relatively limited. For example, much of Los Angeles is fairly densely population, and yet has poor public transportation.

The coefficient on the interaction of commute time and the gasoline price is also somewhat surprising. Vehicles in counties with longer commute times appear to be relatively more responsive. Counties with longer commute times tend to be in the Los Angeles area and in the outskirts of the Bay Area, so these counties must have a combination of characteristics that lead to greater responsiveness. If county-level economic conditions were not being controlled for, we might be concerned that the

[15]Note that in my working paper "Identifying the Elasticity of Driving: Evidence from a Gasoline Price Shock in California," the cluster analysis results suggested that more urban areas were more responsive, but other factors also played a role in the cluster results.

Table 3.20: Heterogeneity in Driving Responsiveness by Geographic Variables

Dependent variable: VMT/month (mean = 1,089 mi/mon)

	(1) base	(2) interact geog vars
avg gasoline price	-69.6***	17.0
	(1.5)	(23.0)
fuel economy	2.6***	2.6***
	(0.1)	(0.1)
zip density	-6.6***	-15.7***
	(0.0)	(0.6)
zip businesses/cap	-29.5***	-71.5
	(8.0)	(105.8)
log(zip population)	-16.8***	-8.4
	(0.4)	(5.6)
zip pop growth rate	3.4***	6.4***
	(0.1)	(1.3)
commute time	5.4***	12.4***
	(0.1)	(0.7)
gaspr*density		3.4***
		(0.2)
gaspr*bus/cap		15.5
		(37.2)
gaspr*log(pop)		-3.1
		(2.0)
gaspr*pop growth rate		-1.1*
		(0.5)
gaspr*commute times		-2.6***
		(0.3)
constant	1,558.8***	1,329.8***
	(17.5)	(65.7)
demographics	Y	Y
veh body & class	Y	Y
vehicle char	Y	Y
month-of-year FE	Y	Y
econ conditions	Y	Y
summer control	Y	Y
R-squared	0.065	0.065
Observations	4.7m	4.7m

Heteroskedasticity-robust standard errors in parentheses
*** indicates significant at 1% level, * significant at 10% level

responsiveness to gasoline price changes is being confounded with the considerable economic malaise that many of these outlying counties felt after the housing bust. However, I believe that these should be adequately controlled for.

To better understand the heterogeneity in how drivers across counties in California respond differently to gasoline price changes, I regress VMT on the gasoline price, while allowing for a separate slope and intercept for each county. In this estimation, I do not control for demographics or vehicle characteristics, for these are characteristics of each county, and I am interested in how the counties differ *inclusive* of all of the characteristics of the counties. However, I can include controls to address certain possible selection biases: a percent summer months control, month-of-the-year fixed effects, and indicators for the length of the interval between smog checks.[16]

The resulting estimates are given in Table 3.21. The columns of this table are slightly different in order to fit all of the counties in California (except Alpine County, for which I do not observe gasoline prices). I run the estimation with Alameda county omitted, so that the coefficients for all of the other counties are relative to Alameda county. For the presentation in the table, I calculate the coefficient indicating the responsiveness for each county and the corresponding elasticity at the means. I calculate the standard errors for each elasticity using the delta method and perform a hypothesis test for statistical significance.

The results in Table 3.21 indicate some degree of heterogeneity across counties, although perhaps less than one might expect. The county with the highest responsiveness (Del Norte) is statistically significantly different from nearly all of the other counties at a one percent level. The same is true for the county with the lowest responsiveness (Colusa). However, when I perform a hypothesis test for the difference between each of the calculated elasticities and the (unweighted) mean responsiveness taken over all of the counties (-0.229), I find that none of the county-level elasticities are statistically significantly different from the mean. If Del Norte and Colusa county are excluded from the analysis, only some counties are statistically significantly different. For example, using a two-sided t-test, Mariposa is statistically significantly

[16] I find not including these additional controls seems to make little difference to the coefficients on the interactions between the county indicators and the gasoline price.

Table 3.21: Summary of Responsiveness in Driving by County

Dependent variable: VMT per month (mean = 1,089 mi/mon)

	Vehicles (000s)	coefficient	s.e.	elasticity at means	s.e.
Alameda	429	-91.42***	(5.82)	-0.226***	(0.014)
Amador	12	-98.85***	(19.26)	-0.244***	(0.047)
Butte	47	-97.87***	(6.48)	-0.242***	(0.016)
Calaveras	17	-101.04***	(11.34)	-0.250***	(0.028)
Colusa	6	-85.40***	(12.41)	-0.211***	(0.030)
Contra Costa	376	-91.31***	(5.08)	-0.226***	(0.012)
Del Norte	5	-109.69***	(11.13)	-0.271***	(0.027)
El Dorado	64	-93.27***	(5.56)	-0.231***	(0.013)
Fresno	207	-90.48***	(5.3)	-0.224***	(0.013)
Glenn	6	-95.54***	(7.75)	-0.236***	(0.019)
Humboldt	29	-92.56***	(6.00)	-0.229***	(0.014)
Imperial	56	-91.22***	(5.63)	-0.226***	(0.013)
Inyo	6	-93.68***	(7.77)	-0.232***	(0.019)
Kern	229	-93.43***	(5.46)	-0.231***	(0.013)
Kings	35	-94.15***	(5.62)	-0.233***	(0.013)
Lake	16	-88.43***	(6.29)	-0.219***	(0.015)
Lassen	7	-90.42***	(6.75)	-0.224***	(0.016)
Los Angeles	3358	-91.74***	(5.52)	-0.227***	(0.013)
Madera	32	-93.87***	(5.67)	-0.232***	(0.014)
Marin	98	-91.61***	(5.58)	-0.226***	(0.013)
Mariposa	5	-101.93***	(6.89)	-0.252***	(0.017)
Mendocino	24	-92.96***	(5.85)	-0.230***	(0.014)
Merced	54	-93.71***	(5.64)	-0.232***	(0.013)
Modoc	2	-86.81***	(9.85)	-0.215***	(0.024)
Mono	4	-92.19***	(6.74)	-0.228***	(0.016)
Monterey	109	-92.96***	(5.63)	-0.230***	(0.013)
Napa	45	-91.82***	(5.66)	-0.227***	(0.013)
Nevada	28	-92.92***	(5.69)	-0.230***	(0.014)
Orange	1357	-91.96***	(5.63)	-0.227***	(0.013)
Placer	124	-92.21***	(5.65)	-0.228***	(0.013)
Plumas	5	-90.05***	(6.31)	-0.223***	(0.015)
Riverside	782	-91.62***	(5.65)	-0.227***	(0.013)
Sacramento	382	-91.34***	(5.65)	-0.226***	(0.013)
San Benito	19	-92.54***	(5.75)	-0.229***	(0.014)
San Bernardino	700	-92.29***	(5.66)	-0.228***	(0.014)
San Diego	1117	-91.47***	(5.66)	-0.226***	(0.014)
San Francisco	192	-91.88***	(5.67)	-0.227***	(0.014)
San Joaquin	186	-92.98***	(5.68)	-0.230***	(0.014)
San Luis Obispo	84	-91.86***	(5.68)	-0.227***	(0.014)
San Mateo	246	-91.04***	(5.68)	-0.225***	(0.014)
Santa Barbara	127	-92.20***	(5.69)	-0.228***	(0.014)
Santa Clara	582	-91.18***	(5.69)	-0.225***	(0.014)
Santa Cruz	69	-91.75***	(5.70)	-0.227***	(0.014)
Shasta	46	-92.09***	(5.71)	-0.228***	(0.014)
Sierra	1	-94.20***	(7.53)	-0.233***	(0.018)
Siskiyou	11	-93.59***	(5.86)	-0.231***	(0.014)
Solano	151	-91.85***	(5.70)	-0.227***	(0.014)
Sonoma	146	-91.23***	(5.71)	-0.226***	(0.014)
Stanislaus	130	-92.49***	(5.71)	-0.229***	(0.014)
Sutter	24	-90.71***	(5.73)	-0.224***	(0.014)
Tehama	12	-90.75***	(5.76)	-0.224***	(0.014)
Trinity	3	-93.16***	(6.10)	-0.230***	(0.015)
Tulare	99	-91.92***	(5.72)	-0.227***	(0.014)
Tuolumne	16	-91.33***	(5.77)	-0.226***	(0.014)
Ventura	350	-91.42***	(5.72)	-0.226***	(0.014)
Yolo	51	-91.12***	(5.73)	-0.225***	(0.014)
Yuba	15	-90.66***	(5.76)	-0.224***	(0.014)

Heteroskedasticity-robust standard errors in parentheses
*** indicates significant at 1% level

different from Yuba county at the one percent confidence level, but is only statistically significantly different from Los Angeles county at the five percent confidence level. Most counties are statistically significantly different from each other at a ten percent confidence level, but some are not even statistically significant at that level (e.g., Los Angeles and Contra Costa).

Thus, we may wish to view the differences across counties cautiously, for many are not statistically significant, and the range of elasticities across counties is limited, so the economic significance is limited. Nevertheless, it may still be useful to see if there is a clear spatial pattern of responsiveness. Figure 3.1 shows the spatial distribution of elasticities. The legend makes clear the relatively limited range of responsiveness across counties. Despite the limited range in elasticities, it remains interesting to note that several of the more responsive counties are in the Central Valley. This is somewhat encouraging for policy since several of these counties are notoriously local air pollution non-attainment zones. However, the additional co-benefits in these counties relative to others may be limited, due to the limited range of elasticities. Note that counties in the far northern California, and counties in eastern California are not identified with as many observations, so we should place less faith in the estimated elasticities for these counties.

It is worth contrasting these results from previous results I had been finding in early work. For example, in the working paper "Identifying the Elasticity of Driving: Evidence from a Gasoline Price Shock in California," I find greater responsiveness in the Bay Area and less responsiveness in the Central Valley. I also find much greater heterogeneity between counties than I do in these results. The difference between the results is the result of including demographics, vehicle characteristics, and a variety of additional controls in the specification where I interacted the gasoline price with the county in the previous working paper. Thus, the county-level elasticities from such a regression can be thought of as the elasticities for each county *already accounting for a variety of characteristics of the counties.* Only the remaining county-level unobserved heterogeneity identifies the interaction coefficients in such a regression, so it is not entirely surprising that I find a considerable degree of heterogeneity.

Given the more recent results, I interpret my older results as indicating that areas

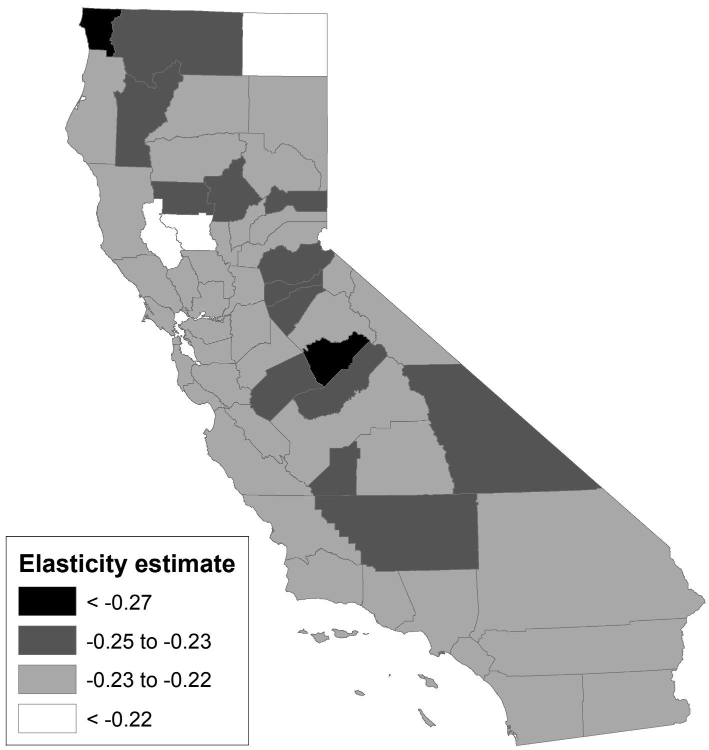

Figure 3.1: The elasticity of driving with respect to the gasoline price differs across counties in California.

such as Bay Area have unobserved heterogeneity influencing the responsiveness above what would be expected given the density, demographic characteristics, and vehicle composition of the counties. Other counties have the opposite. I view the results presented here as a more useful way of thinking about the geographic heterogeneity in the responsiveness of driving to changing gasoline prices.

3.2.6 Summary Discussion

The results shown in this section paint a picture of heterogeneity in the responsiveness of driving to gasoline price changes on several levels. Quantile regressions show that the most responsive decile of the vehicle population are almost twice as responsive as the least responsive decile of the vehicle population. The evidence of heterogeneity in responsiveness by vehicle type suggests that lower fuel economy vehicles are generally more responsive to changing gasoline prices. Large SUVs and pickup trucks are some of the most responsive. Small cars, large cars, and prestige sporty cars are the least responsive. There is also some evidence of heterogeneity in responsiveness by vehicle characteristics: vehicles with larger engine displacement (in liters) tend to be more responsive, while vehicles with larger engine sizes (in cylinders) tend to be less responsive.

Vehicles registered in zip codes with different demographic characteristics appear to display some heterogeneity in responsiveness. For example, vehicles driven in higher median income zip codes tend to be more responsive. Drivers of leased vehicles also appear to be more responsive. There is also evidence that households of different income groups differ in per-vehicle responsiveness. Vehicles owned by relatively low-income households appear to be the most responsive. The responsiveness appears to follow somewhat of a "U-shape" so that responsiveness declines at first and then increases again, although the differences in responsiveness are not dramatic. This relationship is consistent with some current evidence in the literature. At the low end, the higher responsiveness may be due to the tighter budget constraint. At the higher end, the higher responsiveness may be due both to having more vehicles in the household, as well as some lower marginal value trips and attractive substitutes to driving. However, at the very highest income group, the responsiveness declines again, perhaps relating to the more relaxed budget constraint for such high income individuals.

When we look at heterogeneity by location-specific variables, it appears that more densely populated areas are slightly less responsive, while areas with greater commute times are more responsive. These results run counter to my initial intuition and likely relate to the different characteristics of the drivers living in these areas. Another level of geographic heterogeneity is at the county-level. I find that there is some heterogeneity in driving responsiveness across counties in California, with slightly more responsiveness in some areas of the Central Valley than in the Bay Area or Los Angeles. However, the differences in responsiveness are not dramatic. To extent that some of the areas with slightly greater responsiveness have major issues with local air pollution, there may be slightly higher health co-benefits from reduced local air pollution due to this heterogeneity. However, these areas are relatively less congested, so the congestion co-benefits would be slightly less than they otherwise would be. Given the relatively limited range of responsiveness, it is likely that for a policy analysis, this geographic heterogeneity will be a second-order concern.

The evidence presented in this section focuses entirely on new personal vehicles during the first six years. As we saw in the previous section of this chapter, there is also evidence of heterogeneity in responsiveness across model years, so it is very possible that looking at heterogeneity in new personal vehicles only tells part of the story. Further evidence using Department of Motor Vehicles data may help to fill out this story. Knowing the heterogeneity in responsiveness for all drivers is even more important if we are interested in quantifying the local air pollution and congestion co-benefits of policies that reduce driving (or additional costs of policies that increase driving).

3.3 Responsiveness in Vehicle Purchases

When gasoline prices change, consumers who were planning on purchasing a new vehicle may be induced to purchase a vehicle with a different fuel economy. We have already seen initial evidence of this response in Chapter 2. For further evidence of a relationship between gasoline prices and fuel economy, we can model the fuel economy of personal new vehicles in California as a function of the gasoline price and characteristics of the vehicle. Consider a new vehicle purchaser i, who has decided to purchase a vehicle at time t. In the most simple setting, the choice of fuel economy FE_{it} can be thought of as a function of the price of gasoline faced by the purchaser P_{it}, individual- and zip code-specific attributes and demographics $\boldsymbol{D}_i$, and the economic conditions $\boldsymbol{E}_{it}$:

$$FE_{it} = g(P_{it}, \boldsymbol{D}_i, \boldsymbol{E}_{it}).$$

This specification presumes that the purchaser anticipates that the current gasoline price will serve as the future gasoline price. Of course, this is a major simplification, for consumers may consider the previous gasoline price, trends in the gasoline price, and perhaps even the variance in the gasoline price. I will discuss the consumer vehicle purchase decision in greater detail in Chapter 4. To use standard regression approaches, we can assume a linear form for this relationship:

$$FE_{it} = \gamma_0 + \gamma_P P_{it} + \boldsymbol{\gamma}_D \boldsymbol{D}_i + \boldsymbol{\gamma}_E \boldsymbol{E}_{it} + u_{it}, \tag{3.7}$$

where u_i is a mean-zero stochastic error term. I can estimate this model using the entire 12.3 million new personal vehicles dataset.

There may be some identification concerns with this specification. One possible concern is that the increase in fuel economy may just be part of an exogenous trend, such as from the diffusion of hybrids. The subsequent decline in fuel economy after 2008 shown in Figure 2.10 in Chapter 2 provides evidence that this is not entirely the case. However, to help control for this possibility, I can also include time fixed effects or a higher-order polynomial of time. Neither are perfect: time fixed effects have the consequence of restricting the identifying variation to within-time period variation, which reduces much of the time series identifying variation, while a time polynomial imposes a restrictive specification of a possible underlying trend.

A second possible concern is that there is unobserved heterogeneity in preferences for driving that interact with how consumers respond to changes in gasoline prices. For example, if there is substantial heterogeneity across the population in responsiveness (i.e., γ_P should be modeled as a random coefficient), and how the responsiveness varies depends on unobserved preferences for driving, then γ_P could be biased. This issue cannot be easily addressed with this simple specification, but my structural model in Chapter 4 accounts for this issue explicitly.

Third, because fuel economy in miles per gallon is nonlinear in fuel saved with an increase in fuel economy, we may wish to consider a linear specification with fuel consumption in gallons per mile. Most Americans do not recognize the difference between miles per gallon and gallons per mile, but if they were carefully calculating the future savings from reduced fuel costs of higher fuel economy vehicles, then they would implicitly be taking into account the fuel consumption of the vehicle, rather than the fuel economy. Moreover, taking the arithmetic mean is appropriate for average fuel consumption (in gallons per mile) over vehicles, while the harmonic mean is more appropriate for fuel economy (in miles per gallon). Linear regression takes the conditional arithmetic mean, not harmonic mean, so fuel consumption is probably more appropriate for this reason.

Finally, fuel economy standards play an important role in whether there is a response in new vehicle purchases to changing gasoline prices. The national CAFE standard requires vehicle manufacturers to meet a nation-wide sales-weighted harmonic mean fuel economy for both the passenger vehicle and light truck fleets. For many manufacturers, these standards are binding constraints that the manufacturers have an incentive to meet, but not surpass. For some manufacturers, the fleet-wide fuel economy for each fleet easily surpasses the standards. Other manufacturers pay a fee or "gas-guzzler tax" in lieu of meeting the standard. For the manufacturers for which the standard always remains a binding constraint, we would expect changing gasoline prices *not* to change the national fleet-wide fuel economy for each fleet, but rather to allow the manufacturers to re-optimize in meeting the standard. Firm profits would increase, but the achieved national fleet-wide fuel economy would be the same. At the California-level, this may not necessarily be the case, for firms may sell more higher fuel economy vehicles in California when gasoline prices increase in order to allow them to sell more low fuel economy vehicles elsewhere in the United States.[17] Moreover, if the rise in gasoline prices changes demand sufficiently such that the CAFE standards are no longer binding – or if they were never binding – then we may still see a change in the national fleet-wide fuel economy when gasoline prices change. In addition, the average fuel economy for the entire light duty fleet may improve if firms sell more vehicles in the passenger car fleet and fewer in the

[17]This is certainly the case if gasoline prices increase more in California than other states when the oil price rises, which may be the case due to the higher ad velorum sales tax in California. But it may also be the case if the change in demand for higher fuel economy vehicles happens to be greater in California than other states when gasoline prices change. I do not have evidence that this is true, but it remains a theoretical possibility.

light truck fleet.[18] To see the importance of the interaction of CAFE standards are the responsiveness to gasoline prices on the extensive margin, I can examine the responsiveness separately for manufacturers that are subject to a binding standard and those that are not.

In addition to these concerns, there is a important caveat to keep in mind when interpreting the results in this section. In the short-run, when gasoline prices increase, manufacturers can maximize profits by simply increasing the amount of production of higher fuel economy vehicles. There may be capacity constraints, but vehicle manufacturers are quite facile at scaling up production to meet demand for popular vehicles at relatively short notice, often by paying workers overtime wages to keep production lines going longer. However, when a vehicle is in high demand it is unlikely to receive manufacturer and dealer incentives. My vehicle price data appear to incorporate manufacturer incentives, but do not appear to incorporate dealer incentives.[19] Thus, the results in this section can be considered to include the effects of any dealer incentives, and thus include some supply response in addition to a demand response. In the longer-term, design decisions are changed based on consumer demand, which is a function of gasoline prices. To the extent that there were some minor design changes when gasoline prices started increasing in 2005 and 2006, the extensive margin can be considered to include these supply-side changes as well. Given that most manufacturer design decisions are made roughly five years in advance, only very limited supply-side effects from design changes may be influencing the results here.

3.3.1 Results

Miles per gallon results

The results from estimating (3.7) are shown in Table 3.22. These results in Table 3.22 are estimated on the full dataset including all manufacturers. The baseline and county fixed effects results in columns (1) and (2) are quite similar, perhaps largely because most of the variation is time series variation. These suggest that a one dollar per

[18]Under the “reformed” standards, this switching between fleets is no longer a margin of adjustment, for the standards will be based on the vehicle footprint, rather than classification.

[19]Future work with the DMV data can confirm this.

gallon increase in the gasoline price at the time of the vehicle purchase corresponds to a 1.3 to 1.6 miles per gallon increase in new vehicle fuel economy. Given a 1.5 coefficient value, the corresponding elasticity at the means is in the range of 0.21 to 0.23, evaluated at the harmonic or arithmetic mean fuel economy respectively.

When either a third order polynomial of the registration month or year fixed effects are added to attempt to control for an underlying exogenous trend, the coefficient on the gasoline price is greatly reduced to around 0.6. This result is likely due to a combination of capturing an underlying trend and losing some of the identifying time series variation (e.g., within-year variation is being used). Given the 0.6 coefficient, the corresponding elasticity at the means is 0.09 at the means (both the harmonic and arithmetic mean round to 0.09). I view 0.09 as my preferred estimate of the elasticity of fuel economy with respect to the price of gasoline. When the year fixed effects are included, this elasticity is probably best interpreted as a one-year elasticity. Otherwise, an interpretation as a two-year elasticity is probably most appropriate, given the variation in gasoline prices. In either case, this responsiveness allows for some short-term adjustments in manufacturing by firms, but is not long-run enough to allow for the redesign of new vehicles.[20]

Gallons per mile results

When fuel consumption in gallons per mile is used, the results do not change dramatically. In Table 3.23, we can see that the negative coefficient on the price of gasoline indicates that when gasoline prices increase, the average fuel consumption of the new vehicle fleet decreases. The coefficients in Columns (3) through (5) suggest that if gasoline prices increase by one dollar, the average fuel consumption of the vehicle fleet will decrease by just over 0.1 gallons per 100 miles out of a mean of 5.2 gallons per 100 miles. This corresponds to an elasticity of fuel consumption of new vehicles with respect to the price of gasoline in the range of -0.07 at the means. This result corresponds reasonably closely to the result for fuel economy and provides a nice robustness check on that result.

[20]The preparation time for a new model is usually in the order of five years.

Table 3.22: Extensive Margin Regressions: Linear Model

Dependent variable: new vehicle fuel economy (harmonic mean = 19.2 mi/gallon)

	(1) base	(2) county FE	(3) time poly	(4) year FE	(5) county & year FE
gasoline price	1.275***	1.579***	0.672***	0.583***	0.626***
	(0.003)	(0.004)	(0.005)	(0.005)	(0.005)
lease	-0.874***	-0.903***	-0.965***	-0.968***	-0.980***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
zip density	0.041***	0.029***	0.034***	0.034***	0.029***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
zip businesses/cap	-0.101***	-0.215***	-0.167***	-0.164***	-0.212***
	(0.027)	(0.042)	(0.035)	(0.035)	(0.041)
log(zip population)	0.169***	0.128***	0.167***	0.167***	0.136***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
zip pop growth rate	-0.014***	0.000	-0.008***	-0.008***	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
log(zip income)	0.000	-0.614***	-0.253***	-0.237***	-0.606***
	(0.007)	(0.008)	(0.007)	(0.007)	(0.008)
commute time	0.019***	-0.005***	0.019***	0.019***	-0.004***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
county unempl rate	0.005***	0.130***	-0.019***	-0.019***	-0.021***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
consumer conf index	-0.019***	-0.007***	0.002***	-0.004***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
county housing prices	0.000***	-0.002***	0.001***	0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
registration month			6.495***		
			(0.156)		
(registration month)^2			-0.013***		
			(0.000)		
(registration month)^3			0.000***		
			(0.000)		
constant	20.094***	26.763***	-1,072.285***	22.597***	28.333***
	(0.086)	(0.107)	(27.879)	(0.087)	(0.108)
county FE	N	Y	N	N	Y
month-year polynomial	N	N	Y	N	N
year FE	N	N	N	Y	Y
race & age	Y	Y	Y	Y	Y
R-squared	0.040	0.044	0.050	0.050	0.052
Observations	12.3m	12.3m	12.3m	12.3m	12.3m

Heteroskedasticity-robust s.e. in parentheses, clustered on county in (2),(5)

*** indicates significant at 1% level, ** at 5% level, * at 10% level

Table 3.23: Extensive Margin Regressions: GPM Dependent Variable

Dependent variable: new vehicle fuel consumption (mean = 0.052 gallon/mi)

	(1) base	(2) county FE	(3) time poly	(4) year FE	(5) county & year FE
gasoline price	-0.003***	-0.004***	-0.0014***	-0.0012***	-0.0014***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
lease	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
zip density	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
zip businesses/cap	0.000***	0.001***	0.000***	0.000***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
log(zip population)	-0.001***	-0.000***	-0.001***	-0.001***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
zip pop growth rate	0.000***	-0.000***	0.000***	0.000***	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
log(zip income)	-0.000**	0.002***	0.001***	0.001***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
commute time	-0.000***	0.000***	-0.000***	-0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
county unempl rate	0.000***	-0.000***	0.000***	0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
consumer conf index	0.000***	0.000***	-0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
county housing prices	0.000***	0.000***	-0.000***	-0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
registration month			-0.018***		
			(0.000)		
(registration month)^2			0.000***		
			(0.000)		
(registration month)^3			-0.000***		
			(0.000)		
constant	0.053***	0.035***	3.030***	0.047***	0.031***
	(0.000)	(0.000)	(0.063)	(0.000)	(0.000)
county FE	N	Y	N	N	Y
month-year polynomial	N	N	Y	N	N
year FE	N	N	N	Y	Y
race & age controls	Y	Y	Y	Y	Y
R-squared	0.039	0.043	0.049	0.049	0.052
Observations	12.3m	12.3m	12.3m	12.3m	12.3m

Heteroskedasticity-robust s.e. in parentheses, clustered on county in (2),(5)

*** indicates significant at 1% level, ** at 5% level, * at 10% level

Effect of binding CAFE standards

The above results indicate that there is a clear response: when gasoline prices increase, consumers purchase higher fuel economy vehicles. However, since national CAFE standards are not binding on all manufacturers, the effect of gasoline prices on the fleet average fuel economy is likely to be different based on manufacturer. Jacobsen (2010) lists the major firms that face binding CAFE standards and those that do not. Most notably, all of the major US vehicle manufacturers (Ford, Chrysler, GM) are bound by the standards, while many of the major foreign firms (Toyota, Honda, Nissan) are not bound by the standards. Some of the premium European automakers pay the "gas-guzzler tax" for not meeting the standard. In this case there still is a cost to violating the standard, but it is not as high as the cost of complying with the standard.

To examine whether the response to gasoline prices on the extensive margin differs by manufacturer, I run the specification in Table 3.22 if make of the vehicle is from an automaker who is bound by the standard. All of the U.S. automakers are included, as well as Volkswagon and Volvo. I find that removing any single automaker does not tend to change any of the results. Table 3.24 shows that the responsiveness on the intensive margin appears to be less for manufacturers that are bound by CAFE standards than manufacturers that are not. Columns (1) and (4) are identical to Columns (1) and (3) in Table 3.22. Columns (2) and (5) are the same as the previous two columns, only estimated using the subsample of manufacturers that are facing a binding CAFE standard constraint. Columns (3) and (6) estimate the same two specifications using the subsample of manufacturers that are not facing a binding CAFE standard constraint.

The coefficient on the gasoline price at the time of purchase in Table 3.24 suggests that the relationship between gasoline prices and fuel economy is stronger for manufacturers that are not facing a binding CAFE standard constraint. Specifically, in Columns (2) and (5), the coefficient on the gasoline price is smaller than the respective coefficients in Columns (3) and (6). However, the coefficient in Columns (2) and (5) is still positive, statistically significant to the one percent level and economically significant. Interestingly, the coefficients estimated using each of the separate

Table 3.24: Extensive Margin Regressions: Manufacturers With Binding CAFE

Dependent variable: new vehicle fuel economy (harmonic mean = 19.2 mi/gallon)

	(1)	(2)	(3)	(4)	(5)	(6)
		Base			Time polynomial	
	all	bind	not bind	all	bind	not bind
gasoline price	1.275***	0.667***	1.086***	0.672***	0.542***	0.650***
	(0.003)	(0.003)	(0.004)	(0.005)	(0.006)	(0.006)
lease	-0.874***	0.694***	-2.039***	-0.965***	0.577***	-2.098***
	(0.004)	(0.006)	(0.005)	(0.004)	(0.006)	(0.005)
zip density	0.041***	0.039***	0.005***	0.034***	0.030***	-0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
zip businesses/cap	-0.101***	0.053*	-0.492***	-0.167***	-0.011	-0.552***
	(0.027)	(0.025)	(0.088)	(0.035)	(0.013)	(0.097)
log(zip population)	0.169***	0.155***	0.089***	0.167***	0.165***	0.090***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)
zip pop growth rate	-0.014***	-0.017***	-0.019***	-0.008***	-0.000	-0.013***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
log(zip income)	0.000	0.070***	-0.652***	-0.253***	-0.351***	-0.867***
	(0.007)	(0.008)	(0.009)	(0.007)	(0.008)	(0.009)
commute time	0.019***	0.005***	0.020***	0.019***	0.008***	0.022***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
county unempl rate	0.005***	-0.006***	0.044***	-0.019***	0.005***	0.024***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
consumer conf index	-0.019***	-0.012***	-0.016***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
county housing prices	0.000***	-0.001***	0.000***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
registration month				6.495***	8.567***	4.925***
				(0.156)	(0.180)	(0.203)
(registration month)^2				-0.013***	-0.017***	-0.010***
				(0.000)	(0.000)	(0.000)
(registration month)^3				0.000***	0.000***	0.000***
				(0.000)	(0.000)	(0.000)
constant	20.09***	17.47***	28.35***	-1,072.29***	-1,435.47***	-792.07***
	(0.086)	(0.096)	(0.122)	(27.879)	(32.209)	(36.251)
month-year polynomial	N	N	N	Y	Y	Y
race & age	Y	Y	Y	Y	Y	Y
R-squared	0.040	0.032	0.040	0.050	0.049	0.047
Observations	12.3m	4.3m	8.0m	12.3m	4.3m	8.0m

Heteroskedasticity-robust s.e. in parentheses

*** indicates significant at 1% confidence level, ** at 5% level, * at 10% level subsamples are less than the coefficients estimated using the entire sample, although the coefficients using the "not binding" subsample are quite close to those using the entire sample (the elasticity of fuel economy with respect to the price of gasoline using the time polynomial specification is still around 0.09).

In interpreting these results, it is useful to note that there may be something unobserved about consumers who purchase from manufacturers that happen to be bound by CAFE. For example, consumer of vehicles made by Japanese automakers may be different than consumers of vehicles made by U.S. automakers. The finding that the responsiveness is not zero for manufacturers facing a binding CAFE standard may have as much to do with manufacturers optimizing differently in different regions of the U.S. California is known as a state that sells more high fuel economy vehicles so it is possible that manufacturers use sales of high fuel economy vehicles in California simply to offset sales of lower fuel economy vehicles elsewhere. In addition, Chapter 2 shows that there is switching between vehicle classes from lower fuel economy classes to higher fuel economy classes, so there is likely to be switching between the lower fuel economy light truck fleet and the higher fuel economy passenger vehicle fleet. Finally, it is possible that certain manufacturers that I am listing as "binding" may find that the CAFE standards no longer bind for one of the two fleets. Without detailed information about the national sales-weighted average fuel economy for each fleet of each manufacturer, it is impossible to know whether this is the case.

The suggestive result that vehicles from manufacturers facing a binding CAFE standard appear to be less responsive provides useful insight for quantifying the gasoline price elasticity in the future, when CAFE standards are planned to be considerably tightened. Specifically, in the future, the national total gasoline demand elasticity may not include as much of a component from adjustment on the extensive margin due to much tighter CAFE standards. At the California-level, there may still be some adjustment, as manufacturers re-optimize.

3.4 Heterogeneity in Vehicle Purchase Responsiveness

When gasoline prices increase, who is purchasing higher fuel economy vehicles? Examining the heterogeneity in how vehicle choice changes when gasoline prices change gives a sense of who is most responsive, which provides insight into how the vehicle fleet will evolve differently when gasoline prices change.

I look at this question by interacting the demographic and location-specific variables with the gasoline price in the same specifications as in Columns (1) and (5) in Table 3.22. Table 3.25 presents these results. The interaction of the gasoline price and the indicator for whether the vehicle was leased suggests that consumers who lease a new vehicle are much less responsive to changes in gasoline prices. This result may be because those who lease vehicles have a shorter time horizon and do not take into account future fuel savings after the lease term expires. Leased vehicles have slightly higher fuel economy on average, so we know that consumers who lease do not have stronger preferences for low fuel economy vehicles, as would be a plausible explanation. Of course, there may be other unobserved attributes about consumers who lease that could be driving the result.

The coefficients also suggest that consumers in more densely populated areas not only purchase higher fuel economy vehicles, but are also more responsive when gasoline prices change. However, zip codes with a larger population tend to be less responsive. The coefficients on the zip code income variables suggest that consumers who live in wealthier zip codes tend to purchase much lower fuel economy vehicles, but tend to be more responsive when gasoline prices change. This may be a "hybrid effect," for many consumers in wealthy areas, such as the Bay area, chose to purchase hybrids when gasoline prices rose.

The results so far shed light on how different demographic groups of people respond using the full dataset of personal vehicles. But we can also look at how responsiveness varies with income using the income subsample. Specifically, I interact each income category indicator with the gasoline price at the time of purchase. Table 3.26 presents the results. Column (1) is identical to Column (1) in Table 3.22, only restricted to the

Table 3.25: Heterogeneity on Extensive Margin

Dep var: new vehicle fuel economy (harm mean = 19.2 mi/gal)

	(1) interact	(2) interact & FE
gasoline price	-3.836***	-4.431***
	(0.124)	(0.125)
lease	1.960***	1.791***
	(0.016)	(0.016)
zip density	0.028***	0.011***
	(0.002)	(0.002)
zip businesses/cap	0.006	-0.056
	(0.092)	(0.138)
log(zip population)	0.208***	0.203***
	(0.012)	(0.012)
zip pop growth rate	-0.017***	0.010***
	(0.002)	(0.002)
log(zip income)	-1.245***	-1.863***
	(0.026)	(0.027)
zip % pop age 65+	0.004*	0.001
	(0.002)	(0.002)
zip % pop under 18	-0.022***	-0.003
	(0.002)	(0.002)
zip % pop white	-0.040***	-0.036***
	(0.001)	(0.001)
zip % pop black	-0.024***	-0.028***
	(0.001)	(0.001)
zip % pop hispanic	-0.016***	-0.022***
	(0.001)	(0.001)
commute time	0.011***	-0.014***
	(0.002)	(0.002)
gaspr*lease	-1.079***	-1.056***
	(0.006)	(0.006)
gaspr*density	0.005***	0.007***
	(0.001)	(0.001)
gaspr*bus/cap	-0.038	-0.057
	(0.040)	(0.056)
gaspr*log(pop)	-0.014**	-0.026***
	(0.005)	(0.005)
gaspr*poprate	0.001	-0.004***
	(0.001)	(0.001)
gaspr*log(zip inc)	0.479***	0.481***
	(0.010)	(0.010)
gaspr*age 65+	-0.007***	-0.008***
	(0.001)	(0.001)
gaspr*age under 18	-0.019***	-0.020***
	(0.001)	(0.001)
gaspr*white pop	0.008***	0.008***
	(0.000)	(0.000)
gaspr*black pop	0.007***	0.007***
	(0.000)	(0.000)
gaspr*hisp pop	0.003***	0.004***
	(0.000)	(0.000)
gaspr*commute time	0.003***	0.004***
	(0.001)	(0.001)
constant	33.392***	41.565***
	(0.318)	(0.332)
econ conditions	Y	Y
county FE	N	Y
year FE	N	Y
R-squared	0.042	0.054
Observations	12.3m	12.3m

Heteroskedasticity-robust s.e. in parentheses

*** indicates sig at 1% level, ** at 5% level, * at 10% level subsample where I observe income. I then add the year and county fixed effects, so that Column (2) is the same as Column (5) in Table 3.22, but again restricted to the income subsample. Columns (3) and (4) add fixed effects for the income categories. Columns (5) and (6) add the income interactions.

Comparing Columns (1) and (2) with Columns (3) and (4), we can see that adding the income covariates does not change the magnitude of the gasoline price responsiveness very much. Moreover, all of the income categories purchase lower fuel economy vehicles than the first income category, with the wealthiest households purchasing the lowest fuel economy vehicles. Columns (5) and (6) contain the interactions. Again, the responsiveness is relative to the lowest income category, which, as discussed above, may involve unusual consumers (e.g., parents purchasing vehicles for their children or consumers with low income but great wealth). The income interaction results appear to indicate that higher income new vehicle purchasers are relatively more responsive than lower income purchasers. This again may likely relate to the "hybrid" effect, whereby higher income new vehicle purchasers may opt for a hybrid vehicle. Exploring which vehicles higher income consumers versus lower income consumers switch into when gasoline prices change is an interesting topic for future research, with important implications for the evolution of the vehicle stock.

Table 3.26: Heterogeneity on Extensive Margin

Dependent variable: new vehicle fuel economy (harm mean = 19.2 mi/gal)

	(1) base	(2) county & year FE	(3) income FE	(4) income FE	(5) income interact	(6) income interact
gasoline price	1.498***	0.628***	1.525***	0.619***	1.366***	0.578***
	(0.003)	(0.007)	(0.003)	(0.007)	(0.004)	(0.007)
$15k - $20k			-0.127***	-0.113***	-0.934***	0.412*
			(0.013)	(0.013)	(0.199)	(0.203)
$20k - $30k			-0.167***	-0.136***	0.038	0.479***
			(0.010)	(0.010)	(0.121)	(0.131)
$30k - $40k			-0.226***	-0.198***	0.693***	1.171***
			(0.010)	(0.010)	(0.115)	(0.125)
$40k - $50k			-0.419***	-0.378***	-1.368***	-0.764***
			(0.009)	(0.009)	(0.109)	(0.120)
$50k - $75k			-0.634***	-0.610***	-3.752***	-2.940***
			(0.008)	(0.008)	(0.074)	(0.090)
$75k - $100k			-0.773***	-0.790***	-5.408***	-4.478***
			(0.009)	(0.009)	(0.085)	(0.098)
$100k - $125k			-0.916***	-0.966***	-6.640***	-5.749***
			(0.010)	(0.010)	(0.110)	(0.120)
>125k			-1.228***	-1.370***	-8.136***	-6.028***
			(0.010)	(0.010)	(0.107)	(0.116)
gaspr*$15k - $20k					0.134***	-0.088**
					(0.033)	(0.034)
gaspr*$20k - $30k					-0.025	-0.071***
					(0.014)	(0.015)
gaspr*$30k - $40k					-0.080***	-0.117***
					(0.010)	(0.011)
gaspr*$40k - $50k					0.062***	0.025**
					(0.007)	(0.008)
gaspr*$50k - $75k					0.173***	0.130***
					(0.004)	(0.005)
gaspr*$75k - $100k					0.221***	0.176***
					(0.004)	(0.005)
gaspr*$100k - $125k					0.238***	0.199***
					(0.005)	(0.005)
gaspr*>125k					0.258***	0.174***
					(0.004)	(0.004)
constant	20.425***	27.881***	15.703***	22.726***	16.070***	23.331***
	(0.106)	(0.129)	(0.111)	(0.132)	(0.111)	(0.133)
lease control	Y	Y	Y	Y	Y	Y
econ conditions	Y	Y	Y	Y	Y	Y
county FE	N	Y	N	Y	N	Y
year FE	N	Y	N	Y	N	Y
R-squared	0.049	0.061	0.052	0.064	0.052	0.064
Observations	8.7m	8.7m	8.7m	8.7m	8.7m	8.7m

Heteroskedasticity-robust s.e. in parentheses, clustered on county in (2),(5)

*** indicates significant at 1% confidence level, ** at 5% level, * at 10% level

Chapter 4

A Model of Vehicle Choice and Utilization

This chapter develops and estimates a novel two-period model of individual vehicle choice and subsequent vehicle utilization.[1] The motivation for this model is to address the possible selection bias discussed in Chapter 1 in a comprehensive framework that explicitly takes into account the dynamic nature of the decision-making process. In addition, the development of a utility-consistent structural model allows for counterfactual simulation of the welfare implications of different policies. It also allows for an analysis of a more pure form of the rebound effect than any available in previous studies.

The estimation of the model in this chapter is made possible by the same variation in the data that identifies the parameters in the previous chapter. It differs markedly from the previous chapter in the explicit representation of the consumer decision-making process and quantification of the importance of the selection bias. As described in Chapter 1, this selection bias could occur if consumers who expect to drive more (or less) than average will choose to "select into" a vehicle with a different fuel economy in the purchase decision. This would change the cost per mile of driving, and thus the unobserved heterogeneity in consumer types over expected driving

[1] This chapter is based in large part on my job market paper, titled "How Do Consumers Respond to Gasoline Price Shocks? Heterogeneity in Vehicle Choice and Driving Behavior."

would lead to an endogeneity concern about the cost per mile of driving variable.

This chapter begins by discussing the nature of this selection bias and describing in detail why and how it may bias the estimates. I then present the structural model of vehicle choice and utilization itself by first laying out the theoretical model and then describing the econometric specification used to estimate the model. I then present the results of the estimation, including elasticity estimates and an estimate of the rebound effect. Finally, I examine the importance of selection by estimating vehicle choice and utilization separately rather than simultaneously.

4.1 The Nature of the Selection Bias

The possibility that there may be a selection bias in estimating the utilization elasticity of a durable good has been discussed as early as the 1960s (e.g., see Balestra and Nerlove (1966)). The Dubin and McFadden (1984) treatment of the issue from an econometric perspective provided perhaps the first clear exposition of possible solutions to the issue. In this section, I first formally elucidate why this might be an econometric issue and then discuss the solutions that have been used in the literature.

4.1.1 Why a Selection Bias?

To help elucidate the selection bias, I present a stylized static model of household decision-making with some similarities to the model presented in Davis (2008). Consider a household choosing which new vehicle to purchase. Each new vehicle j can be considered a bundle of attributes, represented by the vector Θ_j. The fuel economy of the vehicle MPG_j is one of the elements of the bundle of attributes, i.e., $MPG_j \in \Theta_j$. There are j possible vehicles to choose from in the choice set. The household's choice problem can be considered to be the following:

$$\max_{j \in \{1,\ldots,J\}} V(\Theta_j),$$

where $V(\Theta_j)$ is the conditional indirect utility function. Davis (2008) notes that we can start with the household production function of Becker (1965) and use the insight of Pollak and Wachter (1975) to rephrase the problem as a classical demand problem under the assumption of constant returns to scale for the production technology and no joint production. This household problem can be written in the following form:

$$V(\Theta_j) = \max_{\{VMT,x\}} U(VMT, \Theta_j, x \mid \eta)$$
$$\text{s.t.} \left(\frac{P^g}{MPG_j}\right) VMT + x = wT - r(\Theta_j),$$

where VMT is demand for driving in vehicle-miles-traveled, x is a composite of all other goods and services, η captures the relative value that this particular household places on driving, P^g is the price of gasoline, w is the wage rate, T is the amount of time in the day, and $r(\Theta_j)$ is the capital cost of the vehicle. The price of the composite numeraire good is normalized to unity in this utility maximization problem. P^g/MPG_j is simply the fuel cost per mile of driving, or the marginal (fuel) cost of driving. η is the critical link that leads to the selection bias as we will shortly see.

To solve the optimization problem, we can set up the Lagrangian

$$U(VMT, \Theta_j, x \mid \eta) - \lambda_{inc}\left[\left(\frac{P^g}{MPG_j}\right) VMT + x - wT + r(\Theta_j)\right], \tag{4.1}$$

where λ_{inc} is the shadow price on the constraint or the marginal utility of income. Maximizing with respect to VMT yields the first order condition

$$\frac{\partial U(VMT, \Theta_j, x \mid \eta)}{\partial VMT} - \lambda_{inc}\left(\frac{P^g}{MPG_j}\right) = 0.$$

Rearranging this first order condition for VMT implies that the optimal VMT, assuming an interior solution, is given by

$$VMT^* = \nu\left(\left(\frac{P^g}{MPG_j}\right), \Theta_j, x \mid \eta\right).$$

It is immediately clear that the demand for driving is partly determined by the unobserved η. But clearly $V(\Theta_j)$ is also a function of η, and thus the optimal choice of Θ_j is in part determined by η. Since $MPG_j \in \Theta_j$, we know that MPG_j is also a function of η. So suppose we use ordinary least squares to estimate a linear model for the driving in a particular time period t:

$$VMT_t = \beta_0 + \beta_1 \left(\frac{P_t^g}{MPG_j(\eta)} \right) + \boldsymbol{\beta}_\Theta \Theta_j + \boldsymbol{\beta}_x x + \varepsilon_t(\eta),$$

where β_0 and β_1 are scalar coefficients, $\boldsymbol{\beta}_\Theta$ and $\boldsymbol{\beta}_x$ are vectors of coefficients, and ε_t is a mean zero error term. I write $MPG_j(\eta)$ and $\varepsilon_t(\eta)$ for emphasis. β_1 is the primary coefficient of interest, for it captures the responsiveness of consumers to changes in the marginal cost of driving. This structure clarifies the nature of the endogeneity. Households take into account the unobserved η in the decision of how much to drive, but also take η into account in the decision of what vehicle to purchase. Thus MPG_j is correlated with the error term and OLS yields biased and inconsistent estimates.

We can even anticipate the direction of the bias if we make an assumption about how η interacts with vehicle choice. Suppose a household that has a high value for η is inclined to purchase a higher fuel economy vehicle in order to save on fuel costs. Then MPG_j would be positively correlated with ε_t, so that the marginal cost of driving is negatively correlated with the error term. This implies that $\hat{\beta}_1$ will be biased away from zero – leading us to believe that the responsiveness to changes in the marginal cost of driving is greater than the true value.

In contrast, suppose that a household with a high value for η is inclined to purchase a more comfortable vehicle for all of the long drives. Purchasing a more comfortable bundle of attributes Θ_j will likely involve a lower fuel economy for the vehicle. Thus, the marginal cost of driving would be positively correlated with the error term, so the bias of $\hat{\beta}_1$ would be towards zero. Fundamentally, the direction of the bias is an empirical question.

So far, the discussion has focused on how the fuel economy in the cost per mile of driving is correlated with the error term, leading to the endogeneity issue. What if the specification follows several papers in the literature and many of the estimations in Chapter 3 in either not including MPG_j or separating out P^g from MPG_j and estimating separate coefficients for each? Let's examine each possibility in turn.

Suppose instead of including P^g/MPG_j in the specification, we only include P^g. In this case, we still have an endogeneity concern to the extent that P^g is correlated with MPG_j, for MPG_j would be an omitted variable and would be subsumed by the error term. In this static model, it may seem obvious that P^g is correlated with MPG_j for consumers are more likely to purchase higher fuel economy vehicles if gasoline prices rise, as was demonstrated in Chapter 3. In a dynamic setting, where consumers base the decision of how much to drive on the price of gasoline at the time P_t^g, this correlation may be weaker. To the extent that there is serial correlation in the price of gasoline, there may still be a correlation between P_t^g and P_{t-1}^g, where $t-1$ is the time of purchase. Thus, P_t^g may still be endogenous. Of course, the extent of the correlation would decline with time since the purchase, so endogeneity may not be as important of an issue.

The intuition is similar if we include both P^g and MPG_j in the specification. Then MPG_j is clearly endogenous, for it is correlated with η in the error term. The coefficient on P^g is then also biased as long as P^g correlated with MPG_j. The evidence suggests that a positive correlation exists between P_t^g and MPG_j, with a the Pearson correlation coefficient for P_t^g and MPG_j of 0.08 in the new personal vehicles dataset.

In both cases, the issue of endogeneity remains, but it may not be as severe for identifying the responsiveness of driving to gasoline prices as in a specification including the cost per mile of driving. Importantly, in both cases I discussed the dynamic element of the decision process. The simple household model of vehicle choice and driving laid out here is a static model, but in reality, the decision of what vehicle to purchase and how much to drive is a dynamic decision. The price of gasoline will be different when both decisions are made. The unobserved preference for driving η may even change between the two periods. These issues point to the need for a dynamic model of the vehicle choice and driving decision process in order to fully address the selection issue. Such a model is the goal of this chapter.

4.1.2 Approaches to Address the Selection Bias

Before moving to the model itself, it is worthwhile to briefly review how others in the literature have addressed the selection bias. As mentioned in Chapter 1, Dubin and McFadden (1984) give three potential solutions to address the selection bias, and papers in the literature have used all three. Fundamentally, all three approaches use the estimated choice probabilities from a separate durable good (i.e., vehicle) choice estimation as instruments. The way this is accomplished differs between the three approaches. Dubin and McFadden (1984) title the three approaches as "instrumental variables," "reduced form," and "conditional expectation correction method." In addition to these three approaches, more recent studies have simultaneously estimated vehicle choice and driving so that any interactions between the decisions should be accounted for.

To clarify further, we can first state the Dubin and McFadden utilization equation. I modify the equation to fit the notation and context of this dissertation, but otherwise this is identical to equation (30) in Dubin and McFadden (1984):

$$\begin{aligned} VMT_j = & VMT_j^0 + \sum_{k=1}^{J} \alpha_0^k \mathbb{1}_{\{k=j\}} + \alpha_1 P^g + \alpha_2 p_2 + X'\gamma + \beta \left(y - \sum_{k=1}^{J} OC_k \mathbb{1}_{\{k=j\}} \right) \\ & - \beta_\rho \sum_{k=1}^{J} CC_k \mathbb{1}_{\{k=j\}} + \varepsilon, \end{aligned} \tag{4.2}$$

where VMT_j^0 is the typical driving of a household driving vehicle $j \in \{1, ..., J\}$, $\mathbb{1}_{\{k=j\}}$ is an indicator for vehicle k being the purchased vehicle j, P^g is again the price of gasoline, p_2 is the price of alternatives to driving, X is a vector of household characteristics, y is total household income, OC_k is the operating cost of vehicle k, CC_k is the capital cost of vehicle k, and ε is a mean zero error term. In the way Dubin and McFadden set up the utilization equation, the selection bias stems from the correlation of the indicator variables, OC_k, and CC_k with the error term. Of course, the endogeneity issue is conceptually identical to my treatment above.

The instrumental variables approach is perhaps the most straightforward approach to address the issue. The idea is to first run a vehicle choice model to estimate the choice probabilities of any given vehicle type j in the dataset. Denote these estimated choice probabilities as $\hat{P}_j$. Then, estimate the utilization equation using two-stage-least squares (2SLS) where the variables are instrumented for with P^g, p_2, w, $y - \sum_{k=1}^{J} OC_k \hat{P}_k$, $\sum_{k=1}^{J} CC_k \hat{P}_k$, and $\hat{P}_1, ..., \hat{P}_J$. Goldberg (1998) takes this approach by estimating a nested logit vehicle choice model first and then uses the estimated choice probabilities as instruments in the driving demand equation.

The reduced form approach is slightly more subtle. In the utilization equation laid by Dubin and McFadden, several of the terms, including the term on the annual operating cost (i.e., marginal cost of utilization), contain an indicator variable for the chosen appliance, $\mathbb{1}_{\{k=j\}}$. Rather than directly instrumenting, the reduced form approach simply plugs in $\hat{P}_j$ for the indicator variables. Thus the estimated utilization equation under the reduced form approach is

$$\begin{aligned} VMT_j = & VMT_j^0 + \sum_{k=1}^{J} \alpha_0^k \hat{P}_k + \alpha_1 P^g + \alpha_2 p_2 + X'\gamma + \beta \left(y - \sum_{k=1}^{J} OC_k \hat{P}_k \right) \\ & - \beta_\rho \sum_{k=1}^{J} CC_k \hat{P}_k + \varepsilon. \end{aligned} \tag{4.3}$$

Dubin and McFadden then suggest estimating the model with OLS. This approach is conceptually the same as estimating the utilization equation (4.2) with 2SLS and instrumenting $(\sum_{k=1}^{J} \alpha_0^k \mathbb{1}_{\{k=j\}})$ with $(\sum_{k=1}^{J} \alpha_0^k \hat{P}_k)$, $(y - \sum_{k=1}^{J} OC_k \mathbb{1}_{\{k=j\}})$ with $(y - \sum_{k=1}^{J} OC_k \hat{P}_k)$, and $(\sum_{k=1}^{J} CC_k \mathbb{1}_{\{k=j\}})$ with $(\sum_{k=1}^{J} CC_k \hat{P}_k)$. If one actually simply runs OLS on (4.3), the standard errors would be incorrect without the usual 2SLS correction or bootstrapping. Note that with this approach, the coefficients are just-identified, for the number of instruments is the same as the number of variables being instrumented for. The only paper I am aware of that uses this approach to estimate driving demand is Mannering and Winston (1985).

In the context of the more common framework, where VMT is regressed on the cost per mile of driving and other covariates, both the instrumental variables approach and the reduced form approach are roughly the same idea as instrumenting for the cost per mile of driving in the utilization demand equation with the estimated choice probabilities from a vehicle choice model.

The third approach is the conditional expectation method, which is the same idea as the common Heckman selection model. Rather than instrumenting, the approach adds a control function as another covariate. In the Heckman selection model, the Inverse Mill's ratio is added as another covariate. In the Dubin and McFadden conditional expectation method approach, a slightly different control function is used. The utilization equation using the conditional expectation method is given by

$$\begin{aligned} VMT_j =& VMT_j^0 + \sum_{k=1}^{J} \alpha_0^k \mathbb{1}_{\{k=j\}} + \alpha_1 P^g + \alpha_2 p_2 + X'\gamma + \beta \left(y - \sum_{k=1}^{J} OC_k \mathbb{1}_{\{k=j\}} \right) \\ &- \beta_\rho \sum_{k=1}^{J} CC_k \mathbb{1}_{\{k=j\}} + \sum_{k \neq j}^{J} \gamma_k \left[\frac{\hat{P}_k \ln \hat{P}_k}{1-\hat{P}_k} + \ln \hat{P} \right] + \varepsilon, \end{aligned}$$

As we can see, the only difference between this equation and (4.2) is the second to last term, the control function. The intuition for why this approach addresses the issue is identical to the intuition for why the Heckman selection model: by explicitly modeling the process of selection, we can find the mean of the error term conditional on the selection and then include this as a term in the utilization equation to correct for the issue. Of course, exclusion restrictions are still needed for adequate identification (otherwise identification is entirely from the structure and not the data). Similarly, either corrected or bootstrapped standard errors are needed for inference. In this sense, $\hat{P}_1, ..., \hat{P}_J$ can still be thought of as instruments used to address the endogeneity. West (2004) takes the conditional expectation approach to address the selection problem

An issue with all three of the approaches proposed by Dubin and McFadden is that the vehicle choice model is separately estimated from the utilization equation, potentially leading to inconsistencies between the results of the two equations. More recently, several papers address this concern by simultaneously estimating vehicle choice and utilization (Feng, Fullerton, and Gan 2005; Bento et al. 2009; Jacobsen 2010). Each of these papers fundamentally base the structural model specification on the framework in Dubin and McFadden (1984). Small and Van Dender (2007) arguably may also be addressing the selection issue similarly by simultaneously estimating a structural model of fleet fuel economy choice and driving using aggregate data.

The approach I take in this chapter similarly models vehicle choice and driving, but uses a new framework that for the first time brings in the dynamics, or timing, of each of these decisions.[2]

4.1.3 Suggestive Evidence of a Selection Bias

Since the unobserved preference for driving η is not observed, it is fundamentally impossible to test for whether there is a selection bias or not. However, it is possible to examine some suggestive evidence that consumers who expect to be driving more purchase higher fuel economy vehicles. Since any unobserved preference for driving is likely to be only one of many factors that go into the decision of what type of vehicle to purchase, looking at differences in driving across vehicle classes is not likely to be productive. However, looking at differences in driving by fuel economy within a vehicle class may provide some very cursory suggestive evidence of a selection effect. The suggestive evidence is most believable if all consumers within a vehicle class are the same – except for their unobserved preference for driving, which leads them to sort themselves into different fuel economy vehicles. Of course, there are considerable differences between vehicles even with a vehicle class that may influence how new vehicle buyers sort into vehicles. Moreover, and perhaps even more importantly, vehicles with higher fuel economy within a vehicle class may be driven more because driving them is less expensive. Thus, this evidence must be taken very cautiously, yet it is perhaps the best evidence available on a bias that is inherently due to an

[2]There is a considerable literature of dynamic models of vehicle markets, yet the focus in this literature is on other questions, such as the effects of durability and secondary markets on firm behavior (Esteban and Shum 2007), the identification of transaction costs in vehicle replacement behavior (Schiraldi 2010), the equilibrium resale pattern over the lifetime of vehicles (Stolyarov 2002), and the effect of scrappage subsidies in France (Adda and Cooper 2000).

unobservable.

Given these strong caveats about the interpretation, I find that within a vehicle class there is clear evidence that consumers who drive higher fuel economy vehicles drive more. For example, using the new personal vehicle dataset, I find that full utility vehicles that are in the lower 25 percent based on fuel economy are driven 1,174 miles per month on average, while those in the highest 25 percent based on fuel economy are driven 1,206 miles per month on average. Similarly, small cars that are in the lower 25 percent based on fuel economy are driven 1,096 miles per month on average, while those in the highest 25 percent based on fuel economy are driven 1,134 miles per month on average. Similar trends exist for most of the other vehicle classes as well.

If we interpret these differences as due to differences in the unobserved consumer preference for driving, then these differences can be interpreted as evidence of a selection bias. I view the differences has somewhat suggestive, but perhaps even more likely to be indicative of the rebound effect: consumers who purchase higher fuel economy vehicles may drive them more because they are less expensive to drive. In fact, these differences underlie the much of the variation that is identifying the coefficient on fuel economy in the OLS and fixed effects regressions in Chapter 3. The structural model developed in the remainder of this chapter explicitly addresses the selection bias by simultaneously estimating vehicle choice and driving. If there is no selection bias from an interaction of the two decisions, the results of the simultaneous estimation should theoretically be the same as estimating each equation separately. I view this evidence as the much stronger evidence of the importance of the selection bias.

4.2 Theoretical Model

I now present a stylized two-period model of individual vehicle choice and subsequent vehicle utilization, which will form the basis for the econometric specification and simulation of counterfactuals. I designed this model to take advantage of both the structure, as well as the richness and detail of the new personal vehicle dataset assembled for this dissertation. As mentioned above, the advantage of using a structural approach in this context is the ability to deal with selection by simultaneously modeling both decisions while explicitly taking into account the differing time frame for each decision. It also allows for a more satisfying modeling of the consumer decision process.

In each of the two periods of the model, consumers are assumed to weigh the benefits against the costs of different possible choices. In the first period, consumers optimally choose which new vehicle to purchase, based on the cost of the vehicle, the attributes of the vehicle, the consumer's expected resale price of the vehicle, and the consumer's expected benefit from driving that vehicle in the second period. The expectations here can be considered to be taken over consumer beliefs about the future price of gasoline and future economic conditions. I assume risk-neutral consumers. In the second period, consumers choose how much to drive, conditional on the vehicle purchased in the first time period. The discrete-continuous modeling framework presented here has some similarities to the model used in Einav et al. (2010) in the context of health care plan choice and subsequent utilization.

An important feature of the model structure presented here is how selection is accounted for. Each consumer i is assumed to have an "known utilization type" that captures factors that influence how much the consumer benefits from driving apart from demographics or other observables. These factors may include having a significant other or close friend who lives several hours away, or having proclivity for going on joy-rides. This known utilization type, denoted by η_i^k, is known by the consumer at the time of the vehicle purchase, but is not observed by the econometrician. It can be thought of as a vehicle random effect that enters into both the vehicle choice and utilization decision.

Over a several year period, consumers may also be subject to a variety of shocks that may also influence how much they benefit from driving. For example, a consumer could change jobs or have a death in the family. These shocks, denoted by η_i^u, would not be known to the consumer at the time of the vehicle purchase. At the time of purchase they can be thought of as a mean-zero random variable that is unknown to both the econometrician and the consumer. At the time of driving, these shocks are known to the consumer, yet remain unknown to the econometrician.

4.2.1 Utilization Choice

I begin with the second period, when each new vehicle purchaser i optimally chooses how much to drive conditional on owning a vehicle of type j. In making this decision, consumers face a tradeoff between the benefits of driving and the cost of driving. I assume that the benefit of driving, $b_{ij}(VMT_{ij}, C_i, z_i^d, E_i, \theta_j)$, is a concave function in its first argument, corresponding to a diminishing marginal utility of driving. The benefit of driving is also a function of commuting needs C_i, a vector of demographics z_i^d, a vector of economic conditions E_i, and a vector of the characteristics of the vehicle θ_j.[3] I assume that the fuel economy of the vehicle does not enter θ_j. The cost of driving is in general the sum of the fuel cost, maintenance cost, and time cost of driving. Since the object of interest here is how consumers respond to changes in gasoline prices, I focus entirely on the fuel cost of driving. The fuel cost is defined simply as the price per mile of driving times VMT.

I thus parameterize the second-period utility with a form similar to (4.1):

$$u_2(VMT_{ij}, C_i, z_i^d, \theta_j, MPG_j) = \alpha_{ij}\left(VMT_{ij} - \frac{\lambda}{2}VMT_{ij}^2\right) - \frac{p_i^g}{MPG_j}VMT_{ij}, \quad (4.4)$$

where p_i^g is the retail price of gasoline faced by the new vehicle purchaser and MPG_j is the fuel economy of the vehicle.[4] This specification assumes that the benefits of driving are quadratic in VMT, where λ_{inc} influences the curvature of the function. This specification also normalizes the coefficient on the fuel costs (i.e., the marginal utility of income, or the λ in (4.1)) to unity, so u_2 is a money-metric utility function. The coefficient on the benefits of driving, α_{ij} is a random coefficient that is function of commuting needs, demographics, economic conditions, and characteristics of the

[3] Income and whether the vehicle is leased or not may also affect how consumers value driving, and thus both are included in z_i^d. θ_j may also include model indicator variables to control for unobserved quality, as in ξ_j in Berry, Levinsohn, and Pakes (1995).

[4] This stylized model abstracts from the influence of driving behavior on fuel economy.

vehicle. Specifically, this random coefficient is parameterized as

$$\frac{1}{\alpha_{ij}} = \tilde{\alpha}_{ij} = -(\beta_c C_i + \beta_d z_i^d + \beta_e E_i + \gamma_2 \theta_j + \eta_i), \tag{4.5}$$

where η_i is a stochastic term that captures the unobserved heterogeneity in how new vehicle purchaser i values driving in period two. I assume this term is additively separable with the two components: the *known* component η_i^k, which is known by the consumer in period one, and *unknown* component η_i^u, which is not known to the consumer until period two. Since both components are known in period two, we have

$$\eta_i = \eta_i^k + \eta_i^u.$$

Consumers maximizing (4.4) will optimally choose VMT conditional their vehicle j based on the following first-order condition (assuming an interior solution):

$$VMT_{ij}^* = \frac{1}{\lambda} - \frac{\tilde{\alpha}_{ij}}{\lambda} \left(\frac{p_i^g}{MPG_j} \right). \tag{4.6}$$

With this specification, driving is linear in η_i. Under the anticipated sign of $\tilde{\alpha}_{ij}$, it is increasing in fuel economy and decreasing in the price of gasoline. It is increasing in commute time and demographics when there is a positive coefficient on β_c and β_d respectively. For example, if the coefficient on income is positive, then increasing income would increase driving.

4.2.2 Vehicle Choice

In the first period, the new vehicle purchaser weighs the benefits of owning a particular vehicle against the cost of purchasing that vehicle. The benefits from owning the vehicle accrue from the expected period-two utility, the expected option value from resale at the end of period two, and any prestige or other non-usage value to the consumer from owning a vehicle with the set of vehicle characteristics given by θ_j.[5] The consumer expectations of period-two utility and resale value are taken over

[5] In the second period, I examine specifications with and without the vehicle fuel economy included in θ_j, thus leaving open the possibility that consumers gain non-usage utility just from having a

the joint distribution of consumer beliefs about future gasoline prices and economic conditions.

Following this framework, new vehicle purchasers are assumed to optimally choose a vehicle to maximize

$$u_1(\mathbb{E}[u_2], \theta_j, p_j) = \delta_1 \mathbb{E}[u_2] + \gamma_1 \theta_j - p_j + \delta_2 \mathbb{E}[p_j^R] + \epsilon_{ij},$$

where p_j is the price of the vehicle at the time of purchase, p_j^R is the resale price of the vehicle at the end of the second period, and ϵ_{ij} captures the idiosyncratic unobserved heterogeneity in how consumer i prefers vehicle j. As in the second period utility, the utility at the time of purchase has current period dollar values normalized to unity, so that u_1 is a current period money-metric utility function.

The consumer's expected utility from driving, $\mathbb{E}[u_2]$, is based on the expectation of (4.4), given consumer beliefs about period-two gasoline prices and economic conditions. By construction, η_i^u does not enter into the period two utility.[6] The first period utility is then

$$u_1(\mathbb{E}[u_2], \theta_j, p_j) = \delta_1 \mathbb{E}\left[\frac{1}{\tilde{\alpha}_{ij}}(VMT_{ij} - \frac{\lambda}{2} VMT_{ij}^2) - \frac{p_i^g}{MPG_j} VMT_{ij}\right] + \gamma_1 \theta_j - p_j + \delta_2 \mathbb{E}[p_j^R] + \epsilon_{ij}.$$

This form of the utility is useful for intuition. Consumer utility from purchasing vehicle j is a function of the discounted expected net benefit of driving the vehicle, any non-usage value from owning the vehicle $\gamma_1 \theta_j$, the discounted expected resale price of the vehicle $\delta_2 \mathbb{E}[p_j^R]$, the price of the vehicle p_j, and a term capturing the idiosyncratic preference of consumer i for vehicle j.

To proceed further, I must make an assumptions about the joint distribution of consumer beliefs about future gasoline prices and economic conditions. Specifically, at the time of the vehicle purchase, how do consumers believe gasoline prices and the economy will jointly evolve? One could imagine modeling a distribution over

higher fuel economy vehicle.

[6] One can think of the consumer's expectation of η_i^u as equal to zero.

the joint stochastic processes of these two factors.[7] Yet very little empirical work is available to answer this question. In a recent study, Anderson et al. (2011a) use the Michigan Survey of Consumers to observe how much consumers state that they expect gasoline prices to rise or fall over the next five years. The findings suggest that the nominal forecasts systematically exceed the current gasoline price, but when long-term inflation expectations are taken into account, the time series of the current gasoline price and the stated expectations line up rather closely. Anderson et al. take this as suggestive evidence that consumer beliefs are largely consistent with a random walk for gasoline prices–implying that consumers base their expectation of the future price of gasoline on the current price of gasoline.

An alternative possibility is that gasoline futures prices capture consumer expectations of future gasoline prices (e.g., from the New York Mercantile Exchange (NYMEX)). Alquist and Kilian (2010) find that futures prices do not do any better at forecasting future oil (and gasoline) prices than the current price. Of course, consumers may still use futures prices. Alquist, Kilian, and Vigfusson (2011) review the limited evidence for different views on consumer beliefs about future gasoline prices, and find little evidence for any consumer beliefs other than a no-change forecast (i.e., using the current price of gasoline as the forecast for future prices). Yet the evidence is quite inconclusive and a variety of approaches have been explored in the literature, including ARIMA models, a no-change forecast, and gasoline or oil futures prices (e.g., see Kahn (1986), Allcott and Wozny (2010), and Davis and Kilian (2011a)).

For tractability purposes, I make two key assumptions about the joint distribution of consumer beliefs about future gasoline prices and economic growth. Both of these assumptions are consistent with consumers believing that each of these processes follow a random walk. I assume first that consumer beliefs about future expectations of gasoline prices are independent of consumer beliefs about future economic conditions. This implies that consumers do not anticipate correlated shocks to both gasoline prices and the economic conditions. Second, I assume that consumers use the current gasoline price and economic conditions as their expectation of future realizations of these variables. The intuition behind these assumptions is the idea that

[7] Indeed, future work is planned on this extension of the model.

consumers really do not know what the future gasoline price or business cycle will hold, and thus make a guess about these simply based on the information available on each today. I later perform a sensitivity analysis to examine the robustness of my analysis to these assumptions.

From the interior solution in (4.6), the expected VMT at the time of the purchase conditional on purchasing vehicle j is then

$$\mathbb{E}[VMT_{ij}] = \frac{1}{\lambda} - \frac{\mathbb{E}[\tilde{\alpha}_{ij}]}{\lambda}\left(\frac{\mathbb{E}[p_i^g]}{MPG_j}\right),$$

Since $\tilde{\alpha}_{ij}$ is a function of economic conditions, but not gasoline prices, the consumer's expectation $\mathbb{E}[\tilde{\alpha}_{ij}]$ is only taken over their beliefs about future economic conditions, so that $\mathbb{E}[\tilde{\alpha}_{ij}] = -(\beta_c C_i + \beta_d z_i^d + \beta_e \mathbb{E}[E_i] + \gamma_2\theta_j + \eta_i^k)$, where η_i^k replaces η_i since η_i^k is known to the consumer and the consumer expectation of η_i^u is zero.

In the interior solution, the expected expenditure on fuel conditional on purchasing vehicle j is

$$\mathbb{E}[p_i^g \ VMT_{ij}] = \frac{\mathbb{E}[p_i^g]}{\lambda} - \frac{\mathbb{E}[\tilde{\alpha}_{ij}]}{\lambda}\left(\frac{\mathbb{E}[(p_i^g)^2]}{MPG_j}\right).$$

Rearranging, we have

$$u_1(\mathbb{E}[u_2], \theta_j, p_j) = \frac{\delta_1}{2\lambda}\mathbb{E}[\frac{1}{\tilde{\alpha}_{ij}}] - \frac{\delta_1}{\lambda}\frac{\mathbb{E}[p_i^g]}{MPG_j} + \frac{\delta_1 \mathbb{E}[\tilde{\alpha}_{ij}]}{2\lambda}\frac{\text{var}(p_i^g) + \mathbb{E}[(p_i^g)]^2}{(MPG_j)^2} + \gamma_1\theta_j - p_j + \delta_2\mathbb{E}[p_j^R] + \epsilon_{ij}.$$

Here the second raw moment of the gasoline price is replaced by $\text{var}(p_i^g) + \mathbb{E}[(p_i^g)]^2$. Note that $\text{var}(p_i^g)$ is the variance of the distribution of consumer beliefs about the price of gasoline in the second period. A characteristic of a random walk is that the variance evolves over time and goes to infinity as time goes to infinity. Period two in this model could be conceptualized as either one (six year) period or many identical shorter periods. Following the latter interpretation, VMT_{ij}^* is the VMT per period, and $\text{var}(p_i^g)$ is the period variance of the consumer's belief of the path of the gasoline price. I follow this interpretation, and for consistency my analysis uses the VMT over six months and the observed variance in retail gasoline prices over the previous six months.

The consumer's expected resale price of the vehicle at the end of the period of utilization, $\mathbb{E}[p_j^R]$, remains to be discussed. How much a used car will sell for in six years may be considered by consumers to be a function of the gasoline price (e.g., low fuel economy vehicles sell for less with higher gasoline prices) and economic conditions. To capture the main factors that a consumer may consider in predicting the future resale price of the newly purchased vehicle, I model the consumer's expected resale price at the end of period two as a function of the price of a used similar model vehicle at the time of purchase and the consumer's expected driving:

$$\mathbb{E}[p_j^R] = p_j^{R0} - \mu_j(\mathbb{E}[VMT_{ij}] - BM_j),$$

where p_j^{R0} is the resale price at the time of the vehicle purchase of a used vehicle j with the base mileage BM_j, μ_j is an adjustment factor in the price of a used vehicle for differences between the amount the vehicle has been driven and the base mileage. This specification is also consistent with the assumption that consumer beliefs about future gasoline prices and economic conditions are independent and follow a random walk.

We then have the following final form of the utility in period one:

$$\begin{aligned} u_1 =& \frac{\delta_2}{\lambda} + \frac{\delta_1}{2\lambda}\mathbb{E}[\frac{1}{\tilde{\alpha}_{ij}}] - \delta_2\mu_j BM_j - \frac{\delta_1}{\lambda}\frac{\mathbb{E}[p_i^g]}{MPG_j} + \frac{\delta_1\mathbb{E}[\tilde{\alpha}_{ij}]}{2\lambda}\frac{\mathrm{var}(p_i^g) + \mathbb{E}[(p_i^g)]^2}{(MPG_j)^2} + \\ & \gamma_1\theta_j - p_j + \delta_2 p_j^{R0} - \frac{\delta_2\mathbb{E}[\tilde{\alpha}_{ij}]}{\lambda}\frac{\mathbb{E}[p_i^g]}{MPG_j} + \epsilon_{ij}. \end{aligned} \tag{4.7}$$

This expression captures the intuition that the utility of purchasing a vehicle is a function of the demographics of the consumer and economic conditions (through $\tilde{\alpha}_{ij}$), the characteristics of the vehicle, the price and variance of the price of gasoline, the fuel economy of the vehicle, and the resale price of used vehicles of the same type.

4.3 Econometric Model

4.3.1 Specification

I now move to specifying the stochastic structure. Recall that there are three stochastic terms in the model: ϵ_{ij}, the known "driving type" η_i^k, and the unknown shocks that influence the "driving type" η_i^u.

I begin by assuming that the unobserved heterogeneity in vehicle preference is distributed i.i.d. Type I extreme value, following the classic vehicle choice estimation literature. This allows for a computationally appealing form for the probability of consumer i choosing a particular vehicle of type j from the choice set $\mathcal{J}_i$, with cardinality $|\mathcal{J}_i| = J_i$:

$$\Pr_i(j) = \frac{\exp(V_{ij})}{\sum_{k=1}^{J_i} \exp(V_{ik})}, \tag{4.8}$$

where V_{ij} is the representative utility, given by

$$\begin{aligned} V_{ij} =& \frac{\delta_2}{\lambda} + \frac{\delta_1}{2\lambda}\mathbb{E}[\frac{1}{\tilde{\alpha}_{ij}}] - \delta_2\mu_j BM_j - \frac{\delta_1}{\lambda}\frac{\mathbb{E}[p_i^g]}{MPG_j} + \frac{\delta_1\mathbb{E}[\tilde{\alpha}_{ij}]}{2\lambda}\frac{\mathrm{var}(p_i^g) + \mathbb{E}[(p_i^g)]^2}{(MPG_j)^2} + \\ & \gamma_1\theta_j - p_j + \delta_2 p_j^{R0} - \frac{\delta_2\mathbb{E}[\tilde{\alpha}_{ij}]}{\lambda}\frac{\mathbb{E}[p_i^g]}{MPG_j}. \end{aligned} \tag{4.9}$$

Note that (4.8) holds due to the Type I extreme value assumption and the assumption of independence of the errors.

I next assume that the known driving type η_i^k is i.i.d. Normally distributed with mean zero and an unknown variance σ^2. In other words, $\eta_i^k \sim$ i.i.d $\mathcal{N}(0, \sigma^2)$. Similarly, I assume that the unknown preference for driving η_i^u is also i.i.d. Normally distributed with mean zero and an unknown variance ω^2, i.e., $\eta_i^u \sim$ i.i.d $\mathcal{N}(0, \omega^2)$. With these two assumptions, we have that $\eta_i \sim$ i.i.d $\mathcal{N}(0, \omega^2 + \sigma^2)$. These assumptions underpin the stochastic structure of the random coefficient on the consumer preference for driving α_{ij}. Assuming a mean zero normal distribution for the unobserved heterogeneity leads to a Normal distribution for driving, which is natural considering the quite normal-looking empirical distribution of VMT, as shown in Figure 2.1.

Abstracting from the corner solution, we can rearrange (4.6) as follows:

$$VMT_{ij} = \frac{1}{\lambda} + \left(\frac{p_i^g}{\lambda MPG_j}\right)(\beta_c C_i + \beta_d z_i^d + \beta_e E_i + \gamma_2 \theta_j) + \left(\frac{p_i^g}{\lambda MPG_j}\right)\eta_i \tag{4.10}$$

This expression implies that the conditional distribution for VMT_{ij} is given by $VMT_{ij} \sim \text{i.i.d}\,\mathcal{N}(\zeta_{ij}, \left(\frac{p_i^g}{\lambda MPG_j}\right)^2(\omega^2+\sigma^2))$, with mean $\zeta_{ij} = \frac{1}{\lambda} + \left(\frac{p_i^g}{\lambda MPG_j}\right)(\beta_0 + \beta_c C_i + \beta_d z_i^d + \beta_e E_i + \gamma_2\theta_j)$. Thus, the conditional likelihood of observing a particular amount of driving by consumer i, given the vehicle chosen and the model parameters, is

$$l_i(VMT_{ij}|j \text{ chosen}) = \frac{1}{\sqrt{2\pi(\omega^2+\sigma^2)}} \exp\left(-\frac{(VMT_{ij}-\zeta_{ij})^2}{2(\omega^2+\sigma^2)}\right).$$

The conditional likelihood of both vehicle demand and utilization for consumer i can thus be written as the probability consumer i purchases vehicle j times the probability VMT_{ij} is observed. Since η_i^k is unobserved by the econometrician, we can integrate over the distribution of η_i^k to form the final likelihood:

$$L_i = \int \prod_{j=1}^{J_i} (\Pr_i(j)\ l_i(VMT_{ij}|j \text{ chosen}))^{\mathbb{1}_{ij}}\, dF_{\eta_i^k}, \tag{4.11}$$

where $\mathbb{1}_{ij}$ is an indicator function for whether a vehicle of type j was purchased by consumer i, and $F_{\eta_i^k}$ is the cumulative distribution function of η_i^k.[8] Here J_i is indexed by i to denote that different consumers may face different choice sets depending on when and where they make the vehicle purchase. This likelihood function is conditional on a vehicle being purchased and on the parameter estimates.

4.3.2 Identification

This section briefly discusses the identification of the model. At its core, the structural model developed here is a selection model. Thus, the identification of the model stems from the structure. Specifically, the exclusion restrictions in period two play a

[8]Note we integrate over only η_i^k because η_i^u is implicit in the likelihood and does not enter directly.

crucial role in identifying the coefficients of the model in the same way that exclusion restrictions are critical for identifying the standard Heckman selection model. In the second period, the average gasoline price and economic conditions over the time of driving enter the equation, rather than the gasoline price and economic conditions at the time of purchase (i.e., the expected gasoline price), which enter into the first period equation. These exclusion restrictions, along with a similar exclusion of the price and resale price of the vehicle, are features of the structure that are key for identification.

My dataset also contains considerable variation in the gasoline price that facilitates identification. The data contain both cross-sectional variation from the differences across counties, as well as time series variation, much of it from the striking gasoline price increase from 2006 to 2008. Using the variation from gasoline price spike may be a concern if it is considered an unusual price shock leading to a short-term over-reaction. However, futures prices during the time of high prices remained high, and media reports predicted high gasoline prices long into the future. So it is reasonable that using this variation is still useful for out-of-sample counterfactual simulations. Moreover, using this substantial variation provides a great opportunity for precisely pinning down consumer responsiveness to gasoline price shocks.

Using the variation from this particular gasoline price shock is also advantageous for two additional reasons. First, the shock is only somewhat correlated with economic conditions, unlike many previous gasoline price shocks. So by controlling for economic conditions, I can attempt to disentangle the responsiveness to each of these.[9] Second, the vehicles I observe that have odometer readings were purchased in 2001 to 2004, well before the gasoline price shock. If we assume imperfect foresight, such that consumers in 2001 to 2004 did not make vehicle choice decisions based on the upcoming shocks, then the analysis avoids a second possible selection issue. Specifically, consumers who anticipate much higher gasoline prices may be inclined to purchase a more efficient vehicle, and thus the driving response would be attenuated. This could be a particular issue for purchases of vehicles in 2006, when gasoline prices had

[9]Incidentally, in the descriptive analysis, it does not appear that dropping economic conditions makes much of a difference to the estimated responsiveness. This likely applies to the structural estimation as well.

already started increasing. Fortunately, under the assumption of imperfect foresight, my variation should be free of this possible selection issue.

Identification of the responsiveness in this model also benefits from a few other features of the dataset. I observe many important observables, such as commute times, demographics, and vehicle characteristics. Thus, I can condition to these observables to help avoid possible confounding of the responsiveness by these important factors. The richness of the dataset also permits including model fixed effects in the vector of vehicle characteristics to capture unobserved quality attributes associates with each vehicle model (i.e., to account for the ξ_j as in Berry, Levinsohn, and Pakes (1995)).[10] Finally, different vehicle models have become available over time. Thus the choice set exogenously changes from the staggered timing of new model introductions. Such staggered timing may be particularly useful for identifying the responsiveness to gasoline prices in vehicle choice.

4.3.3 Estimation Strategy

Estimation of the structural model developed here may be possible using a variety of methods. I estimate the model using two methods: Maximum Simulated Likelihood (MSL) and Markov Chain Monte Carlo (MCMC). Consistent with the Bernstein-von Mises Theorem, I find equivalent coefficients using these two methods on a very small subsample of my dataset, and for speed reasons I have opted to use MSL for the coefficient estimates presented in this paper.[11] I will thus describe my MSL approach here.

To find the optimal vector of coefficients, I use the Broyden-Fletcher-Goldfarb-Shanno algorithm alternating with the Newton-Raphson algorithm at every ten iterations.[12] MSL is consistent under the assumption that the number of simulated draws

[10] The results in this dissertation do not contain model fixed effects for computational time reasons, but future work is planned to include these.

[11] Once reasonable starting values are known, I find that MSL is able to converge more quickly than the time it takes to do enough repetitions of MCMC to be reasonably confident that I am drawing from the posterior distribution.

[12] I alternate between these two algorithms to reduce the chance that the conditional likelihood function interacts with the rules of each algorithm in such a way that the algorithm "gets stuck." I use the built-in algorithms in Mata (Stata's interpreted language).

R increases at a faster rate than the number of observations (Train 2003). This is a theoretical concern that suggests it is important to use enough draws. I find that with 5,000 draws, adding more draws seems to make little difference. I have to make a similar modeling choice about the definition of the choice set $\mathcal{J}_i$. I define the choice set for consumer i to be all vehicle types purchased in the same quarter by anyone in the same county. This definition allows for an extremely rich choice set (e.g., in many counties it contains over 2,500 vehicle types) and at the same time the quarter restriction prevents consumers from having the choice of vehicles that were not yet available. The county restriction is included primarily for computational reasons, but it also corresponds to the idea that in more remote areas of California, there are not many dealers and thus not all choices are readily available. Given my research question and dataset, I do not include an outside option in the choice set–the vehicle choice decision is assumed to be made conditional on having already decided to purchase a new vehicle.

Finally, to perform the estimation using the full dataset of 12.6 million observations, I impute missing income and VMT values based on the full set of covariates. For VMT, I focus on vehicles that receive a smog check within two months of six years, to avoid a selection issue based on the time of the smog check. All vehicles with a smog check at other times have VMT coded to missing. I can check the fit of the imputation by basing it on 90% of the observed data and examining how well the remaining 10% are fit. Both the mean and the standard deviation of the imputed 10% appear to match those for the observed 10% quite well, and the fit of the models is high (e.g., an R-squared around 0.7 for both). Of course, imputing data using regression may bias my standard errors for it increases the number of observations without increasing the information available for the estimation. One possible improvement upon this approach is to use a data augmentation technique to impute the missing values simultaneously along with the estimation, either using the method laid out in Erdem, Keane, and Sun (1999) or the standard fixed point data augmentation approach described in detail in Tanner and Wong (1987).

4.4 Results

4.4.1 Parameter Estimates

The estimates of the coefficients in the full structural model are listed in Table 4.1. These results are from an estimation with 1% of the full sample randomly drawn.[13]. Since the purpose of this study is to examine the consumer responsiveness to gasoline price changes and not to identify the discount rate that consumers use in making vehicle purchase decisions, the results here have the discount rate fixed at 7% to facilitate solution of the model.[14] Whether consumers "undervalue" fuel savings relative to other decisions made in the market (which would correspond with studies finding a high implicit discount rate) remains unclear in the literature. As described in Chapter 1, the studies that have attempted to answer this question have come to very different conclusions. For this study, I follow Busse, Knittel, and Zettelmeyer (2010) and Sallee, West, and Fan (2009) in assuming that consumers fully value fuel economy, and save further exploration of a present-bias in automobile purchases for future work.

Most of the coefficient estimates in Table 4.1 are highly statistically significant. The parameter estimates from the structural model do not have a simple interpretation, but the signs of the coefficients and relative magnitude within each type of parameter (e.g., within γ_1 or within γ_2) can be interpreted. I will discuss a few of these coefficients that appear to be of the most interest, beginning with period-one γ coefficients (γ_1), which correspond to the effect on period-one utility just from owning the vehicle with the attribute (i.e., not including the utility from driving the vehicle). Positive attributes of vehicles, such as cylinders and liters lead to a higher valuation of the vehicle simply sitting in the driveway. Convertibles are often considered fun to own, and also have a positive value for γ_1. Hybrids appear to have a very high value relative to other attributes. Imported vehicles also add greatly to the utility of the vehicle purchase. Having a higher safety rating has the same effect, possibly

[13]This keeps the computation time down to under a week on my 2.667 GHz Intel i7 processor (four core) computer

[14]Assuming the total second period is six years, this leads to $\delta_2 = 0.56$.

Table 4.1: Structural Model Parameter Estimates

Period 1 (Purchase) γ coefficients			β coefficients (Driving)		
γ_1 cylinders	80.83***	(9.7)	lease	-0.008**	(0.003)
γ_1 wagon	-1,515.06***	(429.1)	zip density	-0.164***	(0.001)
γ_1 SUV	130.50***	(21.6)	commute	0.196***	(0.002)
γ_1 pickup	859.29***	(5.1)	zip pop growth rate	0.050***	(0.001)
γ_1 convertible	778.30***	(101.5)	zip % age >65	-0.231***	(0.038)
γ_1 turbo	-138.54*	(62.9)	zip % age <18	0.157***	(0.007)
γ_1 luxury	-522.84	(922.1)	zip % white	0.151***	(0.071)
γ_1 roadster	-778.12	(1004.7)	zip % black	-0.098***	(0.012)
γ_1 four wheel drive	-492.08***	(65.2)	zip % hispanic	0.108	(0.096)
γ_1 liters	11.32***	(1.3)	log(zip pop)	0.064***	(0.004)
γ_1 auto	583.77***	(91.4)	log(zip income)	-0.005***	(0.000)
γ_1 gvwr	49.72***	(3.7)	log(zip businesses)	0.041***	(0.001)
γ_1 hybrid	2,772.98***	(70.2)	% summer months	0.004***	(0.000)
γ_1 import	494.56***	(92.5)	$30,000 - $39,999	-0.044	(0.036)
γ_1 safety	384.96***	(75.8)	$40,000 - $49,999	0.090***	(0.008)
γ_1 fuel economy	39.30***	(8.7)	$50,000 - $74,999	0.100***	(0.002)
Period 2 (Driving) γ coefficients			$75,000 - $99,999	0.196***	(0.001)
γ_2 cylinders	-0.026***	(0.00)	$100,000 - $124,999	0.082***	(0.000)
γ_2 wagon	-0.015***	(0.00)	>$125,000	-0.178***	(0.003)
γ_2 SUV	0.104	(0.08)	Period 1 (Purchase) economic conditions		
γ_2 pickup	0.190***	(0.05)	CCI	0.272***	(0.109)
γ_2 convertible	-0.106*	(0.06)	unemployment	-0.132***	(0.014)
γ_2 turbo	-0.041	(0.08)	Period 2 (Driving) economic conditions		
γ_2 luxury	-0.061	(0.04)	average CCI	0.052	(0.064)
γ_2 roadster	-0.019	(0.04)	average unemployment	-0.040***	(0.004)
γ_2 four wheel drive	0.120***	(0.01)	Parameters		
γ_2 liters	0.110***	(0.00)	λ	0.027***	(0.000)
γ_2 auto	-0.150***	(0.00)	ω	0.013	(0.052)
γ_2 gvwr	-0.082***	(0.01)	σ	4.5	
γ_2 hybrid	0.053***	(0.00)			
γ_2 import	0.050***	(0.00)			
γ_2 safety	-0.085***	(0.02)			

*** indicates significant at 1% level, ** at 5% level, * at 10% level
Robust standard errors in parentheses solved for using the Hessian

due to the greater peace-of-mind from having a more safe vehicle. The coefficient on four-wheel-drive is negative, perhaps because conditional on a vehicle being an SUV or pickup, having four-wheel-drive does not add significantly to the utility of just owning the vehicle.

The rest of the coefficients determine the benefits consumer i derives from driving. We can begin with the γ_2 coefficient estimates. The coefficient on convertible suggests that actually driving a convertible lowers the marginal benefits of driving, even though simply the act of owning a convertible provides positive utility. The coefficient on four-wheel-drive suggests that consumers who drive four-wheel-drive vehicles receive a greater marginal utility from driving, even controlling for the density of the zip code of registrations.

For the β coefficients, the signs again appear to make sense. Higher commutes tend to imply a higher marginal value of driving. More dense areas correspond to a lower marginal utility from driving. Interestingly, the coefficients on the income brackets of the household purchasing the new vehicle suggest that there is an increasing marginal utility from driving as income increases, with the exception of the $30,000 to $39,999 range and the highest income bracket. The signs of the coefficients on the economic conditions also correspond with what would be expected. The σ coefficient was fixed in these results to 4.5, a value that was found after performing a grid-search to optimize given several likely values of σ and then choosing the one giving the highest likelihood.[15]

The discussion of these estimates so far has focused simply on the statistical significance and the signs of the coefficients. The next sections provide insight into the economic significance of these coefficients.

4.4.2 Elasticity Estimates

To calculate estimates of the elasticity of driving and fuel economy with respect to the gasoline price, I run a counterfactual analysis of a marginal increase in the price of gasoline. With the coefficient estimates in Table 4.1, I find an average elasticity of

[15]Bootstrapping would allow me to calculated the standard error for σ.

driving with respect to the price of gasoline of -0.15, where the average is taken over all personal vehicle purchases in the dataset. As before, this estimated elasticity is probably best interpreted as a medium-term elasticity, where consumers have time to change the routes taken to work and whether or not to take long driving trips, but do not make larger decisions about where to live.

This elasticity estimate falls within the range of much of the previous literature, as is described in detail in 1. It is useful to put these results in the context of some recent studies using time series data that have suggested that the elasticity is much closer to zero, such as Small and Van Dender (2007) for the utilization elasticity and Hughes, Knittel, and Sperling (2008) for the overall gasoline demand elasticity. Lin and Prince (2009) find a similar result for the utilization elasticity using traffic count data from California. These studies are often are used as evidence supporting the contention that the "rebound effect" from improving CAFE standards is quite small, for the low elasticity suggests that consumers do not respond much to an change in the price of driving. The elasticity estimate from my structural model, while closer to zero than the results in Chapter 3, suggest that when the variation from the 2007-2008 gasoline price spike is used, consumers appear to have responded to the gasoline price changes to a higher degree than these other recent studies would imply. Recall that Knittel and Sandler (2010) use California smog check data from 1996 to 2009 for all vehicles in the vehicle stock older than six years and find that consumers are even more responsive to gasoline price changes than my results suggest (i.e., a utilization elasticity in the range of -0.4). While the specifications differ, one plausible interpretation of this result is that consumers may switch from vehicles that are older to vehicles that are newer (and usually with higher fuel economy).

From looking at the aggregate data, an elasticity in the range of -0.15 appears plausible. Figure 4.1 shows that US gasoline demand has been steadily increasing until the higher prices in 2007 and 2008 made an impact. The decrease off of the trendline is approximately 10%, which, given the magnitude of the gasoline price shock, suggests a gasoline demand elasticity in the range of -0.2 is quite possible, and that elasticity values much larger in absolute value are not likely supported by the data. In the short-run, the gasoline demand elasticity is largely a driving elasticity, so my result provides helpful guidance on what a reasonable bound for the driving elasticity may be.

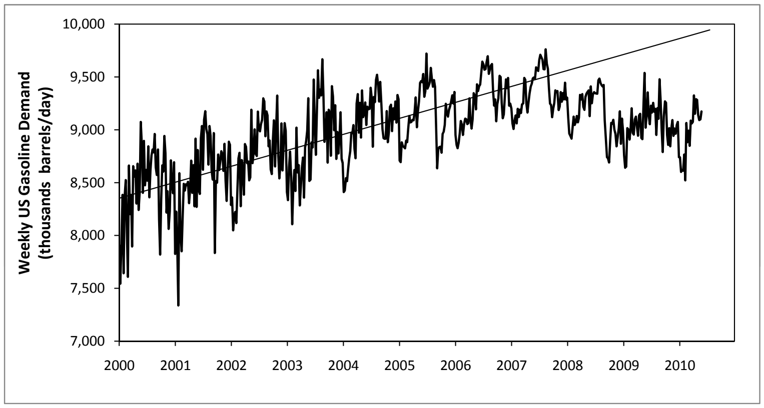

Figure 4.1: Gasoline demand in the US was increasing at a steady pace until the higher prices of 2007 and 2008 made an impact. Source: US Energy Information Administration.

In the longer run, the extensive margin may be quite important for the overall gasoline price elasticity. On the extensive margin, I find an average elasticity of the fuel economy of new personal vehicles with respect to the price of gasoline of 0.10. This elasticity is estimated using data from all manufacturers, including those for which CAFE standards are not a binding constraint at the national level. This elasticity should probably be treated as a short to medium run elasticity, for it abstracts from any shifts by manufacturers to change the characteristics of new vehicles. The estimate is roughly consistent with shifts in the fuel economy of the fleet seen in Busse, Knittel, and Zettelmeyer (2010) and Li, Timmons, and von Haefen (2009). The estimate also corresponds closely to the estimated elasticity of 0.12 in Klier and Linn (2010b) for the period 1970 to 2007, but is less than the estimated elasticity of 0.22 in Austin and Dinan (2005) using data primarily from 2001.

These elasticity estimates provide useful insight into both the short-run and long-run gasoline demand elasticity from personal vehicles, and consequently what the short- and long-run effects of an increased gasoline tax would be. In the short-run, the gasoline demand elasticity should be nearly equal to the driving elasticity. Over time, higher gasoline prices would lead to a shift towards a higher fuel economy fleet, which would reduce the demand for gasoline further. Of course, with a higher fuel economy fleet, the “rebound effect” would cut into some of the gains by inducing some additional driving. This may be offset however by longer term adjustments consumers make in response to higher gasoline prices, such as a move closer to work or to an area with better public transportation. By quantifying each of the components of the gasoline price elasticity, this dissertation helps to fill in the gaps in our knowledge about how the long- and short-run responses to changing gasoline prices occur.

4.4.3 Rebound Effect

The magnitude of the rebound effect is of great interest to policymakers assessing the costs and benefits of tightening fuel economy standards. In Chapter 1, I define of the rebound effect of a policy to improve energy efficiency as the additional energy use due to the decrease in the cost of utilization of the good. I further distinguish between the direct rebound effect, the indirect rebound effect, and the macroeconomic rebound effect. The direct rebound effect is the additional energy use due to increased driving induced by the lower cost per mile of driving. This rebound effect is usually considered to be the more important of the three categories of the rebound effect, and it is the one more commonly categorized.

The structural model estimated in this chapter allows for a unique estimate of the direct rebound effect from a policy that takes into account exactly how consumers will shift vehicles when the prices of vehicles will change. A feebate policy that penalizes vehicles with low fuel economy and provides a rebate for vehicles with a high fuel economy is well-suited for this purpose in this modeling framework. Simplistically, The feebate can be thought of as analogous to a CAFE standard, for the CAFE standard penalizes low fuel economy vehicles with a penalty equal to the shadow price on the constraint. Of course, there are many other details relating to CAFE standards that differentiate it from a feebate, but these are less relevant when we are simply interested in the rebound effect. Chapter 5 discusses many of these details.

There are several ways to think about how to quantify the direct rebound effect from a policy to promote energy efficiency. Ideally, we would like to have a randomized experiment with a two identical groups of drivers: a treatment and control. One way to think about the ideal experiment is simply to increase the fuel economy of all of the vehicles in the treatment group and see how much more the treated population drives. In some sense, this provides a clean measure of the rebound effect. However, this approach may be less useful for policy. What we really want to know in designing a policy is *how driving will increase when the policy is implemented*. Thus, a more useful ideal experiment will implement the policy to promote energy efficiency on the treatment group and then examine how much more the treatment group drives. There is an important distinction between the two experiments, for the policy may have a variety of effects. The policy may raise the price of new vehicles. The consumers who choose to buy different vehicles may be relatively less or more responsive than those who do not. The new vehicles that are purchased contain a bundle of attributes, and households may not want to drive the new vehicles as much.

My approach to quantifying the direct rebound effect incorporates each of these factors. I obtain the estimate of the direct rebound effect through the following steps. First, I estimate the structural model to obtain the primitives. Then, I implement a feebate policy that changes the price of vehicles based on the fuel economy of the vehicle, so that lower fuel economy vehicles become relatively more expensive. I implement such a policy so that it increases the fuel economy of the new personal vehicle fleet by one percent.[16] Then I can examine the resulting driving of the fleet and compare it to the true driving.

I find a direct rebound effect from this small increase in the fuel economy of the fleet of roughly 0.06 or six percent for those vehicle purchasers who made a change of vehicle. This 0.06 value represents the elasticity of driving with respect to the fuel economy of the fleet *when the fuel economy increase is induced by a policy that changes the relative prices of vehicles*. Not all consumers were induced to change a vehicle; many purchased exactly the same vehicle despite the change in vehicle prices. For those who did not change what vehicle they purchased, there is zero rebound effect. However, I find it is more useful to focus only on those consumers who where induced to make a change, rather taking the average over all consumers, in order for the result

[16]Chapter 5 describes the details of implementing a feebate policy.

to be more generalizable. The fraction of consumers that switch vehicles depends on the magnitude of the policy, so each policy would have a separate rebound effect if the average over all consumers is taken.

Of course, this estimated direct rebound effect may not apply for much larger policies, just as any elasticity estimated on the margin may not apply when larger changes are implemented. Yet, this still leaves the question: why does this estimated elasticity of driving with respect to fuel economy differ from the estimated elasticity of driving with respect to the price of gasoline? Both elasticities are estimated from a model that does not include a separate coefficient for both the price of gasoline and the fuel economy, as some of the specifications in Chapter 3.

The difference in the elasticities has two possible explanations. First, when faced with different vehicle prices, consumers will choose a different vehicle. The new chosen vehicle contains a bundle of attributes. The structural model captures how consumers change how much they drive based on the bundle of attributes of the vehicle – and the interaction of this bundle of attributes with the cost per mile of driving. Thus, one way to interpret the difference in elasticities is that consumers would choose to drive the new vehicles less under the policy due to the less desirable bundle of attributes. The second explanation relates to the heterogeneity in responsiveness that was shown in Chapter 3. The 0.15 elasticity of VMT with respect to the price of gasoline is based on the mean over the entire population of new personal vehicles. However, not all new vehicle purchasers would change the vehicle they purchase. If the drivers who are induced to switch to a higher fuel economy vehicle tend to be less responsive in driving than the larger fleet, then I would find a smaller rebound effect.

A combination of both explanations may be at work as well. The first explanation should be invariant to the magnitude of the improvement in fuel economy, while the second may be less relevant once the policy is stringent enough that all new vehicle purchasers are induced to switch to a higher fuel economy vehicle. Exploring the nature of the difference more deeply is an important topic for continued research.

Before moving on, a few caveats are worth mentioning. This estimate of the rebound effect does not take into account any overall increases in the price of new vehicles corresponding with the higher fuel economy – it is entirely based on changing the relative prices of vehicles. This estimate also must be interpreted in the relatively short term, for in the longer term manufacturers may respond to a policy to increase fuel economy in a variety of ways, including introducing entirely new vehicles. Finally, this estimate of the rebound effect does not include any indirect rebound effects or macroeconomic rebound effects. To the extent that these are large, I am underestimating the rebound effect of policies to increase vehicle fuel economy.

4.4.4 Importance of Selection

It is notable that the estimates from the structural model do not exactly match the OLS and fixed effects regression results in Chapter 3. Recall that the results in Chapter 3 suggested that the elasticity of driving with respect to the price of gasoline for new personal vehicles is in the range of -0.17 to -0.25 and the elasticity of the fleet fuel economy with respect to the price of gasoline is in the range of 0.10. The difference between the results may not be too surprising, for the model specifications differ, and in particular, the results in Chapter 3 do not account for selection on the unobserved preference for driving. It is theoretically ambiguous which direction this selection issue works: those who know they are going to drive more may choose to purchase a vehicle with higher fuel economy in order to lower the cost per mile of driving or may choose to purchase a larger, more comfortable vehicle to drive those additional miles in.

Due to the parsimonious nature of the structural model, I can separate the vehicle choice decision from the utilization decision. In particular, (4.12) can be rewritten as follows:

$$VMT_i^* \left(\frac{MPG_i}{p_i^g} \right) = \frac{1}{\lambda} \left(\frac{MPG_i}{p_i^g} \right) + \frac{1}{\lambda} (\beta_c C_i + \beta_d z_i^d + \beta_e E_i + \gamma_2 \theta_i) + \tilde{\eta}_i, \tag{4.12}$$

where $\tilde{\eta}_i = \frac{1}{\lambda} \eta_i$. For simplicity, all variables here are indexed by new vehicle purchaser i, since the specification represents VMT conditional on purchasing vehicle j. This estimation can be performed using ordinary least squares to estimate λ and the β and γ parameters.

The estimation results are given in Table 4.2. By running a counterfactual of a marginal increase in the gasoline price, I find a similar elasticity of driving with respect to the gasoline price as in the descriptive results, albeit slightly lower. The estimated average elasticity over all vehicles in the dataset is -0.21. This result suggests that the structure of the model (and exact variables included) may slightly influence the estimated elasticity, but that explicitly accounting for selection plays an even more important role.

On the extensive margin, the importance of selection appears to be less. This can be observed directly by the small difference between the estimated elasticity in the descriptive results and structural model results. In both sets of results, the elasticity of fuel economy with respect to the gasoline price is in the range of 0.1. To further confirm this observation, I estimate only the vehicle choice equation of the structural model. I again use MSL, where the conditional likelihood function is the same as (4.11) only without the likelihood of driving included. The resulting elasticity from a counterfactual with a marginal increase in driving is again in the range of 0.1, indicating that the joint structural estimation is less important for vehicle choice than for utilization.

4.4.5 Robustness

One of the key assumptions made in the structural model is the nature of consumer beliefs about future gasoline prices and economic conditions. These enter in an important way in (4.7). I perform an estimation where I explore an alternative assumption about these expectations to compare with my baseline assumption of using the current gasoline price and economic conditions as the basis for consumer beliefs about future gasoline prices and economic conditions. Recall that this baseline assumption was chosen because it is consistent with beliefs following a random walk.

The alternative assumption I examine is that consumers use the price of NYMEX futures as the expected future gasoline price. For this estimation, I use the New York Harbor four month NYMEX contract price as the price that all consumers use. I find that this replacement makes very little difference to my results. The coefficients

Table 4.2: Structural Model Coefficients From Intensive Margin-only Estimation

Structural Model Coefficient	Estimate	Standard Error
lease	779.41***	(30.59)
zip density	-308.31***	(16.09)
commute	208.35***	(13.07)
zip pop growth rate	103.26***	(8.73)
zip % age >65	-179.25***	(16.10)
zip % age <18	209.30***	(20.97)
zip % white	223.25***	(18.79)
zip % black	-93.11***	(14.55)
zip % hispanic	78.99***	(21.79)
log(zip pop)	-132.37***	(30.67)
log(zip income)	124.53***	(17.56)
log(zip businesses)	-49.25	(34.91)
% summer months	-125.54***	(15.98)
$30,000 - $39,999	17.48	(54.13)
$40,000 - $49,999	40.72	(50.91)
$50,000 - $74,999	8.96	(39.51)
$75,000 - $99,999	89.93*	(41.68)
$100,000 - $124,999	61.65	(45.03)
>$125,000	-12.76	(46.65)
% summer months	-95.27***	(15.37)
average CCI	-45.60***	(15.09)
average unemployment	653.02**	(311.74)
γ_2 cylinders	-54.06*	(29.89)
γ_2 wagon	-177.27	(285.55)
γ_2 SUV	357.88**	(161.47)
γ_2 pickup	147.15	(183.31)
γ_2 convertible	-2,649.60***	(298.32)
γ_2 turbo	-298.00***	(65.73)
γ_2 luxury	-1,273.59***	(152.83)
γ_2 roadster	-277.28	(191.62)
γ_2 four wheel drive	-112.45***	(29.56)
γ_2 liters	350.44***	(28.69)
γ_2 auto	-422.63***	(90.65)
γ_2 gvwr	-99.24***	(20.00)
γ_2 hybrid	906.88***	(92.72)
γ_2 import	471.47***	(31.10)
γ_2 safety	-55.49***	(14.44)
λ	0.0262***	(0.0004)
Observations	12.3m	

*** indicates significant at 1% level, ** significant at 5% level

Robust standard errors in parentheses calculated using the delta method change slightly, but the elasticity values and welfare calculations remain largely the same. The intuition for this is simple: with the exception of a few short periods, over my time frame the NYMEX futures prices have relatively closely tracked the retail price of gasoline.

There are several other alternative assumptions that can be examined. We could examine the average gasoline price and economic conditions over the previous six months, use an extrapolation of the trend in the past six months (i.e., consistent with a random walk with a drift), or use estimates from a survey of consumer beliefs. One possible survey that includes beliefs is the Michigan Survey of Consumers, used by Anderson et al. (2011a), and incorporating an estimated distribution of beliefs from such a study is a planned future extension of this dissertation work.

Another planned extension for future work is to examine the robustness of the results to separately estimating the model using vehicles from manufacturers who either face binding CAFE standards or do not. This robustness check would be parallel to the check in Chapter 3 section 3.3. We would expect to see much less of a response for vehicles produced by manufacturers who face a binding CAFE standards, as discussed in Chapter 3.

Chapter 5

Policy Implications

The previous chapters of this dissertation quantified the responsiveness to gasoline prices on both the intensive and extensive margins. While quantifying the responsiveness, and the heterogeneity in responsiveness, is interesting in its own right, the primary motivation for such an endeavor is to provide guidance to policymakers. This chapter illustrates what the results imply for policy. It should be viewed as an illustrative exercise in examining the policy implications of the previous results in this dissertation, rather than a full policy analysis. The goal is to illustrate the implications for the energy savings, carbon dioxide emissions reductions, economic efficiency, and distributional consequences of policies to reduce greenhouse gas emissions from the transportation sector.

I perform counterfactual policy simulations for two policies in particular: a tax that increases the gasoline price and a revenue-neutral feebate policy. The feebate policy is chosen to capture many of the key effects that a tightened fuel economy standard would have, for it acts as an implicit tax on low fuel economy vehicles and subsidy for high fuel economy vehicles – changing the relative prices of vehicles. From a broad perspective, these two policies are not strictly comparable, for the gasoline tax policy is a revenue-raising policy that also serves to at least partly internalize environmental, energy security, accident, and congestion externalities from automobile use.[1] The gasoline tax works on both the intensive and extensive margins, by

[1]See Parry and Small (2005) and Harrington, Parry, and Walls (2007) for excellent reviews of the

influencing consumer decisions on vehicle choice *and* driving. In contrast, the feebate policy indirectly aims to internalize the externalities from automobile use by changing the relative prices of vehicles in order to induce consumers to purchase new vehicles with higher fuel economy – thus working only on the extensive margin.[2] In this sense, it is a similar policy to CAFE standards, for the shadow price of the fuel economy standard constraint can be thought of as changing the relative prices of vehicles with different fuel economy. Unlike the gasoline tax, the feebate policy is not generally considered a revenue-raising policy, and the analysis in this chapter examines a revenue-neutral feebate. On the other hand, much like CAFE standards, a feebate policy is often considered to be more politically feasible.

The policy simulations in this chapter are focused on the consumer side of the market and the relatively short-term benefits and costs of the policies. In the long term, decisions made by vehicle manufacturers will be very important in shaping how the vehicle fleet evolves. As shown in Knittel (2010), firms make trade-offs in the choice of the attributes of a vehicle and transportation policies can certainly change this decision-making process. Vehicle manufacturer decisions are particularly important for the analysis of policies targeted at the vehicle choice, for the relative ability of different firms to respond to the policy may have a major impact on the profits of the firms – implying that looking only at the consumer side of the market ignores some important costs. Yet it still provides useful insights into the carbon dioxide emissions reductions and impacts on consumers of such a policy.

The full impacts of a policy on consumers in aggregate depend on both the size of the distortion from altering consumer decisions as well as the avoided damages from the externalities that are reduced. In this chapter I do not attempt to quantify the externalities of driving in California, but rather calculate the deadweight loss to consumers prior to accounting for externalities and compare this to estimates in the literature of the total external damages from driving. Accordingly, throughout the chapter I will refer to the deadweight loss to consumers *absent externalities* to indicate that I am only quantifying the distortion, and not the final social cost of the

externalities in automobile use.

[2] See Sallee (2010) for a very useful overview of issues involved in taxation of fuel economy.

policy.

This chapter uses two methodologies. The first is to use the results of the structural model estimation in Chapter 4 to provide estimates of the change in consumer surplus absent externalities from an increase in the gasoline tax and a feebate policy. The structural model is a utility-consistent framework, so that the utility to consumers in each time period from any policy can be easily calculated. This is most useful for comparing the welfare implications of the gasoline tax policy – a policy that works on both margins – to the feebate policy, which only works on the extensive margin. However, since the structural model was estimated using only the new personal vehicle dataset, it is of limited value for understanding the broader implications of policies for fuel savings and carbon emissions reductions.

Thus, I also develop a simple vintage model of the vehicle fleet to examine how the larger vehicle stock and the driving of each vintage evolve over time with a policy. I use this vintage model primarily to examine the dynamics of how a gasoline tax would affect the fuel economy and driving of different vintages, based on the results of the previous chapters of this dissertation. The vintage model also facilitates the calculation of the energy savings and carbon dioxide emissions reductions from the gasoline tax. To get a rough sense of the distortion from consumers driving less, I can calculate the deadweight loss absent externalities by linearizing the elasticity estimates in the previous chapters of this dissertation and calculating the deadweight loss or Harberger triangle. This can also be easily done accounting for the effect of the pre-existing gasoline tax on the market.

My dataset and previous results are also useful for looking at the geographic heterogeneity in the welfare effects of a gasoline tax policy. Since the results indicate that the responsiveness to gasoline price changes in the amount vehicles are driven is quite inelastic, the primary factor determining the distributional consequences of a gasoline tax policy (prior to redistribution of the revenues) is the amount that households drive. The secondary factor is how households change their driving when gasoline prices change. I directly observe the amount that vehicles are driven and can use my county-level estimates from Chapter 3 for the vehicle-level VMT response to gasoline price changes. From these, I use data on the average number of vehicles per household to calculate household-level estimates of the geographic distributional consequences of a gasoline tax policy.

One finding in this chapter that stems directly from the estimates of the elasticities in the previous chapters is that the price elasticity of gasoline demand is largely a driving elasticity. There is only a small difference between the price elasticity of gasoline demand and the elasticity of driving with respect to the price of gasoline due a response to gasoline price changes in new vehicle fuel economy. This may not seem surprising, for new vehicles only make up a small portion of the entire vehicle fleet, but it does contrast with some of the previous literature.

A second finding in this chapter also is derived from the estimates of the elasticities in the previous chapters. When I examine a policy that leads to a one dollar increase in the price of gasoline, the relatively inelastic response implies that the fuel savings and carbon dioxide emissions reductions are not dramatic. However, the revenue brought in is sizable and the consumer surplus loss absent externalities is quite small. Yet if we examine only the global climate change externality, the cost per tonne of carbon is quite high. Of course, there are other externalities. If we believe the results from the previous literature about the magnitude of all of the other externalities, then the one dollar increase in the gasoline price may still not be economic efficiency-improving, given the currently existing gasoline tax. However it may be a relatively non-distortionary way to raise revenue, and a smaller gasoline tax increase is very likely to be economic-efficiency improving.

A third finding that follows from the estimated elasticities is that the household-level geographic distributional consequences of an increase in the gasoline tax are very strongly driven by the heterogeneity in the amount that households drive in different counties in California. There is a reasonable amount of heterogeneity in driving across counties in California, and this translates to heterogeneity in the distributional consequences of the policy prior to recycling of the revenues. Households in rural areas in California tend to drive more and thus face a heavier burden from a gasoline tax policy. This underscores the importance of carefully redistributing the revenues from any gasoline tax policy. Even if the revenues are returned lump-sum to counties based on the amount paid in taxes, there remains some heterogeneity in household-level distributional consequences across counties. Again, households in rural counties face a relatively heavier burden.

Finally, my exploration of a feebate policy provides an illustration of how the structural model developed in this dissertation can be used to examine how a policy that works only on the extensive margin, such as feebates or CAFE standards, influences consumer welfare. The results also reinforce the finding in Chapter 4 that the rebound effect from a feebate policy is quite small.

This chapter is organized as follows. First, I describe the methodology used to calculate the results. Next, I examine a gasoline tax that increases the price of gasoline by one dollar. I consider both the consumer and government welfare impacts as well as the distributional consequences of the policy. I then examine a revenue-neutral feebate policy and briefly discuss the short-term effects of such a policy. Examining a feebate policy is particularly useful for understanding some of the effects on consumers of fuel economy standards, due to the similarities in how each of the policies work. I discuss these similarities and the relationship of these two policies to gasoline taxes in the final section.

5.1 Methodology

I use both the structural model that was developed in Chapter 4 and a simple vintage model in this chapter. The structural model is a utility-consistent framework that can be used for counterfactual policy simulations to calculate the change in consumer surplus from a transportation policy. In this chapter, I illustrate its use to determine the welfare effects of a gasoline tax and feebate for a particular cohort of new vehicles. To examine the implications of a gasoline tax policy on the entire vehicle fleet, rather than just new personal vehicles, I develop the vintage model.

5.1.1 Structural Model of Vehicle Choice and Driving

The full details of the structural model of vehicle choice and driving are described in Chapter 4. To use the model for policy analysis, I first estimate the coefficients of the model. Once those coefficients are estimated, I can change the values of the parameters based on the policy. For example, I can implement a one dollar gasoline tax policy by increasing the gasoline price that enters into both period one (vehicle choice) and period two (driving) by one dollar. Alternatively, I can implement a feebate policy by changing the price of vehicles according to the fuel consumption (in gallons per mile) of the vehicles. Then I can calculate a variety of results, including the vehicles chosen and the amount driven. This approach to implementing different policies implicitly imposes the historical time frame that my dataset covers as the baseline scenario, and allows for the calculation of counterfactual policy results over that same time frame.

To use the structural model to calculate the welfare changes absent externalities, I use two equations from Chapter 4. Equation (4.4) provides a direct measure of u_2, the utility during period two – the period of driving. It is possible to calculate u_2, since the unobserved preference for driving η can be calculated, for it is simply the residual. The period two consumer surplus change (absent externalities) from a policy is determined by both the amount paid for gasoline and the characteristics of the vehicle driven.

At the time of the vehicle choice, period one, there may also be a consumer surplus change due to a policy, both from the expected utility of driving the vehicle in the future, as well as the cost of the vehicle and its characteristics at the time of purchase. In order to calculate the welfare change (absent externalities) to consumers in period one (i.e., the infinitesimal period of vehicle choice), I use an extension of the approach derived by Small and Rosen (1981). The approach in Small and Rosen (1981) is applicable whenever a Type I extreme value error is assumed in a discrete choice model. The formula I use in the context of my structural model for the change in expected consumer surplus absent externalities is:

$$\Delta\mathbb{E}[CS] = \sum_i \int \left(\log \sum_j \exp(V_{ij}^c) - \log \sum_j \exp(V_{ij}) \right) dF_{\eta_i^k}, \tag{5.1}$$

where V_{ij}^c is the counterfactual representative utility and V_{ij} is the baseline (i.e., calculated directly from my data) representative utility. Both V_{ij}^c and V_{ij} are a function of the known unobserved preference for driving η_i^k, so I integrate over η_i^k to calculate the econometrician's expectation of the change in consumer surplus. V_{ij}^c and V_{ij} can be calculated from equation (4.9) using the parameters of the model and the values of the variables under the counterfactual and baseline.

In order for this calculation to make sense, I must assume that there is a response in vehicle choice to changing gasoline prices, which is only possible if future planned increases in fuel economy standards is not too tightly binding on all manufacturers. This calculation also assumes that there are no pre-existing distortions in the new vehicle market that would interact with the first-period consumer surplus change.

This approach to calculating welfare changes is economically consistent and appealing due to its ability to incorporate several factors that influence consumer utility. I plan to continue using this approach in future work to explore the welfare effects on different cohorts of new vehicles, and perhaps eventually, the entire vehicle fleet.

5.1.2 Vintage Model of the Vehicle Stock

The vintage model is useful to quickly gain insight into what the estimated elasticities in the previous chapters imply for the effects of an increased gasoline tax. Since the structural model is much better suited to looking at a feebate, I do not analyze the feebate with the vintage model, although this would be possible with additional assumptions.

There is a long history of the use of a vintage model to understand the responsiveness of the vehicle fleet to transportation policies, with Sweeney (1979) providing one of the seminal works on the subject. However, most studies in the literature do not consider the vintage nature of the vehicle stock and often simply model the entire fleet fuel economy as a function of the gasoline price (e.g., see Schimek (1996) for a classic example).

The first element of the vintage model is a baseline scenario of driving and fleet fuel economy in California over time. I then use the estimated elasticities from the previous chapters of the dissertation to adjust driving and new vehicle fuel economy as the gasoline price is changed by the policy. I model both scenarios covering the time frame of 2001 out to 2030, with historical data where it is available and extrapolation thereafter.

Model assumptions

The key elements of the vintage model are as follows. For each year in the model, there is an estimate of the number of vehicles by vintage, the amount of driving (in miles per month), and the harmonic mean fuel economy for each vintage (in miles per gallon). I use the year of registration, rather than the model year, as the vintage due to the better data that are available at the year of registration.[3] For the historical number of vehicles by vintage I use three sources: the R.L. Polk data, the smog check data, and the California EMFAC model. The R.L. Polk data on new vehicles (including rental cars, company cars, and government vehicles) provides perhaps the best data available on the number of new vehicles entering the California light-duty vehicle fleet each year. After 2009, I assume that new vehicle sales recover to 2005 levels by 2012 (at the same rate that they declined), and then increase at a rate of 2.5 percent per year, the value that is used in the Argonne National Laboratory VISION model.[4] I assume that there is no response in the number of new vehicle sales to gasoline price changes – only a response in the vehicles chosen. There is a dearth of evidence on whether gasoline price changes influence new vehicle sales, even if there is abundantly evidence that economic conditions change new vehicle sales. I use estimates and forecasts of the entire stock of vehicles in California from the California EMFAC 2007 model to determine the number of older vehicles from

[3]Ideally I could track each model year in each year of registration for several model years are sold in any given. However, this would complicate the analysis and would not likely add much insight for a gasoline tax policy. Future work could expand the model in this direction.

[4]The VISION model is available at: http://www.transportation.anl.gov/modeling_simulation/VISION/.

vintages before 2001 over time.[5]

I use the all vehicle dataset based on the smog check data to get a rough estimate the amount of scrappage of vehicles each year and use the scrappage from the EMFAC model for more recent years. The EMFAC model suggests there is extremely little scrappage in the first few years, most likely only from accidents, so I assume that the fleet decreases by one percent a year for the first four years. After the fifth year, the EMFAC model suggests that stock declines fairly rapidly. I find that both the EMFAC model and my smog check data have a fairly similar path of scrappage of vehicles as they age. The function appears to be roughly quadratic, so I fit a quadratic model and find the estimated model for the fraction of vehicles remaining in the fleet:

$$Fr = 1 - 0.0862Y + 0.0018Y^2,$$

where Fr is the fraction of vehicles remaining, and Y is the number of years after the base year. In my implementation, the base year is the fourth year. After 25 years, the smog check data do not provide any guidance, for smog checks are not required for vehicles with a model year prior to 1975. The EMFAC model suggests that the scrappage of older vehicles continues at a much lower rate. Thus, I assume that the vehicle fleet declines by one percent per year after year 24. Figure 5.1 illustrates this assumed path of scrappage over time for the vintage of vehicles that were first registered in 2001.

I use the same assumed path of scrappage over time for all vintages of vehicles. I also do not model any response to gasoline prices in scrappage behavior in my policy scenario, due to the lack of any empirical evidence on such a response. To the extent that there is a response, such that low fuel economy older vehicles are scrapped more often as gasoline prices rise, the vintage model will underestimate the fuel savings and carbon dioxide emissions reductions from a gasoline tax policy.

The historical VMT estimates come primarily from the smog dataset and are assumed to decline over time following the 2003 data in Table 2.10 in Chapter 2. For

[5] EMFAC 2007 is publicly available at http://www.dot.ca.gov/hq/env/air/pages/emfac.htm. EMFAC is supposed to have historical DMV data over time underlying it, and makes forecasts into the future based on a variety of factors.

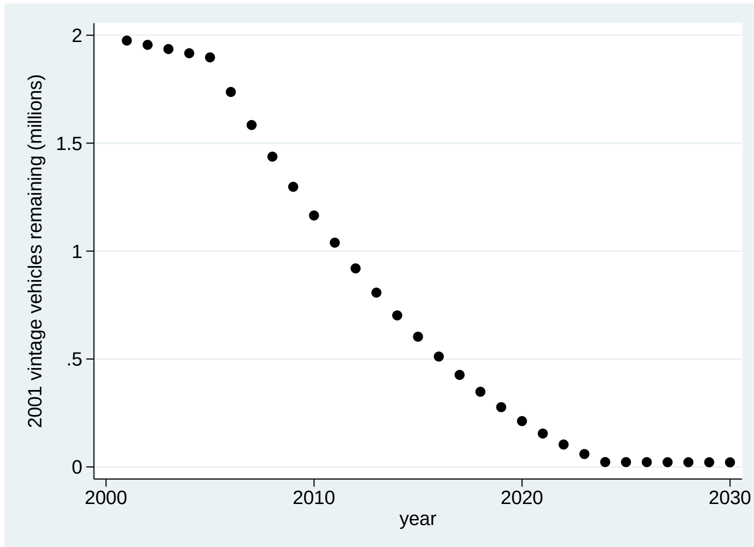

Figure 5.1: This graph indicates the assumed time path of scrappage for vehicles registered in 2001. Other vintages use the same path of decline, only adjusted for the initial value at the year of registration.

the first few years, I use the EMFAC VMT numbers, but adjust them downwards to match the smog data in the fourth year. Figure 5.2 graphically shows the decline over time for the 2001 vintage of vehicles in the model.

After the 2001 vehicle vintage, I assume that driving increases by 0.7 percent each year for the next equivalent vintage, so that, e.g., new vehicles in 2002 drive 1.007 times what new vehicles in 2001 were driven. This increase in driving matches the growth in driving during the early years in my dataset when the gasoline price was flat.

The historical mean fuel economy for each vintage of vehicles comes directly from the R.L. Polk data for new vintages up to 2009. The historical fuel economy of the rest of the fleet comes from the Energy Information Administration (2010). The mean fuel economy in future years depends on assumptions about future policy – in particular future fuel economy standards, which are discussed in the next section.

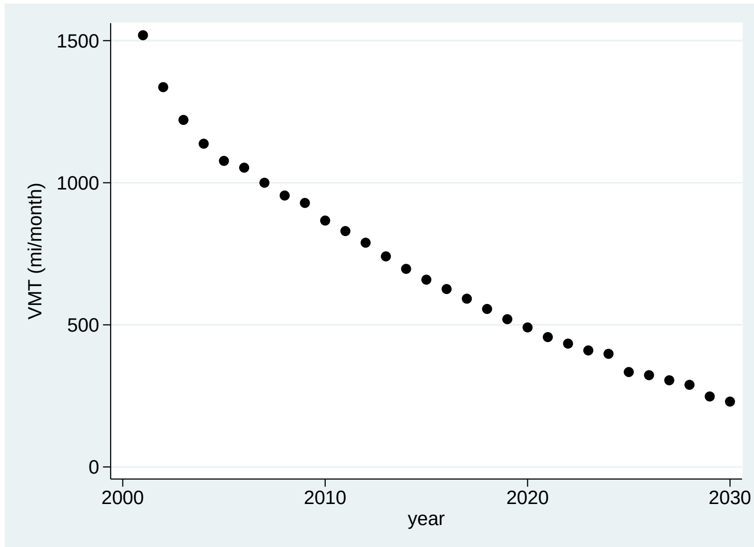

Figure 5.2: This graph shows the assumed path of driving as a vintage of vehicles ages. It is the path used for 2001 vehicles in the vintage model. All but the first four data points are from the smog data. The first four data points are adjusted data from the EMFAC model.

Baseline scenario

I consider the baseline scenario as a best guess of what the future will look like with the current policies that are already planned. Thus, rather than thinking about the baseline as a "no policy whatsoever" scenario, it can be considered a "no additional policy" scenario. The most relevant baseline policy is the increase to the new vehicle performance standards (i.e., CAFE standards and carbon dioxide emissions standards) planned since President Obama's May 19, 2009 initiative. While the exact details of these standards are still in the courts, the planned fleet-wide average fuel economy from the planned regulation is given in Table 5.1.[6] The EPA changed the methodology for rating the fuel economy of new vehicles in 2008 in order to better match common real-world driving conditions. Since CAFE standards are based on the pre-2008 ratings, even though that do not match the achieved fuel economy in real world conditions as closely, Table 5.1 presents the planned standards in both

[6]The 9th Circuit Court of Appeals ruled against the "footprint" based standard, and thus the details are still in the courts.

pre-2008 fuel economy ratings and post-2008 ratings.

Table 5.1: Planned New Vehicle Performance Standards

model year	pre-2008 ratings	post-2008 ratings
2010	25.5	21.3
2011	27.1	22.6
2012	28.5	23.7
2013	29.9	24.9
2014	31.4	26.2
2015	33.0	27.5
2016	34.4	28.6

Source: EPA and DOT (2010)

After 2016, it is unclear what the next standards will be. For the baseline scenario, I assume a fleet-wide average of 35 miles per gallon out to 2020, rather than speculate on what the upcoming rulemakings in both Washington, DC and Sacramento, CA will be. In my data, I also note that California has a slightly higher fleet-wide harmonic average fuel economy than the rest of the nation in the most recent five years. Accordingly, I adjust the California fleet-wide average fuel economy used in the baseline scenario for this difference (a factor of roughly 1.01).

With these quite large increases in fuel economy standards, we can expect there to be a rebound effect. Accordingly, I adjust VMT for new vintages entering the fleet either by my assumptions of the elasticity of VMT with respect to the price of gasoline or by the 0.05 elasticity estimate of VMT with respect to fuel economy in Chapter 3. Given my concerns about my identification of this parameter in Chapter 3, it is quite plausible that the rebound effect is somewhere in between these two values. Fortunately, I find that changing the assumption makes little difference to the counterfactual policy simulation results, largely due to the fact that it just changes the baseline.

I assume that sales of vehicles remain the same, although one could imagine a slight decline in sales if the tighter CAFE standards increase vehicle prices. I consider the assumed 2.5 percent increase in vehicle sales over time to already take this into account.

Given these assumptions, I calculate the baseline total gasoline consumption G in any given year by

$$G = \sum_{i \in \mathcal{I}} S_i \frac{VMT_i}{MPG_i},$$

where i refers to a vintage, $\mathcal{I}$ is the set of all vintages, S_i is the number of vehicles in the stock of vintage i, VMT_i is the average VMT per vehicle for vintage i, and MPG_i is the harmonic mean fuel economy for vintage i. I then calculate the CO_2 emissions from the gasoline consumption using the gasoline carbon coefficient of 8.8 kg CO_2 per gallon of gasoline (US Environmental Protection Agency 2005).

Gasoline tax scenario

The counterfactual gasoline tax scenario is designed to use my elasticity estimates to illustrate how a gasoline tax imposed in 2011 leads to fuel savings and emissions reductions over the baseline in the years following 2011. In the gasoline tax scenario the elasticities enter the vintage model in two places: the VMT for each vintage (intensive margin) and the fuel economy for each vintage (extensive margin).

The fuel economy of each new vintage i is adjusted by the elasticity of fuel economy with respect to the price of gasoline as follows:

$$MPG_i^1 = MPG_i^0 (1 + \beta_{MPG,P^g} \frac{\Delta P^g}{P^g}),$$

where MPG_i^1 is the fuel economy (in miles per gallon) of vintage i under the policy, MPG_i^0 is the fuel economy (in miles per gallon) of vintage i in the baseline, β_{MPG,P^g} is the elasticity of fuel economy with respect to the price of gasoline P^g, and ΔP^g is the change in the price of fuel from the gasoline tax. Thus, if the gasoline price increases, new vintages will have improved fuel economy based on the elasticity of fuel economy.

The higher fuel economy lowers the cost per mile of driving for new vintages, leading to a rebound effect of more driving. The rebound effect works in the opposite direction as the direct effect of higher gasoline prices, which lead to reduced driving.

Thus, the VMT for new vintages has two adjustments under the policy: one for the higher gasoline price decreasing driving, and one for the higher fuel economy increasing driving. The VMT for new vintages after the implementation of the policy in 2011 is given by

$$VMT_i^1 = VMT_i^0(1 + \beta_{VMT,P^g}\frac{\Delta P^g}{P^g})(1 + \beta_{VMT,MPG}\frac{\Delta MPG}{MPG_i^0}),$$

where VMT_i^1 is the VMT under the policy for vintage i, VMT_i^0 is the VMT in the baseline for vintage i, β_{VMT,P^g} is the elasticity of driving with respect to the price of gasoline applicable to vintage i, $\beta_{VMT,MPG}$ is the elasticity of driving with respect to fuel economy, and $\Delta MPG = MPG_i^1 - MPG_i^0$. Different vintages will have a different β_{VMT,P^g}, depending on how old they are and how many years it has been since the policy.

Older vintages before 2011 are not affected by the rebound effect, for their fuel economy remains the same. The VMT for older vintages after the implementation of the policy is thus simply

$$VMT_i^1 = VMT_i^0(1 + \beta_{VMT,P^g}\frac{\Delta P^g}{P^g}).$$

These equations show how the elasticities influence driving and the fuel economy of the vehicle stock. From these, I can solve for cumulative driving, gasoline consumption, and carbon dioxide emissions – and compare these estimates to the equivalent ones in the baseline scenario.

The question of what values to use for the elasticities β_{VMT,P^g}, β_{MPG,P^g}, and $\beta_{VMT,MPG}$ for different vintages remains. Much of this dissertation has focused on β_{VMT,P^g}, so I will begin with the assumptions about this elasticity. I base the assumed responsiveness for different vintages over time on the results in this dissertation, with some smoothing to account for how they change as vehicles age. Table 5.2 shows the assumed elasticities of VMT with respect to the price of gasoline for different vintages over time. In my empirical results in Chapter 4, I find a medium-run elasticity of driving with respect to the price of gasoline for new personal vehicles of about -0.15, and my results in Chapter 3 indicate that the responsiveness is increasing with the age of the vehicle. This motivates my choice of -0.12 and -0.17 for the two-year response for vehicles 0-3 years old and 3-6 years old respectively. I also find in Chapter 3 that the elasticity of driving with respect to the price of gasoline is in the range of -0.3 to -0.5 for older vehicles. Since Chapter 3 does not address the selection issue, which can equally be expected to apply to older vehicles, and Chapter 4 indicates that the selection issue biases the elasticity away from zero, I feel that a reasonable estimate for older vehicles is -0.3 for the two-year elasticity. I fill out the other elasticity estimates accordingly based on my best intuition.

Table 5.2: Assumed VMT Responsiveness

Newer Vehicles (first 3 years):	
1 year response	-0.10
2 year response	-0.12
3+ year response	-0.17
Vehicles 3-6 years old:	
1 year response	-0.15
2 year response	-0.17
3+ year response	-0.25
Rest of the fleet:	
1 year response	-0.2
2 year response	-0.3
3+ year response	-0.4

Assuming a value for the fuel economy elasticity requires even more thought. In Chapters 3 and 4, I estimate a fuel economy elasticity with respect to the price of gasoline of roughly 0.1. However, the estimate was from data covering a time frame when the CAFE standard constraint was less binding in general and not binding at all on some manufacturers. Looking into the future, the CAFE standards have been greatly tightened (as shown in Table 5.1), so it is possible that any change in consumer demand towards higher fuel economy vehicles will just allow firms to re-optimize in meeting the CAFE standards – reducing how tightly they bind. To fully explore this, it would be necessary to develop a model of the supply side of the market.

For this chapter, I assume that the elasticity of fuel economy with respect to the gasoline price either matches my empirical estimate of 0.1 or is zero. For examining the price elasticity of gasoline demand, the 0.1 elasticity case is much more interesting. If the new vehicle fuel economy responsiveness is zero, then the elasticity of gasoline demand will exactly equal the elasticity of VMT demand under my modeling assumption of no scrappage response.

The final assumption to be made is about the magnitude of the rebound effect, or the elasticity of driving with respect to fuel economy $\beta_{VMT,MPG}$. The analytical results in Chapter 1 indicate that in a static sense, if consumers are fully rational, the elasticity of driving with respect to fuel economy should be the same as the elasticity of driving with respect to the cost per mile of driving and the elasticity of driving with respect to the price of gasoline. However, I include a discussion about why there may be a divergence between the rebound effect and the elasticity of driving with respect to the cost per mile of driving or the price of gasoline. For example, one argument is that consumers may respond differently when prices change rapidly than to small changes in the cost of driving – simply because gasoline prices are more salient. The results in Chapter 3 suggest that the elasticity of driving with respect to fuel economy may be less than the elasticity of driving with respect to the price of gasoline, perhaps consistent with this hypothesis. However, in the results in Chapter 3, the fuel economy variable may be endogenous, due to the selection bias discussed in great detail in Chapter 4. The calculations in Chapter 4 of the rebound effect from the small feebate policy also suggest that the rebound effect is small: 0.06. However, this estimate may not be appropriate to apply to all vehicles.

In this chapter, my baseline assumption is a rebound effect equal to the elasticity of driving with respect to the price of gasoline. Given my results suggesting that the rebound effect might be closer to zero, I also examine the effect of assuming a smaller rebound effect, such as an elasticity of driving with respect to fuel economy of 0.1 or 0.05. I find that an assumption of 0.1 makes very little difference to the results and an assumption of 0.05 makes only a minor difference. While the exact numbers may change, the qualitative results remain identical. The intuition for this is simply that for a gasoline tax policy, the direct effect of the higher gasoline price on driving dominates the more indirect effect from the rebound.

5.2 Gasoline Tax

I consider a gasoline tax policy that raises the price of gasoline by one dollar per gallon (in real 2010$). This policy is most easily thought of as a gasoline tax policy, but the policy could equally be a carbon tax that raises the price of gasoline by one dollar per gallon. The actual policy would likely require a gasoline tax increase of greater than one dollar per gallon, for at least some of the burden of the tax would likely be borne by producers. The incidence of gasoline taxes depends on assumptions about the nature of competition in the gasoline market. Previous work by Marion and Muehlegger (2011) indicates that nearly 100 percent of the tax is incident on consumers, suggesting that the supply of gasoline is almost perfectly elastic, which is consistent with a competitive market. Of course, such a high rate of pass-through of the tax also suggests that focusing on the change in consumer surplus and government revenues captures nearly all of the welfare changes. Accordingly, I do not quantify any impacts on producers, but instead focus on the welfare impacts on consumers and government revenue.

In this chapter I also assume that consumers respond to increasing retail gasoline prices the same regardless of whether the source is a tax or price change. I believe this is a very reasonable assumption, although there is some recent work suggesting that there may be a differing response. Specifically, both Davis and Kilian (2011b) and Li, Linn, and Muehlegger (2011) find results suggesting that consumers may respond more to permanent changes in gasoline taxes more than to changes in gasoline prices. However, identifying such a difference is extremely difficult due to many other changes occurring at the same time as increased gasoline taxes. While neither paper currently gives a story for why this might occur, if it is the case, I posit that it could be due to consumers responding to long-term constant changes more than to short-term, possibly ephemeral changes in gasoline prices. This is not what I tend to see in my data, so I am comfortable assuming that the response is the same until stronger evidence on the subject emerges.

To examine the implications of my elasticity estimates for a gasoline tax policy, I perform three analyses. First, I examine how the elasticities of VMT and fuel economy with respect to the price of gasoline are combined into a total price elasticity of gasoline demand. Second, I calculate the consumer surplus change, government revenue, fuel savings, and carbon dioxide savings from the gasoline tax policy. Third, I examine the geographic heterogeneity in several of these estimates in order to shed light on the distributional consequences of the policy. Most of the results in this section are direct implications of the elasticity estimates from the previous chapters, although I also use results from the structural model in Chapter 4 for comparison purposes.

5.2.1 Price Elasticity of Gasoline Demand

In Chapters 3 and 4, I obtain estimates of both the elasticity of driving with respect to the price of gasoline and the elasticity of fleet-wide fuel economy with respect to the price of gasoline. I begin by using these estimates to examine how the two effects may be combined to give a total price elasticity of gasoline demand in order to shed light on the dynamics of fuel savings from a gasoline tax. As mentioned before, this analysis assumes that there is no effect of changing gasoline prices on the old vehicle scrappage decision.

There are two ways to use the estimated elasticities to calculate the total price elasticity of gasoline demand. The first is to derive it straight from the elasticities using the mathematical relationship between the elasticities.[7] The second is to use the vintage model to calculate it. Both provide identical results if given the same inputs. The mathematical relationship is very useful for clearly seeing how price elasticity of gasoline demand is determined. The vintage model is useful for seeing the effects of gasoline price changes on different vintages of vehicles and for calculating the total fuel savings and carbon dioxide emissions from the gasoline tax policy. Both are based on the same fundamental relationships.

I first describe the mathematical relationship between the price elasticity of gasoline demand and both the elasticity of driving with respect to the cost of driving and the elasticity of fuel economy with respect to the price of gasoline. Let G_t be the

[7] I have Jim Sweeney to thank for pointing this out to me.

total gasoline consumption of the fleet in a given time period t. Then $G_t = \frac{VMT_t(C_t)}{MPG_t(P^g)}$, where $VMT_t(C_t)$ is the fleet average VMT at time t, C_t is the average cost per mile of driving at time t, $MPG_t(P_t^g)$ is the harmonic average fleet fuel economy at time t, and P_t^g is the price of gasoline at time t. The cost per mile of driving can also be written as $C_t = \frac{P_t^g}{MPG_t(P^g)}$. For notational convenience, I drop the t subscripts, but it should be recognized that the mathematical relationship refers to the relationship at a particular time t.

Next let β_{G,P^g} refer to the elasticity of gasoline consumption with respect to the price of gasoline, $\beta_{VMT,C}$ is the elasticity of driving with respect to the cost per mile of driving, and β_{MPG,P^g} is the elasticity of fuel economy with respect to the price of gasoline for the entire stock of vehicles in the fleet. β_{MPG,P^g} can be derived from the elasticity of the each vintage of vehicles by

$$\beta_{MPG,P^g} = \sum_{i \in \mathcal{I}} \frac{S_i}{S} \beta^i_{MPG,P^g},$$

where $\mathcal{I}$ is the set of all vintages, S_i is the number of vehicles in the stock in vintage i, S is the total number of vehicles in the light-duty vehicle stock, and β^i_{MPG,P^g} is the elasticity of fuel economy with respect to the price of gasoline the newest vintage of vehicles.

We then can derive the relationship between these three elasticities as follows:

$$
\begin{aligned}
\beta_{G,P^g} &= \frac{dG}{dP^g}\frac{P^g}{G} \\
&= \left[\frac{1}{MPG}\frac{dVMT}{dC}\frac{dC}{dP^g} - \frac{VMT}{(MPG)^2}\frac{dMPG}{dP^g}\right]\frac{P^g}{G} \\
&= \left[\frac{1}{MPG}\frac{dVMT}{dC}\left(\frac{1}{MPG} - \frac{P^g}{(MPG)^2}\frac{dMPG}{dP^g}\right) - \frac{G}{P^g}\beta_{MPG,P^g}\right]\frac{P^g}{G} \\
&= \left[\frac{1}{(MPG)^2}\frac{dVMT}{dC}\left(1 - \beta_{MPG,P^g}\right)\right]\frac{P^g}{G} - \beta_{MPG,P^g} \\
&= \left(1 - \beta_{MPG,P^g}\right)\frac{P^g}{MPG}\frac{1}{(G)(MPG)}\frac{dVMT}{dC} - \beta_{MPG,P^g} \\
&= \left(1 - \beta_{MPG,P^g}\right)\beta_{VMT,C} - \beta_{MPG,P^g} \\
&= \beta_{VMT,C} - \beta_{MPG,P^g} - \beta_{MPG,P^g}\beta_{VMT,C}.
\end{aligned}
$$

This mathematical relationship shows that the elasticity of price gasoline demand is equal to the elasticity of driving with respect to the cost of driving (a negative number), minus the elasticity of fuel economy with respect to the price of gasoline (a positive number), and minus the interaction of these two terms, which can be considered a rebound effect term that captures how higher fuel economy induces more driving.

Recall that in Chapter 1, I showed that the elasticity of driving with respect to the cost of driving $\beta_{VMT,C}$ is mathematically equivalent to the elasticity of driving with respect to the price of gasoline β_{VMT,P^g}. My results in Chapter 3 suggest that the responsiveness to the cost of driving may be less than to the price of gasoline – but these results may be confounded by selection into higher fuel economy vehicles. If the change in the cost per mile of driving is coming about due to changes in gasoline prices, it seems very reasonable to assume that $\beta_{VMT,P^g} = \beta_{VMT,C}$. Thus, I make this assumption when examining how the price elasticity of gasoline demand is determined by the change in driving and change in fuel economy.

For comparing my vintage model to the prediction of the mathematical model, I assume that the rebound effect is the same as the elasticity of driving with respect to the price of gasoline. Then the rebound effect that directly enters into the vintage model has the exact same effect as the rebound term $\beta_{MPG,P^g}\beta_{VMT,C}$. However, the discussion and results in Chapter 3 suggest that the rebound effect may be closer to zero. To the extent that the rebound effect is closer to zero than the price elasticity of driving, then the rebound effect (as an elasticity) should enter the rebound term rather than $\beta_{VMT,C}$.

As described in the previous section, the elasticity of driving with respect to the price of gasoline β_{VMT,P^g} and the elasticity of fuel economy with respect to the price of gasoline β_{MPG,P^g} directly enter into the vintage model. To perform the analysis, I implement a one dollar gasoline tax in 2011. I assume a 2011 gasoline price of $4 per gallon (in real 2011 dollars). The vintage model then provides estimates of the change in gasoline consumption in the policy case, which can be compared to the baseline gasoline consumption in order to calculate the price elasticity of gasoline demand.

I find that the price elasticity of gasoline demand is just a little greater (in absolute value) than the price elasticity of VMT demand, a result that follows intuitively from the mathematical relationship derived above. Using either the mathematical relationship or the vintage model, I can examine the time path of the price elasticity of gasoline demand based on a change in the gasoline price in 2011 and the responsiveness estimates in Table 5.2. Table 5.3 shows the computed time path of elasticities for ten years following 2011. The results are almost exactly the same regardless of whether I use the vintage model of the mathematical relationship.[8] Column (1) shows the fuel economy elasticity *for the entire fleet* over time. This shows how the 0.1 fuel economy responsiveness affects the fuel economy of the entire fleet and helps to make the calculations more transparent. Column (2) shows the fleet-wide price elasticity of gasoline demand with the fuel economy response zeroed out in the vintage model. Column (2) is equivalent to the elasticity of VMT with respect to the price of gasoline since the only response that is occurring is the driving response. Column (3) presents the price elasticity of gasoline demand including the fuel economy response. Column (4) indicates the difference between Columns (2) and (3) (i.e., $-\beta_{MPG,P^g} - \beta_{MPG,P^g}\beta_{VMT,C}$). This difference asymptotes to just under -0.07

[8] Any differences between the two I attribute to rounding error and would not be noticeable in Table 5.3.

as $t \to \infty$. Column (5) shows the fraction of the total response that can be attributed to the VMT responsiveness to gasoline price changes.

Table 5.3: Price Elasticity of Gasoline Demand Over Time

year	(1) fleet FE elast	(2) no FE response	(3) with FE response	(4) difference	(5) fraction VMT response
2011	0.01	-0.16	-0.17	-0.00	0.97
2012	0.01	-0.21	-0.22	-0.01	0.96
2013	0.02	-0.27	-0.29	-0.01	0.95
2014	0.02	-0.26	-0.28	-0.02	0.94
2015	0.03	-0.26	-0.28	-0.02	0.92
2016	0.04	-0.25	-0.28	-0.03	0.90
2017	0.04	-0.25	-0.28	-0.03	0.89
2018	0.05	-0.25	-0.29	-0.04	0.87
2019	0.06	-0.25	-0.30	-0.04	0.86
2020	0.06	-0.25	-0.30	-0.05	0.85
2021	0.07	-0.25	-0.30	-0.05	0.84

Estimates calculated from the vintage model, VMT elasticities in Table 5.2, a 0.1 new vehicle fuel economy elasticity with respect to the gasoline price, and a rebound effect equal to the elasticity of driving with respect to the price of gasoline.

The results in Table 5.3 indicate how the change in the gasoline price initially leads to a more limited responsiveness, followed by increasing responsiveness over time as consumers make more adjustments to driving patterns and higher fuel economy vehicles come into the fleet. There is a small drop in responsiveness after 2015, but this is largely because the fuel economy in the baseline is so rapidly increasing that the fuel savings from the change in gasoline prices is brought down. Without such a steep increase in baseline fuel economy in the first several years, the response is monotonically increasing.

The difference between Columns (2) and (3) is interesting for it indicates the relative importance of the VMT response in the total gasoline price response. The result indicates that the VMT response makes up all but a small part the total response. The difference, shown in Column (4), can be attributed to the response in fuel economy to the gasoline price change. This difference rises over time as the new higher fuel economy vehicles make up a larger fraction of the fleet.

The fraction of the VMT response in Column (5) is included in order to compare my results to the literature. It also shows that the price elasticity of gasoline demand is primarily a response in driving. The fraction begins near one and declines over time. However, even after another ten years, the fraction does not drop much below 0.8. This finding corresponds closely with the result in Bento et al. (2009). It differs from the estimates in Austin and Dinan (2005) and Lin and Prince (2009), who both suggest that the fraction of the response on the extensive margin is greater (i.e., a lower fraction). Similarly, Parry and Small (2005) use an estimate of 0.4 based on a review of the literature.

One difference between the estimate in this dissertation and previous estimates may be that some studies simply use two different data sources to estimate an elasticity of VMT with respect to the price of gasoline and an elasticity of fuel economy with respect to the price of gasoline. For example, this is the case in Lin and Prince (2009). Another explanation may be that some papers attempt to capture the longer term supply-side responses to changing gasoline prices. For example, firms may respond to higher gasoline prices by developing higher fuel economy vehicles in the longer-term, so that the long term elasticity of fuel economy with respect to the price of gasoline is much larger in absolute value. Thus, the fuel economy responsiveness would be even more important in the longer term. Austin and Dinan (2005) and Bento et al. (2009) both attempt to include these supply-side responses using calibrated marginal cost curves from two different sources, which explains the difference in estimates in these two papers. An important caveat, which only Bento et al. (2009) discuss in any detail, is that the response on the extensive margin depends in part on the absence of tightly binding fuel economy standards on all manufacturers.

5.2.2 Effects of Gasoline Tax Policy

Fuel Savings and CO_2 Emissions Reductions

The vintage model of the California vehicle stock allows me to easily calculate the fuel savings and CO_2 emissions reductions from the policy over time. I follow the assumption in the previous section of a $4 gasoline price in 2011 (in 2011 dollars).

Unlike the previous section, I assume a zero fuel economy response in this section. This corresponds to the assumption that the much tightened baseline CAFE standards are sufficiently binding on all manufacturers that even a one dollar gasoline price increase does not leave some manufacturers unconstrained. In order for there to be any response in fuel economy to gasoline prices, some manufacturers must be unconstrained, and this appears unlikely.[9] All of the estimates are based on my California vintage model and thus only apply to California. I assume there is no leakage that would substantially change the results.

The results of the policy leading to a one dollar increase in the gasoline price are given in Table 5.4. I present results for the first 15 years after the policy. Columns (1) and (3) give the gasoline consumption and CO_2 emissions in the policy scenario in millions of gallons per year and millions of metric tons of CO_2 per year. Column (2) shows the difference in gasoline consumption between the baseline and policy scenarios. Column (4) shows the difference in CO_2 emissions between these two scenarios. For comparison, the total 2004 California greenhouse gas emissions were 492 million metric tons of CO_2 equivalent, with 116 of that amount coming directly from gasoline consumption California Energy Commission (2006).

Table 5.4 shows the decline in fuel use over time as the tighter nationwide fuel economy standards take effect and the gasoline tax induces consumers to reduce driving. The fuel savings increase over time at first due to the increasing responsiveness, but then decline slightly due to the tighter fuel economy standards. By 2025, I am assuming that the fuel economy standards are not tightened as much each year, and thus the savings begin slightly increasing again. The fuel savings are entirely from a change in driving, so the results suggest that the gasoline tax policy leads to a decrease in driving of about five to six percent.

The CO_2 emissions reductions follow a similar pattern: increasing and then tailing off over time as the baseline fuel economy standards are tightened. The emissions reductions are not insubstantial, although not very large: they start at roughly five percent of the light duty fleet emissions and increase to closer over eight percent.

[9] However, when I include a small response on the extensive margin, the results barely change. A large response on the extensive margin (e.g., an elasticity of fuel economy with respect to the price of gasoline of 0.1 or greater) does have the potential to change the results.

Table 5.4: Fuel and CO_2 Savings from Gas Tax in California

year	(1) Policy Fuel Use (mil gal/yr)	(2) Fuel Savings (mil gal/yr)	(3) Policy CO_2 Emissions (MMT CO_2/yr)	(4) CO_2 Savings (MMT CO_2/yr)
2011	12,317	564	108.4	5.0
2012	11,962	737	105.3	6.5
2013	11,593	928	102.0	8.2
2014	11,450	905	100.8	8.0
2015	11,316	885	99.6	7.8
2016	11,179	867	98.4	7.6
2017	10,985	875	96.7	7.7
2018	10,834	878	95.3	7.7
2019	10,736	883	94.5	7.8
2020	10,693	889	94.1	7.8
2021	10,704	898	94.2	7.9
2022	10,765	908	94.7	8.0
2023	10,877	922	95.7	8.1
2024	11,042	940	97.2	8.3
2025	11,253	961	99.0	8.5
2026	11,500	986	101.2	8.7

MMT = millions of metric tons. These estimates are calculated with a zero fuel economy response and thus are entirely from a VMT response.

Of course, a one dollar increase in the gasoline price is a substantial increase in the gasoline price due to a tax, so the emissions reductions are particularly not large relative to the size of the policy.

Welfare Effects on Consumers: Background

The welfare implications from the gasoline tax policy are the result of several factors. First, there is the revenue transferred from drivers to the government. Next, there is the distortion from reduced driving. Third, the new vehicle purchase choice may be changed, leading to some consumer surplus loss both at the time of purchase and the time of driving. Fourth, there may be an additional distortions due to the existence of the pre-existing gasoline tax. Last, but most certainly not least, there are the externalities from driving, including global climate change, congestion, accidents and local air pollution externalities. Quantifying these externalities is extremely tricky, for estimates of the damages from global climate change are highly controversial and the external cost of driving from congestion and local air pollution are highly location-specific. Thus, I follow the approach of Bento et al. (2009) and other studies by first quantifying the welfare effects on consumers *absent externalities*, and then discussing how these estimates compare the values in the literature of the external costs of gasoline consumption and driving.

Figure 5.3 provides the basic insight behind quantifying the revenue transferred to the government and the distortion from reduced driving. The graph assumes that the supply of gasoline is perfectly elastic, an assumption consistent with 100 percent passthrough. Rectangle A in Figure 5.3 represents both the additional financial cost to the consumer for the driving they continue to do under the policy, as well as the government revenue. It is simply the size of the tax (e.g., one dollar per gallon) times the number of gallons consumed when the tax is implemented. The smaller triangle B captures the loss to consumers from driving less due to the policy. Absent externalities, this triangle B is the deadweight loss from the policy if there are no pre-existing distortions. Of course, there may be an additional distortion from the tax policy if the government revenue is used for non-economically productive means, rather than returned lump-sum to consumers or used for economically efficient expenditures (e.g., providing public goods).

Among the many assumptions, Figure 5.3 makes a major simplification by assuming that there are no pre-existing distortions in the economy. Externalities are one important pre-existing distortion in the economy. In addition, there is already a gasoline tax, as well as taxes on labor and property. I focus on the pre-existing gasoline tax for it most directly influences the gasoline market.[10] With a pre-existing gasoline tax, the current equilibrium gasoline price and consumption are already different than the pre-tax values, which would be the socially optimal values if there are no externalities. Of course, with externalities, whether consumption is greater than or less than optimal depends on whether the marginal externalities are greater than or less than the pre-existing gasoline tax.

Figure 5.4 indicates how to think about the distortion due to a gasoline tax when

[10]Note that there is a literature about how the other taxes interact with the excess burden from gasoline taxes, which in part depends on whether gasoline is a substitute or complement for leisure (Goulder and Williams 2003; West and Williams 2007).

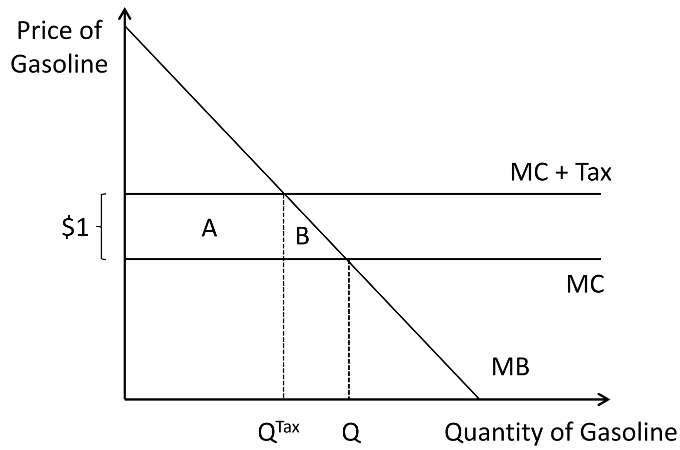

Figure 5.3: This stylized diagram indicates the welfare effects to consumers from the gasoline tax that increases the price of gasoline by $1 (ignoring the externalities of driving). This graph assumes perfectly elastic supply for the ease of interpretation. Rectangle A represents the government revenue and triangle B represents the deadweight loss ignoring the externalities of driving. MC = marginal cost (supply) and MB = marginal benefit (demand)

there is a pre-existing gasoline tax and we are ignoring the external costs. In this case, the additional gasoline tax exacerbates the distortion. "Q" represents what the quantity of gasoline consumed would be if there were no gasoline taxes at all (ignoring all other distortions in the economy), "$Q^{\text{Pre-tax}}$" represents the quantity of gasoline consumed given the pre-existing gasoline tax, and "Q^{Tax}" represents the quantity of gasoline consumed given the additional $1 per gallon gasoline tax. Rectangle A is still the revenues to the government from the tax. However, the deadweight loss absent externalities is now the trapezoid B+C, since the additional tax adds to the distortion from the previous tax (triangle D).

Figures 5.3 and 5.4 provide a very simple view of how to think about and how to calculate the change in consumer welfare from decreased driving absent externalities. Calculating the revenue is straightforward. Calculating the distortion (triangle B or trapezoid B+C) is somewhat more nuanced, for the demand curve may not be linear and the pre-existing tax contains an ad velorum tax component.

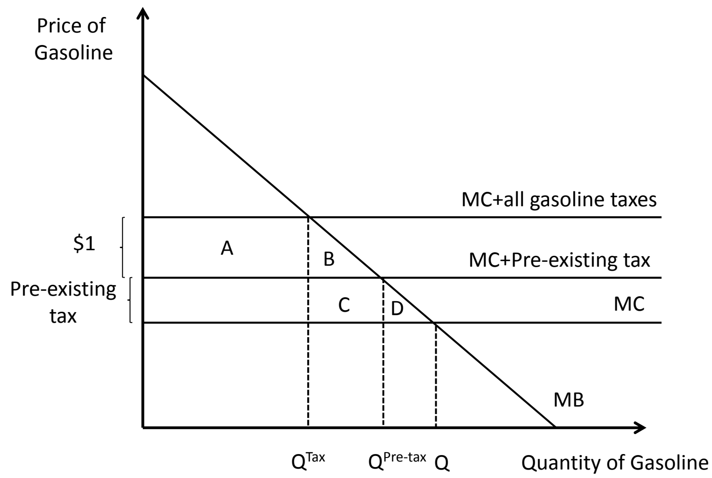

Figure 5.4: This stylized diagram indicates the welfare effects to consumers from the one dollar gasoline tax when there is a pre-existing gasoline tax. This graph assumes perfectly elastic supply for the ease of interpretation. Rectangle A represents the government revenue and the trapezoid B+C represents the deadweight loss from the additional gasoline tax ignoring the externalities of driving. MC = marginal cost (supply) and MB = marginal benefit (demand)

I use two approaches to calculate the size of the triangle. In the first approach, I linearize the elasticity and calculate the triangle manually. Specifically, the triangle B is calculated simply as

$$B = 0.5(\Delta P^g)(\Delta Q) = 0.5(\Delta P^g)(Q\beta_{Q,P^g}\frac{\Delta P^g}{P^g}),$$

where ΔP^g is the change in the price of gasoline (e.g., one dollar per gallon), ΔQ is the change in consumption of gasoline (in gallons), β_{Q,P^g} is the elasticity of gasoline consumption with respect to the price of gasoline (only allowing for an adjustment in driving), and P^g is simply the price of gasoline. This approach can be used for any vehicles in the fleet, assuming that the relevant elasticity is used.

In the second approach, I use the results of the my structural model to examine what the consumer surplus change is for new personal vehicles. The two approaches are not entirely comparable. The structural model inherently allows for a response in new vehicle purchasing to gasoline price changes, for it is modeling both vehicle choice and driving decisions. Accordingly, it provides an estimate of the welfare loss in both period one (i.e., when the vehicle choice is made) and in period two (i.e., during the time of driving). In period one, the consumer surplus loss stems from consumers being incentivized to purchase a less desirable vehicle due to the policy. In period two, the consumer surplus loss is from A+B in Figures 5.3 or 5.4 *and* any loss in consumer surplus from driving a less desirable vehicle. The structural model does not currently include any pre-existing distortions.

Welfare Effects on Consumers: Results

In describing the welfare implications, I begin by discussing the results using my vintage model and then describe some results from the structural model. The structural model results are intended to be illustrative of the capabilities of the model. At the end, I briefly discuss the overall welfare implications when externalities are also considered.

Using the vintage model, I find that the revenue raised from a policy that increases the gasoline price by one dollar is roughly \$500 per vehicle per year. This corresponds to roughly \$11 billion per year from the entire California vehicle fleet. This estimate assumes that 100 percent of the revenues are raised from consumers, corresponding to a 100 percent pass-through. Marion and Muehlegger (2011) suggest that the pass-through is close to 100 percent, but if it is less than 100 percent, some revenue would be raised by producers, and this estimate can be considered a lower bound. For comparison, in 2008 California had a fixed \$0.18 per gallon excise tax and a 7.25 percent sales tax, and these combined taxes brought in approximately \$0.48 per gallon, for a total of \$5.6 billion per year based on California total gasoline consumption from the US Energy Information Administration Energy Information Administration (2010). Thus, my revenue calculations for the one dollar per gallon increase in the gasoline price make sense in light of previous revenue estimates and also suggest that the pass-through to consumers of gasoline tax is near 100 percent.

In the vintage model, I use the linearization approach for calculating the deadweight loss triangle. Thus, the elasticity estimates from the previous chapters directly determine the size of the deadweight loss. I find that the triangle B of the deadweight loss is quite small and varies by vintage over time, along with the variation in responsiveness. In 2011, the deadweight loss triangle is in the range of \$9 to \$16 per vehicle per year, depending on the vintage. In later years, the loss ranges between \$18 to \$27 per vehicle per year. It is important to note that these are vehicle-level estimates, rather than household-level estimates. In 2000-2001, survey results indicate that households in California had almost exactly two vehicles on average (California Department of Transportation 2002). Thus, a very rough estimate of the deadweight loss triangle at the household level is double the vehicle-level estimates.

Before moving to the results for the structural model, it is worthwhile to first calculate an estimate of additional excess burden from the pre-existing gasoline tax distortion. Calculating rectangle C in Figure 5.4 requires assumptions about the gasoline price and the exact sales tax that consumers are currently paying. In 2011, California has a fixed \$0.18 per gallon excise tax and a base 8.25 percent state sales tax. Many cities and counties in California add an additional sales tax, so that local tax-inclusive sales tax rates range from 8.25 percent to above 10 percent.[11] For convenience, I choose an ad velorum sales tax rate of 9 percent and a pre-tax gasoline price of \$3.5 per gallon, which implies an ad velorum tax of \$0.32 per gallon (in 2010 dollars). This implies that the current gasoline tax is \$0.50 per gallon, although this would change with changing gasoline prices.

Accordingly, the height of rectangle C in Figure 5.4 is \$0.50 per gallon. I use the linearization approach to calculate the width. Conveniently enough, since \$0.50 per gallon is exactly one half of one dollar per gallon, the calculations work out such that the area of rectangle C is exactly the same as the area of triangle B. The deadweight loss absent externalities is the sum of triangle B and rectangle C when the previous distortion from the pre-existing gasoline tax is taken into account, so the value of this deadweight loss is just double the value calculated for triangle B. This implies a deadweight loss absent externalities in the range of \$18 to \$32 per vehicle per year, depending on the vintage. In later years the loss ranges between \$36 to \$54

[11] See the California State Board of Equalization for details on sales tax rates by city and county at http://www.boe.ca.gov/sutax/sutprograms.htm.

per vehicle per year. Converting these to household estimate would require roughly doubling them.

The structural model provides a different set of counterfactual policy simulation estimates. While the vintage model is using a forecast starting in 2011, the structural model is currently designed to perform a counterfactual policy analysis over a historical time frame, just like most structural econometric models. Essentially, the question it is designed to answer is: "how would things have been different over the time frame of my dataset if an additional policy had been in place?" Thus, using the structural model, I implement a policy by increasing the price of gasoline for a particular vintage of vehicles, and then observe the change in utility.

For tractability, I examine a single vintage of new vehicles, although I find that other vintages give similar results. I find that over the six-year driving period, the loss in consumer surplus (e.g., u_2) for the 2002 cohort of new vehicles is roughly \$510 per vehicle per year. This includes both the transfer to the government as well as the deadweight loss. It does not include any pre-existing distortions. The per-vehicle revenue is around \$480 per year based on the structural model estimation, so the deadweight loss per vehicle is about \$30 per year. This welfare loss stems primarily from the decreased driving, but also includes any welfare loss from driving a less desirable vehicle that was purchased due to the higher gasoline price. This \$30 per year estimate is slightly larger than the size of the deadweight loss triangle calculated using the vintage model for most years. The difference may be in part due to capturing the additional consumer surplus loss from driving a less desirable vehicle. However, it may also be due to the different time frame and assumptions going into the model. This is a useful area of research to pursue further.

The structural model also allows me to calculate the change in consumer surplus in period one – the time of vehicle choice. This is an infinitesimal period and can be considered a one-time shock to consumer welfare due to purchasing a less desirable new vehicle and having a lower expected welfare in the future driving period. The change in consumer surplus is quantified by equation (5.1). In this calculation, I assume that there are no pre-existing distortions in the new vehicle market that would interact with the first-period consumer surplus change.

I find that for the 2002 cohort of vehicles, the one-time period-one consumer surplus loss is roughly \$3.70 per vehicle on average. Other cohorts are similar. This one-time shock to welfare is quite small relative to the period two consumer surplus loss, for it occurs only once, while the period two welfare loss occurs over the entire six-year period. I attribute this relatively small consumer surplus loss at the time of purchase to the rich choice set in the structural model, which allows consumers to find a slightly more fuel-efficient vehicle that is still quite attractive.

The results so far do not provide a full welfare analysis of the gasoline tax policy, for they do not include externalities in the welfare calculation. Driving imposes global climate change, local air pollution, energy security, congestion, and accident externalities – all of which would lead to more driving than is socially optimal. The currently existing tax on gasoline helps to address these externalities, but may not entirely address them (Parry and Small 2005). Quantifying each of these externalities is fraught with difficulty. To gain some sense of the what values for the externalities we would need to assume for the policy to be economic efficiency-improving, I calculate the cost of carbon dioxide by dividing the deadweight loss by the carbon dioxide emissions savings. I can then compare this to other estimates in the literature of the cost of carbon dioxide and the external costs of the other driving externalities.

I use the linearization approach to calculate the deadweight loss, for it includes all vehicles, rather than just new personal vehicles. By simply dividing the deadweight loss by the carbon dioxide emissions savings, I find that the implicit carbon price is in the range of \$50 to \$60 per tonne of CO_2 during the first ten years of the policy (in real 2011 dollars). If the distortion from the pre-existing gasoline tax is included, these implicit carbon prices would roughly double to range from \$100 to \$120 per tonne of CO_2. This estimate is only appropriate as a carbon price if there are no other externalities. However, the best estimates in the literature suggest that the global warming externality makes up less than half of the total external cost of driving (Parry and Small 2005; Harrington, Parry, and Walls 2007; Litman 2005). Thus, if we are examining only the global warming externlity, the actual cost per tonne of carbon is likely to be less than half of the \$100 to \$120 per tonne of CO_2 range.

While the social cost of carbon is highly controversial, some recent estimates from the economics literature put the social cost of carbon on the order of $20 per tonne of CO_2 (Anderson et al. 2011a; Aldy et al. 2010; Tol 2009). I view this as evidence that a one dollar increase in the gasoline tax may be too much to be economic-efficiency improving – but perhaps a $0.50 increase in the gasoline tax would indeed be economic efficiency-improving, as long as the revenues are returned lump-sum or spent wisely.[12] The revenues could even be used to lower other distortionary taxes, potentially resulting in a double-dividend.

Even if the increased gasoline tax is too high relative to the socially optimal gasoline tax, the highly inelastic response of driving to gasoline price changes suggests that if California is looking to raise revenue, then an increase in the gasoline tax would be a relatively non-distortionary approach, for it would bring in significant revenue at the cost of at most only a small distortion. A full policy analysis, including a sensitivity analysis on the primary assumptions may provide additional insights.

5.2.3 Distributional Consequences

The distributional consequences of the gasoline tax policy are quite important for political economy reasons, especially since the gasoline tax is widely considered to affect rural areas much more than urban areas. This contention is largely borne out the following analysis. I begin by examining the heterogeneity in the burden to consumers ignoring externalities, and then briefly discuss how including externalities could change the story about the degree of heterogeneity in the overall welfare implications.

As discussed above, the change in consumer surplus absent externalities can be roughly thought of as a larger rectangle (e.g., rectangle A in either Figure 5.3 or 5.4) and a smaller triangle (e.g., triangle B in Figure 5.3) or trapezoid when there is a pre-existing tax distortion (e.g., B+C in Figure 5.4). The rectangle, representing the revenues from the tax, is determined directly from the amount that consumers drive (with the policy) and the magnitude of the policy. The triangle is determined by the

[12] Parry and Small (2005) suggest that the optimal gasoline tax for the U.S. is one dollar, so an increase in the California gasoline tax by $0.50 may be in the right range.

responsiveness of consumers to the change in gasoline prices and accordingly is based on my elasticity estimates.

Thus, the starting point for looking at the geographic distributional consequences of a gasoline tax policy is to simply look at the heterogeneity in the amount households drive across counties. Since my data are at the vehicle-level, I can only examine the amount that vehicles are driven. To the extent that households in different counties have different numbers of vehicles, the heterogeneity at the household-level will differ from the heterogeneity at the vehicle-level. Figure 5.5 shows the geographic heterogeneity in vehicle-level driving in California. The per-vehicle monthly VMT varies from less than 700 miles per month to over 875 miles per month.

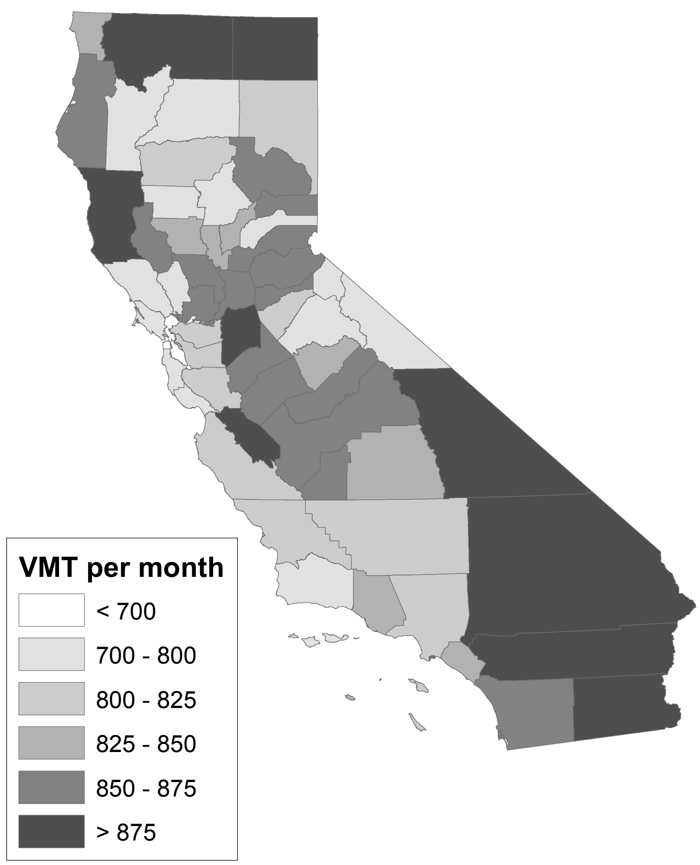

Figure 5.5: Vehicles in California are driven considerably different amounts based on the county.

Figure 5.5 shows that vehicles are driven more in rural areas in California, underscoring that without a careful recycling of the revenues the burden of the policy will be primarily borne by the vehicles in these regions. However, we may be more interested in how households are affected and the number of vehicles per household may vary across counties.

The CalTrans 2000-2001 Statewide Household Travel Survey contains 17,040 households spread across all counties in California, and includes a survey question for the total number of vehicles used by the household (Calfornia Department of Transportation 2002). The county-level mean average number of vehicles per household in California in 2000-2001 is 2.01 and the standard deviation is 0.18. With the exception of San Francisco, which displays an average of 1.3 vehicles per household, all other counties fall in the range from 1.7 to 2.5 average vehicles per household. It appears that more rural counties tend to have more vehicles per household: the Pearson correlation coefficient between the total number of vehicles in the county and the population density of the county (from the 2000 Census) is -0.63. Figure 5.6 more clearly shows the relationship between the number of vehicles per household and the density of the county. There is some heterogeneity in the number of vehicles per household across less populated counties, and an overall very slight downward trend.

The downward trend is seen more clearly in a graph of the county average number of vehicles and the log of the population density (Figure 5.7). Figures 5.6 and 5.7 provide some evidence that more rural (i.e., less densely populated) counties have more vehicles per household, even if the relationship is fairly noisy.

This evidence reinforces the finding that the burden of a gasoline tax policy is likely to be heavier on rural households, for rural households appear to not only face more burden per household, but also have more vehicles per household. Figure 5.8 uses the estimates of the county-level mean average number of vehicles per household to calculate the average amount of driving by household in counties in California. It is clear from the map that households in more rural areas in California drive substantially more than households in less urban areas.

In the legend in Figure 5.8, I also calculate a rough estimate of the average amount each household pays to the government based on all counties having a county fleet

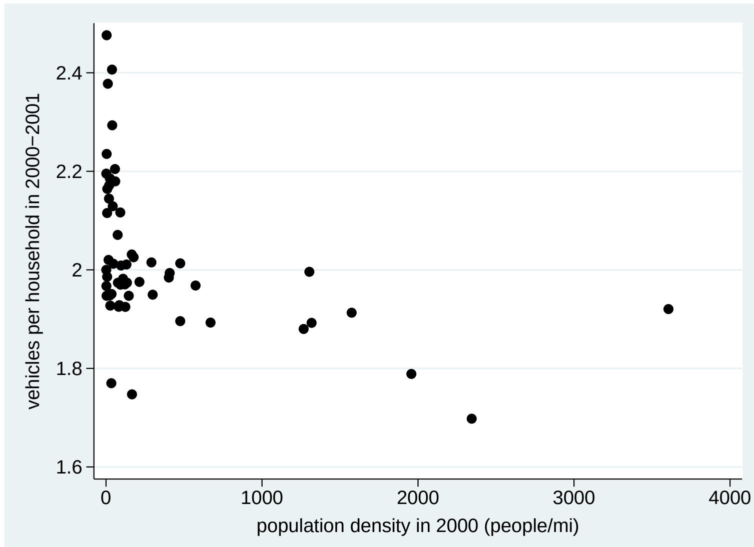

Figure 5.6: The county average number of vehicles per household plotted with the density of the county. All counties in California except San Francisco, which has much higher density and much fewer vehicles per household, are included in this graph. The county with the highest density in this graph is Orange County. Sources: 2000-2001 Caltrans Statewide Household Travel Survey for the vehicles per household and 2000 census for the county-level population density

average fuel economy of 20 miles per gallon. The exact estimate for each county will certainly differ from this estimate due to county-level differences in fuel economy, but the spatial pattern on the map remains identical.

The government revenues make up the largest component of the burden on consumers from the gasoline tax, with the deadweight loss triangle or trapezoid (ignoring externalities) making up the secondary component. The distributional impacts of the gasoline tax policy depend importantly on how the government recycles the revenues. If the government recycles the revenues lump-sum in a way such that all counties receive exactly the revenues raised from them, then the triangle or trapezoid is the only component of the burden to drivers remaining. Since this component is largely based on the slope of the demand curve, quantified at the margin by the elasticity, Figure 3.1 in Chapter 3 provides a sense of the heterogeneity in the triangle per vehicle.

Figure 5.9 indicates how the household-level the deadweight loss trapezoid (absent externalities) varies across counties in California. I use the linearization method to

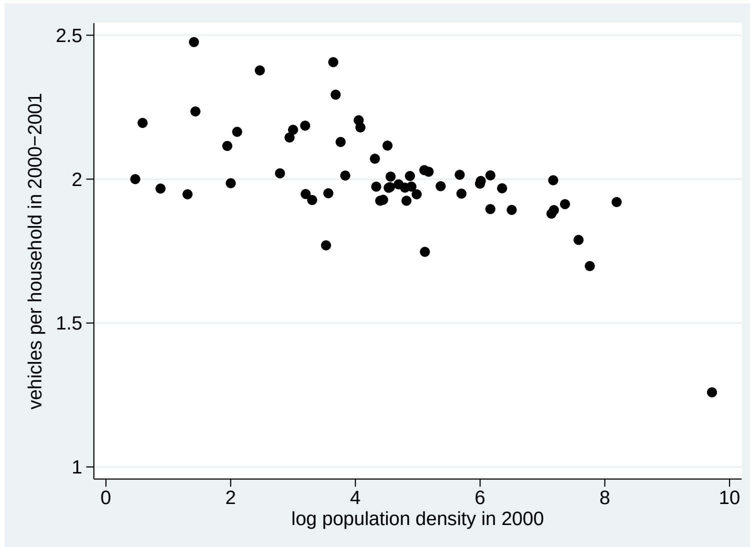

Figure 5.7: The county average number of vehicles per household can more clearly be seen to decrease as the density of the county increases when the log of the density is used. All counties are included in this graph. San Francisco is the outlier with the highest density. Sources: 2000-2001 Caltrans Statewide Household Travel Survey for the vehicles per household and 2000 census for the county-level population density

first calculate the triangle and then add on the small rectangle to find the trapezoid. This approach ignores the welfare effects of any distortion due to consumers being incentivized to purchase a less desirable vehicle, although the previous results in this chapter indicate that this is very much a secondary concern. The elasticity estimates also are based on the new personal vehicle dataset, so to the extent that the elasticities are greater (in absolute value) for all vehicles, I will be underestimating the size of the trapezoid. However, unless the heterogeneity across counties differs for newer vehicles and older vehicles, the spatial pattern of the relative size of the trapezoid should not change in Figure 5.9. Figure 5.9 shows that even if the gasoline tax revenue is recycled to return all tax revenues to the counties from which they are collected, rural counties still tend to be affected the most. The differences across counties are based primarily on differences in vehicles per household, but are partly based on the differences in elasticities across counties.

Comparing the magnitude of the welfare changes in Figures 5.8 and 5.9 also shows

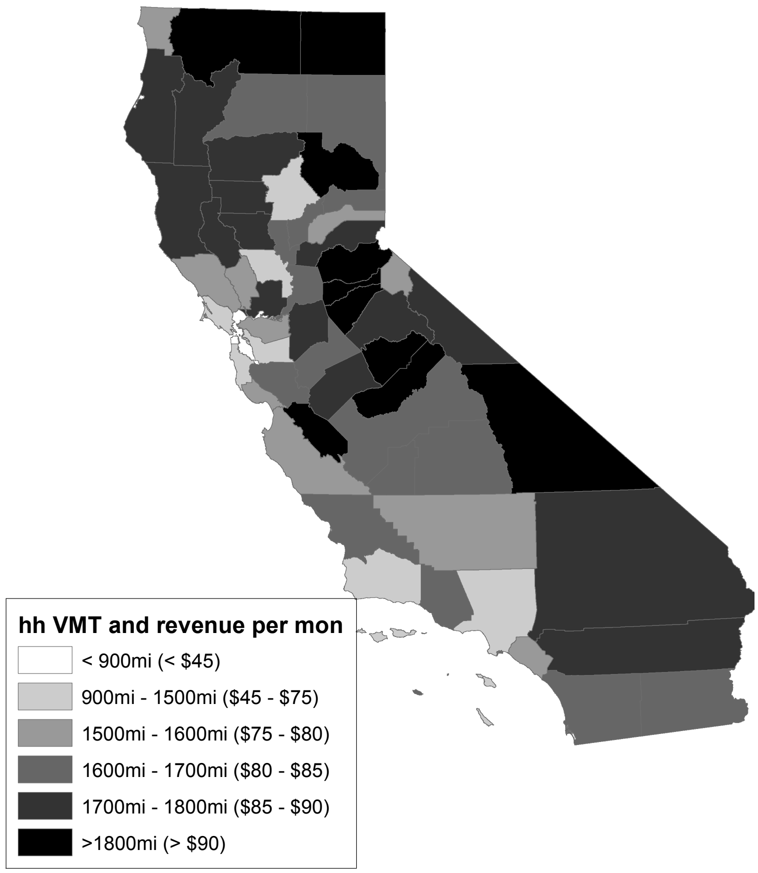

Figure 5.8: Households in rural areas in California drive more on average and thus pay more in gasoline taxes. The legend includes the county-level household driving per month and the average revenues to the government from households in that county from a one dollar gasoline tax (assuming an fuel economy of 20 mi/gal).

that if there is no revenue recycling and the revenues are ignored, the burden on households from the transfer to the government far exceeds the burden due to the trapezoid. Combining these two maps together yields a map nearly identical to Figure 5.8, with only very minor differences.

Up to this point, I have been careful to emphasize that the estimates are all based on the consumer surplus change *ignoring externalities*. For the full welfare implications by county of a gasoline tax, we should also be interested in how the external costs of driving vary by county. For example, one of the important externalities from

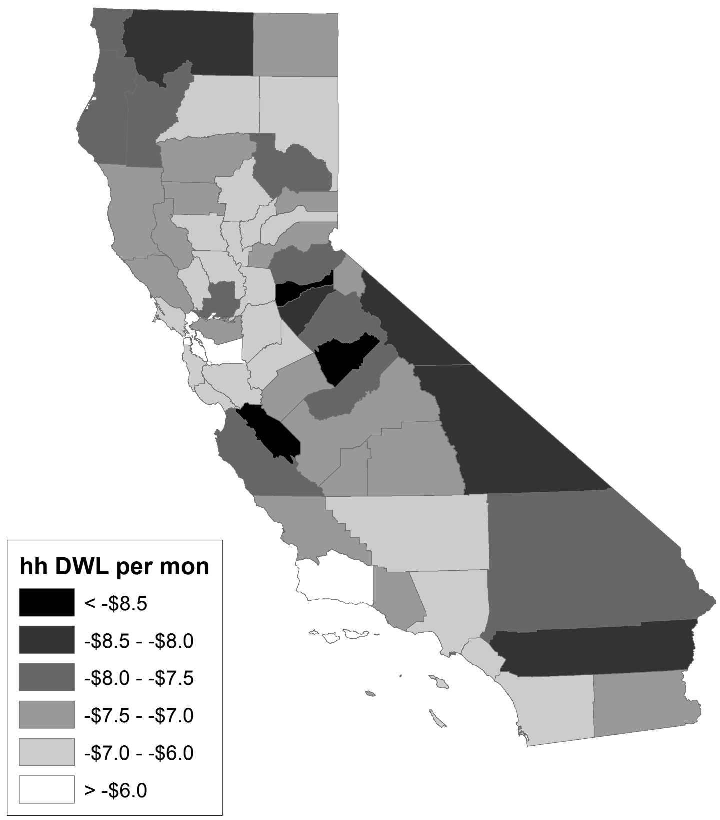

Figure 5.9: The deadweight loss trapezoid (absent externalities) for households varies across counties based on the vehicles per household and the differences in elasticities.

driving is the congestion externality. While a per mile driving charge (i.e., congestion pricing) is the preferred policy instrument to address this externality, the gasoline tax can help to internalize this externality. However, the externality is likely to be far more significant in urban areas than rural areas, implying that the actual deadweight loss *including externalities* may indeed be higher in rural areas than urban areas. Local air pollution externalities depend on where exactly in California the county is, for some of the worst air quality in the United States is in urban Los Angeles County and the more rural San Joaquin County. Greenhouse gases are a global pollutant, so the global warming externality can be considered the same across counties in California. Quantifying all of the externalities (and pre-existing distortions) for every county in

California along with my above results would provide a picture of the overall welfare implications by county.

In addition to differing distributional consequences of the gasoline tax policy by geography, we may also be very interested in how the distributional consequences vary by household income. However, given the considerable difference between the number of vehicles owned by households with different incomes, I view any analysis using only vehicle-level data as less than satisfactory. Thus, I will save this analysis for future work when I have access to household-level data, such as from the Department of Motor Vehicles records.

5.3 Feebate Policy

A feebate policy consists of a tax added to the price of low fuel economy new vehicles and a rebate given to purchasers of high fuel economy new vehicles. There are many different ways that a feebate policy can be structured. All feebate structures must include some "pivot point" fuel economy that marks the change from penalties to incentives. A straightforward feebate would base the size of the tax and rebate on the difference in the rate of fuel consumption (in gallons per mile) between the purchased vehicle and the pivot point. The use of fuel consumption in setting the feebates is preferable to fuel economy because fuel savings are linear in fuel consumption, but non-linear in fuel economy.

The formulation of a feebate based on the difference in fuel consumption from a pivot point is often given with the following simple functional form:

$$F_j = R\left(\frac{1}{MPG_p} - \frac{1}{MPG_j}\right),$$

where F_j is the size of the rebate or tax (in dollars) for a vehicle of type j, R is the *rate* that sets the stringency of the policy (in dollars per gallons per mile), and MPG_p is the pre-defined pivot point.

In practice, one could imagine a feebate that does not change so continuously based on fuel consumption. For example, the feebate may be more of a "doughnut" feebate, where only the most and least efficient vehicles are not penalized or incentivized. The tax credits for hybrids under the Energy Policy Tax Act of 2005 can be thought of as the incentive side of a doughnut feebate: they provided up to $3,400 to hybrid vehicles based on the fuel economy of the vehicles.[13] If there was a tax on very low fuel economy vehicles that penalized vehicles more based on fuel economy, then the combination of the two policies could be considered a doughnut feebate. A doughnut feebate may be easier to administer, but would provide a more limited incentive for consumers to switch to higher fuel economy vehicles.

The pivot point for a feebate can be set based on the expected sales in the fleet so that the policy brings in a specified amount of revenue. For example, it could be set so that the revenue brought in from the penalties exactly offsets the rebates paid out. In this case, the feebate would be revenue-neutral. In some respects, a revenue-neutral feebate policy acts similarly in the short-term to fuel economy standards. The final section of this chapter discusses the similarities and differences.

5.3.1 Effects of a Feebate Policy

My analysis of a feebate policy here is intended to illustrate how the structural model developed in Chapter 4 can be used to provide estimates of the change in consumer surplus of a policy that changes prices on the vehicle choice margin. It is not intended as a full policy analysis, which would calculate estimates of fuel savings, emissions reductions, and welfare impacts on producers. Yet it does provide useful guidance for the costs to consumers of a policy to improve fuel economy on the vehicle choice margin.

To implement the policy, I use the estimated coefficients of the structural model and run a scenario where I change the price of the new vehicle, p_j, in the vehicle choice model. Then I examine the new chosen vehicles, the amount these vehicles are driven, and the welfare implications. An important assumption in this analysis is that the feebate policy is additive with the current CAFE standards, so that implementing a feebate will not just allow manufacturers to re-optimize, but still just meet the

[13] This tax credit expired at the end of 2010.

binding CAFE standard. For manufacturers where the fleet-wide fuel economy for each fleet exceeds the CAFE standard, this is a reasonable assumption. For the other manufacturers this is less reasonable, even if the California fleet-wide fuel economy is above the CAFE standard. This is an important caveat since the estimated values in the structural model are consistent with the historical experience in which a CAFE standard has been in existence and binding for many manufacturers. A full analysis of the interactions between feebates and CAFE standards is a promising area of research, but out of the scope of this dissertation.

Following Greene et al. (2005), I examine a policy of \$50,000 per gallons per mile. This implies that if the pivot point is 25 miles per gallon and the vehicle has a fuel economy of 20 miles per gallon, we would have $F_j = -\$500$. Similarly, if the new vehicle has a fuel economy of 30 miles per gallon, this formulation would suggest a rebate of $F_j = \$333.33$. I set the pivot point at 21 miles per gallon, which brings in only a very small amount of revenue and can be considered largely revenue-neutral.

The feebate policy works by incentivizing consumers to purchase higher fuel economy vehicles. Not all consumers make the switch. For some consumers, the feebate incentive is not enough to change the new vehicle choice. For other consumers, the feebate incentive is sufficient and a different vehicle is purchased. I find that the overall harmonic mean new vehicle fleet fuel economy increases by 15 percent. The resulting decrease in the cost per mile of driving leads to an average increase in driving in period two of about 1 percent for all new vehicles, including those consumers who chose the same vehicle. This corresponds to a rebound effect of about 0.07 for all of these new vehicles. Of course, it also makes sense to examine those consumers who changed the vehicle purchased because of the policy. For these consumers, the harmonic mean fuel economy increases by about 22 percent and driving increases by 3 percent. This corresponds to a rebound effect of about 0.14.

To give a sense of the welfare implications of a feebate policy, I examine a particular vintage of vehicle purchases: all vehicles purchased in 2002. The period-one change in consumer surplus from the policy is again calculated from equation (5.1). The result suggests that the loss in consumer surplus from the feebate policy is \$8.7 million, or roughly \$5.6 per vehicle on average. As before, this can be thought of as a shock to consumer welfare during the time of the vehicle choice. This captures the loss in welfare from choosing a different vehicle than the consumers would have otherwise preferred as well as any expected loss from driving that less preferred vehicle in the future. The fact that this welfare loss is relatively small suggests that there are close enough substitutes with higher fuel economy that the loss to consumers from switching is relatively small.

Over the six-year period of driving (i.e., period two), the change in consumer surplus can be calculated directly from u_2 in the structural model. The result suggest that the consumer surplus change is +\$18 million per year, or roughly +\$11 per vehicle per year on average. This consumer surplus calculation includes several factors: a negative factor from driving a less desirable vehicle, and positive factors from spending less on fuel and driving more. The result is the net of these three factors. Since the result is positive, it suggests that the savings from spending less on fuel and extra utility from driving more overtake the loss in utility from driving a less preferred vehicle. Note that the consumer surplus would differ depending on the six year period the consumers face, and is relatively large and positive for the 2002 cohort at least in part because of the unexpectedly higher gasoline prices in 2007 and 2008. In effect, the results capture the possibility that consumers in 2002 may have had higher ex post utility by being induced into a higher fuel economy vehicle because of the gasoline price increase in 2007 and 2008. Had the gasoline price increase not occurred, the period-two change in consumer surplus would likely still have been positive, but the discounted net present value of the change in consumer surplus would necessarily be negative.[14] Future work can examine the welfare implications when the gasoline price is kept constant at the current price. I anticipate that the period-two change in consumer surplus is likely to be negative if the low gasoline price in 2002 was used rather than the higher gasoline price that actually occurred.

[14]Note that this feature of the model is based on the assumption that consumers trade off consumption in different periods appropriately. This may not necessarily be the case if consumers exhibit a present bias and undervalue fuel economy, as suggested in Allcott and Wozny (2010) and Kilian and Sims (2006).

5.4 Implications for CAFE Standards

While gasoline taxes are a common policy instrument in the United States used to raise revenue, feebates to incentivize higher fuel economy vehicles have not yet been implemented. Instead, the United States has implemented fuel economy standards. Since 1978, when CAFE standards were introduced, manufacturers have met the separate nation-wide sales-weighted harmonic average fuel economy for the passenger vehicle and light duty truck fleets – or paid the "gas-guzzler tax." Figure 5.10 indicates how CAFE standards increased in the late 1970s and early 1980s, but have been largely flat since then. The nation-wide fleet average fuel economy also increased along with the CAFE standards, and more recently along with gasoline prices.

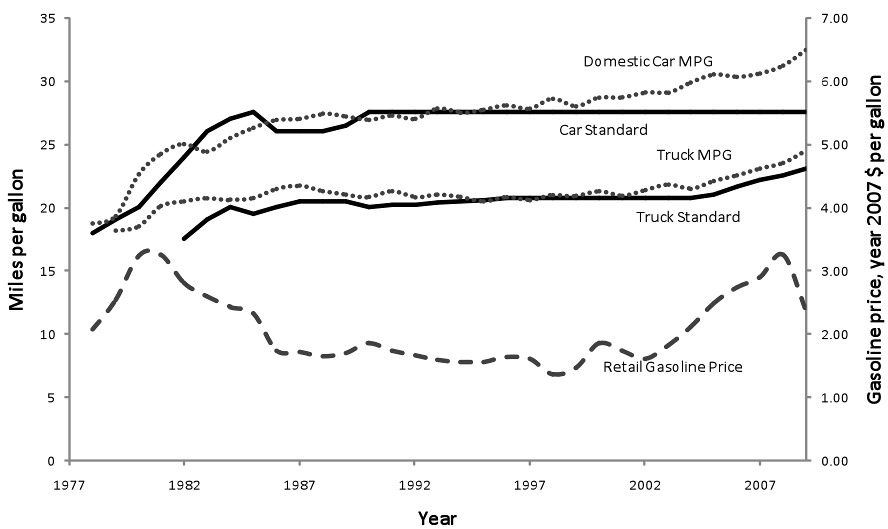

Figure 5.10: CAFE standards increased in the 1970s and then remained largely flat since then. The new passenger vehicle and light truck fleet fuel economy followed a very similar path. Source: Anderson et al. (2011b).

In this section, I focus the discussion on the implications my results have for CAFE standards. For more detailed summaries covering the history of how CAFE standards have been implemented, see Agras and Chapman (1999), Jacobsen (2010), Anderson et al. (2011b) and Anderson and Sallee (2011).

Feebates and CAFE standards are similar in many respects, particularly in the short-run. CAFE standards place a binding constraint on most manufacturers. The shadow price on the constraint penalizes manufacturers for lower fuel economy vehicles based on the relative fuel economy of that vehicle type to the fuel economy of all of the other vehicles in the passenger vehicle or light truck fleet. In an idealized setting, where the CAFE standard is a constraint on the total fleet, and trading of fuel economy credits is permitted across firms, CAFE standards and feebates are nearly identical policy instruments. Both incentivize manufacturers to sell higher fuel economy vehicles in a very similar way. The shadow price on the CAFE standard constraint changes the relative prices of new vehicles in the same way that the feebate changes relative prices. In this sense, analyzing a feebate policy provides insight into the effects of a similar CAFE standard.

However, there are some key differences between CAFE standards and a feebate applying to all vehicles based the difference in fuel consumption from a pivot point. The most striking is that CAFE standards in practice have a separate standard for the passenger vehicle and light truck fleet. Moreover, the next iteration of CAFE standards may base the standards on the vehicle footprint.[15] In both cases, the CAFE standards would have different incentives than a feebate. Similarly, until the 2011 model year, there were no credit trading provisions in CAFE. Without credit trading, the marginal cost of compliance across manufacturers could vary substantially. Under a feebate, all manufacturers face the same incentives and thus the marginal cost of compliance should be equal across all manufacturers.

Another major difference between feebates and CAFE standards is that feebates can be thought of as additive with other price policies in the absence of a CAFE standard. When gasoline prices increase (e.g., from an increased gasoline tax), manufacturers have even more of an incentive to improve fuel economy if there is a feebate. Under a binding CAFE standard, this is not the case. An increase in gasoline prices (or a feebate) may just lead to a lower cost of compliance with CAFE standards, rather than any real increase in fleet-wide fuel economy.

A third difference relates to cost uncertainty versus benefits uncertainty. This

[15] As mentioned above, the 9th Circuit Court of Appeals ruled against the "footprint" based standard, and thus it is unclear whether this type of standard will come to pass.

is analogous to the prices versus quantities argument for policy instruments that dates back to Weitzman (1974). Feebates are a price instrument and thus provide more flexibility to manufacturers. When compliance costs are high, manufacturers can simply produce more low fuel economy vehicles and pay the fees and forgo the subsidies. The opposite is true when compliance costs are low (Anderson et al. 2011b). CAFE standards are a quantity instrument that guarantees that the fleet fuel economy meets a certain standard. Of course, CAFE standards do not entirely guarantee a particular environmental benefit, for the amount of emissions is a function of both the fuel economy and the amount driven. As long as the amount driven does not dramatically change (i.e., the direct rebound effect is small), CAFE standards can provide more certainty about the environmental benefit than feebates. On the other hand, CAFE standard provide no certainty about the cost of compliance.

To the extent that the cost of compliance is higher for CAFE standards than an equivalent feebate, the price of new vehicles may also be higher as manufacturers pass on at least part of the additional cost. This may be a more important factor in the longer term, as the differing design decisions largely determine the differences in the cost of compliance. In the longer term, feebates and CAFE standards may also provide slightly different incentives relating to technological change. Unless CAFE standards are continually changed along with technology improvements, they may provide a weakened incentive over time as technology improves and it becomes increasingly easy to meet the standard. Feebates provide continual incentives for improving fuel economy that do not decrease as technology improves (Anderson et al. 2011b).

Despite all of these differences, CAFE standards and feebates still work in a similar manner by emphasizing only on the new vehicle choice margin. Thus, the results in my feebate analysis at least give a very rough sense of the consumer welfare changes we might expect from a CAFE standard. In contrast to both the feebate and CAFE standards, the gasoline tax works on both the vehicle choice and driving margins, even if it may have a muted effect on the new vehicle choice margin when CAFE standards are present. One could extend the framework I have developed in this chapter to compare the cost-effectiveness of the feebate and gasoline tax by including a supply side so the full welfare implications are known. Any longer run analysis would require a detailed modeling of design decisions and technology change in the auto industry.

One of the arguments given for CAFE standards over gasoline taxes is that they create a stable environment for innovation by providing clear expectations of the value fuel economy improvements. Gasoline prices are highly uncertain, so even with a gasoline tax, automakers investing in technology to improve fuel economy face the risk of low gasoline prices. Feebates perform the similarly to CAFE standards on this measure, by again providing continual incentives for improving fuel economy. The downside of providing this assurance to innovators though a feebate or fuel economy standard is simply that a policy that only works on one margin (i.e., vehicle choice) will inherently be more expensive in the short run than a policy that addresses several margins like the gasoline tax. One would have to believe that the induced-innovation effect is very strong. Quantifying these effects is an important area of future research.

Chapter 6

Conclusions

This dissertation addresses two important questions. First, when gasoline prices change, how much do consumers change the new vehicles they purchase and the amount they drive? Second, what does this imply for policies to reduce emissions from the transportation sector? This work makes several contributions to the literature.

I begin in the introduction by providing a quite comprehensive literature review to provide the context for my results. I also include a discussion to clarify how to think about the rebound effect and how the elasticity estimates in the literature – and in this dissertation – can be interpreted as estimates of the rebound effect.

Chapters 2 and 3 provide evidence on both the magnitude and heterogeneity in the responsiveness of consumers to gasoline price changes in both vehicle choice and driving. I was able to take full advantage of a unique dataset by first showing suggestive evidence of the responsiveness to gasoline prices on both margins, and then using ordinary least squares and fixed effects estimations to provide initial quantitative estimates of the elasticity of both driving and fuel economy with respect to the gasoline price. The choice of the time period of the research study is particularly important, for my data cover a time when gasoline prices increased rapidly, but the economy was still robust – a somewhat unusual occurrence that is extremely useful for separately identifying the effect of gasoline prices from the effect of economic conditions. I also provide evidence indicating the degree of heterogeneity in responsiveness in driving along several characteristics, including income, geography, and demographics. This evidence is new to the literature. I consider the choice of specification very carefully when presenting this regression evidence, yet the driving elasticity results are still subject to the selection bias first noted in Dubin and McFadden (1984). The selection bias implies that the fuel economy (or cost per mile of driving) may be endogenous in a regression where the dependent variable is driving, for consumers may have an unobserved preference for driving that influences the fuel economy of the chosen vehicle.

This selection bias motivates the development of a structural model of both vehicle choice and subsequent driving in Chapter 4. This model has several innovations. It is a two period model that accounts for the dynamic nature of decision-making, so that consumers make decisions based on the price of gasoline at the time of driving and expectations about the price of gasoline. The same is true for economic conditions. The structural model also is the first in the literature to bring in the effect of used vehicle prices on how vehicle choice is affected by gasoline prices. The model is also designed to allow me to take advantage of the rich vehicle-level data that contains revealed preference data on the amount vehicles are driven.

Chapter 5 examines the implications of the results of the previous chapters for policy. I develop a vintage model to convert the estimated elasticities into the total elasticity of gasoline consumption with respect to the price of gasoline. I can then use this vintage model to calculate the fuel savings and carbon dioxide emissions reductions from a gasoline tax policy. I can also calculate the revenue brought in from the gasoline tax and the welfare implications to consumers of the policy. The results in the previous chapters also allow me examine the differing distributional consequences of a gasoline tax by geography, an analysis with important implications for the political feasibility of policies. Finally, Chapter 5 includes a cursory analysis of the welfare effects on consumers of a feebate policy to illustrate how the structural model developed in Chapter 4 can be used for welfare analysis. This chapter brings out just a few of the many possible policy implications that can be examined using the results and framework developed in the previous chapters.

6.1 Summary of Findings

The results of this dissertation first indicate that consumers respond to changes in gasoline prices in both vehicle choice and utilization decisions. However, while there is a clear response, it is still quite inelastic on both margins. I will begin with the intensive margin, i.e., driving decisions. In my empirical results in Chapter 3, which do not account for the selection bias, I find a medium-run elasticity of driving with respect to the price of gasoline in the range of -0.17 to -0.25 for vehicles in the first six years of life. The -0.17 estimate is for all vehicles registered in 2001 to 2004 that had an observed smog check before 2010, while the -0.25 estimate is for only those who had a smog check within a few months of the normal six years. These results appear relatively insensitive to the choice of specification and a variety of robustness checks.

In contrast, I find that older vehicles are much more responsive. Using the smog check data that contains the older vehicles in the fleet who have the required biennial smog check, I find an elasticity of driving with respect to the price of gasoline around -0.5. When year fixed effects are included, the elasticity is around -0.3. In order to roughly compare the elasticity estimated using the six year smog check data on the new personal vehicles and the elasticity estimated using the biennial smog check data on the older vehicles, I examine the elasticity based the odometer readings from the first and last test in the biennial smog check dataset. By using the first and last readings, most tests are several years apart, and in many cases six years apart. I find that using the longer time between tests does not dramatically change the results, and depending on the specification it leads to either a larger or smaller elasticity.

Thus, I view the evidence in Chapter 3 as indicating that the medium-run elasticity of driving with respect to the price of gasoline is around -0.2 for newer vehicles and in the range of -0.3 to -0.5 for older vehicles. Moreover, I find that the elasticity of driving with respect to the cost per mile of driving, rather than the price of gasoline, is closer to zero for the newer vehicles where fuel economy is observed. Similarly, the elasticity of driving with respect to fuel economy is even closer to zero: 0.05.

Importantly, these estimates do not account for the selection bias that could imply that fuel economy, and thus the cost per mile of driving, is endogenous. Thus, the elasticities estimated in Chapter 3 may be biased and inconsistent. When I account for the selection bias using my structural model estimated on the dataset of new personal vehicles, I find a medium-run elasticity of driving with respect to the price of gasoline of about -0.15. I can quantify the importance of selection by estimating the vehicle choice and driving decisions separately using the identical specification as the structural model. The results correspond closely with the fixed effects regression results in Chapter 3, providing evidence that the elasticity of driving with respect to the price of gasoline is biased away from zero if selection into different fuel economy vehicles based on unobserved driving preferences is not accounted for. This is consistent with the story that consumers who know that they are going to drive more will purchase higher fuel economy vehicles to lessen the total cost of driving.

This bias can also be expected to occur for drivers of used vehicles. My structural model is only estimated on the new vehicle dataset, due to the additional richness of the dataset, but one might expect the same bias away from zero to characterize the estimates for older vehicles. Thus, I hypothesize that a reasonable estimate for the medium-run elasticity of driving with respect to the price of gasoline is closer to -0.3.

How do these estimates compare to the literature? Chapter 1 contains a quite comprehensive review of the literature, including detailed tables listing the elasticity estimates in previous studies. Such estimates vary substantially by study, with the elasticity of VMT with respect to the price of gasoline ranging from -0.05 to -0.23 in the short-run and -0.04 to -0.34 in the long-run. If the elasticity of VMT with respect to the cost per mile of driving is examined, the short-run range is largely the same, but the long-run range is widened substantially to as high as -0.87. My results suggesting that the medium-run elasticity of VMT with respect to the price of gasoline is in the range of -0.2 to -0.3 (once the selection bias is accounted for) are certainly in line with the literature over the past three decades.

My results are not consistent with the claim in some recent studies that the elasticity of driving – or gasoline demand – has been decreasing over time as consumers become wealthier and spend more time in traffic (Small and Van Dender 2007; Hughes, Knittel, and Sperling 2008; Hymel, Small, and Van Dender 2010). Each of these recent studies have results suggesting that the short-run driving elasticity with respect to the price of gasoline is already close to zero and perhaps will even more closely approach zero over time as consumer incomes rise and congestion worsens.

There are several ways to view the difference between my results and the results of these recent studies. Besides important differences in methodology and dataset, the time frame in this dissertation was also selected to cover a period when there was considerable variation in the gasoline price, while at the same time controlling for changing economic conditions. The time frame covered in each of the recent studies suggesting that the elasticity is close to zero cover a period of time when gasoline prices were low and not highly variable. Thus, it is plausible that the VMT elasticity results near zero in these recent studies are correct only for time periods when gasoline prices are low and not highly variable. One explanation for lower responsiveness when gasoline prices are low and not highly variable could be that gasoline prices are simply more salient to consumers when they are high and changing rapidly. Unfortunately my time frame does not include period when gasoline prices are high, but not variable, so I cannot disentangle the effect of high prices from the effect of prices that do not change rapidly.

A second difference between the results in this dissertation and those of other recent studies showing estimates of the VMT elasticity with respect to the price of gasoline is the interpretation of the time of adjustment for the elasticities. The elasticities in this dissertation can best be considered medium-run, allowing for a roughly two year time for a response. Other recent studies have shown that the short-run elasticity is very near zero. Thus, the time frame of the elasticity may partly help explain the differences between my results and other studies. However, it is unlikely to entirely explain the difference, for studies such as Small and Van Dender (2007) and Hymel, Small, and Van Dender (2010) also have estimates of the long-run elasticity of either driving or gasoline demand that are also closer to zero than the estimated elasticities in this dissertation.

What do the results in this dissertation imply for the rebound effect of a policy to improve fuel economy of the vehicle fleet? One of the primary uses of these elasticity estimates in the policy sphere is to determine the magnitude of the rebound effect.

Ideally we would like to know the elasticity of driving with respect to a change in fuel economy. In the past, policy analysts have inferred the magnitude of the rebound effect from estimates of the elasticity of driving with respect to the cost per mile of driving, the elasticity of driving with respect to the gasoline price, and even the elasticity of gasoline demand with respect to the gasoline price. If consumers respond to changes in fuel economy in exactly the same way that they respond to changes in gasoline prices, then using the elasticity of driving with respect to the cost per mile of driving is entirely appropriate. However, consumers may not respond the same way, for consumers may respond asymmetrically to increases in the cost per mile of driving, such that there is less response to decreases in the cost of driving than to increases in the cost of driving. Similarly, consumers may find that the slight decrease in the total cost of filling up a gasoline tank of a vehicle from improved fuel economy is less salient than the sign at the gasoline station showing higher prices along with increase in the cost of filling up a tank when gasoline prices increase. The results in this dissertation suggest that such a difference in consumer response may be the case in California, but it is very likely that the fuel economy is endogenous due to the selection bias.

In Chapter 4, I show that the rebound effect from a small feebate policy is around 6 percent for those purchasers who purchased a different vehicle, i.e., the elasticity of driving with respect to fuel economy is 0.06 at the means. This rebound effect is quite small, which may be due in part to the fact that different vehicles have different attributes, and thus are driven different amounts. It may also relate to the heterogeneity in responsiveness to driving.

The evidence of heterogeneity in responsiveness is clear in Chapter 3. Quantile regression results show that there is evidence of considerable differences in responsiveness across the population of consumers. In particular, there is strong evidence that the responsiveness in driving to changes in gasoline prices varies considerably by vehicle class. This may come about due to both within-household switching of driving to a higher fuel economy vehicle and consumers choosing to drive low fuel economy vehicles less when gasoline prices increase. There is also evidence that the responsiveness in driving to gasoline price changes varies by household income. The results suggest that the responsiveness to gasoline price changes is higher for lower incomes and higher incomes than for middle income households, but is the lowest for the wealthiest households (i.e., those who earn more than $125,000 per year). There is also some evidence of heterogeneity in the driving response to gasoline prices along several other dimensions: geography, other demographics, and other vehicle characteristics.

On the extensive margin, I find a short-run to medium-run elasticity of fuel economy with respect to the price of gasoline around 0.1. This response can be interpreted primarily as a demand response, for some manufacturer incentives that influence the price of the vehicle are included in the price variable in the dataset. On the other hand, some supply-side effects may also play a role, for zero percent financing programs and dealer incentives are not included in the price variable and thus these supply effects would be included in the 0.1 estimate. Hence the 0.1 elasticity can be considered an lower bound for the demand response, for some of these incentive programs on the supply side may disappear for high fuel economy vehicles when gasoline prices rise. This result does not change noticeably if I use fuel economy or fuel consumption as the dependent variable. It also does not appear to change appreciably if I used the structural vehicle choice model or fixed effects regressions.

How might this response come about? When gasoline prices change and high fuel economy vehicles sell more quickly, manufacturers can make adjustments to production in a matter of weeks to meet the demand. For example, they can run manufacturing lines for popular high fuel economy models longer by paying workers overtime. Similarly, some manufacturing lines are designed to produce more than one vehicle model, allowing the manufacturer to switch production between the two models. The 0.1 estimate takes into account the very limited design changes that occur over the period of a year or two, but would not take into account longer-term design changes.

The highly inelastic response on the extensive margin is at least in part due to the influence of CAFE standards. For some manufacturers, the CAFE standards are a binding constraint that has an important influence on pricing and design decisions. Others have such high fuel economy vehicles that the CAFE standard does not apply. Still others choose to pay the gas-guzzler tax to allow them to violate the constraint and still sell low fuel economy vehicles. For manufacturers who face a binding CAFE standard constraint, there may be no response in fleet fuel economy to changing gasoline prices at all at the national level. This is simply because if gasoline prices rise and sales of low fuel economy vehicles increase, the manufacturer may simply be able to re-optimize pricing to just meet the CAFE standard constraint. At the California level, this may not be the case, and this may not be the case to the extent that consumers switch from purchasing light trucks to passenger cars, due to the separate standard for each fleet.[1] On the other hand, there may be a response to gasoline prices on the extensive margin for vehicle manufacturers who either have a sufficiently high fuel economy fleet that the standard is not binding or choose to pay the gas-guzzler tax.

The estimate in this dissertation of the responsiveness in fuel economy to gasoline prices corresponds well with the few estimates available in the existing literature. As shown in Chapter 4, the estimates of the short-run elasticity of fuel economy with respect to the price of gasoline range from nearly zero to 0.2. Estimates for the long-run elasticity range from 0.22 to 0.60. Some of these elasticities were estimated over periods that partly cover time before CAFE standards were implemented, and thus they would be expected to be larger. Given this, the medium-run elasticity in this dissertation of 0.1 appears to fit in well with the literature. Whether this result will apply in the future is questionable given that the planned future CAFE standards are sufficiently tight that they very likely will apply to all vehicle manufacturers.

In Chapter 5, I use the estimated elasticities from the previous chapters to show that the price elasticity of gasoline demand is dominated by the elasticity of VMT with respect to the price of gasoline, even if there is a response to gasoline prices on the extensive margin consistent with my 0.1 result. As additional vintages of new vehicles enter the fleet, the elasticity of fuel economy with respect to the gasoline price begins to have a slightly larger effect, although it remains less than 20 percent of the price elasticity of gasoline demand. This result is similar to the result in Bento et al. (2009), but differs from the results in Lin and Prince (2009) and Austin and Dinan (2005), although for different reasons in each case.

[1] With footprint-based standards, there may be even more possibility for a response.

The elasticity findings in the preceding chapters largely determine the policy implications in Chapter 5. The fact that the response to changing gasoline prices is quite inelastic implies that the deadweight loss ignoring externalities to consumers from a gasoline tax is relatively small. I find that a gasoline tax that raises the gasoline price by one dollar per gallon leads to a welfare loss during the time of driving in the range of \$9 to \$16 per vehicle per year in 2011 without accounting for the exacerbated distortion from the pre-existing gasoline tax. In later years this rises to \$18 to \$27 per vehicle per year. When the additional distortion from the pre-existing gasoline tax is included, the deadweight loss absent externalities is roughly doubled, and thus can be considered in the range of \$36 to \$54 per vehicle per year in later years. These results are calculated under the assumption that the planned CAFE standards over the next decade will be tightened sufficiently that all manufacturers will face a binding constraint and thus there will be no appreciable response on the extensive margin.

There are not many studies to compare these results to. One of the few is Bento et al. (2009), who find a welfare loss of roughly \$30 per year per household (in 2001 dollars) for a 25 cent gasoline tax policy with the revenues redistributed lump-sum. Converting this estimate to 2010 dollars and extrapolating to a one dollar gasoline tax implies a welfare loss of just over \$144 per month per household. Since the results in this dissertation are given in per vehicle terms and households in California own just over two vehicles on average, I view the Bento et al. (2009) estimate as slightly above the range of my estimates. Part of the reason for this difference may relate to the fact that the estimated elasticity of driving with respect to the price of gasoline in Bento et al. (2009) is larger (in absolute value) than the estimate in this dissertation.

The deadweight loss estimates presume that there are no externalities from driving. However, driving is associated with many externalities, including damages from global climate change, energy security, accidents, congestion, and local air pollution. If we only consider the global climate change externality, the implied cost of carbon is in the range of \$100 to \$120 per tonne of CO_2. While this estimate is above many estimates of the social cost of carbon, this comparison is not very meaningful without accounting for the other externalities. The literature suggests that the other externalities from driving may be more than half of the total external cost of driving. These other externalities may or may not be sufficient to justify this gasoline tax policy on economic efficiency grounds. If the estimates of Parry and Small (2005) are correct, a one dollar gasoline tax on top of the current gasoline tax of approximately $0.50 per gallon would be above the optimal gasoline tax. The framework developed in this dissertation could be used to more carefully analyze this contention.

A clear result of my analysis is that a policy that increases the price of gasoline by one dollar raises a large amount of revenue for a relatively small distortion. The policy raises roughly $11 billion in revenue per year from the entire vehicle stock or around $500 per vehicle per year. This estimate is roughly double what California raises in revenue from the current roughly $0.50 per gallon gasoline taxes. Thus, if California is interested in raising a significant amount of revenue in a relatively non-distortionary way, the gasoline tax may be a policy to consider.

However, the political feasibility of the gasoline tax policy is an important challenge. Part of the reason why increased gasoline taxes are considered politically unacceptable is that gasoline taxes may affect different regions very differently. Rural areas are generally considered to face a greater burden than urban areas. I find that the geographic heterogeneity in both VMT and the responsiveness of VMT to changing gasoline prices leads to a spatial pattern of distributional consequences of the gasoline tax policy that largely corresponds to this contention. I find that rural areas generally tend to be affected more by the gasoline tax policy than urban areas, but that there is some variation that does not appear to be based entirely on urban-rural differences. This result provides guidance for redistributing the revenues from a gasoline tax in order to assure that different geographic regions are not affected disproportionately.

Unlike the gasoline tax, a feebate policy works only on the extensive margin, and only under the assumption that CAFE standards are not tightly binding. Under this assumption, it saves fuel by increasing fuel economy, but also leads to more driving. I illustrate use of the structural model for welfare analysis by examining a $50,000 per gallon per mile revenue-neutral feebate under the assumption that CAFE standards are not so tightly binding that there is a response on the extensive margin to higher gasoline prices. I find that the feebate increases the average fuel economy of the 2002 new vehicle fleet by 15 percent. Driving increases on average by 1 percent for all vehicles. However, many new vehicle purchasers were not influenced to buy a different vehicle, and thus had no change in driving. When I examine only those vehicles from buyers who were incentivized to purchase a different (i.e., higher fuel economy) vehicle, I find that driving increases by 3 percent. For vehicles purchased in 2002, the consumer surplus loss from the policy is only $5.6 per vehicle on average at the time of purchase, indicating a relatively small welfare loss from the policy. This is a result both of the relatively small magnitude of the policy and the rich choice set that allows consumers to switch to slightly higher fuel economy vehicles with only a limited loss in consumer surplus.

These results enrich our understanding of the effects of a variety of policies, including gasoline taxes, feebates, and CAFE standards. The elasticity and welfare calculations from the gasoline tax provide additional impetus for using the gasoline tax over other measures that only address the vehicle choice margin. On the other hand, the relatively low estimate of the rebound effect suggests that the loss from a rebound from using feebates and CAFE standards may be less important, at least in the medium-run, in a cost-benefit analysis of these policies than has been suggested by some authors. The results also have important implications for a carbon dioxide cap-and-trade system that includes the transportation sector, for the reductions in carbon dioxide emissions that can be expected from the transportation sector very much depend on the elasticity of gasoline demand. A relatively inelastic gasoline demand suggests that achieving significant cuts in carbon dioxide emissions from the transportation sector will likely involve a high carbon price.

All of the analysis performed in this dissertation is focused on the state of California. To what degree do these results apply to the rest of the United States? California is the most populated state in the United States and has a wide variety of demographic and geographic conditions. In some sense, California can be considered a microcosm of the United States. There are certainly some differences though: California generally has a higher fuel economy fleet, and one may argue that there are some idiosyncratic differences between Californians and other Americans due to the regulatory and economic climate that is similar to only a handful of other states in the United States. My view is that an analysis of the responsiveness in driving to gasoline price changes in California can provide very useful insights for policy in the rest of the United States. On the vehicle choice margin, there may be somewhat more important differences due to the nationwide CAFE standard. Despite this, I believe that there are many of the insights gained from the analysis of response to gasoline price on the extensive margin still have broad applicability to the rest of the United States as long as they are taken in context.

6.2 Future Research

There are numerous directions that this research can be taken in the future. Some of these involve relatively minor adjustments or additions to the model or additional analysis. Others are larger areas to tackle that flow logically from the developments in this dissertation. I will begin by discussing planned improvements to the structural model and then will discuss future paper ideas.

The structural model provides a new methodology for addressing the selection bias, and the estimation results appear reasonable and robust. Nevertheless, there are several important improvements and additional analyses that could help bolster the estimation. First, a simple instrumental variables approach, in line with one of the three approaches suggested in Dubin and McFadden (1984), could be used to provide further intuition for the effects of correcting the selection bias. The instruments in this estimation would be the estimated choice probabilities from a vehicle choice model, perhaps even the vehicle choice model developed in the structural model. In addition, this vehicle choice model could be explored further, with calculations of own-price and cross-price elasticities of vehicle shares. Further analysis could also be done to elucidate the features of the structural model that most importantly drive the results. For example, how important are the dynamics? How important is it that used car prices are included in the analysis?

The vehicle choice model could also use several other improvements. Currently, I do not account for the possibility of model-specific unobserved attributes, such as unobserved quality, that may be correlated with the price of the vehicle. Including model fixed effects would entirely alleviate this concern, at a substantial computational time cost. Another specification improvement would be to include an interaction of the price of the vehicle and income, so that the marginal utility of income can vary with income.[2] Another specification improvement is to explicitly incorporate a joint distribution of expectations over gasoline prices and economic conditions into the vehicle choice model, rather than plugging in the current gasoline price (or another price in the robustness checks). The data used by Anderson et al. (2011a) would be perfect for this purpose. A larger and more difficult task would be to use data augmentation to estimate the missing VMT and income as parameters, rather than impute them before estimation of the model. This can be done in a Bayesian framework or perhaps by using the Expectation-Maximization (EM) Algorithm. The Erdem, Keane, and Sun (1999) approach also holds promise. Another similar change that would require considerable work is to include an outside option in the choice set of vehicles. One way to approach this would be to estimate the choice of either a new vehicle or outside option based on data on the total sales in a first stage prior to estimating the choice of new vehicle in a second stage.

A major caveat in the structural model analysis is that it is only estimated using the dataset of new personal vehicles. The selection bias may certainly also arise in an estimation with the dataset that includes the larger vehicle stock. It would be excellent to be able to estimate the model with the larger dataset, perhaps using a randomly drawn sub-sample to keep the analysis tractable. The primary reason for not estimating the structural model on the larger dataset is that the number of variables available is much more limited. With the addition of the rest of the variables, perhaps through the purchase of a VIN decoder, I would have the detailed vehicle attributes and could match a fuel economy to each vehicle. Another key component that would be necessary to estimate the structural model on the larger dataset is detailed knowledge of when vehicles changed title and what the transaction price of the sale was. This is important because the goal of the structural model is to

[2] I have Michael Keane of Arizona State University to thank for this suggestion.

address selection into different vehicles at the time of purchase. With data from the California Department of Motor Vehicles (DMV), this would be possible. However, there may be other important interactions with the used car market that would need to be considered.

The California DMV data would also allow for several additional analyses that are logical extensions of this dissertation. The first is to change the structural model from a vehicle-based model to a household-based model, where the household has a fleet of vehicles and can choose which vehicles to drive. Such an analysis is the next logical step, for it would allow me to directly estimate within-household switching behavior, rather than infer it from my results. Equally important, it would allow me to examine the importance of state dependence in vehicle choice decisions. For example, if a household already has an SUV, how likely is it to purchase another SUV versus a small car? This has obvious marketing implications, as well as implications for policies to promote the diffusion of new vehicle technologies.

The richer DMV data would also allow for a household-level analysis of heterogeneity in responsiveness, which is more useful for understanding the distributional consequences of policy than a vehicle-level analysis. For example, it would be particularly interesting to see whether there is still a U-shaped responsiveness by income by household, so that within-household switching of vehicles will not be affecting the results. The DMV data also has the address of the owner of a vehicle, allowing for spatially explicit analysis of driving behavior. One interesting analysis would be to examine how driving – and the responsiveness in driving to gasoline price changes – would change along with each household's "Walkscore" that captures how walkable a particular neighborhood is. Similarly, it would also be useful to explore in greater detail the effect of proximity to local public transportation options.

Many of the future research topics listed above can lead to separate journal articles. I have several additional ideas as well. One idea is to tackle the issue of aggregation in the estimation of energy demand elasticities. Most of the previous work in the literature is based on highly aggregated data. The dataset used in this dissertation is highly disaggregated. One would imagine that the richer disaggregated data has advantages for estimating elasticities beyond the obvious benefit of a larger sample size. But what are these advantages? How important are they?

Another idea, contingent on access to the DMV data, is to develop a fully dynamic model of vehicle choice in order to better quantify whether and when consumers undervalue future fuel savings from purchasing a higher fuel economy vehicle relative to other decisions in their lives. Such a model would have consumers solving a dynamic stochastic optimization problem where the expected utility of durable good utilization is maximized given the realized gas prices and the expectations of future gasoline prices. With detailed information about what vehicles consumers buy, how much they drive, and when they sell their vehicles, I may be able to model the possible undervaluation of fuel economy better than any of the previous articles in the literature.

Finally, there is much room for extending the framework and results in this dissertation to perform a more expansive policy analysis. A critical addition would be to bring in a supply side into the model to quantify the effects of policies on producers, perhaps in a similar framework to Jacobsen (2010). This would allow for an explicit comparison of the economic efficiency of higher gasoline taxes and higher CAFE standards. Feebates could be examined as well. Equally importantly, my framework holds promise to fully disentangle the differing distributional consequences of gasoline taxes and CAFE standards. This would be a valuable contribution to the literature and at the same time may be of great interest to policy-makers.

Appendix A

Counties in Smog Check Program

There are 58 counties in California, 40 of which are covered by the biennial smog check program. Non-exempt vehicles in all counties are required to have a smog check after four model years, but only vehicles in the 40 covered counties are required to have a biennial smog check after six model years. The covered counties are by far the most populous counties and cover nearly 98% of the population of California. Of the covered counties, six counties do not require smog certifications in select rural zip codes. Below is a list of the counties covered and not covered.

- Counties fully covered: Alameda, Butte, Colusa, Contra, Costa, Fresno, Glenn, Kern, Kings, Los Angeles, Madera, Marin, Merced, Monterey, Napa, Nevada, Orange, Sacramento, San Benito, San Francisco, San Joaquin, San Luis Obispo, San Mateo, Santa Barbara, Santa Clara, Santa Cruz, Shasta, Solano, Stanislaus, Sutter, Tehama, Tulare, Ventura, Yolo, Yuba.

- Counties where not all zip codes are covered: El Dorado, Placer, Riverside, San Bernardino, San Diego, and Sonoma.

- Counties not covered: Alpine, Amador, Calaveras, Del Norte, Humboldt, Imperial, Inyo, Lake, Lassen, Mariposa, Mendocino, Modoc, Mono, Plumas, Sierra, Siskiyou, Trinity, Tuolumne.

Appendix B

Data Merging and Cleaning

B.1 Personal Vehicles During First Six Years

This section describes the merging and cleaning of the personal vehicle dataset in more detail. The foundation for the dataset is the new vehicle registration data at the VIN-level from R.L. Polk. All of the other data sources are merged into this dataset by one or more of the variables. For this dataset, I begin by restricting the dataset to personal vehicles, so that vehicles purchased by rental car companies, other firms, or government entities are not included. For the years 2001 to 2004, just over 80 percent of new vehicles are purchased as personal vehicles. Similarly, the dataset is restricted to vehicles that run on gasoline. Fortunately 97.7 percent of the new vehicles in California run on gasoline in the years the dataset covers, with nearly all of the remainder running on diesel fuel.

B.1.1 Merging registration and smog check data

The most important merge is between the R.L. Polk data and the smog check data from the California Bureau of Automotive Repair (BAR). I have smog check data for all vehicles that received a smog check in California during the years 2003 to 2010. Each vehicle in the smog check dataset is identified by the 17 digit VIN, and the data include such details as the license plate (state and number), test station details, zip code of registration at the time of the test (for vehicles tested after 2007), make of the vehicle, model of the vehicle, vehicle body type, engine cylinders, engine displacement (liters), gross vehicle weight rating,[1] transmission type (automatic or manual), fuel type (gasoline, diesel, natural gas, or electric), odometer reading, pollutant readings (e.g., carbon monoxide, nitrous oxide, carbon dioxide), and overall test result (pass or fail).

The smog check data required a significant amount of cleaning. I corrected many misspellings, used a VIN decoder to check the make implied by the VIN against the make in the dataset, and confirmed that the odometer reading never decreased between consecutive tests. In some cases, the VIN was incorrect in one digit, but could easily be corrected when all of the other digits of the VIN matched the characteristics of the vehicle. In other cases, the VIN was completely incorrect and the observation was discarded. Much of the cleaning was facilitated by converting the dataset to have one observation for each VIN, rather than one observation for each test occurrence as in the raw data.

The matching between vehicles in the smog check data and R.L. Polk data is accomplished through a series of merges. The first merge is based on both VIN and vehicle make. Roughly 60 percent of the R.L. Polk data for 2001-2003 were matched based on VIN and vehicle make. The next merge is based on VIN alone. Another roughly 10 percent of the sample is matched on VIN. Finally, I merge based on vehicle make, model, and county (i.e., either registration county or test county at the time of test must match the registration county). This final merge matches another roughly 6 percent of the full sample. In total, 76 percent of the 2001-2003 personal new vehicle registrations in the R.L. Polk dataset are matched with smog check odometer readings. The remaining unmatched vehicle registrations in the R.L. Polk data can be considered to be either have miscoded VINs or are vehicles that were no longer in existence in California by the time of the first required smog check. The latter vehicles may have moved out of a county that requires a smog check, moved out of California, or were involved in an accident and were scrapped. In addition, roughly 20

[1]The gross vehicle weight rating is the maximum allowable total weight of the vehicle when loaded (i.e., the weight of the vehicle plus the weight of the load).

percent of the 2004 new vehicle registration data are also matched with smog check data from vehicles that were given early smog checks due to a transfer of title outside of the family.

B.1.2 Vehicle characteristics

To facilitate data cleaning, I collapse the R.L. Polk data to create a dataset of all unique "vehicle types" that exist in the new vehicle registration data. A "vehicle type" here is defined by the following characteristics: make, model, model year, series, subseries, engine displacement (liters), engine cylinders, drive type (four-wheel drive/all-wheel drive or two-wheel drive), transmission type (automatic or manual), hybrid electric drivetrain, turbo or super-charger. Other characteristics of the vehicles, such as gross vehicle weight rating, fuel economy, safety rating, body type, and number of doors, are uniquely identified by this classification of vehicle type. I clean the R.L. Polk vehicle types to assure there are no duplicates due to incorrect spelling or reclassifications (e.g., Chrysler being reclassified DiamlerChrysler for some of the years). There are 17,147 different vehicle types in the dataset, covering 55 different vehicle makes, 545 models, and 1,519 series.

All of the vehicle characteristics mentioned above are included in the R.L. Polk data except for the gross vehicle weight rating, transmission type, fuel economy, and safety rating. The gross vehicle weight rating and transmission type are available in the smog check data. To determine the gross vehicle weight rating for each vehicle type, I first aggregate the matched R.L. Polk and smog check data to the vehicle type level, taking the mean weight rating for each vehicle type (with a check performed first to catch outliers). For the remaining vehicle types that are missing a gross vehicle weight, I either use the weight rating from a different subseries with the same make-model-model year-series or manually look up the weight rating from manufacturer websites (this was done at the make-model-model year-series level).

For the transmission type variable, which is coded as an indicator variable for the transmission type being an automatic transmission, the merged R.L. Polk-smog check data are used where available. In these data there is a clear trend whereby fewer vehicles are sold in California with a manual transmission each year. In fact, many vehicles in the dataset are only available with automatic transmissions, and this is more common in the more recent years. This analysis currently assumes all vehicles that have a missing value for the transmission variable (e.g., all vehicles after 2004) are coded as having an automatic transmission, except for models that have only ever been available with a manual transmission. This assumption adds measurement error to any estimates of the coefficient on the transmission type, and thus any coefficient on the transmission type may be viewed as biased. In addition, the fuel economy for vehicles with manual transmissions is often slightly greater than the fuel economy for vehicles with automatic transmissions, so some of the vehicles incorrectly marked may have a slightly greater fuel economy. I deem incurring these minor biases as preferable to omitting the transmission variable altogether. Future work is underway to use a data augmentation technique to base the estimation on only the observed values of transmission type.

The fuel economy data are from the US Environmental Protection Agency's Fuel Economy Guides, issued once a year for each model year. In 2008 EPA changed how the test fuel economy ratings are reported in order to more accurately reflect the fuel economy achieved under real-world conditions (the new ratings are roughly 20 percent less than the pre-2008 ratings). My dataset includes both the pre-2008 and post-2008 ratings, and for the analysis I use the adjusted post-2008 ratings. The EPA fuel economy data are aggregated differently than the R.L. Polk data, and thus an iterative matching process is used to match a fuel economy to each vehicle type in the R.L. Polk data. After significant cleaning, the EPA fuel economy data are matched to the vehicle type data by merging on increasingly aggregated data. The first merge is on "make, model, model year, series, subseries, automatic transmission, drive-type, liters, cylinders, turbocharger, body type, hybrid." This merge matches a fuel economy to roughly 40 percent of the R.L. Polk. The fuel economy data are then aggregated and matched in several iterations. 80 percent of the vehicles are matched by the aggregation "make, model, model year, drive-type, liters, cylinders, body type, hybrid." By the final aggregation "make, model, body type," 100 percent of the vehicle types are assigned a fuel economy. For a sample of the last 20 percent of matches, I look up the fuel economy of the exact vehicle on the manufacturer's website to check the fuel economy. In all cases, the fuel economy was very close to the manufacturer's advertised fuel economy.

The safety rating data used in this study are from the National Highway Traffic and Safety Administration (NHTSA) Safercar.gov website. These data provide an overall safety rating of one to five. This safety rating is analogous to the Consumer Reports and Insurance Institute for Highway Safety (IIHS) ratings, and in fact appears to correspond closely to these ratings from my brief comparison. The NHTSA data list the safety ratings of vehicles covering all years in my sample at the make-model-model year aggregation. Similar to fuel economy, I first match each vehicle type in the R.L. Polk dataset at the most disaggregated level possible and then perform matches at higher levels of aggregation. At this aggregation, 72 percent of the vehicle types are matched. The next aggregation is at the make-model level. After this merge, 88 percent of the vehicle types are matched. To complete the match, I finally merge only by vehicle make. This is final aggregation is rough, but it still allows for vehicles by certain makes, such as Volvo, to have better safety ratings than vehicles by other makes that have poorer safety ratings on average.

B.1.3 Used vehicle prices

Used vehicle prices from the National Automotive Dealers Association (NADA) are available aggregated by "make, model, model year, series, cylinders, body type, region." For this study I use the California average retail transaction price for used vehicles. In matching these data with the R.L. Polk data, I match a vehicle type in the R.L. Polk data with the closest vehicle available that is six years older (in order to capture the used car price that consumers would be expecting to get for their vehicle in six years). The matching is performed incrementally, on increasingly aggregated data, just as was done for the fuel economy data. The first merge by "model year, make, model, series, cylinders, body type" successfully matches about 4 percent. This is largely because the series coding is quite different between R.L. Polk and NADA. The match is much better when the prices are aggregated over series, with 42 percent matched. After aggregating over the cylinders for each model, another 27 percent of the sample is matched. Nearly all of the remaining sample is matched when the used vehicle prices are aggregated over make and model.

In addition to used vehicle prices, the NADA data also includes data on how the price of a used vehicle is adjusted by the odometer reading. I find that there is exceedingly little difference in this adjustment in percentage terms by different makes and models, so all vehicles are given the same adjustment factor, calculated as the average over all vehicles types.

B.1.4 Economic conditions

Finally, I bring in three variables to capture the economic conditions that purchasers and drivers are facing. The Bureau of Labor Statistics provides data on the monthly unemployment in each county in California. These are merged into the full dataset by county and month. Similarly, the Conference Board puts out a national "consumer confidence index" (CCI), which is merged in at the monthly level. The California Association of Realtors has publicly available data on monthly county-level average housing prices 1990 to present.[2] There are many missing values in the dataset, particularly before 2000 and in some of the less populated counties. The data are complete for all of the highly populated counties. Roughly 20 percent of the dataset is missing. For the missing counties, I use the housing values for an adjacent county with similar housing values (which I ascertain from a quick web search to get housing values for at least one year). I adjust the housing values by the ratio of the housing values in the two counties for the year I find. I deem this approximation quite reasonable given that housing values tend to follow regional patterns and that all of the highly populated counties have complete data.

[2] These data can be accessed at http://www.car.org/marketdata/data/housingdata/.

B.2 All Vehicles

This section describes the details of the merging and cleaning process for the larger dataset of all vehicles that have received a smog check. As mentioned in Chapter 2, the advantage of these data is the coverage of the entire stock of vehicles. The disadvantage is the relative lack of detailed information about the vehicle and buyer. For this study, I bring in smog check data from January 2002 through December 2009. The raw data start with on the order of 900,000 smog test instances per month.

I begin the data cleaning process by performing several checks on the VINs. After 1980, VINs have been standardized, such that the ninth position in the VIN is a "check-digit" that can be used along with a hash code and the rest of the first ten positions of the VIN (i.e., the VIN prefix) to check whether a VIN is valid.[3] Similarly, after 1980, the letters "I", "O", and "Q" are not permitted in any position in a VIN, so I can check for these. In addition to the VIN check code validity check, I also drop obviously miscoded VINs that are not picked up, such as "11111111111111111." These VIN validity checks lead me to drop another roughly 6,000 tests per month.

I then fix any miscoded model years, by using the model year indicated by the VIN (I do not have any model year 2011 vehicles, so all vehicles after 1980 are uniquely identified by tenth position in the VIN). I drop all tests with a model years before 1975 (roughly 6,000 per month) for the smog check program does not require vehicles with model years older than 1975 to be tested. I thus infer that an older test is somehow miscoded. I also drop vehicles that are coded with a model year more than two years ahead of the year of the test, for such an occurrence is simply not possible. Less than 500 tests are dropped from this cleaning in each month. Next I clean the many miscoded vehicle makes and some of the miscoded models in the dataset. The dataset was replete with misspellings, so I use the VIN prefix to fix these misspelled or miscoded observations.

In many cases, I found that there would be multiple tests within a month, perhaps because the vehicle failed several of the test, or because the vehicle title was transferred multiple times. There is little additional information from these tests, so I

[3]The details of this calculation are found online in many places, including wikipedia: http://en.wikipedia.org/wiki/Vehicle_Identification_Number#Check_digit_calculation.

drop all tests that have identical odometer readings to the previous test or are within 31 days of a previous test for the same vehicle. This cleaning drops just over 2 million tests per year for most years in the dataset (i.e., roughly 175,000 per month). I then clean the data further by catching situations where the odometer reading goes down in a subsequent tests. This is a clear red flag that something has been miscoded. This cleaning drops roughly 20,000 tests per year (around 1,666 per month).

I also clean the cylinders and liters by making sure the cylinders, liters, and automatic transmission are correct and possible and take the mode within each VIN, breaking ties with the lower one. I do not create a "vehicle type" as in the new personal vehicles dataset, for I only have cleaned R.L. Polk data for a few years. Ideally, I will eventually have access to a full VIN decoder, such as is commercially available from companies like Chrome Systems.

After this cleaning, I merge the smog check data with the R.L. Polk new vehicle data to get additional periods between tests for the same vehicle. This merge is done by VIN. Then I reshape the dataset from one where an observation is a single test, to one where an observation is a time between two subsequent tests for the same vehicle. I then know where and when the vehicle was tested in both tests, and how much the vehicle was driven in between those two tests. Given this form of the dataset, I drop all vehicles that only have one test. This leads to a dataset with roughly 49.7 million observations for tests between 2002 and 2009. As in the "new personal vehicle" dataset, I drop extremely high odometer reading differences that would be extremely unlikely and almost certainly indicates a miscoding. Roughly 4,000 of these extremely high odometer reading differences are dropped per year.

I finally merge in the county of the test and the previous test, the average gasoline price over the time of the test, and the average economic conditions (unemployment rate, housing prices, and Consumer Confidence Index) over the time of the test. The final cleaned dataset has 49.7 million observations.

Appendix C

Vehicle Class Definitions

This section provides a brief description of each vehicle class and lists the vehicle models that are included in each. These vehicle classes are only used in the "personal vehicles over the first six years" dataset, for such detail is not available in the "all vehicles" dataset. These can be considered to be the universe of light duty personal vehicle makes and models sold in California in 2001 to 2004.

Small Car - a small, inexpensive car. Included makes and models: Nissan Versa, Hyundai Elantra, Toyota Scion, Mitsubishi Lancer, Mitsubishi Mirage, Toyota Prius, Nissan Sentra, Toyota Corolla, Ford Focus, Saturn Ion, Volkswagen Jetta, Toyota Yaris, Mini Mini Cooper, Subaru Impreza, Honda Civic, Chevrolet Cobalt, Chevrolet Cavalier, Hyundai Accent, Mazda 3, Honda Fit, Chevrolet Aveo, Kia Spectra, Volkswagen New Beetle, Dodge Neon, Kia Rio, Kia Sephia, Chevrolet Malibu, Pontiac Grand Am, Plymouth Neon, Toyota Echo, Suzuki Aerio, Volkswagen Golf, Suzuki Verona, Pontiac Sunfire, Saturn Sl, Ford Escort, Mazda Protege, Pontiac G6, Smartcar Fortwo, Kia Optima, Suzuki Forenza, Volkswagen Rabbit, Volkswagen Gli, Chevrolet Geo Prizm, Pontiac Vibe, Honda Insight, Pontiac G5, Suzuki Reno, Saturn Astra, Daewoo Nubira, Daewoo Leganza, Daewoo Lanos, Suzuki Esteem, Kia Rondo, Mazda 5, Suzuki Swift, Suzuki Sx4, Volkswagen R32, Ford Contour, Saturn Sw, Chevrolet Geo Metro, Chrysler Cirrus, Chevrolet Classic, Pontiac G3, Plymouth Breeze, Mercury Mystique, Kia Forte.

Large Car - a full-sized non-luxury passenger sedan. Included makes and models:

Honda Accord, Toyota Camry, Nissan Altima, Hyundai Sonata, Acura Integra, Buick Lesabre, Chrysler 300, Dodge Charger, Ford Five Hundred, Ford Taurus, Chevrolet Impala, Subaru Legacy, Chrysler Sebring, Toyota Avalon, Chevrolet Monte Carlo, Volkswagen Passat, Mercury Grand Marquis, Buick Lacrosse, Mitsubishi Galant, Saturn Lw, Pontiac Grand Prix, Ford Fusion, Dodge Stratus, Subaru Outback, Saturn Aura, Saturn L, Saturn Ls, Hyundai Azera, Mercury Sable, Oldsmobile Alero, Mazda 626, Ford Crown Victoria, Buick Century, Chrysler Concorde, Buick Lucerne, Pontiac Bonneville, Lexus Gs, Mazda 6, Mercury Montego, Oldsmobile Intrigue, Hyundai Xg350, Pontiac G8, Buick Regal, Dodge Challenger, Mercury Milan, Mercury Marauder, Dodge Intrepid, Kia Amanti, Saturn Sw, Chevrolet Lumina.

Sporty Car - a car designed for speed. Included makes and models: Chrysler Pt Cruiser, Ford Mustang, Volkswagen Gti, Mitsubishi Eclipse, BMW M3, Chevrolet Corvette, Nissan 350Z, Chrysler Sebring, Acura Cl, Toyota Celica, Volvo C70, Hyundai Tiburon, Chevrolet Camaro, Mazda Rx8, Honda Prelude, Acura Rsx, Toyota Mr2, Chrysler 300M, Dodge Caliber, Chrysler Crossfire Sport, Volkswagen Cabrio, Mercury Cougar, Saturn Sc, Mazda Miata, Lexus Ls, Pontiac Firebird, Pontiac Gto, Mazda Protege, Volkswagen Eos, Dodge Avenger, Mazda Mx5, Nissan 370Z, Ford Gt, Pontiac Solstice.

Prestige Sporty Car - a high-end luxury car designed for speed and show. Included makes and models: Nissan 350Z, Jaguar Xk8, Honda S2000, BMW 650, Mercedes-Benz Slk-Class, Mazda Mx5, Maserati Gransport, Ferrari 430 Modena, Lexus Is-F, Porsche Boxster, Audi Tt, Porsche 911, Mercedes-Benz Sl-Class, Ford Thunderbird, BMW Z4, Pontiac Solstice, BMW M3, BMW M5, BMW M6, Jaguar Xk, Mazda Miata, BMW 645, Jaguar Xkr, Porsche Cayman, Cadillac Xlr, BMW Z8, Ferrari 599 Gtb, Chevrolet Corvette, Aston Martin Vantage, BMW M Roadster, BMW Z3, Saturn Sky, Aston Martin Db9, Audi R8, Ferrari 360, Lamborghini Gallardo, Lotus Elise, BMW M Coupe, Maserati Coupe, Aston Martin Vanquish, Maserati Quattroporte, Dodge Viper, Ferrari 575 Maranello, Lotus Exige, Lotus Esprit, Ferrari 612 Scaglietti, Acura Nsx, Maserati Granturismo, Chrysler Prowler, Plymouth Prowler, Maserati Spyder, Lamborghini Murcielago, Nissan Gt-R, Aston

Martin Db7, Ferrari 550, Porsche Carrera Gt, Ferrari F550 Maranello, Mercedes-Benz Slr, Aston Martin Dbs, Lamborghini Diablo, Ferrari Enzo, Ferrari 456, Bugatti Veyron

Luxury Car - a luxury car. Included makes and models: BMW 330, BMW 328, BMW 525, Acura Tl, Mercedes-Benz Clk-Class, Jaguar S-Type, Nissan Maxima, BMW 550, Mercedes-Benz C-Class, Lexus Es, Infiniti G35, Audi A4, Jaguar X-Type, Lexus Is, Cadillac Cts, BMW 325, Lincoln Ls, Infiniti I35, BMW 530, Infiniti G37, BMW 535, Lexus Gs, Audi S4, Audi Tt, Saab 93, Volvo S40, Acura Tsx, Cadillac Seville, Volvo V70, Hyundai Genesis, Cadillac Sts, Volvo S60, Audi Rs 4, BMW 128, Lincoln Town Car, Audi A3, Mitsubishi Diamante, BMW 135, BMW 528, Cadillac Deville, Volkswagen Cc, Acura Rl, Jaguar Xf, BMW 335, Infiniti I30, Volvo V50, Audi S8, Volvo S80, Lincoln Mks, Lincoln Mkz, Infiniti M45, Audi A6, Mazda Millenia, Cadillac Eldorado, Cadillac Catera, BMW 540, Cadillac Dts, Infiniti G20, BMW 545, Buick Park Avenue, Audi A5, Volvo V40, Lincoln Zephyr, Saab 95, Chrysler Lhs, Audi S6, Volvo C30, Infiniti M35, Saab 92, Lexus Sc, Oldsmobile Aurora, Volvo S70, Lincoln Continental, BMW 323, Audi S5

Prestige Luxury Car - an extremely high end luxury car. Included makes and models: Lexus Sc, Mercedes-Benz E-Class, Lexus Ls, Jaguar Xj8, Mercedes-Benz Cl-Class, Mercedes-Benz S-Class, Infiniti Ex, Rolls Royce Silver Seraph, Bentley Continental, Infiniti M35, Volkswagen Phaeton, BMW 745, BMW 750, Mercedes-Benz Cls-Class, Infiniti Q45, Audi Allroad, Infiniti M45, Jaguar Super V8, BMW 740, Audi A8, Jaguar Vanden Plas, Jaguar Xjr, BMW Alpina B7, BMW 540, Jaguar Xjl, Audi S8, Bentley Arnage, Rolls Royce Phantom, Jaguar Xjs, BMW 760, Rolls Royce Corniche, Bentley Azure, Maybach 57, Bentley Brooklands, Maybach 62, Maybach G 57S, Rolls Royce Parkward Limousine.

Pickup - a standard pickup truck. Included makes and models: Nissan Frontier/Xe, Dodge Dakota, Toyota Tacoma, Chevrolet Colorado, Ford Ranger, Chevrolet S10, Gmc Canyon, Honda Ridgeline, Gmc Sonoma, Mitsubishi Raider, Mazda B2300, Chevrolet Ssr, Mazda B3000, Hummer H3T, Mazda B4000, Isuzu I290, Mazda B2500, Isuzu I280, Isuzu I350, Isuzu I370, Suzuki Equator, Isuzu Hombre.

Full Pickup - a large, full-sized pickup truck. Included makes and models:

Chevrolet Avalanche, Ford F Series, Gmc Sierra, Dodge Ram, Toyota Tundra, Chevrolet Silverado, Nissan Titan, Chevrolet C/K, Lincoln Mark Lt, Lincoln Blackwood.

Sport Utility Vehicle - a standard SUV. Included makes and models: Toyota Fj Cruiser, Toyota Highlander, Ford Explorer, Nissan Xterra, Chevrolet Trail Blazer, Kia Sportage, Hyundai Santa Fe, Bmw X5, Mazda Cx-9, Jeep Grand Cherokee, Honda Pilot, Gmc Envoy Xl, Toyota 4 Runner, Acura Mdx, Honda Cr-V, Subaru Forester, Chrysler Pt Cruiser, Nissan Armada, Jeep Wrangler, Lincoln Aviator, Porsche Cayenne, Jeep Liberty, Lexus Rx, Toyota Rav4, Chrysler Pacifica, Toyota Scion, Ford Escape, Mitsubishi Outlander, Saturn Vue, Gmc Acadia, Ford Freestyle, Honda Element, Nissan Murano, Saturn Outlook, Mazda Tribute, Kia Sorento, Volkswagen Touareg, Infiniti Fx35, Audi Q7, Bmw X3, Mercury Mountaineer, Chevrolet Tracker, Volvo Xc90, Jeep Compass, Dodge Magnum, Hyundai Tucson, Dodge Durango, Gmc Envoy Denali, Land Rover Discovery, Gmc Envoy, Jeep Patriot, Chevrolet Equinox, Lincoln Mkx, Mitsubishi Montero, Acura Rdx, Ford Sport Trac, Nissan Rogue, Chevrolet S10 Blazer, Mercedes-Benz M-Class, Nissan Pathfinder, Honda Passport, Dodge Nitro, Ford Edge, Jeep Cherokee, Cadillac Srx, Chevrolet Hhr, Mercedes-Benz Glk, Mazda Cx-7, Subaru Baja, Isuzu Rodeo, Subaru B9 Tribeca, Pontiac Aztek, Infiniti Fx, Saab 9-7X, Buick Rainier, Dodge Journey, Chevrolet Trailblazer Ext, Infiniti Qx56, Pontiac Torrent, Ford Flex, Buick Rendezvous, Gmc S10 Jimmy, Chevrolet Trailblazer Ss, Isuzu Axiom, Land Rover Lr3, Suzuki Xl7, Infiniti Qx4, Suzuki Grand Vitara, Mitsubishi Endeavor, Mercury Monterey, Isuzu Ascender, Subaru Tribeca, Mercury Mariner, Land Rover Freelander, Land Rover Lr2, Suzuki Vitara, Oldsmobile Bravada, Isuzu Trooper/Trooper, Buick Enclave, Hyundai Veracruz, Gmc Envoy Xuv, Ford Taurus X, Volvo Xc70, Bmw X6, Toyota Venza, Infiniti Fx45, Audi Q5, Chevrolet Traverse, Volkswagen Tiguan, Gmc Envoy Xl Denali, Kia Borrego, Nissan Cube, Volvo Xc60, Kia Soul, Isuzu Vehicross, Infiniti Fx50, Isuzu Amigo.

Full Utility Vehicle - a full-sized (i.e., massive) SUV. Included makes and models: Toyota Sequoia, Chevrolet Suburban, Cadillac Escalade Ext, Cadillac Escalade Esv, Chevrolet Tahoe, Land Rover Range Rover, Ford Expedition, Gmc Yukon Denali Xl, Cadillac Escalade, Gmc Yukon Denali, Land Rover Range Rover Sport, Lincoln

Navigator, Mercedes-Benz G-Class, Jeep Commander, Mercedes-Benz Gl-Class, Gmc Yukon, Gmc Yukon Xl, Ford Excursion, Hummer H2, Lexus Gx, Toyota Land Cruiser, Hummer H3, Lexus Lx, Cadillac Srx, Chrysler Aspen, Hummer Sut, Mercedes-Benz R-Class, Hummer 4-Psgr Canvas Top, Hummer 4-Psgr Wagon, Hummer 4-Psgr Hard Top, Gmc Suburban, Hummer 4-Psgr Slant Back.

Minivan - a standard passenger minivan. Included makes and models: Chrysler Town & Country, Honda Odyssey, Dodge Caravan, Toyota Sienna, Chevrolet Astro, Chevrolet Venture, Saturn Relay, Volkswagen Eurovan, Kia Sedona, Chrysler Voyager, Mazda Mpv, Mercury Villager, Nissan Quest, Pontiac Montana, Ford Windstar, Gmc Safari, Oldsmobile Silhouette, Ford Freestar, Chevrolet Uplander, Buick Terraza, Mercury Monterey, Hyundai Entourage, Volkswagen Routan, Plymouth Voyager.

Bibliography

Adda, Jerome and Russell Cooper. 2000. "Balladurette and Juppette: A Discrete Analysis of Scrapping Subsidies." *Journal of Political Economy* 108(4):778–806.

Agras, Jean and Duane Chapman. 1999. "The Kyoto Protocol, CAFE Standards, and Gasoline Taxes." *Contemporary Economic Policy* 17(3):296–308.

Aldy, Joseph, Alan Krupnick, Richard Newell, Ian Parry, and William Pizer. 2010. "Designing Climate Mitigation Policy." *Journal of Economic Literature* 48(4):903–934.

Allcott, Hunt and Nathan Wozny. 2010. "Gasoline Prices and the Fuel Economy Discount Puzzle." *MIT Working Paper,* Cambridge, MA.

Alquist, Ron and Lutz Kilian. 2010. "What Do We Learn from the Price of Crude Oil Futures?" *Journal of Applied Econometrics* 25:539–573.

Alquist, Ron, Lutz Kilian, and Robert Vigfusson. 2011. "Forecasting the Price of Oil." In *Handbook of Economic Forecasting.* Elsevier:forthcoming.

Anderson, Michael. 2008. "Safety for Whom? The Effects of Light Trucks on Traffic Fatalities." *Journal of Health Economics* 27:973–989.

Anderson, Soren, Ryan Kellogg, James Sallee, and Richard Curtin. 2011a. "Forecasting Gasoline Prices Using Consumer Surveys." *American Economic Review Papers and Proceedings* forthcoming.

Anderson, Soren, Ian Parry, James Sallee, and Carolyn Fischer. 2011b. "Automobile Fuel Economy Standards: Impacts, Efficiency, and Alternatives." *Review of Environmental Economics and Policy* forthcoming.

Anderson, Soren and James Sallee. 2011. "Using Loopholes to Reveal the Marginal Cost of Regulation: The Case of Fuel-Economy Standards." *American Economic Review* forthcoming.

Austin, David. 2008. "Effects of Gasoline Prices on Driving Behavior and Vehicle Markets." *Congressional Budget Office Report 2883,* Washington, DC.

Austin, David and Terry Dinan. 2005. "Clearing the Air: The Costs and Consequences of Higher CAFE Standards and Increases in Gasoline Taxes." *Journal of Environmental Economics and Management* 50(3):562–582.

Balestra, Pietro and Marc Nerlove. 1966. "Pooling Cross-section and Time-series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas." *Econometrica* 34(3):585–612.

Barker, Terry, Paul Ekins, and Tim Foxon. 2007. "Macroeconomic Effects of Efficiency Policies for Energy-Intensive Industries: The Case of the UK Climate Change Agreements, 2000-2010." *Energy Economics* 29(4):760–778.

Basso, Leonardo and Tae Hoon Oum. 2007. "Automobile Fuel Demand: A Critical Assessment of Empirical Methodologies." *Transport Reviews* 27(4):449–484.

Becker, Gary. 1965. "A Theory of the Allocation of Time." *Economic Journal* 75(299):493–517.

Bento, Antonio, Lawrence Goulder, Mark Jacobsen, and Roger von Haefen. 2009. "Distributional and Efficiency Impacts of Increased US Gasoline Taxes." *American Economic Review* 99(3):667–699.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63(4):841–890.

———. 2004. "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market." *Journal of Political Economy* 112(1):68–105.

Blum, Ulrich, Gertraud Foos, and Marc Gaudry. 1988. "Aggregate Time Series Gasoline Demand Models: Review of the Literature and New Evidence for West Germany." *Transportation Research A* 22(2):75–88.

Bresnahan, Timothy. 1981. "Departures from Marginal Cost Pricing in the American Automobile Industry: Estimates for 1977-1978." *Journal of Econometrics* 17(2):201–227.

Brookes, Leonard. 1979. "A Low Energy Strategy for the UK by G. Leach et al.: A Review and Reply." *Atom* 269.

Busse, Meghan, Christopher Knittel, and Florian Zettelmeyer. 2010. "Pain at the Pump: The Differential Effect of Gasoline Prices on New and Used Automobile Markets." *UCEI Working Paper,* Berkeley, CA.

Calfornia Department of Transportation. 2002. "2000-2001 California Statewide Household Travel Survey." *Final Report,* June 2002.

California Energy Commission. 2006. "Inventory of Greenhouse Gas Emissions and Sinks: 1990 to 2004." *CEC Staff Report CEC-600-2006-013-SF,* Sacramento, CA.

Copeland, Adam, Wendy Dunn, and George Hall. 2011. "Inventories and the Automobile Market." *RAND Journal of Economics* 42(1):121–149.

Dahl, Carol. 1995. "Demand for Transportation Fuels: A Survey of Demand Elasticities and Their Components." *Journal of Energy Literature* 1(2):3–27.

Dahl, Carol and Thomas Sterner. 1991. "Analysing Gasoline Demand Elasticities: A Survey." *Energy Economics* 13(3):203–210.

Davis, Lucas. 2008. "Durable Goods and Residential Demand for Energy and Water: Evidence from a Field Trial." *RAND Journal of Economics* 39(2):530–546.

Davis, Lucas and Matthew Kahn. 2011. "International Trade in Used Vehicles: The Environmental Consequences of NAFTA." *American Economic Journal: Economic Policy* 2(4):58–82.

Davis, Lucas and Lutz Kilian. 2011a. "The Allocative Cost of Price Ceilings in the US Residential Market for Natural Gas." *Journal of Political Economy* 119(2):212–241.

———. 2011b. "Estimating the Effect of a Gasoline Tax on Carbon Emissions." *Journal of Applied Econometrics* 26(3):forthcoming.

de Jong, Gerard and Hugh Gunn. 2001. "Recent Evidence on Car Cost and Time Elasticities of Travel Demand in Europe." *Journal of Transport Economics and Policy* 35(2):137–160.

Dimitropoulos, John. 2007. "Energy Productivity Improvements and the Rebound Effect: An Overview of the State of Knowledge." *Energy Policy* 35(12):6354–6363.

Dix, Michele and Phil Goodwin. 1982. "Elasticities of Road Traffic and Fuel Consumption with Respect to Price and Income: A Review." *Transport Policy and Decision Making* 2(2):179–195.

Dubin, Jeffrey. 1985. *Consumer Durable Choice and the Demand for Electricity.* New York, NY: North-Holland Publishing Company.

Dubin, Jeffrey and Daniel McFadden. 1984. "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption." *Econometrica* 52(2):345–362.

Einav, Liran, Amy Finkelstein, Paul Schrimpf, Stephen Ryan, and Mark Cullen. 2010. "Selection on Moral Hazard in Health Insurance." *Stanford University Working Paper,* Stanford, CA.

Energy Information Administration. 2010. "Annual Energy Outlook 2011." *EIA Report DOE/EIA-0383,* Washington, DC.

EPA and DOT. 2010. "Light-Duty Vehicle Greenhouse Gas Emission Standards and Corporate Average Fuel Economy Standards; Final Rule." *Federal Register,* May 7, 2010.

Erdem, Tulin, Michael Keane, and Baohong Sun. 1999. "Missing Price and Coupon Availability Data in Scanner Panels: Correcting for the Self-selection Bias in Choice Model Parameters." *Journal of Econometrics* 89(1-2):177–196.

Espey, Molly. 1998. "Gasoline Demand Revisited: An International Meta-Analysis of Elasticities." *Energy Economics* 20(3):233–345.

Esteban, Susanna and Matthew Shum. 2007. "Durable-goods Oligopoly with Secondary Markets: The Case of Automobiles." *RAND Journal of Economics* 38(2):332–354.

Feng, Ye, Don Fullerton, and Li Gan. 2005. "Vehicle Choices, Miles Driven, and Pollution Policies." *NBER Working Paper 11553,* Cambridge, MA.

Gately, Dermot. 1992. "Imperfect Price Reversibility of U.S. Gasoline Demand: Asymmetric Responses to Price Increases and Declines." *Journal of Transportation and Statistics* 2(1):1–17.

———. 1993. "The Imperfect Price Reversibility of World Oil Demand." *Energy Journal* 14(4):163–182.

Goldberg, Pinelopi. 1998. "The Effects of the Corporate Average Fuel Efficiency Standards in the US." *The Journal of Industrial Economics* 46(1):1–33.

Goodwin, Phil, Joyce Dargay, and Mark Hanly. 2004. "Elasticities of Road Traffic and Fuel Consumption with Respect to Price and Income: A Review." *Transport Reviews* 24(3):275–292.

Goulder, Lawrence and Roberton Williams. 2003. "The Substantial Bias from Ignoring General Equilibrium Effects in Estimating Excess Burden, and a Practical Solution." *Journal of Political Economy* 111(4):898–927.

Graham, Daniel and Stephen Glaister. 2002. "The Demand for Automobile Fuel: A Survey of Elasticities." *Journal of Transport Economics and Policy* 36(1):1–26.

Greene, David. 1992. “Vehicle Use and Fuel Economy: How Big is the ‘Rebound’ Effect.” *Energy Journal* 13(1):117–143.

———. 2011. “Rebound 2007: Analysis of U.S. Light Duty Vehicle Travel Statistics.” *Energy Policy* forthcoming.

Greene, David and P Hu. 1984. “The Influence of the Price of Gasoline on Vehicle Use in Multivehicle Households.” *Transportation Research Record* 988:19–23.

Greene, David, James Kahn, and Robert Gibson. 1999. “Fuel Economy Rebound Effect for U.S. Household Vehicles.” *Energy Journal* 20(3):1–31.

Greene, David, Philip Patterson, Margaret Singh, and Jia Li. 2005. “Feebates, Rebates and Gas-guzzler Taxes: A Study of Incentives for Increased Fuel Economy.” *Energy Policy* 33(6):757–775.

Greening, Lorna, David Greene, and Carmen Difiglio. 2000. “Energy Efficiency and Consumption - the Rebound Effect - A Survey.” *Energy Journal* 28(6-7):389–401.

Harrington, Winston, Ian Parry, and Margaret Walls. 2007. “Automobile Externalities and Policies.” *Journal of Economic Literature* 45:374–400.

Haughton, Jonathan and Soumodip Sarkar. 1996. “Gasoline Tax as a Corrective Tax: Estimates for the United States, 1970-1991.” *Energy Journal* 17(2):103–126.

Hausman, Jerry and Whitney Newey. 1995. “Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss.” *Econometrica* 63(6):1445–1476.

Hubbard, Thomas. 1998. “An Empirical Examination of Moral Hazard in the Vehicle Inspection Market.” *RAND Journal of Economics* 29(2):406–426.

Hughes, Jonathan, Christopher Knittel, and Daniel Sperling. 2008. “Evidence in a Shift in the Short-Run Price Elasticity of Gasoline Demand.” *Energy Journal* 29(1):113–134.

Hymel, Kent, Kenneth Small, and Kurt Van Dender. 2010. “Induced Demand and Rebound Effects in Road Transport.” *Transportation Research B* forthcoming.

Jacobsen, Mark. 2010. "Evaluating U.S. Fuel Economy Standards in a Model with Producer and Household Heterogeneity." *UCSD Working Paper,* San Diego, CA.

Jenkins, Jesse, Ted Nordhaus, and Michael Shellenberger. 2011. "Energy Emergence: Rebound and Backfire as Emergent Phenomena." *Breakthrough Institute Report,* Oakland, CA.

Jevons, William Stanley. 1865. *The Coal Question; An Inquiry Concerning the Progress of the Nation, and the Probable Exhaustion of Our Coal Mines.* London, UK: Macmillian and Company.

Jones, Clifton. 1993. "Another Look at U.S. Passenger Vehicle Use and the Rebound Effect from Improved Fuel Efficiency." *Energy Journal* 14(4):99–110.

Kahn, James. 1986. "Gasoline Prices and the Used Automobile Market." *Quarterly Journal of Economics* 101(2):323–339.

Kahn, Matthew. 1996. "New Evidence on Trends in Vehicle Emissions." *RAND Journal of Economics* 27(1):183–196.

Kayser, Hilke. 2000. "Gasoline Demand and Car Choice: Estimating Gasoline Demand Using Household Information." *Energy Economics* 22(3):331–348.

Khazzoom, Daniel. 1980. "Economic Implications of Mandated Efficiency Standards for Household Appliances." *Energy Journal* 1(4):21–39.

Kilian, Lutz and Eric Sims. 2006. "The Effects of Real Gasoline Prices on Automobile Demand: A Structural Analysis Using Micro Data." *University of Michigan Working Paper,* Ann Arbor, MI.

Klier, Thomas and Joshua Linn. 2010a. "New Vehicle Characteristics and the Cost of Corporate Average Fuel Economy Standard." *Resources for the Future Discussion Paper 10-50,* Washington, DC.

———. 2010b. "The Price of Gasoline and the Demand for Fuel Efficiency: Evidence from Monthly New Vehicle Sales Data." *American Economic Journal: Economic Policy* 2(3):134–53.

Knittel, Chris. 2011. "Automobiles on Steroids: Product Attribute Trade-offs and Technological Progress in the Automobile Sector." *American Economic Review* forthcoming.

Knittel, Christopher. 2010. "Automobiles on Steriods: Product Attribute Trade-offs and Technological Progress in the Automobile Sector." *American Economic Review* forthcoming.

Knittel, Christopher and Ryan Sandler. 2010. "Carbon Prices and Automobile Greenhouse Gas Emissions: The Extensive and Intensive Margins." *NBER Working Paper 16482,* Cambridge, MA.

Koenker, Roger. 2005. *Quantile Regression.* London, UK: Cambridge University Press.

Koenker, Roger and Kevin Hallocks. 2001. "Quantile Regression." *Journal of Economic Perspectives* 15(4):143–156.

Krupnick, Alan, Margaret Walls, and Carol Collins. 1993. "Global Warming and Urban Smog: Cost-Effectiveness of CAFE Standards and Alternative Fuels." *Energy Journal* 14(4):75–97.

Kydes, Andrew. 1997. "Sensitivity of Energy Intensity in US Energy Markets to Technological Change and Adoption." In *Issues in Midterm Analysis and Forecasting. DOE/EIA-060797.* Washington, DC: US Department of Energy, 1–42.

Li, Shanjun, Joshua Linn, and Erich Muehlegger. 2011. "Gasoline Taxes and Consumer Behavior." *Harvard University Working Paper,* Cambridge, MA.

Li, Shanjun, Christopher Timmons, and Roger von Haefen. 2009. "How Do Gasoline Prices Affect Fleet Fuel Economy." *American Economic Journal: Economic Policy* 1(2):113–137.

Lin, Cynthia and Lea Prince. 2009. "The Optimal Gas Tax for California." *Energy Policy* 37(12):5173–5183.

Litman, Todd. 2005. "Efficient Vehicles Versus Efficient Transportation: Comparing Transportation Energy Conservation Strategies." *Transport Policy* 12(2):121–129.

Mannering, Fred and Clifford Winston. 1985. "A Dynamic Empirical Analysis of Household Vehicle Ownership and Utilization." *RAND Journal of Economics* 16(2):215–236.

Manzan, Sebastiano and Dawit Zerom. 2011. "A Semiparametric Analysis of Gasolie Demand in the US: Reexamining the Impact of Price." *Econometric Reviews* forthcoming.

Marion, Justin and Erich Muehlegger. 2011. "Fuel Tax Incidence and Supply Conditions." *Journal of Public Economics* forthcoming.

Mayo, John and John Mathis. 1988. "The Effectiveness of Mandatory Fuel Efficiency Standards in Reducing the Demand for Gasoline." *Applied Economics* 20(2):211–219.

Nevo, Aviv. 2000. "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand." *Journal of Eocnomics and Management Strategy* 9(4):513–548.

NHTS. 2009. "National Highway Transportation Survey 2009." *US Department of Transportation* Washington, DC.

Parry, Ian and Kenneth Small. 2005. "Does Britian or the United States Have the Right Gasoline Tax?" *American Economic Review* 95:1276–1289.

Petrin, Amil. 2002. "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy* 110(4):705–729.

Pickrell, Don and Paul Schimek. 1999. "Growth in Motor Vehicle Ownership and Use: Evidence from the Nationwide Personal Transportation Survey." *Journal of Transportation and Statistics* 2(1):1–17.

Polk, R.L. 2010. "Consumers Continuing to Hold Onto Vehicles Longer." *R.L. Polk Report,* Southfield, MI.

Pollak, Robert and Michael Wachter. 1975. "The Relevance of the Household Production Function and Its Implications for the Allocation of Time." *Journal of Political Economy* 83(2):255–278.

Poterba, James. 1989. "Lifetime Incidence and the Distributional Burden of Excise Taxes." *American Economic Review* 79(2):325–330.

———. 1991. "Is the Gasoline Tax Regressive." In *Tax Policy and the Economy, Vol 5*, edited by David Bradford. Cambridge, MA: MIT Press, 145–164.

Puller, Steven and Lorna Greening. 1999. "Household Adjustment to Gasoline Price Change: An Analysis Using Nine Years of U.S. Survey Data." *Energy Economics* 21(1):37–52.

Sallee, James. 2010. "Taxation of Fuel Economy." *NBER Working Paper 16466,* Cambridge, MA.

Sallee, James, Sarah West, and Wei Fan. 2009. "Consumer Valuation of Fuel Economy: A Microdata Approach." *Proceedings of the National Tax Association Annual Conference on Taxation,* Denver, CO.

Santos, Georgina and Tom Catchesides. 2005. "Distributional Consequences of Gasoline Taxation in the United Kingdom." *Transportation Research Record* 1024:103–111.

Saunders, Harry. 1992. "The Khazzoom-Brookes Postulate and Neo-classical Growth." *Energy Journal* 13(4):131–148.

———. 2011. "Historical Evidence for Energy Consumption Rebound in 30 US Sectors and a Toolkit for Rebound Analysis." *Working Paper,* Danville, CA.

Schimek, Paul. 1996. "Gasoline and Travel Demand Models Using Time Series and Cross-section Data from the United States." *Transportation Research Record* 1558:83–89.

Schiraldi, Pasquale. 2010. "Automobile Replacement: A Dynamic Structural Approach." *London School of Economics Working Paper,* London, UK.

Schmalensee, Richard and Thomas Stoker. 1999. "Household Gasoline Demand in the United States." *Econometrica* 67(3):645–662.

Small, Kenneth and Harvey Rosen. 1981. "Applied Welfare Economics with Discrete Choice Models." *Econometrica* 49(1):105–130.

Small, Kenneth and Kurt Van Dender. 2007. "Fuel Efficiency and Motor Vehicle Travel: The Declining Rebound Effect." *Energy Journal* 28(1):25–51.

Sorrell, Steven. 2007. "The Rebound Effect: An Assessment of the Evidence for Economy-wide Energy Savings From Improved Energy Efficiency." *UK Energy Research Center Working Paper,* Sussex, UK.

Sorrell, Steven and John Dimitropoulos. 2008. "The Rebound Effect: Microeconomic Definitions, Limitations and Extensions." *Ecological Economics* 65(3):636–649.

Sperling, Daniel and Deborah Gordon. 2008. "Advanced Passenger Transport Technologies." *Annual Review of Environment and Resources* 33:63–84.

Sterner, Thomas and Carol Dahl. 1992. "Modeling Transport Fuel Demand." In *International Energy Economics*, edited by Thomas Sterner. London, UK: Chapman and Hall, 65–79.

Stolyarov, Dmitriy. 2002. "Turnover of Used Durables in a Stationary Equilibrium: Are Older Goods Traded More?" *Journal of Political Economy* 110(6):1390–1413.

Sweeney, James. 1979. "Effects of Federal Policies on Gasoline Consumption." *Resources and Energy* 2(1):3–26.

Tanner, Martin and Wing Hung Wong. 1987. "The Calculation of Posterior Distributions by Data Augmentation." *Journal of the American Statistical Association* 82(398):528–540.

Todorova, Aleksandra. 2007. "Timing a Car Purchase." *Wall Street Journal,* August 26.

Tol, Richard. 2009. "The Economic Effects of Climate Change." *Journal of Economic Perspectives* 23(2):29–51.

Train, Kenneth. 2003. *Discrete Choice Methods with Simulation.* London, UK: Cambridge University Press.

US Environmental Protection Agency. 2005. "Emission Facts: Greenhouse Gas Emissions from a Typical Passenger Vehicle." *Report EPA 420-F-05-004,* Washington, DC.

———. 2011. "Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2009." *Report EPA 430-R-11-005,* Washington, DC.

Wadud, Zia, Daniel Graham, and Robert Noland. 2009. "Modelling Fuel Demand for Different Socio-economic Groups." *Applied Energy* 86(12):2740–2749.

———. 2010a. "Gasoline Demand with Heterogeneity in Household Responses." *Energy Journal* 31(1):47–74.

———. 2010b. "A Semiparametric Model of Household Gasoline Demand." *Energy Economics* 32(1):93–101.

Weitzman, Martin. 1974. "Prices vs. Quantities." *Review of Economic Studies* 41(4):477–491.

West, Sarah. 2004. "Distributional Effects of Alternative Vehicle Pollution Control Technologies." *Journal of Public Economics* 88:735.

West, Sarah and Roberton Williams. 2004. "Estimates From a Consumer Demand System: Implications for the Incidence of Environmental Taxes." *Journal of Environmental Economics and Management* 47(3):535–558.

———. 2005. "The Cost of Reducing Gasoline Consumption." *American Economic Review Papers and Proceedings* 95(2):294–299.

———. 2007. "Optimal Taxation and Cross-Price Effects on Labor Supply: Estimates of the Optimal Gas Tax." *Journal of Public Economics* 91(3-4):593–617.

White, Michelle. 2004. "The "Arms Race" on American Roads: The Effect of Sport Utility Vehicles on Traffic Safety." *Journal of Law and Economics* 47(2):333–355.

Whitefoot, Katie, Meredith Fowlie, and Steven Skerlos. 2011. "Product Design Response to Industrial Policy: Evaluating Fuel Economy Standards Using an Engineering Model of Endogenous Product Design." *Energy Institute at Haas Working Paper 214*, Berkeley, CA.

Yatchew, Adonis. 2001. "Household Gasoline Demand in Canada." *Econometrica* 69(6):1697–1709.