How Do Consumers Respond to Gasoline Price Shocks? Heterogeneity in Vehicle Choice and Driving Behavior*

Kenneth Gillingham†

Draft Version: July 1, 2011

Abstract

This paper develops a structural econometric model of vehicle choice and subsequent driving decisions to examine the consumer responsiveness to gasoline price changes on both margins. Consumer decisions are modeled in a dynamic setting that explicitly accounts for selection on unobserved driving preference. The model leverages a unique and extremely rich dataset of all vehicle registrations in California in 2001-2009, which are matched at the vehicle-level with smog check data that include odometer readings at the time of the test. Results suggest that consumers are responsive to gasoline prices in both vehicle choice and driving decisions, with a medium-run (roughly two years) gasoline price elasticity of fuel economy and driving for personal vehicles around 0.09 and -0.15 respectively. These responses vary by income, geographic, and demographic groups. Counterfactual policy simulations of a gasoline tax and feebate policy illustrate the use of the model. The results have key implications for the effectiveness and consequences of policies to reduce emissions from the transportation sector.

*I would like to thank the Precourt Energy Efficiency Center, the Shultz Graduate Student Fellowship in Economic Policy through the Stanford Institute for Economic Policy Research, the US Environmental Protection Agency STAR Fellowship program, a grant through the Stanford Institute for Economic Policy Research from Exxon-Mobil, and Larry Goulder for the funding that made this research possible. I would also like to thank Jim Sweeney, Tim Bresnahan, Jon Levin, Larry Goulder, Liran Einav, and Matt Harding for extremely useful suggestions, Zach Richardson of the California Bureau of Automotive Repair for providing the smog check data, Ray Alvarado of R.L. Polk for his help with the registration data, Mark Mitchell of OPIS for his help with the retail gasoline price data, and Hunt Allcott for providing the raw data on EPA fuel economy ratings.

†Yale University, School of Forestry & Environmental Studies, 195 Prospect St, New Haven, CT 06511, Phone: (203) 436-5465; Email: `kenneth.gillingham@yale.edu`.

1. Introduction

Nothing seems more noticeable to the average American than drastic increases in gasoline prices, as evidenced by Google Trends in 2007 and 2008 when "high gas prices" was one of the hottest search terms on Google. 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. Yet there is surprisingly little empirical work on the elasticity of driving with respect to the gasoline price. Most recent work on the price elasticity of gasoline consumption suggests extremely small elasticity values in the short term, contrary to the apparent changes in 2007 and 2008. In addition, there is remarkably little evidence on the heterogeneity in consumer responsiveness to gasoline price changes, even if we might expect very different responses by different income, demographic, and geographic groups. Yet understanding how consumers respond to changing gasoline prices on both the intensive (driving) margin and extensive (vehicle choice) margin is critical to understanding the welfare consequences of a variety of policies intended to address externalities in the transportation sector, including gasoline taxes, carbon policies, feebates, and Corporate Average Fuel Economy (CAFE) standards.

This paper uses a rich and novel dataset to demonstrate that consumers responded to the substantial recent changes in gasoline prices in both driving and vehicle choice decisions, and that there is substantial heterogeneity in the response. The dataset includes all personal new vehicle registrations in California 2001 through mid-2009, and the subsequent smog check odometer readings. I develop a structural model of new vehicle choice and driving decisions that takes advantage of this unique dataset and allows me to explicitly model the dynamics of the decision process and the interactions between the two decisions.

My approach has several innovations. First, it jointly models vehicle choice and driving in a two-period setting that allows consumers to base the vehicle choice decision on the relevant expectation of gasoline prices and the driving decision on the gasoline prices during the period of driving. Second, it explicitly accounts for how adjustments in used vehicle prices due to changes in the gasoline price may affect vehicle choice. Third, the model allows for counterfactual simulations of a variety of policies that affect either or both the vehicle choice and driving margin.

I find a medium-run (i.e., roughly two year) elasticity of driving with respect to the price of gasoline in the range of -0.15. In vehicle choice, I find a medium-run elasticity of fuel economy with respect to the price of gasoline in the range of 0.09. I conduct two counterfactual simulations to illustrate the use of the model: a gasoline tax that raises the price of gasoline by $1 per gallon (in real 2010$) and a revenue-neutral feebate policy that penalizes vehicles with low fuel economy and rewards vehicles with high fuel economy by $500 per a 0.01 gallon per mile difference from the pivot point.[1] On average, the gasoline tax decreases driving by 5% and increases the fuel economy of the new vehicle fleet by 3%. Absent externalities, this policy leads to an average loss in consumer surplus to new vehicles purchased in 2002 by roughly $1.80 per vehicle per day from more expensive gasoline and reduced driving, but brings in $1.70 per vehicle per day in tax revenues from consumers. The loss in consumer surplus at the time of purchase is $0.37 per vehicle on average. The feebate policy increases fuel economy by 15% on average and increases driving by 3% on average for those consumers who were induced to purchase a different vehicle. Absent externalities, the average welfare change for the same 2002 cohort is a loss of $0.56 per vehicle at the time of purchase, but a gain of $0.03 per vehicle per day on the intensive margin from increased driving and lower fuel costs.

The motivation for jointly modeling vehicle choice and subsequent utilization is the concern that selection may lead to biased and inconsistent estimates. Specifically, for a durable good, such as a vehicle, a consumer who plans to use the good more may select into a more efficient product in the purchase decision (Dubin and McFadden 1984). In the context of vehicles, the selection issue may occur because a consumer who plans to drive more may select into a different vehicle. As suggested by Davis (2008), the consumer may choose a more fuel efficient vehicle in order to save money on fuel costs. Alternatively, the consumer may choose a larger, more comfortable vehicle to make the long drives more pleasant (West 2004). Either way, the selection could confound the estimated coefficients for it changes the price of subsequent utilization.

Much of the previous literature on vehicle choice and utilization uses the two-step approach in Dubin and McFadden (1984) and instruments for the price of utilization using estimated choice probabilities from a first-stage discrete choice estimation (West 2004; Goldberg 1998; Berkowitz, Gallini, Miller, and Wolfe 1990; Mannering and Winston 1985). Several more recent studies jointly estimate vehicle choice and utilization in a static model (Feng, Fullerton, and Gan 2005; Bento, Goulder, Jacobsen, and von Haefen 2009; Jacobsen 2010). My approach similarly accounts for selection, but is the first approach to bring in the dynamics of the vehicle choice and utilization

[1] The "pivot point" is a benchmark set to determine the revenue the policy raises. Note that feebate policies tend to be based on fuel consumption (in gallons per mile), rather than fuel economy (in miles per gallon), for an improvement of 1 mile per gallon (mpg) saves much more fuel for a low mpg vehicle than a high one (e.g., see Larrick and Soll (2008)).

decision.[2] Moreover, it allows for a relatively straightforward test of the importance of selection in confounding the estimated coefficients. I find that if I estimate vehicle choice and utilization separately, the average elasticity of driving with respect to the gasoline price is -0.21, rather than -0.15, and the average elasticity of new vehicle fuel economy is again 0.09, as it was in the joint model. This provides a sense of the quantitative importance direction of the bias due to selection.

My rich dataset also allows for an exploration into the heterogeneity in responsiveness to a degree not possible in previous studies.[3] I find that conditional on purchasing a new vehicle, wealthier households appear to reduce per-vehicle driving more in response to a gasoline price increase than less-wealthy households. This may be due to switching of vehicles within a household, more discretionary driving by wealthier households, and selection based on who purchases new vehicles. The only exception to this is the lowest-income category of households (i.e., those who earn less than $30,000 a year). This least-wealthy category displays the highest elasticity of driving. I also find evidence of heterogeneity in the responsiveness of driving geographically across counties in California, a result with implications for the effect of the policy in reducing congestion and local air pollution.

I examine counterfactual policy simulations of a gasoline tax and feebate policy to illustrate how the model can be used to examine the effects and welfare implications of policies to reduce emissions from the transportation sector. The gasoline tax policy simulation results follow from the estimated elasticities: assuming a competitive supply-side in the gasoline market, the excess burden from the gasoline tax is relatively small, even if the pre-existing distortions are taken into account. However the distributional consequences are important, for there is substantial heterogeneity in the burden of the policy across counties in California. The counterfactual analysis of the feebate policy illustrates how the welfare effects to consumers from being incentivized to purchase a different vehicle can be quantified.

This paper is structured as follows. The second section describes the dataset assembled for this study and includes some brief descriptive evidence illustrating the sources of variation in

[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).

[3]West (2004) and Bento, Goulder, Jacobsen, and von Haefen (2009) are two important studies that use survey data to explore some types of heterogeneity in consumer responsiveness.

the data.[4] The third section presents the structural model and discusses identification and the estimation strategy. The fourth section presents the estimation results and discusses the importance of accounting for selection in modeling vehicle choice and utilization decisions. The fifth section presents the gasoline tax and feebate policy counterfactual simulation results. The final section concludes.

2. Data and Descriptive Evidence

2.1. Data

The dataset assembled for this study is unprecedented in its breadth and detail of vehicle choices and driving behavior. The focus of the study 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 2006-2008. 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.

The primary foundation of the dataset contains all 12.6 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.[5] 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% of

[4] A complementary paper, Gillingham (2010), goes into much more detail in providing ordinary least squares and fixed effects estimation evidence to explore the responsiveness on the intensive margin for *all* vehicle purchases, including firm and government purchases.

[5] The buyer type (i.e., personal, firm, or government) is also observed, and Gillingham (2010) examines responsiveness on the intensive margin in much more detail for each of these buyer types.

the categorical income variable is observed and in most years over 85% is observed). Table 1 shows the number of personal vehicles registered by year broken down by vehicle class.

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. For this study, the 2008 ratings are used throughout the entire time frame of the study for a consistent measure that more closely reflects on-road fuel economy.[6] Next, 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 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 reported to the California Bureau of Automotive Repair during the mandatory smog check. Since 1984, every vehicle in California that is covered by the smog check program must be in compliance in order to be registered with the DMV.[7] To be in compliance, vehicles must meet the California criteria air pollution standards for several local air pollutants. Since 1998, these tests have been required at the seventh registration renewal (usually at the end of six years of vehicle life), and then biennially thereafter.[8] If a vehicle of more than four model years old is sold, it also is required to be smog tested at the time of the sale. Earlier 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 my study, there is no obvious incentive

[6]Future work is planned to examine whether the switch in the posted fuel economy ratings in January 2008 had an influence on vehicle sales.

[7]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. Interestingly, many hybrids seem to have been smog checked anyway in my dataset. Appendix A lists the counties covered by the smog check program.

[8]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, and is examined in the empirical section.

for mechanics to falsely report the odometer readings, and these readings are perhaps the best revealed preference measure of how much people actually have been driving.

For this study, I use smog check data from 2005 to mid-2010. 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 1 shows the distribution of vehicle miles traveled (VMT) per month for personal vehicles. The mean of VMT per month is 1,090, with a surprisingly high variance of 465. This high variance provides the first evidence that there is substantial heterogeneity in driving behavior.

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). To address economic conditions that may affect driving, I bring in data on the unemployment rate in California from the Bureau of Labor Statistics (BLS) and the national-level Consumer Confidence Index (CCI) from the Conference Board. Figure 3 shows that the gasoline price shock preceded when the recession began to have a major impact on employment in late 2008. Finally, I add zip code-level demographics and county-level commute times from the US Census Bureau. Table 2 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. A detailed description of the data cleaning process is included in Appendix B.

2.2. **Descriptive evidence of responsiveness on the intensive margin**

The amount Californians drive appears to have been on an increasing trend for the past 30 years. Figure 4 shows this upward trend based on traffic count data on state highways taken by the California Department of Transportation. Figure 4 also provides suggestive evidence that gasoline prices influence how much consumers drive, with the clear drop-off in driving in 2007 and 2008. This same decrease in driving is also evident in my dataset. Figure 5 illustrates that the average VMT over the first six years of the vehicle life has decreased along with the average gasoline price over that same time frame. Of course, the evidence in both Figure 4 and Figure 5 is suggestive of the inverse relationship between gasoline prices and driving, for economic conditions may also play an important role, and are often correlated with gasoline prices.

For further evidence, we can examine a simple regression of VMT on the gasoline price, while conditioning on the characteristics of the vehicle purchased, demographics, and economic conditions. Specifically, for each new personal vehicle purchase i, we have

$$VMT_i = \beta_0 + \beta_p P_i + \beta_V V_i + \beta_L L_i + \beta_D D_i + \beta_E E_i + \varepsilon_i, \tag{1}$$

where P_i is the average gasoline price over the time between the registration and smog check, V_i contains vehicle characteristics, L_i is an indicator for whether the vehicle is leased, D_i contains demographics of the zip code the vehicle is registered in, E_i contains economic conditions over the time between registration and the smog check, and ε_i is a mean-zero stochastic term. I include all of the vehicle characteristics and demographics that are listed in Table 2. Here VMT_i is the average monthly vehicle-miles-traveled during the time between the first vehicle registration and the first smog check. In this specification, I focus on vehicles registered in 2001-2004 for which I observe a smog check. The relationship between VMT and the gasoline price is primarily identified with time series variation in the average gasoline price, but some cross-sectional variation also plays a role. Figure 6 indicates the substantial variation in average gasoline prices in my sample.

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.[9] Second, if the patterns of where consumers moved over the time between registration and the smog check are somehow correlated with gasoline prices,

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

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 on only consumers who did not move to determine if this is a concern.

Next, selection may confound the estimates in four different ways. First, 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 shock 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 second selection issue could be a concern if there is unobserved heterogeneity in consumer preferences for driving. Consumers who know they are going to be driving more may purchase a more efficient vehicle, thus reducing the cost per mile of driving and leading them to drive more. If this selection is correlated with cross-sectional differences in gasoline prices, then there may be an endogeneity concern. My structural analysis addresses this issue directly. For this analysis, I can include county fixed effects to rely entirely on time series variation.

A third selection issue may confound the estimation if consumers of different unobserved driving preferences selected into an early or late smog check. Figure 7 shows that roughly 40% 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. 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, possibly leading to a spurious correlation between the gasoline price and driving. To address the possibility of this selection issue, I can focus entirely on consumers who had a smog check within a few months of the standard six years and compare the results.

A fourth selection issue may be 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.

Finally, we may be concerned that economic conditions could be influencing the decrease in driving. Fortunately, in the time frame of my study, the gasoline price shock 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.[10] By conditioning on the county-level unemployment rate and the CCI, I can control for changes in macro-level economic factors that could influence driving.

The results from estimating (1) by ordinary least squares, fixed effects estimation, and quantile estimation are given in Table 3. The coefficient on the average gasoline price is surprisingly robust across specifications, indicating that an increase in the gasoline price by \$1 (in 2010\$) corresponds to a decrease in per-vehicle driving by around 100-120 miles a month. A change of 110 miles per month implies a medium-run elasticity in the range of -0.25 at the mean.[11] Using a log-log specification also gives an elasticity estimate in this range. The lack of difference between columns (1) and (2) demonstrates that any selection based on whether the test was early or late does not influence the coefficient of interest much at all. The same check performed on the fixed effects estimations has a similar result. County, month of year, and model fixed effects all increase the estimated coefficient slightly, and using indicator variables to combine several of these increases the estimated coefficient somewhat more. Year fixed effects in column (6) have the beneficial effect of controlling for time trends, but rely entirely on within-year variation. Using within-year variation implies that the estimated coefficient is a shorter-term response (perhaps why the coefficient is slightly less), and it also may exacerbate a selection issue from early or late smog checks, for there are different numbers of early or late smog checks in different years.[12] To make sure this is not an issue, I include indicator variables for the very early (<5 years), early (5-5.8 years), late (6.2-7 years), and very late (>7 years) smog checks. Without such controls, the coefficient on the gasoline price is less than -200. Out of the other coefficients, the most interesting one is the coefficient on fuel

[10]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.

[11]I consider this a medium-run elasticity because the primary identifying variation is time series variation over roughly two years.

[12]For example, in 2004, I do not observe any late smog checks, but do observe more early smog checks.

economy. To the extent that the selection issues described above do not confound this evidence, the coefficient on fuel economy can be considered an estimate of the "rebound effect," for it indicates how an increase in fuel economy corresponds with an increase in driving. In specifications without model fixed effects, the fuel economy coefficient suggests that a 1 mile per gallon increase in fuel economy corresponds to a roughly 3 miles per month increase in driving–quite a small rebound effect.

Finally, the 0.25 and 0.75 quantile results indicate that there is quite a substantial heterogeneity in this responsiveness, with the decrease in driving from a $1 gasoline price increase ranging from 256 miles per month at the 0.25 quantile to 63 miles per month at the 0.75 quantile. The determinants and details of this heterogeneity will be explored further based on the structural model results. None of these regression results appear to change if restricted to the subset of vehicles that did not change county between the time of the vehicle registration and the time of the test.

2.3. **Descriptive evidence of responsiveness on the extensive margin**

Until 2006, the fuel economy of the average new vehicle 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 choice at least partly due to low gasoline prices (Knittel 2010). Yet in 2006, the average fuel economy of new vehicles began inching upwards, and peaked in 2008 along with the gasoline price peak, before dropping again as gasoline prices returned to lower levels (Figure 8).

For further evidence of a relationship between gasoline prices and fuel economy, we model the fuel economy of personal new vehicles in California as a function of the gasoline price and characteristics of the vehicle. For each new vehicle purchase i we have

$$FE_i = \gamma_0 + \gamma_P P_i + \gamma_L L_i + \gamma_D D_i + \gamma_E E_i + u_i,$$

where FE_i is the fuel economy of the vehicle, P_i is the price of gasoline faced by the purchaser at the time of purchase, L_i is an indicator for whether the vehicle is leased, D_i contains demographics of the zip code the vehicle was registered in, E_i contains economic conditions at the time of purchase, and u_i is a mean-zero stochastic error term. Again, I include all demographics contained in Table 2. This specification can be run using the full dataset of 12.6 million new vehicle registrations.

There may be some identification concerns with this specification as well. 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 8 provides evidence that this is not the case. To provide further evidence, 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 the next section accounts for this issue explicitly.

The results from this estimation are shown in Table 4. 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 \$1 per gallon increase in the gasoline price corresponds to a 1.3 to 1.4 miles per gallon increase in new vehicle fuel economy. 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. Given the 0.6 coefficient, the corresponding elasticity at the mean is in the range of 0.10. As in the case on the intensive margin, since the primary identifying variation is over two years, this elasticity is probably best interpreted as a medium-run elasticity. 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.[13] The structural model developed in the next section explicitly accounts for selection and allows for exploring the heterogeneity in how gasoline prices affect the responsiveness on the extensive margin.

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

3. **A Model of Vehicle Choice and Utilization**

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. The advantage of 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. In each of the two periods, 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, Finkelstein, Schrimpf, Ryan, and Cullen (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.

3.1. **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.[14] 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 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 as

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

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.[15] This specification assumes that the benefits of driving are quadratic in VMT, where λ influences the curvature of the function. This specification also normalizes the coefficient on the fuel costs 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 vehicle. Specifically, this random coefficient is parameterized as

$$\frac{1}{\alpha_{ij}} = \tilde{\alpha}_{ij} = -(\beta_c C_i + \beta_d z_i^d + \beta_e E_i + \gamma_2\theta_j + \eta_i), \tag{3}$$

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

[14] 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).

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

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

(4)
$$VMT_{ij}^* = \frac{1}{\lambda} - \frac{\tilde{\alpha}_{ij}}{\lambda}\left(\frac{p_i^g}{MPG_j}\right).$$

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.

3.2. **Vehicle choice**

In the first period, the new vehicle purchaser weighs the 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.[16] The consumer expectations of period-two utility and resale value are taken over the joint distribution of consumer beliefs about future gasoline prices and economic conditions, which is denoted as G.

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

$$u_1(u_2, \theta_j, p_j) = \int \left(\delta_1 u_2 + \gamma_1 \theta_j + \delta_2 p_j^R - p_j + \epsilon_{ij}\right) dG,$$

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

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.

Since only u_2 and p_j^R contain stochastic terms, we can rewrite the first period utility in a more transparent form as follows:

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

The consumer's expected utility from driving, $\mathbb{E}[u_2]$, is based on the expectation of (2), where the expectation is taken over the joint distribution of consumer beliefs about period-two gasoline prices and economic conditions. Importantly, η_i^u does not enter into the expected period two utility by construction.[17] Plugging in (2), we find that the first period utility is

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

This form of the first period 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 the joint stochastic processes of these two factors.[18] Yet very little empirical work is available to answer this question. In a recent study, Anderson, Kellogg, Sallee, and Curtin (2011) 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

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

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

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 (2011)).

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 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), 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:

$$u_1 = \frac{\delta_2}{\lambda} + \frac{\delta_1}{2\lambda}\mathbb{E}[\frac{1}{\tilde{\alpha}_{ij}}] - \delta_2 \mu_j BM_j - \frac{\delta_1}{\lambda}\frac{\mathbb{E}[p_i^g]}{MPG_j} + \frac{\delta_1 \mathbb{E}[\tilde{\alpha}_{ij}]}{2\lambda}\frac{\mathrm{var}(p_i^g) + \mathbb{E}[(p_i^g)]^2}{(MPG_j)^2} + \gamma_1\theta_j - p_j + \delta_2 p_j^{R0} - \frac{\delta_2 \mathbb{E}[\tilde{\alpha}_{ij}]}{\lambda}\frac{\mathbb{E}[p_i^g]}{MPG_j} + \epsilon_{ij}. \tag{5}$$

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. Econometric Model: Identification and Estimation

4.1. 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 $|\mathcal{J}_i| = J_i$:

$$\mathrm{Pr}_i(j) = \frac{\exp(V_{ij})}{\sum_{k=1}^{J_i}\exp(V_{ik})}, \tag{6}$$

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

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

Note that (6) 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 1.

Abstracting from the corner solution, we can rearrange (4) as follows:

$$VMT_{ij} = \frac{1}{\lambda} + \left(\frac{p_i^g}{\lambda MPG_j}\right)(\beta_c C_i + \beta_d z_i^d + \beta_e E_i + \gamma_2 \theta_j) + \left(\frac{p_i^g}{\lambda MPG_j}\right)\eta_i \tag{7}$$

This expression implies that the conditional distribution for VMT_{ij} is given by $VMT_{ij} \sim$ 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:

$$L_i = \int \prod_{j=1}^{J_i} \left(\Pr_i(j)\ l_i(VMT_{ij}|j \text{ chosen})\right)^{\mathbb{1}_{ij}} dF_{\eta_i^k}, \tag{8}$$

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.[19] 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

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

vehicle purchase. This likelihood function is conditional on a vehicle being purchased and on the parameter estimates.

4.2. **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 crucial role in identifying selection. 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. My 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.[20] 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

[20]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.

purchases of vehicles in 2006, when gasoline prices had 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 associated with each vehicle model (i.e., to account for the ξ_j as in Berry, Levinsohn, and Pakes (1995)).[21] Finally, different vehicle models 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. Estimation Strategy

Estimation of the structural model developed here may be possible using a variety of methods. For this paper, I perform the estimation using Maximum Simulated Likelihood (MSL).[22] To find the optimal vector of coefficients, I use the Broyden-Fletcher-Goldfarb-Shanno algorithm alternating with the Newton-Raphson algorithm at every 10 iterations.[23] MSL is consistent under the assumption that the number of simulated draws 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

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

[22]I also reformulate the problem to be able to use a Bayesian Markov Chain Monte Carlo technique (Gibbs Sampler with adaptive Metropolis-Hastings) and find very similar coefficients using a very small subsample of my dataset. I opt for MSL for the results shown here because I found MSL was faster once reasonable starting values are known.

[23]I alternate between these two algorithms to reduce the chance that the 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).

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.[24] 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. A next step is currently underway 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).

5. Results

5.1. Parameter estimates

The estimates of the coefficients in the full structural model are listed in Table 5.[25] Since the purpose of this study is to examine the consumer responsiveness to gasoline price shocks 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.[26] Whether consumers "undervalue" fuel savings relative to other decisions made in the market (which would correspond

[24]I plan to perform a sensitivity analysis by allowing the choice set to include all vehicle types purchased in that quarter, but this will likely take weeks of computing time.

[25]These results are from an estimation with 1% of the full sample randomly drawn, which keeps the computation time down to under a week on my 2.667 GHz Intel i7 processor (four core) computer.

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

with studies finding a high implicit discount rate) remains unclear in the literature. 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). For this study, I follow Busse, Knittel, and Zettelmeyer (2010) and Sallee, West, and Fan (2009), and save further exploration of a present-bias in automobile purchases for future work.

Most of the coefficient estimates in Table 5 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 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.

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.

5.2. **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 5, I find an average elasticity of 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, which finds a U.S. short-run utilization elasticity with respect to the gasoline price in the range of -0.10 to -0.16 and a long-run elasticity in the range of -0.26 to -0.31 (Austin 2008). Several studies, such as West (2004), indicate that consumers are much more elastic, but these studies tend to be based on relatively poor quality survey data and usually rely primarily on cross-sectional variation, so that the estimates are probably best interpreted as long-run elasticities. More importantly, some recent studies using time series data 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 descriptive results, 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. Knittel and Sandler (2010) use California smog check data from 1996 to 2009 for *all* vehicles in the vehicle stock (regardless of age) 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 younger (and usually with higher fuel economy).

From looking at the aggregate data, an elasticity in the range of -0.15 appears plausible. Figure 9 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.

In the longer run, the extensive margin is 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 0.09. 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 (2010) 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 the drivers of 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 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 study helps to fill in the gaps in our knowledge about how the long- and short-run responses to changing gasoline prices occur.

5.3. **Importance of Selection**

It is notable that the estimates from the structural model do not exactly match the descriptive results, which suggested a driving elasticity in the range of -0.25 and a fuel economy elasticity in the range of 0.10. This may not be too surprising, for the model specifications differ, and in particular, the descriptive evidence does 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.

The parsimonious nature of the structural model permits separating the vehicle choice decision from the utilization decision. In particular, (9) can be rewritten as follows:

$$VMT_i^*\left(\frac{MPG_i}{p_i^g}\right) = \frac{1}{\lambda}\left(\frac{MPG_i}{p_i^g}\right) + \frac{1}{\lambda}(\beta_c C_i + \beta_d z_i^d + \beta_e E_i + \gamma_2 \theta_i) + \tilde{\eta}_i, \tag{9}$$

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 6. 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 (8) 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.

5.4. **Heterogeneity in Elasticities**

How each of the elasticities varies across different income, demographic, and geographic groups has important implications for the distributional consequences of a policy that changes the gasoline price. In addition, it can provide insights into likely improvements in congestion and local air pollution from such a policy if more congested or polluted regions have a higher or lower elasticity. In this section, I focus on the driving elasticity, for it has greater immediate policy relevance.[27] I break down the driving elasticity by income groups, selected demographics, and geography by examining the mean responsiveness in each of these groups calculated in my counterfactual analysis of a marginal increase in the gasoline price. I reserve the discussion of the geographic heterogeneity for the discussion of the gasoline tax policy, when I discuss the distributional consequences of the policy.

Table 7 shows the heterogeneity in the mean elasticity of driving by new vehicles purchased by each of the household income groups in my dataset. There appears to be somewhat of a "U-shape," with the highest elasticities in the lowest income bracket and highest income bracket. Consumers in the lowest income bracket may have a higher elasticity simply due to the tighter budget constraint, but perhaps also partly due to better availability of public transit. Consumers in the highest income bracket may have less need to respond to gasoline prices, yet at the same time may have a lower marginal utility from an additional mile driven. At the same time, they may own several vehicles, so the higher elasticity may be partly capturing wealthy consumers switching vehicles to a newer and more efficient vehicle. Similarly, they may be more inclined to switch to air travel for longer trips. A somewhat similar "U-shape" pattern was also been observed in the driving elasticity with respect to the operating cost of driving by West (2004), and in the gasoline demand elasticity by Wadud, Graham, and Noland (2009). Both of these studies use survey data and include all vehicles–not just new vehicles.

Table 7 also presents the mean elasticity of driving by several zip code demographic groups. Vehicles registered in the most densely populated zip codes in California (i.e., greater than the 75th percentile) appear to have a much higher elasticity than those in the least densely populated zip codes, which may relate to greater access to public transportation. Vehicles registered in counties with higher average commute times tend to be much less elastic than those in counties with lower commute times. This may capture the fact that consumers who commute longer distances (e.g., from counties in the Central Valley just outside of the Bay Area, or counties in the suburbs of Los

[27]Similar patterns emerge in the fuel economy elasticity.

Angeles) often do not have many viable alternatives to driving. Vehicles registered in zip codes with a higher percentage of the population greater than 65 years old tend to have a much higher elasticity than those in zip codes with a much lower percentage, possibly reflecting more flexibility by retirees to change driving habits. 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 have a lower elasticity. This may reflect the less flexible driving needs of families with children. Finally, there is no clear pattern of a difference in elasticities by the percentage of the population being of different races.

5.5. **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 (5). 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 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. Work is underway to use the average gasoline price and economic conditions over the 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), and use estimates from a survey of consumer beliefs. One possible survey that includes beliefs is the Michigan Survey of Consumers, used by Anderson, Kellogg, Sallee, and Curtin (2011).

6. **Counterfactual Simulations**

I perform counterfactual simulations of a gasoline tax policy and feebate policy to illustrate the advantage of a utility-consistent joint model of vehicle choice and subsequent driving. These policy simulations should not be considered a complete policy analysis, but rather as demonstrative of the usefulness of the model for the welfare analysis of policy. In this section I calculate the welfare effects *before* accounting for the external costs of driving. Thus, the estimates of the excess burden can be compared to magnitudes in the literature of the external costs of driving.

The gasoline tax policy and feebate 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, while the feebate will actually increase some of these externalities from driving.[28]

The reason for the difference is simply that the gasoline tax works on both the intensive and extensive margins, by influencing consumer decisions on vehicle choice *and* driving. In contrast, the feebate policy indirectly aims to internalize the externalities from automobile use by improving the fuel economy of the new vehicle fleet–thus working only on the intensive margin.[29] In this sense, it is a similar policy to CAFE standards. Both lead to the rebound effect. Unlike the gasoline tax, the feebate policy is not generally considered a revenue-raising policy, and the analysis here examines a revenue-neutral feebate. Also unlike an increase in the gasoline tax, a feebate policy is often considered to be more politically feasible.

6.1. **Gasoline Tax Policy**

I consider a gasoline tax policy that raises the price of gasoline by $1 per gallon (in real 2010$) over the entire time frame of the study. Note that the actual policy would require a gasoline tax increase of greater than $1 per gallon, for some of the burden of the tax would be borne by producers. For this analysis, I choose not to take a stand on the incidence of the gasoline tax, but instead focus on the welfare impacts on consumers. In particular, I find it most constructive to focus on the welfare impacts on a particular cohort of vehicle purchasers, such as all new vehicle purchasers in California in 2002. The welfare effects of a gasoline tax policy would occur through two channels: the welfare impact in the first period vehicle choice decision, and the welfare impact in the second period driving decision.

I find that the gasoline tax policy that raises the price of gasoline by $1 per gallon leads to a 4.9% decrease in driving on average. For the 1.6 million new vehicles purchased in the 2002 cohort, the gasoline tax revenue that would be raised from this policy from consumers (i.e., ignoring the

[28]See Parry and Small (2005) and Harrington, Parry, and Walls (2007) for excellent reviews of the externalities in automobile use.

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

amount that would be raised from producers) would be \$2.85 million per day, or roughly \$1.70 per vehicle per day on average. If the elasticity of demand is equal to the elasticity of supply and thus the incidence of the tax falls equally on producers and consumers (i.e., 50% pass-through), then the policy would raise roughly \$5.7 million per day from a \$2 per gallon tax (\$1 on producers and \$1 on consumers).[30] If pass-through is much closer to 100%, as suggested by Marion and Muehlegger (2011), then the policy would raise closer to \$2.85 million, and the deadweight loss burden would be borne nearly entirely by consumers.

The welfare effects on the 2002 cohort of vehicles are calculated for period two directly from (2). 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.

On the vehicle choice margin, the gasoline tax policy leads to a 3.4% increase in the average fuel economy of the fleet, averaged over vehicles purchased in all cohorts. To compute the welfare change to consumers, I use the approach derived by Small and Rosen (1981) that is applicable when a Type I extreme value error is assumed in the discrete choice model. In my case, the change in expected consumer surplus is calculated as:

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

where V_{ij}^c is the counterfactual 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.

[30] For comparison, California currently has a fixed \$0.18 per gallon excise tax and a 7.25% sales tax, so in 2008, the tax brought in approximately \$0.48 per gallon, or a total of \$15.3 million per day based on California total gasoline consumption from the US Energy Information Administration. Based my smog check data on the total stock of vehicles, there appear to be around 30 million registered light duty vehicles in California at any one time. Thus, \$5.7 million for the 1.6 million 2002 cohort of vehicles appears to be within the range that would be expected.

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. 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.

With a model of the incidence of gasoline taxation, an assumption about discount rates, and values for the magnitude of each of the externalities of driving, one could calculate the net welfare impacts of the policy. However, even without such assumptions, the primary findings are clear: the gasoline tax is a relatively non-distortionary tax when compared to the revenues it brings in. This can be largely attributed to a still-low elasticity of driving, even if I find that consumers are more elastic than suggested by some previous studies.

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, and lower-income households more than higher-income households. The distributional consequences depend importantly on how the revenue is recycled. Suppose there is no revenue recycling – the extreme case where the revenues are considered to be “lost.” Then the distributional consequences of the policy contain two components. The primary component is the transfer from consumers to the government, which is a function of the (post-policy) amount of driving and the distortion to consumer decisions from the policy. The secondary component is the excess burden. However, if the revenue is returned to consumers lump-sum based on the amount raised from them, then the distributional consequences are based on the differing excess burden.

In this section, I focus on the geographic distributional consequences at the household level. I use estimates from Calfornia Department of Transportation (2002) of the number of vehicles per household in each county in California to convert the per-vehicle results to per-household results. 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. This exacerbates the differing distributional consequences between more urban and more rural households.

Figure 10 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 10, I also calculate a rough estimate of the average amount each household pays to the government based on all counties having a county fleet 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.

Figure 11 indicates how the household-level deadweight loss varies across counties in California. Here I also include the deadweight loss from the distortion due to pre-existing gasoline taxes.[31] Figure 11 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 10 and 11 also shows 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 10, 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 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)

[31] The pre-existing gasoline tax in 2008 was roughly \$0.50 when the fixed and ad velorum sales gasoline taxes are added together.

for every county in California along with my above results would provide a picture of the overall welfare implications by county.

6.2. 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.[32] 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

[32] This tax credit expired at the end of 2010.

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.

My analysis of a feebate policy here is intended to illustrate how the structural model 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 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, Patterson, Singh, and Li (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 just over 1 percent for all new vehicles, including those consumers who chose the same vehicle. This corresponds to a direct 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 direct 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 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.[33] Future work can examine the welfare implications when the gasoline price is kept constant at the current price. I

[33]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).

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.

7. Conclusions

This paper develops a new utility-consistent framework for jointly modeling vehicle choice and utilization decisions and applies this framework to analyze both a gasoline tax policy and a feebate policy. The framework takes advantage of a massive and novel dataset that includes the vehicles chosen and subsequent amount driven in California over a period with considerable variation in the gasoline price. It explicitly accounts for selection on the unobserved driving type and allows for a clean analysis of the importance of such selection. The framework is based on a two period setting, where consumer vehicle choice and utilization decisions are modeled as based on the gasoline price and economic conditions at the time of each decision.

The findings suggest that consumers respond to changes in gasoline prices in both vehicle choice and utilization decisions. The decrease in driving from higher gasoline prices provides a short to medium term effect that initially has a far greater impact on the demand for gasoline than the change in new vehicle choice, with an average elasticity of driving with respect to the gasoline price of -0.15 for all vehicles in the first six years of life. However, the average elasticity of new vehicle fuel economy with respect to the price of gasoline is estimated to be 0.09, which indicates that a shock in the gasoline price today will have an effect on the stock of vehicles that will last long into the future. These two estimates help to break down the components of the gasoline price elasticity, improving our understanding of the nature of the response to an increase in gasoline prices. Moreover, they indicate that in the medium-run there is still a clear response, while other recent evidence suggests that, at least in the short-run, the response is very small (Small and Van Dender 2007; Hughes, Knittel, and Sperling 2008).

Yet the response estimated in this paper indicates that gasoline demand is still quite inelastic–a point with important ramifications for policy. The counterfactual policy simulations reflect this result. I examine a gasoline tax policy that raises the price of gasoline by $1 per gallon. The policy brings in a substantial amount of revenue with a quite small welfare loss. However, the distributional consequences from such a policy are likely to be important and substantial – with ramifications for the political feasibility of the policy.

I also use the structural model to examine a feebate policy that incentivizes consumers to purchase higher fuel economy vehicles. The increased fuel economy from the different new vehicles purchased leads consumers to drive these vehicles more with a fleet-wide elasticity of driving with respect to the fuel economy of the fleet of 0.07. This is in some respects is a more pure estimate of the "direct rebound effect" of a policy to improve the fuel economy of vehicles than in previous literature.

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 in the past. 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 depend very much 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.

This work lends itself to several avenues of future research. One promising avenue is to quantify consumer beliefs about future gasoline prices and economic conditions. If a reasonable joint distribution of these can be found based on information available to the consumer, then these stochastic processes can be explicitly modeled in the framework developed in this paper. Another promising avenue for further research is to examine the possibility of present-bias in consumer decisions about the fuel economy of the vehicle. The dataset collected in this project has significant potential to shed light on this issue. Finally, one could imagine a wide variety of policy counterfactuals using this dataset and framework in order to provide guidance for policymakers interested in reducing gasoline demand and emissions from the transportation sector.

References

Adda, J., and R. Cooper (2000): "Balladurette and Juppette: A Discrete Analysis of Scrapping Subsidies," *Journal of Political Economy*, 108(4), 778–806.

Allcott, H., and N. Wozny (2010): "Gasoline Prices and the Fuel Economy Discount Puzzle," *MIT Working Paper,*, Cambridge, MA.

Alquist, R., and L. Kilian (2010): "What Do We Learn from the Price of Crude Oil Futures?," *Journal of Applied Econometrics*, 25, 539–573.

Alquist, R., L. Kilian, and R. Vigfusson (2011): "Forecasting the Price of Oil," in *Handbook of Economic Forecasting*, ed. by G. Elliott, and A. Timmerman. Elsevier:forthcoming.

Anderson, S., R. Kellogg, J. Sallee, and R. Curtin (2011): "Forecasting Gasoline Prices Using Consumer Surveys," *American Economic Review Papers and Proceedings*, forthcoming.

Austin, D. (2008): "Effects of Gasoline Prices on Driving Behavior and Vehicle Markets," *Congressional Budget Office Report 2883,*, Washington, DC.

Austin, D., and T. 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.

Bento, A., L. Goulder, M. Jacobsen, and R. von Haefen (2009): "Distributional and Efficiency Impacts of Increased US Gasoline Taxes," *American Economic Review*, 99(3), 667–699.

Berkowitz, M., N. Gallini, E. Miller, and R. Wolfe (1990): "Disaggregate Analysis of the Demand for Gasoline," *Canadian Journal of Economics*, 23(2), 253–275.

Berry, S., J. Levinsohn, and A. 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.

Bresnahan, T. (1981): "Departures from Marginal Cost Pricing in the American Automobile Industry: Estimates for 1977-1978," *Journal of Econometrics*, 17(2), 201–227.

Busse, M., C. Knittel, and F. 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.

Copeland, A., W. Dunn, and G. Hall (2011): "Inventories and the Automobile Market," *RAND Journal of Economics*, 42(1), 121–149.

Davis, L. (2008): "Durable Goods and Residential Demand for Energy and Water: Evidence from a Field Trial," *RAND Journal of Economics*, 39(2), 530–546.

Davis, L., and L. Kilian (2011): "Estimating the Effect of a Gasoline Tax on Carbon Emissions," *Journal of Applied Econometrics*, 26(3), forthcoming.

Dubin, J., and D. McFadden (1984): "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52(2), 345–362.

Einav, L., A. Finkelstein, P. Schrimpf, S. Ryan, and M. Cullen (2010): "Selection on Moral Hazard in Health Insurance," *Stanford University Working Paper,*, Stanford, CA.

Erdem, T., M. Keane, and B. 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.

Esteban, S., and M. Shum (2007): "Durable-goods Oligopoly with Secondary Markets: The Case of Automobiles," *RAND Journal of Economics*, 38(2), 332–354.

Feng, Y., D. Fullerton, and L. Gan (2005): "Vehicle Choices, Miles Driven, and Pollution Policies," *NBER Working Paper 11553,*, Cambridge, MA.

Gillingham, K. (2010): "Identifying the Elasticity of Driving: Evidence from a Gasoline Price Shock in California," *Stanford University Working Paper*.

Goldberg, P. (1998): "The Effects of the Corporate Average Fuel Efficiency Standards in the US," *The Journal of Industrial Economics*, 46(1), 1–33.

Greene, D., P. Patterson, M. Singh, and J. Li (2005): "Feebates, Rebates and Gas-guzzler Taxes: A Study of Incentives for Increased Fuel Economy," *Energy Policy*, 33(6), 757–775.

Harrington, W., I. Parry, and M. Walls (2007): "Automobile Externalities and Policies," *Journal of Economic Literature*, 45, 374–400.

Hubbard, T. (1998): "An Empirical Examination of Moral Hazard in the Vehicle Inspection Market," *RAND Journal of Economics*, 29(2), 406–426.

Hughes, J., C. Knittel, and D. Sperling (2008): "Evidence in a Shift in the Short-Run Price Elasticity of Gasoline Demand," *Energy Journal*, 29(1), 113–134.

Jacobsen, M. (2010): "Evaluating U.S. Fuel Economy Standards in a Model with Producer and Household Heterogeneity," *UCSD Working Paper,*, San Diego, CA.

Kahn, J. (1986): "Gasoline Prices and the Used Automobile Market," *Quarterly Journal of Economics*, 101(2), 323–339.

Kahn, M. (1996): "New Evidence on Trends in Vehicle Emissions," *RAND Journal of Economics*, 27(1), 183–196.

Kilian, L., and E. 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, T., and J. Linn (2010): "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, C. (2010): "Automobiles on Steriods: Product Attribute Trade-offs and Technological Progress in the Automobile Sector," *American Economic Review*, forthcoming.

Knittel, C., and R. Sandler (2010): "Carbon Prices and Automobile Greenhouse Gas Emissions: The Extensive and Intensive Margins," *NBER Working Paper 16482,*, Cambridge, MA.

Larrick, R., and J. Soll (2008): "The MPG Illusion," *Science*, 320(5883), 1593–1594.

Li, S., C. Timmons, and R. von Haefen (2009): "How Do Gasoline Prices Affect Fleet Fuel Economy," *American Economic Journal: Economic Policy*, 1(2), 113–137.

Lin, C., and L. Prince (2009): "The Optimal Gas Tax for California," *Energy Policy*, 37(12), 5173–5183.

Mannering, F., and C. Winston (1985): "A Dynamic Empirical Analysis of Household Vehicle Ownership and Utilization," *RAND Journal of Economics*, 16(2), 215–236.

Marion, J., and E. Muehlegger (2011): "Fuel Tax Incidence and Supply Conditions," *Journal of Public Economics*, forthcoming.

Parry, I., and K. Small (2005): "Does Britian or the United States Have the Right Gasoline Tax?," *American Economic Review*, 95, 1276–1289.

Petrin, A. (2002): "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 110(4), 705–729.

Sallee, J. (2010): "Taxation of Fuel Economy," *NBER Working Paper 16466,*, Cambridge, MA.

Sallee, J., S. West, and W. Fan (2009): "Consumer Valuation of Fuel Economy: A Microdata Approach," *Proceedings of the National Tax Association Annual Conference on Taxation,*, Denver, CO.

Schiraldi, P. (2010): "Automobile Replacement: A Dynamic Structural Approach," *London School of Economics Working Paper,*, London, UK.

Small, K., and H. Rosen (1981): "Applied Welfare Economics with Discrete Choice Models," *Econometrica*, 49(1), 105–130.

Small, K., and K. Van Dender (2007): "Fuel Efficiency and Motor Vehicle Travel: The Declining Rebound Effect," *Energy Journal*, 28(1), 25–51.

Stolyarov, D. (2002): "Turnover of Used Durables in a Stationary Equilibrium: Are Older Goods Traded More?," *Journal of Political Economy*, 110(6), 1390–1413.

Tanner, M., and W. H. Wong (1987): "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82(398), 528–540.

Train, K. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press, London, UK.

Wadud, Z., D. Graham, and R. Noland (2009): "Modelling Fuel Demand for Different Socio-economic Groups," *Applied Energy*, 86(12), 2740–2749.

West, S. (2004): "Distributional Effects of Alternative Vehicle Pollution Control Technologies," *Journal of Public Economics*, 88, 735.

Appendix A. California counties in the smog check program

There are 58 counties in California, 40 of which are covered by the smog check program. 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 Cleaning

This section describes the cleaning and merging of the 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 study, 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. Similarly, the dataset is restricted to vehicles that run on gasoline. Fortunately in the years my dataset covers, 97.7% of the new vehicles in California run on gasoline, with nearly all of the remainder running on diesel fuel. Then I perform the merges.

B.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,[34] 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.

[34]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).

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% 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% 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% of the full sample. In total, 76% 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% 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.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 56 different vehicle makes, 545 models, and 1,525 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% 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% of the R.L. Polk. The fuel economy data are then aggregated and matched in several iterations. 80% 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% of the vehicle types are assigned a fuel economy. For a sample of the last 20% 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% of the vehicle types are matched. The next aggregation is at the make-model level. After this merge, 88% 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.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%. 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% matched. After aggregating over the cylinders for each model, another 27% 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.4. **Economic Conditions**

Finally, I bring in two variables to capture 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.

TABLE 1. Personal New Vehicle Registrations in California

	Counts of Vehicles (thousands)								
	2001	2002	2003	2004	2005	2006	2007	2008	2009*
Small Car	289	265	270	290	302	317	318	288	76
Large Car	265	239	224	227	233	230	215	169	46
Sporty Car	80	67	53	48	48	48	35	21	6
Prestige Sporty	19	20	20	21	24	26	20	14	4
Luxury	149	150	155	164	165	164	156	125	38
Prestige Luxury	34	31	36	36	38	36	28	22	6
Pickup	121	103	97	90	87	80	64	40	10
Full Pickup	205	200	210	231	220	179	142	75	18
Sport Utility	301	311	327	351	342	330	308	222	72
Full Utility	129	135	139	131	110	102	77	41	10
Minivan	93	83	75	81	80	70	50	33	10
Total	1,687	1,605	1,607	1,672	1,648	1,581	1,413	1,052	295
	Fraction of Vehicles								
Small Car	0.17	0.17	0.17	0.17	0.18	0.20	0.23	0.27	0.26
Large Car	0.16	0.15	0.14	0.14	0.14	0.15	0.15	0.16	0.16
Sporty Car	0.05	0.04	0.03	0.03	0.03	0.03	0.02	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	0.06	0.05	0.05	0.05	0.05	0.04	0.04
Full Pickup	0.12	0.12	0.13	0.14	0.13	0.11	0.10	0.07	0.06
Sport Utility	0.18	0.19	0.20	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

Table 2. Summary Statistics

Variable	Mean	Std Dev	Min	Max	N
cylinders	5.81	1.56	2	12	12,559,783
liters	3.34	1.3	0.4	8.4	12,559,783
automatic transmission	0.96	0.21	0	1	12,559,783
gross vehicle weight rating	5,344	1,217	440	14,050	12,559,783
hybrid	0.02	0.15	0	1	12,559,783
import	0.66	0.47	0	1	12,559,783
safety rating	4.3	0.44	1	5	12,559,783
convertible	0.03	0.16	0	1	12,559,783
turbo	0.03	0.17	0	1	12,559,783
4WD or AWD	0.18	0.39	0	1	12,559,783
fuel economy (2008 ratings)	20.49	5.71	8	50	12,559,783
vehicle MSRP (2010$)	29,723	12,522	9,034	1,500,000	12,559,783
months to smog test	69.29	10.9	13	107	4,747,751
VMT	1,089.54	465.82	0	4,987	4,747,751
income	5.86	2.29	1	9	9,033,689
resale price of same model 6 yrs old	10,793.02	4,853.08	1,362	641,609	12,559,783
gas price at purchase (2010$)	2.59	0.63	1.25	5	12,559,783
county unemployment rate	5.97	2	2.8	27.1	12,559,783
consumer confidence index	93.56	18.44	25.3	118.9	12,559,783
zip density (000/mi^2)	5.07	5.49	0	52.18	12,559,783
commute time 2000 (minutes)	27.09	4.28	13.4	43.1	12,559,783
zip businesses 2000	1,514	960	1	6,521	12,559,783
zip population 2007	41,440	20,466	1	109,549	12,559,783
zip pop growth rate 00-07	1.77	3.1	-32.5	199.2	12,559,783
zip median hh income 2007	70,627	27,366	0	375,000	12,559,783
zip % pop age 65+	11.14	5.29	0	100	12,559,783
zip % pop under 18	25.73	6.06	0	41.3	12,559,783
zip % pop white 2007	59.67	18.54	4.4	100	12,559,783
zip % pop black 2007	5.14	7.43	0	86.60	12,559,783
zip % pop hispanic 2007	31.93	21.45	0	97.8	12,559,783
zip lawn & garden SPI	118.47	55.89	0	486	12,559,783

Table 3. Descriptive Evidence on the Intensive Margin

Dependent variable: vehicle-miles-traveled per month (mean = 1,090 miles per month)

	(1) base	(2) six yr sample	(3) county FE	(4) month of year FE	(5) model & month FE	(6) year FE	(7) 0.25 quantile	(8) 0.50 quantile	(9) 0.75 quantile
average gas price	-106.3***	-106.3***	-114.6***	-115.2***	-126.6***	-97.5***	-256.1***	-146.7***	-63.2***
	(1.4)	(2.6)	(1.5)	(1.5)	(8.1)	(2.8)	(1.7)	(1.6)	(2.0)
lease	19.0***	41.4***	19.2***	18.9***	25.2***	6.0***	-96.4***	17.4***	-24.3***
	(0.5)	(0.9)	(0.6)	(0.5)	(3.5)	(0.6)	(0.7)	(0.7)	(0.9)
liters	-11.1***	-4.3***	-11.2***	-10.6***	-25.9***	-9.6***	53.7***	-43.8***	-4.0***
	(0.6)	(0.8)	(0.6)	(0.6)	(5.8)	(0.6)	(0.7)	(0.6)	(0.8)
cylinders	-6.8***	-11.5***	-6.4***	-6.9***	-4.6	-8.0***	6.3***	-18.6***	-7.4***
	(0.4)	(0.6)	(0.4)	(0.4)	(4.9)	(0.4)	(0.4)	(0.4)	(0.5)
fuel economy	2.3***	3.5***	2.6***	2.4***	-0.0	3.5***	11.6***	-13.3***	4.0***
	(0.1)	(0.2)	(0.1)	(0.1)	(3.2)	(0.1)	(0.2)	(0.2)	(0.2)
turbo	-37.8***	-30.8***	-36.8***	-37.6***	-20.8**	-42.2***	223.5***	-14.6***	-26.8***
	(1.2)	(1.9)	(1.2)	(1.2)	(7.0)	(1.2)	(1.9)	(1.6)	(2.0)
auto	-39.3***	-48.2***	-38.8***	-39.2***	-40.5***	-37.0***	323.3***	-58.5***	-33.2***
	(0.8)	(1.1)	(0.8)	(0.8)	(8.4)	(0.8)	(0.8)	(0.9)	(1.1)
gross veh weight	-0.0***	-0.0***	-0.0***	-0.0***	-0.0	-0.0***	-0.0***	-0.0***	-0.0***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)	(0.0)
four wheel drive	4.1***	10.9***	6.6***	4.7***	20.7***	4.5***	221.7***	-31.1***	4.2***
	(0.7)	(1.0)	(0.7)	(0.7)	(3.7)	(0.7)	(0.9)	(0.8)	(1.0)
safety rating	-7.9***	-9.1***	-7.9***	-8.2***	29.6*	-13.4***	-147.3***	-28.7***	-23.9***
	(0.5)	(0.8)	(0.5)	(0.5)	(13.3)	(0.5)	(0.7)	(0.6)	(0.8)
hybrid	-20.3**	-33.4**	-20.2**	-19.2**	-16.4	-51.9***	-237.0***	164.6***	-50.4***
	(6.8)	(12.2)	(6.8)	(6.8)	(28.7)	(6.8)	(6.7)	(6.9)	(8.7)
import	9.2***	25.3***	10.2***	9.4***		12.3***	194.1***	48.9***	1.3
	(0.6)	(0.9)	(0.6)	(0.6)		(0.6)	(0.7)	(0.7)	(0.9)
log(zip pop)	1.0*	0.4	-5.0***	0.9	-0.4	1.7***	150.9***	13.3***	-7.0***
	(0.5)	(0.7)	(0.5)	(0.5)	(0.9)	(0.5)	(0.5)	(0.5)	(0.7)
zip pop growth 00-07	4.1***	3.9***	2.2***	4.0***	4.0***	3.6***	24.7***	5.1***	0.9***
	(0.1)	(0.1)	(0.1)	(0.1)	(0.2)	(0.1)	(0.1)	(0.1)	(0.1)
density (000/mi^2)	-6.9***	-7.1***	-5.3***	-6.9***	-6.7***	-6.9***	0.0	-1.7***	-7.7***
	(0.0)	(0.1)	(0.1)	(0.0)	(0.2)	(0.0)	(0.1)	(0.1)	(0.1)
log(zip businesses)	-18.8***	-21.0***	-15.0***	-18.9***	-18.1***	-18.9***	8.9***	-14.0***	-22.8***
	(0.4)	(0.5)	(0.4)	(0.4)	(0.8)	(0.4)	(0.4)	(0.4)	(0.5)
log(zip income)	-61.1***	-55.2***	-35.4***	-61.0***	-55.0***	-57.1***	164.7***	6.9***	-85.4***
	(0.8)	(1.2)	(1.1)	(0.8)	(3.4)	(0.8)	(1.0)	(0.9)	(1.2)
zip % pop age 65+	-3.8***	-4.6***	-3.7***	-3.8***	-3.5***	-3.7***	-0.5***	-0.1	-4.3***
	(0.1)	(0.1)	(0.1)	(0.1)	(0.2)	(0.1)	(0.1)	(0.1)	(0.1)
zip % pop under 18	3.5***	3.3***	3.2***	3.5***	3.3***	3.5***	-15.9***	9.7***	4.2***
	(0.1)	(0.1)	(0.1)	(0.1)	(0.2)	(0.1)	(0.1)	(0.1)	(0.1)
zip % pop white	0.9***	0.9***	0.6***	0.9***	0.9***	0.7***	2.4***	1.4***	1.0***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(0.0)	(0.0)	(0.0)	(0.0)
zip % pop black	0.1*	-0.0	0.2***	0.1*	0.1	-0.0	2.6***	2.2***	-0.1
	(0.0)	(0.1)	(0.0)	(0.0)	(0.1)	(0.0)	(0.0)	(0.0)	(0.1)
zip % pop hispanic	0.1***	0.2***	0.1***	0.1***	0.1*	0.1***	4.9***	0.9***	-0.1*
	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(0.0)	(0.0)	(0.0)	(0.0)
commute time	5.6***	6.0***	6.9***	5.6***	5.6***	5.5***	0.3***	3.8***	8.7***
	(0.1)	(0.1)	(0.1)	(0.1)	(0.3)	(0.1)	(0.1)	(0.1)	(0.1)
unemployment rate	0.4***	-1.2***	1.4***	0.2	0.2	-0.2	33.8***	-2.7***	0.1
	(0.1)	(0.2)	(0.3)	(0.1)	(0.3)	(0.1)	(0.2)	(0.1)	(0.2)
cons conf index	-0.1***	-0.2***	-0.1***	-0.4***	-0.4***	0.4***	3.2***	-1.1***	0.2***
	(0.0)	(0.0)	(0.0)	(0.0)	(0.1)	(0.0)	(0.0)	(0.0)	(0.0)
constant	1,892.2***	1,680.4***	1,605.3***	1,976.6***	1,995.4***	1,727.1***	-3,749.2***	1,841.6***	2,387.3***
	(13.3)	(22.5)	(15.9)	(13.9)	(83.4)	(14.9)	(15.1)	(14.8)	(18.5)
county FE	N	N	Y	N	N	N	N	N	N
month of year FE	N	N	N	Y	N	N	N	N	N
model FE	N	N	N	N	Y	N	N	N	N
year & smog time FE	N	N	N	N	N	Y	N	N	N
veh body & class	Y	Y	Y	Y	Y	Y	Y	Y	Y
summer controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	4.75m	2.24m	4.75m	4.75m	4.75m	4.75m	4.75m	4.75m	4.75m

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

Table 4. Descriptive Evidence on the Extensive Margin

Dependent variable: new vehicle fuel economy (mean = 20.5 miles per gallon)

	(1) base	(2) county FE	(3) time poly	(4) year FE	(5) county & year FE
gas price at purchase	1.318***	1.412***	0.675***	0.606***	0.612***
	(0.003)	(0.050)	(0.005)	(0.005)	(0.042)
lease	-0.862***	-0.862***	-0.966***	-0.970***	-0.964***
	(0.004)	(0.102)	(0.004)	(0.004)	(0.103)
log(zip pop)	0.128***	0.133**	0.129***	0.128***	0.139**
	(0.004)	(0.041)	(0.004)	(0.004)	(0.041)
zip pop growth 00-07	-0.014***	-0.001	-0.015***	-0.015***	-0.002
	(0.001)	(0.003)	(0.001)	(0.001)	(0.004)
density (000/mi^2)	0.043***	0.030*	0.039***	0.039***	0.030*
	(0.000)	(0.013)	(0.000)	(0.000)	(0.013)
log(zip businesses)	0.036***	-0.008	0.023***	0.023***	-0.007
	(0.003)	(0.032)	(0.003)	(0.003)	(0.032)
log(zip income)	0.050***	-0.605***	-0.016*	-0.010	-0.597***
	(0.006)	(0.135)	(0.006)	(0.006)	(0.132)
zip % pop age 65+	-0.015***	-0.019*	-0.016***	-0.016***	-0.019*
	(0.000)	(0.008)	(0.000)	(0.000)	(0.008)
zip % pop under 18	-0.071***	-0.055***	-0.071***	-0.071***	-0.054***
	(0.000)	(0.009)	(0.000)	(0.000)	(0.009)
zip % pop white	-0.021***	-0.015***	-0.021***	-0.021***	-0.015***
	(0.000)	(0.003)	(0.000)	(0.000)	(0.003)
zip % pop black	-0.004***	-0.009*	-0.004***	-0.004***	-0.009*
	(0.000)	(0.004)	(0.000)	(0.000)	(0.004)
zip % pop hispanic	-0.007***	-0.012***	-0.006***	-0.006***	-0.011***
	(0.000)	(0.003)	(0.000)	(0.000)	(0.003)
commute time	0.019***	-0.004	0.017***	0.017***	-0.004
	(0.000)	(0.008)	(0.000)	(0.000)	(0.008)
unemployment rate	0.001	0.151***	-0.051***	-0.049***	-0.009
	(0.001)	(0.038)	(0.001)	(0.001)	(0.012)
cons conf index	-0.019***	-0.012***	0.004***	-0.004***	-0.003***
	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)
registration month			6.437***		
			(0.154)		
(registration month)^2			-0.013***		
			(0.000)		
(registration month)^3			0.000***		
			(0.000)		
constant	19.612***	25.433***	-1,070.820***	20.854***	27.122***
	(0.081)	(1.631)	(27.583)	(0.082)	(1.528)
county FE	N	Y	N	Y	
month-year polynomial	N	N	Y	N	N
year FE	N	N	N	Y	Y
Observations	12.6m	12.6m	12.6m	12.6m	12.6m

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

Table 5. Structural Model Parameter Estimates

Period 1 γ coefficients			β coefficients		
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 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 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 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

TABLE 6. Parameter Estimates from Structural Model Run on Only Intensive Margin

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	12.6m	

*** indicates significant at 1% level, ** significant at 5% level
Robust standard errors in parentheses calculated using the delta method

Table 7. Heterogeneity in the Elasticity of Driving

	elasticity of driving	observations
\$0 - \$29,999	-0.212	2,202,019
\$30,000 - \$39,999	-0.122	835,623
\$40,000 - \$49,999	-0.153	920,101
\$50,000 - \$74,999	-0.144	1,058,655
\$75,000 - \$99,999	-0.135	2,330,401
\$100,000 - \$124,999	-0.169	1,416,895
>\$125,000	-0.177	1,645,893
density > 75th percentile	-0.189	306,350
density < 25th percentile	-0.087	306,358
commute > 75th percentile	-0.101	306,356
commute < 25th percentile	-0.161	306,351
% pop 65+ > 75th percentile	-0.202	306,358
% pop 65+ < 25th percentile	-0.078	306,353
% pop under 18 > 75th percentile	-0.057	306,356
% pop under 18 < 25th percentile	-0.227	306,353
% pop white > 75th percentile	-0.160	306,359
% pop white < 25th percentile	-0.149	306,350
% pop black > 75th percentile	-0.135	306,352
% pop black < 25th percentile	-0.147	306,356

Estimates are from a counterfactual with a marginal increase in the gasoline price

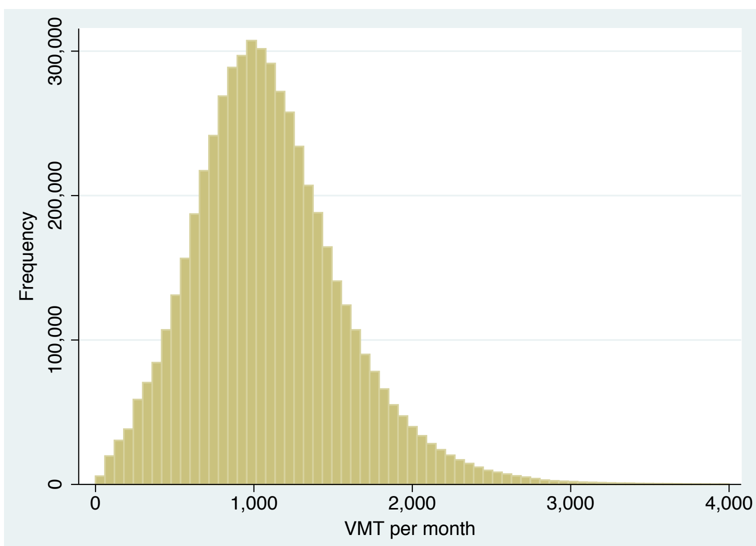

Figure 1. Driving per month by vehicles during their first six years of use in California has been remarkably single modal.

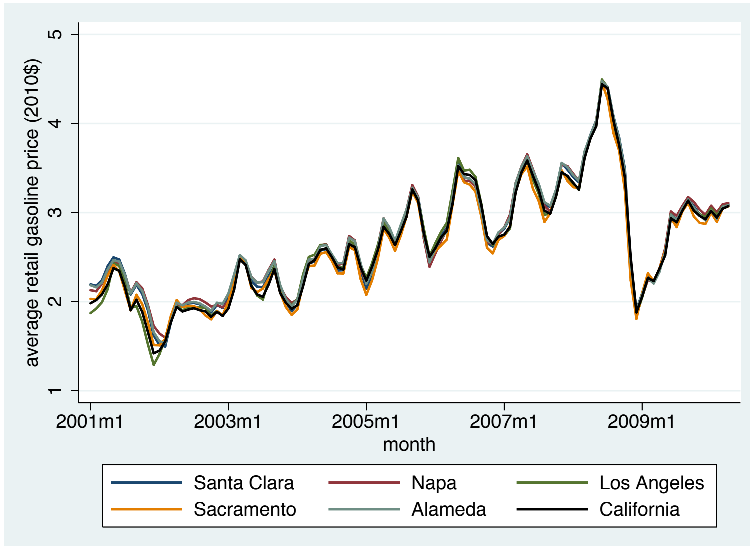

FIGURE 2. Retail gasoline prices in California were relatively flat and then rose substantially until 2008, providing substantial time series variation in addition to some cross-county variation. Four representative counties are shown here. Sources: Oil Price Information Service for the county time series and US Energy Information Administration (EIA) for the California average.

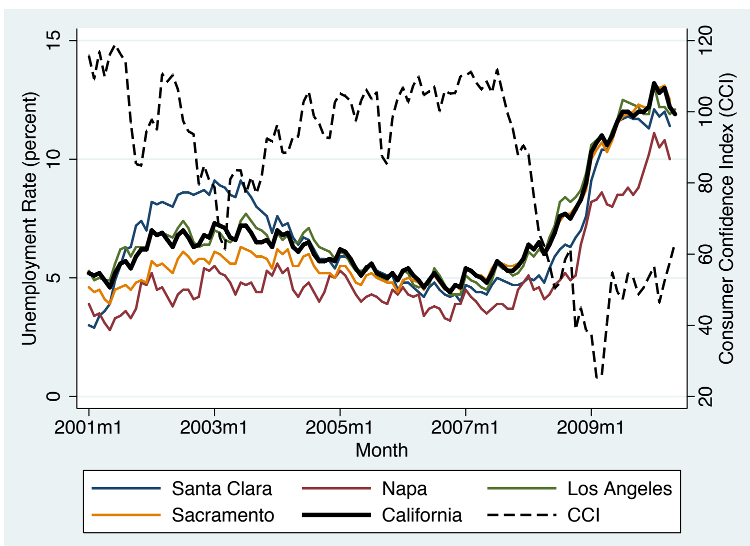

FIGURE 3. Unemployment and the consumer confidence index show that the economy was doing quite well until 2008, when the CCI plummeted and unemployment began increasing. Sources: Bureau of Labor Statistics for unemployment data and The Conference Board for CCI data.

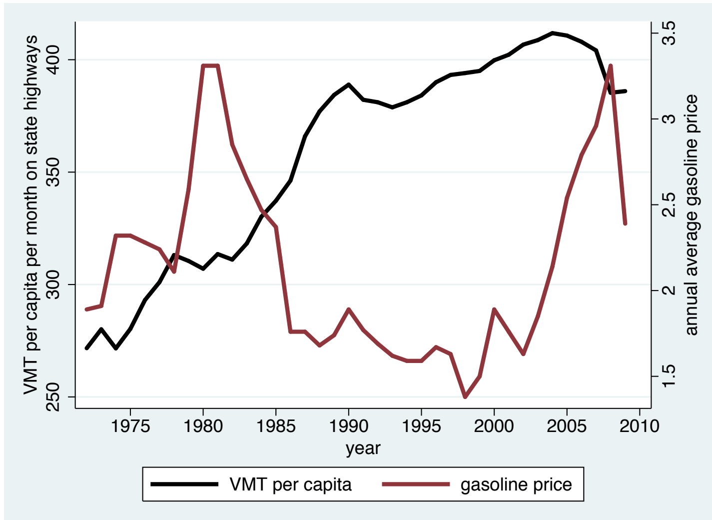

FIGURE 4. Vehicle miles traveled per capita has been increasing steadily, with the only deviations from this trend occurring during times of high gasoline prices (in real 2010$). Sources: California Department of Transportation state highway traffic counts and US Energy Information Administration.

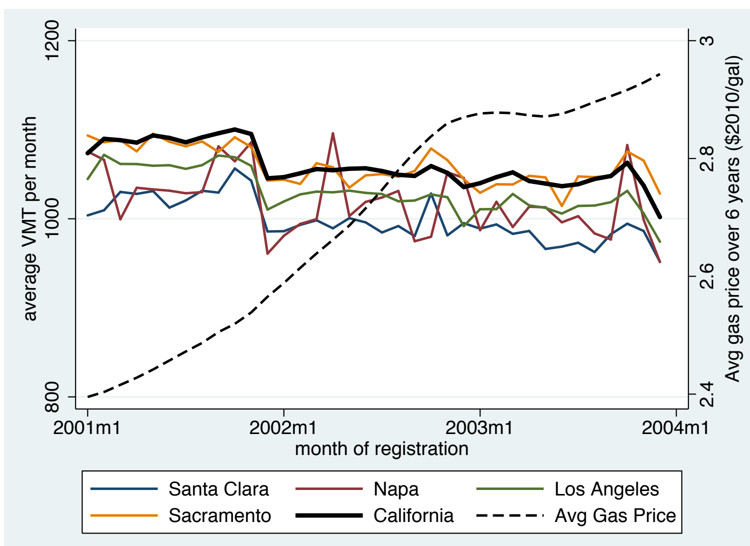

FIGURE 5. Average vehicle miles traveled per month has dipped slightly over the time frame of the study. The average in this graph is taken for all personal vehicles that received a smog check within two months of six years after registration (over 70% of the data), and is the average over the six years between the registration and smog check. The average gasoline price shown is the average over those same six years.

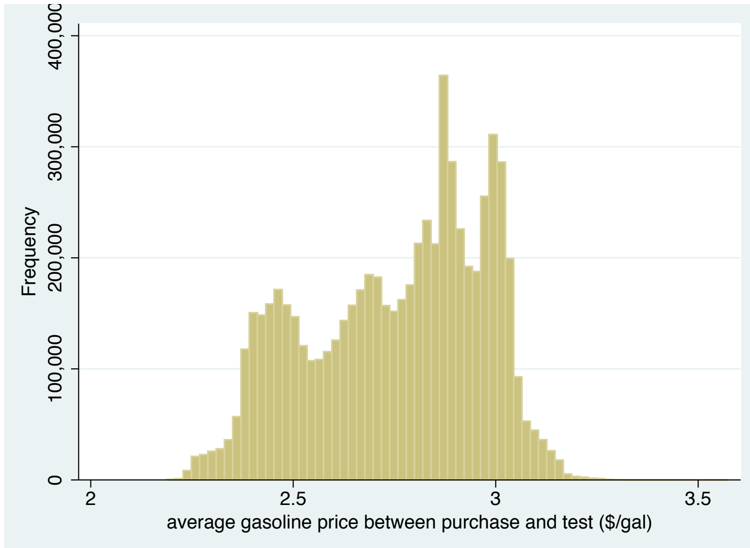

FIGURE 6. The histogram of the average gasoline price over the time between registration and the first smog check shows substantial variation, which reflects mostly time-series variation, but includes some cross-county variation.

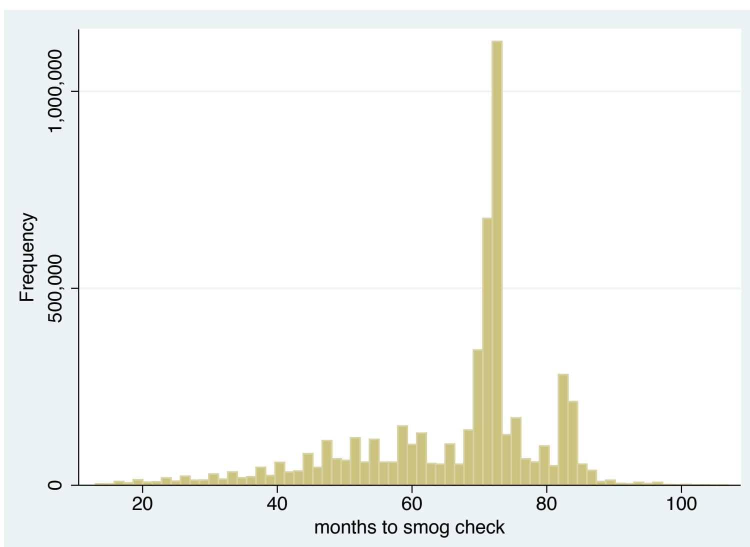

FIGURE 7. The histogram of the months between the first registration of a new car and the first smog check shows that roughly 40% had an early or late smog check.

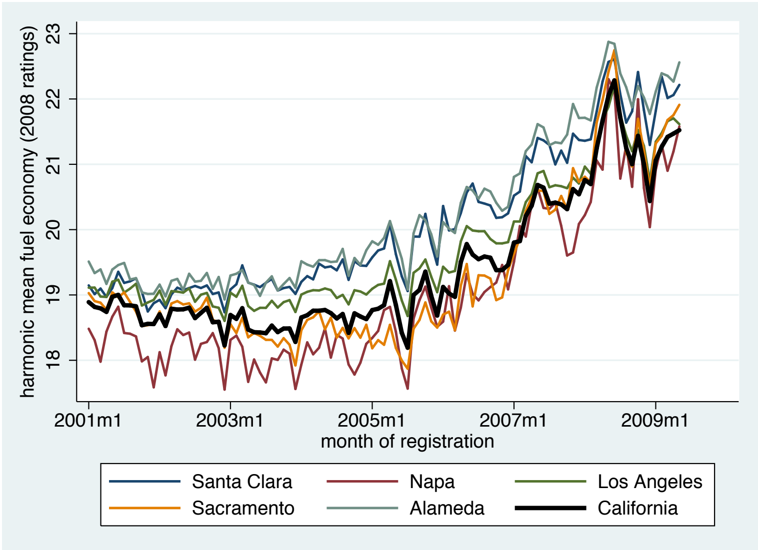

Figure 8. The average fuel economy of the new fleet in California was flat and then peaked at the same time as the gasoline price peaked.

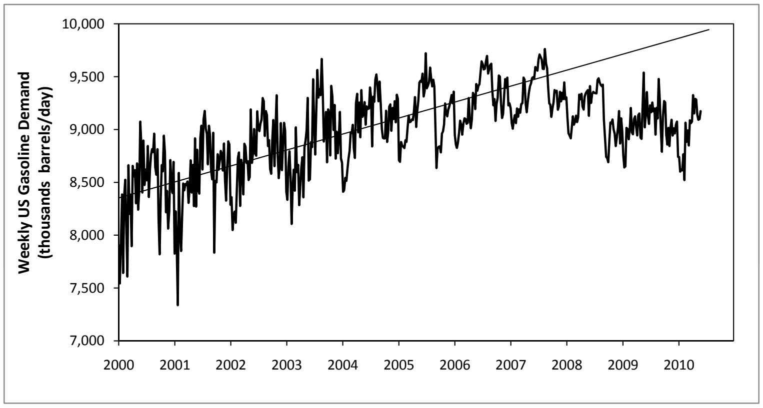

Figure 9. Gasoline demand in the US was increasing at a steady pace until the higher prices of 2007 and 2008 made an impact. Source: US Energy Information Administration.

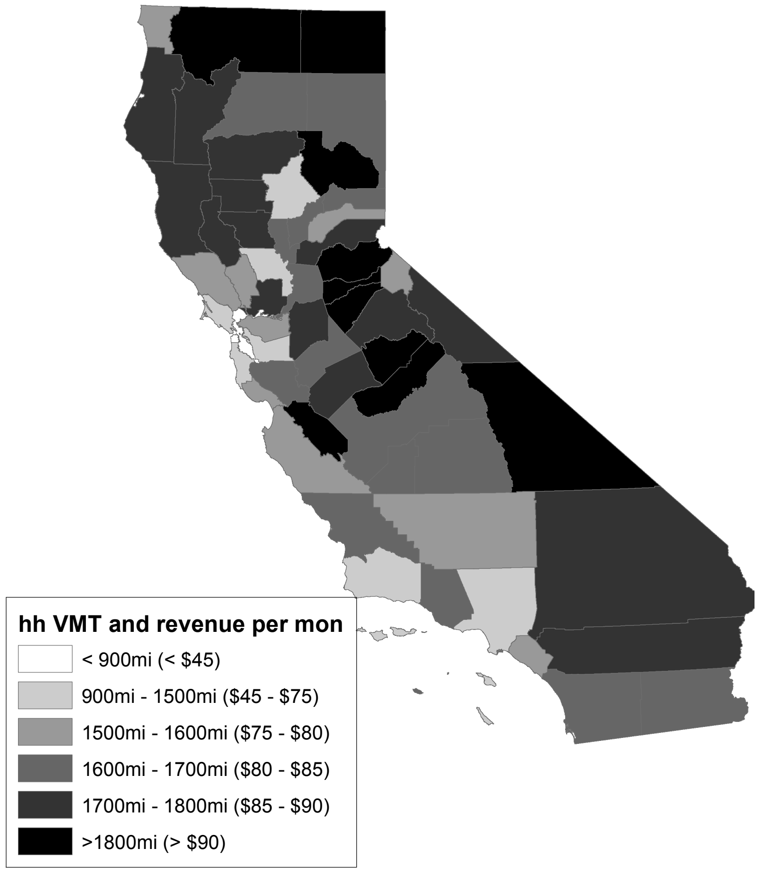

FIGURE 10. Households in rural areas in California drive more on average and thus pay more in gasoline taxes. The legend includes the county-level household driving per month and the average revenues to the government from households in that county from a one dollar gasoline tax (assuming an fuel economy of 20 mi/gal).

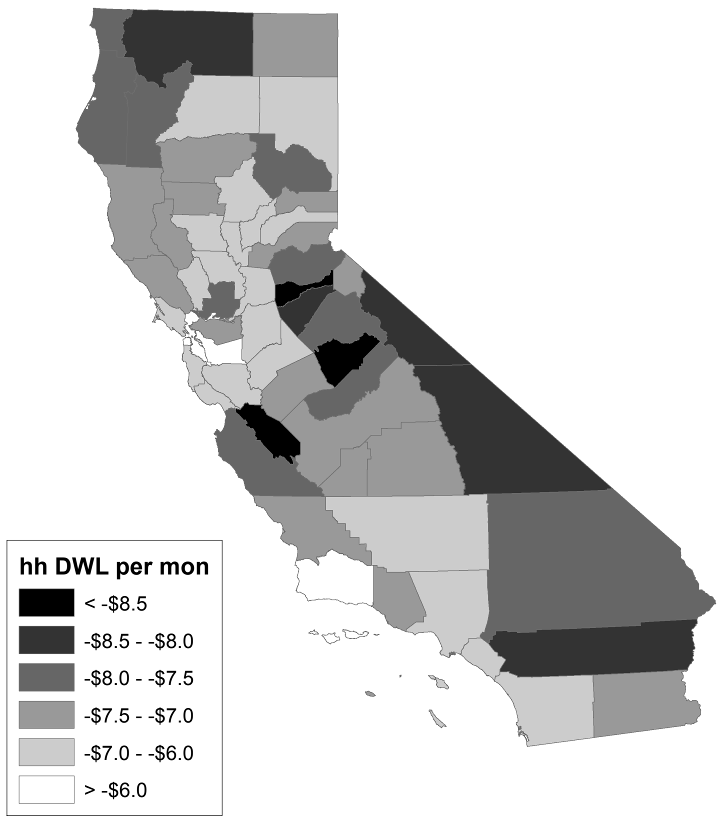

Figure 11. The deadweight loss for households varies across counties based on the vehicles per household and the differences in elasticities.