Identifying the Elasticity of Driving: Evidence from a Gasoline Price Shock in California

Kenneth Gillingham

Yale University
195 Prospect Street, New Haven, CT 06511
kenneth.gillingham@yale.edu

Abstract

There have been dramatic swings in retail gasoline prices over the past decade, along with reports in the media of consumers changing their driving habits – providing a unique opportunity to examine how consumers respond to changes in gasoline prices. This paper exploits a unique and extremely rich vehicle-level dataset of all new vehicles registered in California in 2001-2003 and then subsequently given a smog check in 2005-2009, a period of steady economic growth but rapidly increasing gasoline prices after 2005. The primary empirical result is a medium-run estimate of the the elasticity of vehicle-miles-traveled with respect to gasoline price for new vehicles of -0.22. There is evidence of considerable heterogeneity in this elasticity across buyer types, demographics, and geography. Surprisingly, the vehicle-level responsiveness is increasing with income, perhaps due to within-household switching of vehicles. The estimated elasticity has important implications for the effectiveness of price policies, such as increased gasoline taxes or a carbon policy, in reducing greenhouse gases. The heterogeneity in the elasticity underscores differing distributional and local air pollution benefits of

policies that increase the price of gasoline.

Keywords: Urban transportation, heterogeneity, vehicles, gasoline taxes
JEL Codes: R2, R4, Q4, Q5

1. Introduction

Starting in 2004 and early 2005, retail gasoline prices in the United States began creeping upwards, culminating in an increase of over 100% by 2008, the largest since the Mideast oil supply interruptions in the 1970s. Consumers can respond to gasoline price shocks on both the intensive and extensive margins by changing driving behavior, purchasing a more fuel-efficient vehicle or scrapping an old gas guzzler. Quantifying exactly how consumers respond has been a research topic of great interest to economists for decades, yet it remains just a relevant as ever for policy analysis of price policies to reduce greenhouse gas emissions from the vehicle fleet.

This paper focuses on the intensive margin of consumer response to the recent gasoline price shock by providing new evidence on the utilization elasticity for vehicles, i.e., the elasticity of vehicle-miles-traveled (VMT) with respect to the cost of driving. This study estimates the response in driving to changes in gasoline price by bringing together a novel vehicle-level dataset in which vehicle characteristics, vehicle purchaser characteristics, the odometer reading several years later, the location at both registration and at time of the odometer reading, and the relevant gasoline prices over the time period are all observed. This rich dataset, along with a careful research design, helps to overcome key identification challenges, while at the same time allowing for a closer look into the heterogeneity in consumer response at the demographic and geographic levels.

The empirical literature on the responsiveness of consumers to changes in gasoline prices has a long history going back to studies of traffic counts, a few empirical studies estimating a utilization elasticity, and an extensive literature estimating the price elasticity of gasoline demand. Austin (2008) reviews the older literature and finds that the utilization elasticity has been estimated to range from -0.10 to -0.16 in the short run and -0.26 to -0.31 in the long run.[1] Several of these empirical studies, such as Goldberg (1998) and West (2004), use cross-sectional survey data for VMT and estimate a utilization elasticity along with vehicle choice using the framework developed in Dubin and McFadden (1984).

More recent evidence suggests that the utilization elasticity is becoming more inelastic over time. One of the more notable studies in this literature is Small and Van Dender (2007), who simultaneously estimate a system of equations capturing the choice of aggregate VMT per capita, the size of the vehicle stock, and the fuel efficiency of the fleet. Small and Van Dender estimate this system on panel data of US states over the period 1966-2001 and find that at the sample averages of the variables the preferred short-run and long-run utilization elasticities are -0.045 and -0.222 respectively. They find less responsiveness in the period 1997-2001, a result they attribute to the growth of income and lower fuel prices over the time frame of their study. Hymel et al. (2010) use the same methodology with more recent data to reach similar conclusions. These results indicating a declining utilization

[1] The National Academies of Sciences report on CAFE standards had a similar range (National Research Council, 2002).

elasticity (in absolute value) are then interpreted as evidence of a declining "rebound effect," which can most intuitively be thought of as the elasticity of driving with respect to fuel economy improvements.[2] The idea behind this interpretation is that the cost per mile of driving would change with both a change in gasoline prices and a change in fuel economy, so a rational consumer would treat both the same.

Recent evidence on the gasoline price elasticity of demand also indicates that the responsiveness may be declining over time. Specifically, Hughes et al. (2008) find that the short-run gasoline price elasticity is in the range -0.21 to -0.34 for 1975-1980, but by 2001-2006 consumers are significantly less responsive, with an elasticity in the range of -0.03 and -0.08. Small and Van Dender (2007) also calculate a gasoline price elasticity and find evidence that it too is becoming closer to zero over time, with a preferred short-run estimate of -0.087 over 1966-2001 and -0.066 over 1997-2001. Older estimates of the gasoline price elasticity from the 1970s and 1980s often indicate much greater price responsiveness (e.g., see the survey by Graham and Glaister (2002)).

There is surprisingly limited empirical evidence quantifying the heterogeneity in consumer responsiveness across any dimension for either the gasoline price or utilization elasticity. West (2004) uses the 1997 Consumer Expenditure Survey to find that the lowest income decile has over a 50% greater responsiveness to gasoline price than the highest income decile, along with a U-shaped pattern of responsiveness. West and Williams (2004) also find that

[2] It is called a "rebound effect" due to the idea that people driving more is a "rebound" that reduces the benefits of fuel economy improvements.

the lowest expenditure quintile (elasticity of -0.7) has over three times the responsiveness of the highest expenditure quintile (elasticity of -0.2), but with no U-shaped pattern. Bento et al. (2009) find that families with children and owners of trucks and sport utility vehicles are more responsive to changes in gasoline prices, but the differences are relatively minor. Wadud et al. (2009) use Consumer Expenditure Survey data to find a U-shaped pattern of price elasticities of gasoline demand across income quintiles, while Wadud et al. (2010) use a different empirical strategy to find that the responsiveness to gasoline prices declines with income.

This paper uses observed VMT to provide new evidence on the gasoline price elasticity of driving and explore the heterogeneity in this elasticity. A primary result is an estimate of the medium run elasticity of VMT with respect to the price of gasoline for vehicles in their first several years of use in the range of -0.22. This result suggests that consumer responsiveness to the gasoline price shock of recent years was not inconsequential. Moreover, the results point to important heterogeneity by income group and geography, and some degree of heterogeneity by buyer type and demographics.

The paper is organized as follows. Section 2 describes the unique dataset assembled for this study. Section 3 discusses the estimation strategy and the basis for identification. Section 4 presents the empirical results, and Section 5 concludes.

2. Data

2.1. Data sources

The dataset used in this study was assembled from several sources. I begin with data from R.L. Polk's National Vehicle Population Profile on all new vehicle registrations in California from 2001 to 2003. An observation in this dataset is a vehicle, identified by the 17 digit vehicle identification number (VIN). There are roughly two million vehicles registered in each of these years, and the date of registration is observed. The dataset includes a variety of characteristics of the vehicle such as make, model, model year, trim, transmission type, fuel, doors, body type, engine size (liters), engine cylinders, and the presence of a turbo- or super-charger. The dataset also includes some details about the purchaser of the vehicle, such as the zip code in which the vehicle is registered, the transaction type (personal, firm, rental or government), and whether the vehicle was leased. The income of the purchaser is observed on a subsample of the dataset, primarily based on forms filled out at the dealer relating to a financing agreement and supplemented by R.L. Polk with data from marketing companies. I match this dataset with vehicle safety ratings from the National Highway Transportation Safety Administration (NHTSA) Safercar.gov website. The Safercar.gov safety ratings are based on a 5 star rating scheme, and can be considered comparable to the ratings from Insurance Institute for Highway Safety and Consumer Reports.

Since 1984, every vehicle in most counties in California is required to get a smog check to ensure the effectiveness of the vehicle emissions control system. The smog check program has been subsequently updated several times, and currently requires vehicles to get a smog check within six years of the initial registration of the vehicle.[3] If the ownership of a vehicle is transferred, any vehicle more than four years old is required to have a smog check unless the transfer is between a family member or a smog check has already been submitted within 90 days of the transfer date. Thus, some smog checks are observed as early as four years after the initial registration, and as late as seven years or more years if the owner is violating the law.

The California smog check program currently covers 40 of the 58 counties in California and all but a few percent of the population (see the Appendix for a list of counties covered by the program). At the time of the smog check, the VIN, characteristics of the vehicle (e.g., make, model, model year, transmission type), odometer reading, pollutant readings, and test outcome (pass or fail) are all sent to the California Bureau of Automotive Repair (BAR) and Department of Motor Vehicles to ensure compliance. I observe the smog check results for all vehicles in 2005 to 2009, and match these by VIN to the vehicle registrations in the R.L. Polk dataset. There are some VIN miscodings, so to ensure a perfect match, I only retain tests for VIN matches where the make and model also match. Approximately 0.1 million tests each year are not included due to miscoding, exemption from the smog check, were scrapped, violated the law, or were in a location that

[3]Exempted vehicles are: hybrids, electric vehicles, motorcycles, trailers, gasoline powered vehicles 1975 or older, diesel vehicles 1998 or older or greater than 14,000 lbs, or natural gas vehicles greater than 14,000lbs. Interestingly, many of these vehicles, particularly hybrids, show up in my dataset anyway, perhaps because the vehicles had ownership transferred to a dealer who performed the smog check, or because the owners were not aware of the exception.

does not require a smog check at the time of the smog check. Furthermore, I am interested in only those vehicles for which I can match a gasoline price that the consumers observe over the entire period for which the odometer reading is taken. Thus, I only retain VIN matches where the county in which the vehicle was registered when it was new matches either the county of registration at the time of the smog test or the county of the smog test station (if the registration county is unavailable). An additional 0.1 to 0.2 million vehicles are dropped from each year for this reason. This leaves me with a dataset with the first smog test matched by VIN for over 85% of the new registrations each year.

The primary question of this study is how consumers respond to changing gasoline prices, so the gasoline price data is of great importance. The Oil Price Information Service (OPIS) has retail gasoline prices throughout the United States based on credit card transactions at gas stations. For this analysis, I acquire county-level, monthly average retail gasoline prices in California for 2000 to 2010.[4] All counties except Alpine County are fully represented in the data. Figure 1 shows how retail gasoline prices (in real terms) in California were relatively flat until around 2005, after which time they steadily increased to a peak in mid-2008 before a substantial crash. Most of the variation in gasoline prices is time series variation, but there is

[4] I also use Cushing, OK (WTI) oil prices from the Energy Information Administration for a robustness check. Given the extremely high correlation between oil prices at different locations during this time period, WTI prices can be taken as a good measure of the global oil price.

also some cross-county variation.[5] The gasoline price data are matched with each vehicle in the dataset by finding the average gasoline price over the time period until the first odometer reading. For example, if a car is registered in Santa Clara County in October 2001 and tested or registered in Santa Clara County in June 2007, the average gasoline price that the vehicle faced over the time until odometer reading is the average of the prices in Santa Clara County between October 2001 and June 2007. The variable constructed this way is intended to be the best estimate possible of the average gasoline price that the consumer faced over the time period when he or she was making the driving decisions that led to the observed odometer reading. All dollars are adjusted by the Consumer Price Index to 2010 real dollars.

Next, I merge in zip code-level demographics from the American Factfinder (i.e., Census and American Community Survey). The variables included are population, population density, population growth rate from 2000 to 2007, number of businesses, median household income, percentage of population over 65, percentage of the population under 18, and percentage of the population of different ethnic groups. All of these demographics are for 2007, except businesses, which was only available for 2000. The median household income is used to account for the wealth of the area in which the vehicle is registered. I also include county-level unemployment over the time frame of the study from the Bureau of Labor Statistics to control for changing economic conditions. Similarly, I include the national monthly "consumer confidence index"

[5]The cross-county variation most likely stems from differences in transportation costs from refineries, and perhaps differences in retail markups depending on the density of gas stations in a county.

(CCI) from the Conference Board. Finally, I include county-level monthly median house prices from the California Realtor's Association. These last three variables control for the dramatic economic downturn beginning in 2008.

The full sample contains 5.86 million vehicles over the years 2001 to 2003. I code a very small number of odometer readings (1,730 vehicles) as zero for vehicles that have an average miles driven per month greater than 5,000, for these are obviously miscoded. Similarly, I code zero odometer readings for vehicles that do not run on gasoline (918 vehicles) or are commercial trucks (1,228 vehicles). This yields a dataset of 5.04 million vehicles, 36% of which are registered in 2001, 34% in 2002, and 29% in 2003. The smog tests for this sample occurred between January 2006 and December 2009.

2.2. Summary statistics

The dependent variable in this analysis is the log of the amount driven per month over the time frame between the date of registration when the vehicle was new, and the date of the smog test when the odometer reading was taken. Figure 2 shows the distribution of VMT per month in my sample. It is remarkably sharp peaked, with a mean at 1,130 and median at 1,081 miles per month. The 99% percentile is 2,585 miles per month and there are very few vehicles that drive more than 4,000 miles per month or are almost never driven. Figure 3 shows that there is considerable identifying variation in the average gas prices faced by the vehicles in the sample. The mean is \$2.67/gal and the median is \$2.69/gal. There are two prominent peaks, which are due to the seasonality of vehicle sales.

One important aspect of the odometer readings is that the time between registration and the first smog test varies considerably (Figure 4). Most vehicles have a smog check within a few months of the required smog check after six years. Some have smog checks earlier or later due to a title transfer, for all vehicles over four years of age are required to get a smog check with a title transfer. Some of the earliest smog checks appear to be fleet vehicles. Perhaps most interestingly, approximately 10% have a smog check within a few months of seven years, rather than six years. These 10% appear to have been randomly assigned by VIN to the later test, perhaps to lessen the burden of the smog check policy on Californians.[6] To confirm that these 10% as "as-good-as randomly assigned," I compare the vehicle characteristics and demographics of tests within two months of six years with tests within two months of seven years. I find that these seven year tests are not statistically different than the six year sample in the means of any of the major observables, such a vehicle characteristics or demographics. Several of the estimation results will restrict the sample to only "normal" tests within two months of six or seven years, as will be discussed below.

Looking back at Figure 1, there is a very noticeable seasonality in gasoline prices, with higher prices every summer. The mean price in the summer over all counties and the entire time frame is \$2.84/gal, while the mean price during the rest of the year is \$2.51/gal, a difference in means that is statistically significant with a t-statistic of -12.65. This seasonality could influence my estimates if those who face more time in the summer over the time frame between the new vehicle registration and the odometer reading

[6] While I have not been able to back out the algorithm used for the randomization, I have been told that this is the case by others working with these data.

also face a higher average price (from the raised summer prices), but respond differently because it is summer. To control for this possibility, I first create a variable for the percentage of months during the odometer reading time period that are summer months. The mean of this variable is 25.0%, with a standard deviation of 1.0 percentage points. I then also create variables for the counts of each month in the time between purchase and the first test (e.g., the number of Januarys during the driving period).

Table 1 provides summary statistics for all of the non-demographic variables used in this study, including VMT per month and the average gasoline price. Most new vehicles are personal vehicles (87%). Table 2 provides the same summary statistics, but only on the restricted sample of vehicles that received a test within two months of six or seven years. The summary statistics are similar, although with slightly more personal vehicles and slightly fewer leased vehicles. While not shown in the tables, the data also classify vehicles into 22 vehicle classes.

Table 3 gives summary statistics for the zip code-level demographic variables, taken over observations (vehicles) in the dataset. The top half of the table provides demographics for the full 5.03 million sample, while the lower half restricts the sample to vehicles that received a test within two months of six or seven years. One of the most interesting demographic variables for this study is the household income variable reported by R.L. Polk (Table 4). The subsample of my final dataset that includes income is 2.1 million with 19% in 2001, 26% in 2002, and 55% in 2003. Of the observations with VMT in 2001, 21% have income reported. For 2002, it is 32%, and for 2003 it is 77%. Fortunately, the distribution of income is nearly the same across the three years, so I believe that missing-at-random is a reasonable assumption, with the caveat that the subsample is more heavily weighted towards 2003. The distribution of income in Table 4 appears as would be expected for those who purchased new vehicles and financed them.[7]

3. Estimation Strategy

3.1. Model

The demand for driving (vehicle-miles-traveled) by a vehicle owner i at time $t \in \{1, ..., T\}$, VMT_{it}, can be thought of as a function of the retail price of gasoline the owner faces P_{it}, the characteristics of the vehicle being driven X_i, individual-specific demographics D_i, economic conditions E_{it}, county-level fixed effects θ_c, and purchase month-of-the-year fixed effects η_m:

$$VMT_{it} = f(P_{it}, X_i, D_i, E_{it}, \theta_c, \eta_m).$$

The county-level fixed effects capture time-invariant differences in the necessity of driving across counties. Month-of-the-year fixed effects based on when the vehicle is purchased may be important due to the possibility that different types of consumers purchasing vehicles at different times of year. Copeland et al. (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 more budget-conscious consumers may be purchasing vehicles at a different time than consumers who value the newest model.

[7]Comparing these values to the California distribution of income from 2000 Census shows very similar figures, but with more wealthy households purchasing new vehicles, as would be expected.

Suppose that this demand relationship takes the following form at each time period t:

$$VMT_{it} = (P_{it})^{\gamma} \exp(\beta_0 + \beta_X X_i + \beta_D D_i + \beta_E E_{it} + \theta_c + \eta_m) \exp(\varepsilon_i), \quad (1)$$

where ε_i is a mean-zero stochastic error term. This would then imply the following log-log specification:

$$\log(VMT_{it}) = \beta_0 + \gamma \log(P_{it}) + \beta_X X_i + \beta_D D_i + \beta_E E_{it} + \theta_c + \eta_m + \varepsilon_i, \quad (2)$$

where γ can be interpreted as the elasticity of VMT with respect to the price of gasoline–exactly the desired coefficient.[8] Given the time frame of the gasoline price shock within my dataset, a reasonable interpretation of this elasticity is a medium run elasticity – perhaps a 2 year elasticity.[9] This is long enough for people to make some adjustments to their working hours, driving routes, schedules, and planned trips, but not long enough for many people to make larger decisions (e.g., moving or changing jobs). I can examine heterogeneity in γ by restricting the dataset to subsamples

[8]To directly apply this specification using my dataset, with averaged VMT and averaged gasoline price over the driving period, t must be the driving period. If the period is shorter, then taking the average will take the sum of log driving. Due to Jensen's inquality, this may not be the same as the log of the sum of driving. In my robustness check section, I also examine a linear specification and a specification using the average of the log of driving.

[9]Recall that the gasoline price was relatively flat for all but the last 2 or so years of the odometer reading period for most vehicles.

for discrete variables and by interacting the gasoline price with continuous variables.

3.2. Identification Strategy

In this analysis, there is abundant variation in average gasoline prices and VMT per month. To understand the identification strategy, it is useful to focus on the variation in average gasoline prices. Since the time between the new vehicle registration and first smog test ranges from under four years to over eight years, many vehicles purchased in the same month have a different time period between registration and first test. Accordingly, vehicles purchased in the same month can have a different average price over the period between the registration and first test. This may be problematic, for vehicles that receive very early tests may be leased vehicles, fleet vehicles, or otherwise used for very different purposes than vehicles that had a standard six year smog test. On the other hand, this also presents an opportunity: the exact date that consumers receive the mailing notifying them of the requirement to get a smog test varies, and the exact date when the vehicle owner gets the test is plausible exogenous. Whether the vehicle is tested two months before six years, exactly six years, or one month after six years (there is a two month grace period) is likely to depend on when exactly the notification arrives and how busy the vehicle owner happens to be at that particular time. The same logic holds for the 10% of the sample who were randomly assigned seven years until the smog test: when exactly the vehicle is tested within two months of seven years is plausibly exogenous.

The identification strategy of this analysis uses both sources of plausibly exogenous variation. The primary identifying assumption is that the variation within a few months of the standard six/seven year time to test is exogenous to the driving decision and that the seven year tests are randomly assigned, as discussed above. To implement the model with this assumption, I restrict the sample to vehicles that received a smog test within two months of either six or seven years of the date of registration. However, if I add the identifying assumption that the date of a title change is also exogenous, then I can use the full sample. Specifically, by including indicator variables for early tests, tests within two months of six years, tests between six and seven years, tests within two months of seven years, and very late tests, I can consistently estimate the coefficients under this additional identifying assumption.

I also need several other identifying assumptions for this research design to yield consistent estimates of the desired elasticity. A common assumption in the literature using disaggregated data from a particular region is that individual drivers are price-takers with respect to the price of gasoline (e.g., Knittel and Sandler (2013)). The intuition for this assumption is that gasoline prices are determined based on supply and demand forces around the world, rendering any demand shocks in California imperceptible. However, the assumption rules out localized VMT demand shocks that are correlated across drivers, which could lead to localized gasoline price increases.

Are localized VMT demand shocks likely? The very limited variation in gasoline prices across counties suggests not. Nevertheless, one could relax the assumption of no localized VMT demand shocks by instrumenting for the county-level average gasoline price. Strong and valid instruments are never easy to find, but one plausible instrument in this context is the global oil price.[10] The crude oil price is unquestionably the primary determinant of the gasoline price (the Pearson correlation coefficient is 0.97), and thus can be expected to be a strong instrument. It seems plausible that the only way oil prices influence driving in California is indirectly through the local gasoline price, in which case the exclusion restriction would hold. On the other hand, it is also possible that oil prices influence driving by influencing fuel prices for other transportation options, such as aviation. Given this, I follow the literature and assume away localized VMT demand shocks in my preferred specification, while presenting the instrumental variables regression results as a robustness check.

Identification of γ also depends on consumers not forecasting future gasoline price increases and modifying their purchase decision accordingly. I observe consumers who purchased vehicles in 2001-2003 and then faced a gasoline price increase several years later, beginning in 2005 (recall Figure 1). Oil futures prices remained relatively close to oil spot prices during 2001-2003, and it is highly unlikely that consumers could forecast the impending increase in gasoline prices. This is a very useful element of my empirical setting, for it suggests that consumers are not purchasing more efficient vehicles in response to the expected gasoline price increases. Consumers with perfect foresight would be expected to purchase more efficient vehicles, thus potentially leading them to drive more over the observed period of driving, the effect popularly known as the "rebound effect." This would pose a clear

[10]I have an anonymous reviewer to thank for this suggestion. Other studies, such as Hughes et al. (2008), have used global oil supply disruptions, which have a similar intuition as the global oil price.

endogeneity concern, since expectations of future gasoline prices are unobserved.[11]

Finally, I include several critical controls: the percentage of summer months, the counts of months in the time period, county-level monthly unemployment and housing prices, 2000 to 2007 rate of population growth, and the CCI. As discussed above, the percentage of summer months and counts of months addresse the seasonality of driving demand and gasoline prices. The monthly unemployment and housing price variables control for local macroeconomic forces that could influence driving demand. The 2000 to 2007 rate of population growth in the zip code the vehicle is registered in should also help provide cross-sectional data to capture local macroeconomic forces. The CCI also helps control for the nationwide decline in consumer confidence in 2008 and 2009.

4. Results

4.1. Primary Evidence of Responsiveness

The primary results from estimating equation (2) are given in Table 5. Column (i) begins with the 2.99 million observation restricted sample that contains only vehicles with a smog test within two months of either six years or seven years. All of the fixed effects and controls discussed above are

[11] Addressing this endogeneity would require simultaneously modeling vehicle choice and driving decisions based on forecasted gasoline prices. Gillingham (2012) performs such an estimation, but is focusing on the endogeneity of fuel economy, as in Dubin and McFadden (1984).

included.[12] Consistent estimation of the gasoline price elasticity relies on the identifying assumption that exactly when the vehicle owner performs the smog check within the legal window is random. Column (ii) provides results from the same estimation performed on the entire 5.04 million observation sample. This estimation includes the indicator variables for the number of months between registration and the date of the first test that are described above. Consistent estimation in this specification relies on exogeneity in the date of any changes of title.

The results in columns (i) and (ii) both show a highly statistically significant medium run elasticity of VMT with respect to the price of gasoline of -0.22. I find that other similar specifications, such as including zip code fixed effects, result in a very similar estimate. Column (iii) presents the results from a instrumental variables (IV) estimation where the log oil price is used as an instrument for the log gasoline price. The estimation is performed on the full dataset and contains all of the control variables used in column (ii). Interestingly, the estimated elasticity is -0.23, which is only 0.05 different than the estimated coefficient in column (2). This is a comforting result, for if we believe the validity of the instruments, it provides evidence against localized supply shocks as a confounding factor. Given this finding, my preferred specification is given by column (ii) and I use this for my exploration into the heterogeneity in responsiveness.

[12]The vehicle characteristics in these specifications include engine size, engine displacement, turbocharger/supercharger, automatic transmission, gross vehicle weight rating, all-wheel-drive, safety rating, and an import indicator. Incidentally, if I also include a variable for fuel economy, the coefficient on the log gasoline price barely changes at all.

To put these primary results into context, Figure 5 shows the gasoline demand in the United States over the years of the gasoline price shock. A linear trendline is fit through the data up until 2005, when gasoline prices started rising, in order to provide a rough baseline to give a sense of the magnitude of the decline in 2007 and 2008. Interestingly, there is little noticeable decline in gasoline demand until mid-2007, a feature of the data that may partly explain the low elasticity estimates in Small and Van Dender (2007) and Hughes et al. (2008), which were both base on datasets that do not include 2007 and 2008. The decrease in gasoline demand after mid-2007 is very noticeable, and continues through 2010 as the economy sputtered into a recession. The aggregate gasoline demand change relative to the trendline appears to be just above 10%. Since this change accounts for the utilization choices of the entire vehicle stock across the US, as well as a few other end uses of gasoline, it is reasonably consistent with my estimation results indicating that the medium run elasticity for new vehicles in their first six years of life is in the range of -0.22.

4.2. Heterogeneity in responsiveness

The richness of my dataset allows for further exploration into how new vehicle owners differ in the responsiveness to gasoline price changes. I examine heterogeneity in several ways: (1) using a quantile regression approach to examine heterogeneity in the responsiveness over all consumers, (2) estimating the model on the subsample of each buyer type, (3) estimating the model on the subsample of each income group, (4) interacting the demographic variables with the gasoline price, (5) estimating the model on the subsample of each county, and (6) performing a k-means clustering analysis.

4.2.1. Quantile regressions

Quantile regression is useful for examining the heterogeneity across the quantiles of response. While linear regression estimates the conditional mean of the dependent variable given values of covariates, quantile regression estimates the conditional median or other quantiles of the dependent variable given values of covariates (Koenker and Bassett, 1978). In this context, I am interested in the quantiles of the VMT elasticity with respect to the gasoline price. I estimate (2) using the full 5.04 million observation dataset and the specification presented in (ii) in Table 5. The results given in Table 6.

Column (i) presents the 0.25 quantile regression result. The coefficient on the log average gasoline price indicates that at the 0.25 quantile of response, the elasticity is -0.33, considerably more responsive than the median (0.5 quantile) regression result of -0.24 in Column (ii). Not surprisingly, the median regression result is very close to the mean regression result given above in Table 5. The 0.75 quantile regression result in Column (iii) is somewhat less responsive, with an estimated elasticity of -0.17. All of these estimated elasticities at highly statistically significant. This first look at heterogeneity in responsiveness belies several further differences that are explored below.

4.2.2. Buyer types

I next examine the heterogeneity in driving responsiveness by estimating equation (2) on the subsample for each buyer type: personal, firm, rental, and government. Recall from Table 1 that in the full dataset, 87% of the vehicles are personal vehicles, so personal vehicles dominate the results described above. Table 7 illustrates the heterogeneity in responsiveness across vehicle buyer types. Columns (i) through (iv) are estimated using the full dataset and all covariates included in (ii) in Table 5.

The coefficient on the log of the average gasoline price in Column (i) indicates that for personal vehicles the elasticity of driving with respect to the price of gasoline is -0.21, very much in line with the results in Table 5. This is not surprising given that most vehicles are personal vehicles. Column (ii) indicates that company vehicles as responsive as personal vehicles, although the coefficient is not statistically significant. Column (iii) shows that the responsiveness for rental vehicles is also very similar to the responsiveness for all personal vehicles, with an elasticity of -0.20, which is statistically significant at the 10% confidence level. Gasoline in rental vehicles is paid for by personal drivers so it follows that the responsiveness may be similar. However, some rental car drivers on business trips have their gasoline paid for, so one would expect less responsiveness for this population. On the other hand, many rental car drivers are on vacation, so they may be more sensitive to gasoline price changes. Perhaps these opposing effects nearly cancel out. Columns (iv) does not have a statistically significant coefficient on the log of the gasoline price, so not much can be said about government workers. Given that the sample is much smaller, it is not surprising that significance is lost with the extensive controls used in the estimation.

4.2.3. Household Income

To examine the heterogeneity in responsiveness by income, I rely on the 2.1 million subsample of observations where R.L. Polk reports the household income of the vehicle purchaser. This subsample consists entirely of new personal vehicles. Table 8 presents the results of estimation 2 using a subsample for each income bracket. Results for consumers with income less than \$20,000 per year are not reported for there are fewer observations in these categories and new vehicle buyers who have income less than \$20,000 are likely to be unusual.[13]

The coefficients on the log of the average gasoline price in Columns (i) through (vii) in Table 8 indicate that the responsiveness generally *increases* with income, although it levels off at the highest income brackets. The estimated elasticity for households earning \$20,000 to \$30,000 is -0.22, which is exactly what the estimated elasticity is for the entire sample, including non-personal vehicles. As we move up income brackets, the responsiveness generally increases, until it reaches -0.45 at the \$75,000 to \$100,000 range. The wealthiest category, with annual income greater than \$125,000, is estimated to have a slightly lower elasticity of -0.40. All of these estimated elasticities are highly statistically significant.

This positive relationship between income and responsiveness contrasts with the result in Small and Van Dender (2007) and does not quite match with the "U-shaped" relationship found in West (2004) and Wadud et al. (2009). However, it is consistent with the findings in Hughes et al. (2008). There are a few possible explanations for the results in Table 8. The increase in the elasticity at higher incomes may relate to wealthier households having more discretionary driving. Alternatively, wealthier households may be more likely to switch from driving to flying for trips. The least wealthy households who purchase new vehicles may have a very strong preference for driving, thus resulting in a lower responsiveness. A perhaps even more important factor for

[13]The estimated coefficients for these groups are just slightly closer to zero than those of the \$20,000 to \$30,000 group.

explaining the results is that wealthier households tend to own more vehicles. Since each observation is a vehicle and I do not observe households, it is possible that within-household switching of vehicles to other more efficient vehicles in the household may account for the greater responsiveness at higher income levels. Knittel and Sandler (2013) find some evidence of within-household switching using California registration data, although they do not explore this particular issue with wealthier households, leaving it a promising area for future research.

4.2.4. Demographics

Table 9 presents results with the zip code demographic variables interacted with the log of the gasoline price. These estimations are performed on the full 5.04 million observation sample. Column (i) repeats column (ii) in Table 5 for reference. Column (ii) adds the interactions. Despite the large sample size, only a few of these interaction coefficients are statistically significant. It is worth noting that to the extent to which there is variation within zip codes, we would expect measurement error in these coefficients, possibly leading to a standard attenuation bias. Nevertheless, given the unavailability of individual demographic data, I proceed further.

The first of these interaction coefficients that is highly statistically significant is the zip code density interaction. It indicates that higher density zip codes are slightly less responsive. However, given the scaling of the density variable, the interaction coefficient is not very economically significant. The population rate of increase and commute time interactions are also statistically significant, but neither are highly economically significant. Both indicate a slight increase in responsiveness. However, if I estimate the model without county fixed effects, the county-level commute time interaction coefficient becomes positive. This accords with economic intuition: counties where the average commute time is longer will tend to be less responsive, due to the necessity of commuting and the lack of public transport in the outskirts of cities in California.

4.2.5. Counties

To examine the heterogeneity by county, I estimate equation (2) for each county in California. Some counties either do not require a smog check or have a small number of observations, due to the low population in the county. Accordingly, I only estimate the model for counties with greater than 2,000 observations in the full 5.04 million observation dataset. The model estimated on each county separately is identical to the model given in column (ii) of Table 5.

The results for selected counties are given in Table 10 and the results for all counties are graphically illustrated in Figure 6. Many of the analyzed counties have a statistically significant coefficient on the log of the average gasoline price. All of the statistically significant coefficients are negative. The statistically significant coefficients range quite widely from -0.01 to -0.50, although many are closer to the primary results. Some large and populated counties appear to have a slightly lower responsiveness, such as Orange County with an estimated elasticity of -0.16. Some of the wealthier counties, such as Marin and Santa Barbara, appear to have a higher per-vehicle elasticity. This result again may stem from the number of vehicles the households own. It may also relate to the fact that the vehicle stock in both Marin and Santa Barbara is considerably less fuel efficient on average than the California mean (although at the same time, the Prius is a popular vehicle in Marin).

4.2.6. Cluster analysis

A k-means cluster analysis is a common approach used in statistics to partition observations into groups of similar observations. Observations can be thought of as vectors of variable values. The k-means clustering approach partitions N observations into K disjoint sets $S = \{S_1, S_2, ..., S_K\}$ in order to minimize the sum of squared differences between each observation and the nearest mean. Since each observation is usually a vector, the sum of squares is taken over a particular norm, most commonly the L^2 (Euclidean) norm. More specifically, a k-means cluster analysis involves solving the following minimization problem for the optimal partition:

$$S^* = \arg\min_S \sum_{i=1}^{K} \sum_{x_j \in S_i} \|x_j - \mu_i\|^2, \tag{3}$$

where x_j is a vector of variable values for each observation j (i.e., vehicle) and μ_i is the vector of means across all of the observations in set S_i. Just as in principal components analysis, the attributes of resulting clusters can then be examined and interpreted. While there are many ways to cluster, for the purposes of my analysis, I estimate clusters based on zip code density and zip code median household income to give a sense of the heterogeneity across income and degree of urbanization (so $x_j, \mu_i \in \mathbb{R}^2$). I choose 6 clusters for this analysis and use the L^2 norm.

Table 11 presents summary statistics for the density and zip code income in each cluster, along with a description given to each cluster. Just as for the county results, I estimate the same specification as in column (ii) of Table 5 on the subsample of observations in each cluster. Table 12 shows the results of these estimations, where each column refers to the cluster number in Table 11. The elasticity of VMT with respect to gasoline price varies across clusters, ranging from -0.12 for the semi-rural upper class cluster to -0.30 for the rural wealthy cluster. The greater responsiveness displayed by the vehicles in the rural wealthy cluster may again reflect the number of vehicles each household owns and within-household switching as gasoline prices rose. Interestingly, the urban low income cluster elasticity is -0.26, which is more responsive than several of the other clusters. This slightly higher elasticity may relate to better access to public transportation.

4.3. Robustness

I perform several robustness check to examine the sensitivity of the results to different assumptions. None of these robustness checks change the general results substantially, even if they may change the exact quantitative estimate by a small amount. First, I truncate the sample at different times. If I truncate the sample so that none of the sample faces the large spike of gasoline prices in 2008, I find a slightly lower elasticity in the range of -0.15 to -0.2. I see this result as indicative of a changing elasticity over time depending on the salience of the gasoline price changes. This dataset with a limited number of years is not well-suited for a full exploration of changing elasticities over time, but this is a promising area for future research.

Second, I add the vehicle purchaser's household income to the primary specification in (2) in order to explicitly control for household income, rather than the average household income in the zip code of the purchaser. To perform this test, I am limited to the much smaller income subsample. I find that adding the income controls (and restricting the sample) makes only a small change to the estimated VMT elasticity value (increases it slightly).

Finally, since the model of household driving decisions is based on the log of driving in time period t, it is prudent to check whether the choice of a log-log specification is problematic, for t could be a much shorter period than the time between registration and the first test. I examine this in two ways. I first create a variable for the average of the log gasoline price over the years of each vehicles odometer reading. Using this variable instead of the log of average driving yields nearly identical results. Next, I run a specification that is linear in average VMT and gasoline price, rather than the preferred log-log specification. With this specification, Jensen's inequality would not be an issue. The estimated coefficient on the gasoline price is -94.84. At the mean VMT of 1,100 per month and average gasoline price of $2.69, this estimated coefficient implies an elasticity of VMT with respect to the gasoline price of -0.23. These final robustness checks are encouraging and imply that a misspecification bias due to the log-log specification is not a concern.

5. Conclusions

This study uses an extremely rich vehicle-level dataset where both vehicle characteristics and actual driving are observed to provide new evidence on the responsiveness of new vehicle purchasers to the gasoline price shock of 2006-2008. I find evidence for a medium run elasticity of driving with respect to the gasoline price of -0.22 for drivers in California during the first six years of their vehicle's lifespan. While this subset of vehicles is not the full vehicle stock, it represents a fairly large portion of the vehicle stock. Even more importantly, it represents the portion of the vehicle stock that is driven the most, for vehicles in the first several years of their lifespan are known to be driven much more than older vehicles. Thus, this point estimate has important implications for policy.

The most striking implication is that policies that quickly increase the price of gasoline can influence consumer behavior with corresponding reductions in the demand for oil and greenhouse gas emissions. Economists have long been convinced that this is the case, but recent work by Small and Van Dender (2007) and Hughes et al. (2008) called this intuition into question. This study indicates that when gasoline prices increase as much as they did in 2006-2008, then the vehicle utilization choice – at least for newer vehicles – is inelastic, but not quite as inelastic as in recent studies. To the extent that consumers respond equally to a change in fuel economy as a change in gasoline prices, this implies that the rebound effect from energy efficiency policy may be slightly larger than previous estimates.[14]

A variety of explanations can be posited to reconcile the results of this study with the previous results. One is simply that newer vehicles may be more elastic than older vehicles. I believe this is unlikely, for older vehicles tend to be less fuel-efficient, so consumers are more likely to switch some driving to their newer vehicle, implying that the newer vehicle would appear *less* elastic. Another explanation is that any gasoline price variations up

[14]Standard economic theory would suggest an equal response, yet there is some suggestive evidence that consumers may respond less to fuel economy changes than changes in gasoline prices, perhaps because gasoline prices are more salient (Gillingham, 2011).

until 2006 had been so limited that consumers largely ignored them, but the gasoline price shock of 2006-2008 was large enough that it could not be ignored. Given the media reports of the past few years, it is evident that the gasoline price shock was quite salient to consumers. Thus, it may not be surprising that the elasticity appears to be much larger over this period than in the previous period when gasoline prices remained low.

What does appear to be somewhat surprising is the degree of heterogeneity in the elasticity. The quantile regression results provide a first glimpse into this heterogeneity, showing that the range over the 0.25 to 0.75 quantiles is from -0.33 to -0.17. There is a striking pattern of increasing vehicle-level responsiveness with income, with a dramatic difference in responsiveness between the high income categories and the lower income categories. This pattern may be in part due to within-household switching to even newer and higher fuel economy vehicles. There is some limited evidence of heterogeneity by demographics, and there is even more evidence of heterogeneity in responsiveness across counties in California. The k-means cluster analysis underscores the heterogeneity across income and density. Quantifying this heterogeneity is crucial to being able to quantify the distributional consequences of polices that increase the price of gasoline–as well as the potential co-benefits of those policies from reduced local air pollution.

Future work is warranted to extend the analysis to all vehicles. Similarly, a valuable extension would be to explore the implications of the heterogeneity found in this paper for the efficiency and equity impacts of increased gasoline taxes or a carbon cap-and-trade that includes the transportation sector, such as is planned under California's Global Warming Solutions Act (AB 32).

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.

Acknowledgments

This paper is based in part on my Ph.D. dissertation at Stanford University. I 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 also thank Jim Sweeney, Tim Bresnahan, Jon

Levin, Larry Goulder, Liran Einav, Matt Harding, and two anonymous reviewers 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. Finally, I would like to thank the Lincoln Institute for Land Policy for support for finishing this paper and the participants of the October 2012 Lincoln conference in honor of John Quigley. Any errors are solely the responsibility of the author.

References

Austin, D., 2008. Effects of gasoline prices on driving behavior and vehicle markets. Congressional Budget Office Report 2883, Washington, DC.

Bento, A., Goulder, L., Jacobsen, M., von Haefen, R., 2009. Distributional and efficiency impacts of increased us gasoline taxes. American Economic Review 99(3), 667–699.

Copeland, A., Dunn, W., Hall, G., 2011. Inventories and the automobile market. RAND Journal of Economics 42(1), 121–149.

Dubin, J., McFadden, D., 1984. An econometric analysis of residential electric appliance holdings and consumption. Econometrica 52(2), 345–362.

Gillingham, K., 2011. The consumer response to gasoline price changes: Empirical evidence and policy implications. Stanford University Ph.D. Dissertation .

Gillingham, K., 2012. Selection on anticipated driving and the consumer response to changing gasoline prices. Yale University Working Paper New Haven, CT.

Goldberg, P., 1998. The effects of the corporate average fuel efficiency standards in the us. The Journal of Industrial Economics 46(1), 1–33.

Graham, D., Glaister, S., 2002. The demand for automobile fuel: A survey of elasticities. Journal of Transport Economics and Policy 36(1), 1–26.

Hughes, J., Knittel, C., Sperling, D., 2008. Evidence in a shift in the short-run price elasticity of gasoline demand. Energy Journal 29(1), 113–134.

Hymel, K., Small, K., Van Dender, K., 2010. Induced demand and rebound effects in road transport. Transportation Research B 44(10), 1220–1241.

Knittel, C., Sandler, R., 2013. The welfare impact of indirect pigouvian taxation: Evidence from transportation. MIT Working Paper .

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46(1), 33–50.

National Research Council, 2002. Effectiveness and Impact of Corporate Average Fuel Economy (CAFE) Standards. National Academy Press, Washington, DC.

Small, K., Van Dender, K., 2007. Fuel efficiency and motor vehicle travel: The declining rebound effect. Energy Journal 28(1), 25–51.

Wadud, Z., Graham, D., Noland, R., 2009. Modelling fuel demand for different socio-economic groups. Applied Energy 86(12), 2740–2749.

Wadud, Z., Graham, D., Noland, R., 2010. Gasoline demand with heterogeneity in household responses. Energy Journal 31(1), 47–74.

West, S., 2004. Distributional effects of alternative vehicle pollution control technologies. Journal of Public Economics 88, 735.

West, S., Williams, R., 2004. Estimates from a consumer demand system: Implications for the incidence of environmental taxes. Journal of Environmental Economics and Management 47(3), 535–558.

Table 1: Summary Statistics for the Full Sample

Variable	Mean	Std Dev	Min	Max
VMT per month	1,100.78	472.27	0	4,996
Avg gasoline price (2010$/gal)	2.673	0.186	2.187	3.375
Personal	0.870	0.336	0	1
Firm	0.035	0.183	0	1
Rental	0.085	0.279	0	1
Government	0.01	0.1	0	1
Lease	0.146	0.353	0	1
Engine size (liters)	33.421	12.436	4	83
Engine cylinders	5.857	1.534	2	12
Turbocharged	0.026	0.158	0	1
Automatic transmission	0.91	0.286	0	1
Gross vehicle weight rating	5.418	1.227	0.44	14.1
All-wheel drive	0.161	0.367	0	1
Safety rating	4.119	0.472	1	5
Import	0.562	0.496	0	1
County unemployment rate	5.881	1.408	3.783	19.293
Consumer conf. index	93.331	4.191	84.285	101.876
County housing prices	530.016	164.874	172.88	1,044.27
% summer months	0.25	0.011	0.182	0.308

Notes: All variables contain 5,038,554 non-missing observations. County-level monthly median house prices are in units of hundreds of thousands of 2010 dollars.

Table 2: Summary Statistics for Six or Seven Year Tests

Variable	**Mean**	**Std Dev**	**Min**	**Max**
VMT per month	1065.79	457.59	0	4,996
Avg gasoline price (2010$/gal)	2.692	0.179	2.307	3.372
Personal	0.896	0.305	0	1
Firm	0.028	0.166	0	1
Rental	0.071	0.256	0	1
Government	0.005	0.067	0	1
Lease	0.122	0.328	0	1
Engine size (liters)	33.3	12.285	4	83
Engine cylinders	5.84	1.524	3	12
Turbocharged	0.023	0.15	0	1
Automatic transmission	0.917	0.277	0	1
Gross vehicle weight rating	5.394	1.199	0.44	14.1
All-wheel drive	0.157	0.364	0	1
Safety rating	4.125	0.472	1	5
Import	0.583	0.493	0	1
County unemployment rate	5.894	1.38	4.114	19.293
Consumer conf. index	92.831	4.273	84.333	98.540
County housing prices	532.528	166.134	185.94	1,009.12
% summer months	0.25	0.006	0.229	0.26

Notes: All variables contain 2,991,120 non-missing observations. County-level monthly median house prices are in units of hundreds of thousands of 2010 dollars.

Table 3: Demographic Summary Statistics

Variable	Mean	Std Dev	Min	Max
Full Sample				
zip density	5.173	5.252	0	52.182
zip businesses per capita 2000	0.058	1.031	0	108.519
zip population 2007	41,676.171	19,512.001	1	109,549
zip pop growth rate 00-07	1.554	2.682	-32.5	199.2
zip median hh income 2007	70,694.172	26514.254	0	375,000
county commute time (min)	27.01	4.038	13.4	43.1
zip % pop age 65+	11.172	5.023	0	100
zip % pop under 18	25.42	5.97	0	41.3
zip % pop white 2007	58.896	18.011	4.4	100
zip % pop black 2007	5.698	7.566	0	86.600
zip % pop hispanic 2007	31.437	20.713	0	97.8
Six & Seven Year Sample				
zip density	5.216	5.253	0	52.182
zip businesses per capita 2000	0.053	0.886	0	108.519
zip population 2007	42,008.156	19,576.236	1	109,549
zip pop growth rate 00-07	1.513	2.54	-8.4	199.2
zip median hh income 2007	71,099.214	26,478.019	0	375,000
county commute time (min)	27.013	4.019	13.4	43.1
zip % pop age 65+	11.166	4.968	0	100
zip % pop under 18	25.498	5.881	0	41.3
zip % pop white 2007	58.588	18.161	4.4	100
zip % pop black 2007	5.575	7.546	0	86.600
zip % pop hispanic 2007	31.475	20.837	0	97.8

Notes: The zip code population density is in units of thousand people/mi^2. In the full sample, all demographic variables contain 5,038,576 non-missing observations and in the six and seven year sample, all demographic variables contain 2,991,120 non-missing observations.

Table 4: Tabulations of income 2001-2003

Income Category	Observations	Percent
<$15,000	144,052	7
$15,000 - $19,999	58,921	3
$20,000 - $29,999	157,172	8
$30,000 - $39,999	179,990	9
$40,000 - $49,999	203,592	10
$50,000 - $74,999	512,110	25
$75,000 - $99,999	345,910	17
$100,000 - $124,999	178,044	9
>$125,000	293,910	14
Total	2,073,701	100%

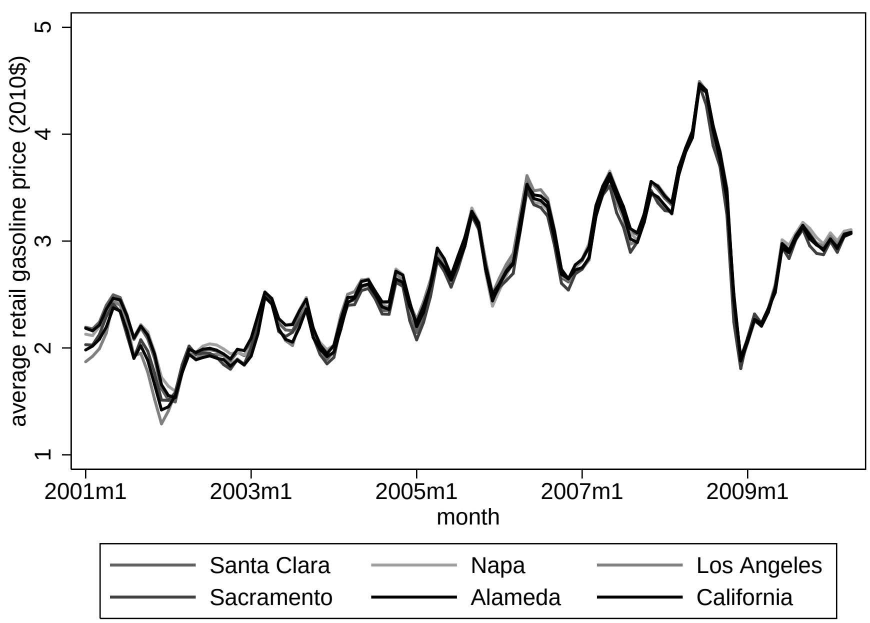

Figure 1: Retail gasoline prices in California were relatively flat and then rose substantially until 2008, providing substantial time series variation. Five representative counties are shown here.

Table 5: Primary log(VMT) results

	OLS **(i)**	OLS **(ii)**	IV **(iii)**
log(avg gasoline price)	-0.22***	-0.22***	-0.23***
	(0.06)	(0.03)	(0.03)
lease	0.08***	0.05***	0.05***
	(0.00)	(0.00)	(0.00)
firm	0.15***	0.14***	0.14***
	(0.01)	(0.01)	(0.01)
rental	0.19***	0.15***	0.15***
	(0.01)	(0.01)	(0.01)
government	0.01	-0.06	-0.06
	(0.05)	(0.04)	(0.04)
constant	5.95***	6.64***	6.17***
	(0.84)	(0.07)	(0.08)
Vehicle segment indicators	X	X	X
Vehicle characteristics	X	X	X
Purchase month indicators	X	X	X
Zip code % age brackets	X	X	X
Zip code % race brackets	X	X	X
County indicators	X	X	X
Model fixed effects	X	X	X
Economic conditions	X	X	X
% summer months	X	X	X
Counts of months covered	X	X	X
Time to test indicators		X	X
R-squared	0.031	0.033	0.088
N	2.99m	5.04m	5.04m

Notes: Regressions of log(VMT) on covariates using the six and seven year subsample in column (1) and the full sample in columns (2) and (3). Robust standard errors are in parentheses, clustered on vehicle model. Column (3) is performed using two-stage least squares with ln(avg gasoline price) instrumented for with ln(world oil price). The F-test rejects the null of weak instruments with a p-value of 0.000. Economic conditions include county-level unemployment, national-level consumer confidence index, and county-level median house prices. Purchase month indicators are indicators for the month of the year that the vehicle was purchased. The counts of months covered refers to a variable for each month of the year (Jan-Dec) counting the number of times a that month is included in the months to test. The indicators for the number of months to test are: <58 months, >57 and <63 months, >62 and <70 months, >69 and <74 months, >73 and <82 months, >81 and <86 months, and >85 months. *** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level.

Table 6: log(VMT) quantile regression results

	0.25 **(i)**	0.50 **(ii)**	0.75 **(iii)**
log(avg gas price)	-0.33***	-0.24***	-0.17***
	(0.01)	(0.01)	(0.01)
lease	0.07***	0.02***	-0.00***
	(0.00)	(0.00)	(0.00)
firm	0.13***	0.15***	0.16***
	(0.00)	(0.00)	(0.00)
rental	0.18***	0.13***	0.09***
	(0.00)	(0.00)	(0.00)
govt	-0.08***	0.03***	0.07***
	(0.00)	(0.00)	(0.00)
constant	6.84***	7.30***	7.73***
	(0.06)	(0.04)	(0.04)
Vehicle segment indicators	X	X	X
Vehicle characteristics	X	X	X
Purchase month indicators	X	X	X
Zip code % age brackets	X	X	X
Zip code % race brackets	X	X	X
County indicators	X	X	X
Model fixed effects	X	X	X
Economic conditions	X	X	X
% summer months	X	X	X
Counts of months covered	X	X	X
Time to test indicators	X	X	X
N	5.04m	5.04m	5.04m

Notes: Quantile regressions of log(VMT) on covariates using the full sample. (1) is the 0.25 quantile, (2) the 0.5 quantile (median), and (3) is the 0.75 quantile. A fitted nonparametric density estimator is used to compute the variance-covariance matrix and the standard errors. All other variables are the same as in (ii) in Table 5. *** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level.

Table 7: log(VMT) regressions by buyer type

	Personal **(i)**	Firm **(ii)**	Rental **(iii)**	Govt **(iv)**
log(avg gas price)	-0.21***	-0.21	-0.20*	-0.50
	(0.03)	(0.11)	(0.09)	(0.37)
lease	0.04***	0.03***		
	(0.00)	(0.01)		
constant	7.39***	6.67***	6.86***	9.26***
	(0.09)	(0.37)	(0.29)	(1.10)
Vehicle segment indicators	X	X	X	X
Vehicle characteristics	X	X	X	X
Purchase month indicators	X	X	X	X
Zip code % age brackets	X	X	X	X
Zip code % race brackets	X	X	X	X
County indicators	X	X	X	X
Model fixed effects	X	X	X	X
Economic conditions	X	X	X	X
% summer months	X	X	X	X
Counts of months covered	X	X	X	X
Time to test indicators	X	X	X	X
R-squared	0.069	0.100	0.050	0.123
N	4,382,537	174,568	429,234	50,637

Notes: Regressions of log(VMT) on covariates using subsamples of the full sample based on buyer type: personal, firm, rental, and government. Robust standard errors are in parentheses, clustered on vehicle model. All other variables are the same as in (ii) in Table 5. *** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level.

Table 8: log(VMT) regressions by income group

	20-30k **(i)**	30-40k **(ii)**	40-50k **(iii)**	50-75k **(iv)**	75-100k **(v)**	100-125k **(vi)**	>125k **(vii)**
log(avg gas price)	-0.22***	-0.34***	-0.31***	-0.41***	-0.45***	-0.38***	-0.40***
	(0.07)	(0.08)	(0.05)	(0.04)	(0.05)	(0.07)	(0.05)
lease	0.03***	0.04***	0.03***	0.03***	0.02**	0.03***	0.04***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
constant	7.95***	7.59***	7.56***	7.48***	7.53***	7.83***	8.14***
	(0.17)	(0.18)	(0.15)	(0.11)	(0.13)	(0.18)	(0.17)
Vehicle segment indicators	X	X	X	X	X	X	X
Vehicle characteristics	X	X	X	X	X	X	X
Purchase month indicators	X	X	X	X	X	X	X
Zip code % age brackets	X	X	X	X	X	X	X
Zip code % race brackets	X	X	X	X	X	X	X
County indicators	X	X	X	X	X	X	X
Model fixed effects	X	X	X	X	X	X	X
Economic conditions	X	X	X	X	X	X	X
% summer months	X	X	X	X	X	X	X
Counts of months covered	X	X	X	X	X	X	X
Time to test indicators	X	X	X	X	X	X	X
R-squared	0.057	0.061	0.059	0.065	0.074	0.092	0.118
N	157,172	179,990	203,592	512,110	345,910	178,044	293,910

Notes: Regressions of log(VMT) on covariates using the subsamples based on income brackets from the sample of personal vehicles that have non-missing individual-level income. Robust standard errors are in parentheses, clustered on vehicle model. All other variables are the same as in (ii) in Table 5. *** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level.

Table 9: Demographic interactions

	(i)	(ii)
log(avg gas price)	-0.22***	-1.26***
	(0.03)	(0.37)
lease	0.05***	0.04
	(0.00)	(0.03)
log(zip population)	-0.02***	-0.04***
	(0.00)	(0.01)
zip pop growth rate	0.00***	0.00*
	(0.00)	(0.00)
zip density	-0.00***	-0.01***
	(0.00)	(0.00)
zip businesses/cap	-0.02***	-0.08
	(0.01)	(0.06)
commute time	0.01***	0.01***
	(0.00)	(0.00)
lgaspr*lease		0.01
		(0.03)
lgaspr*lpop		0.02*
		(0.01)
lgaspr*poprate		-0.00
		(0.00)
lgaspr*density		0.01***
		(0.00)
lgaspr*buspercap		0.06
		(0.06)
lgaspr*commute		-0.00**
		(0.00)
constant	6.64***	8.37***
	(0.07)	(0.38)
Vehicle segment indicators	X	X
Vehicle characteristics	X	X
Purchase month indicators	X	X
Zip code % age brackets	X	X
Zip code % race brackets	X	X
County indicators	X	X
Model fixed effects	X	X
Economic conditions	X	X
% summer months	X	X
Counts of months covered	X	X
Time to test indicators	X	X
R-squared	0.033	0.068
N	5.04m	5.04m

Notes: Regressions of log(VMT) on covariates. Column (1) replicates column (2) in Table 5 for reference. Column (2) contains interactions with a variety of demographic variables. Specification includes interactions for age brackets, race brackets, and zip code average income, but these were not statistically significant and are thus not reported. Robust standard errors are in parentheses, clustered on vehicle model. All other variables are the same as in (ii) in Table 5. *** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level.

Table 10: Responsiveness for selected counties

	Coefficient	Std Error	N
Butte	-0.55**	0.282	16,878
Orange	-0.16*	0.098	533,545
Placer	-0.50***	0.134	46,872
Sacramento	-0.23***	0.075	186,487
San Benito	-0.69**	0.400	7,255
San Diego	-0.18***	0.058	445,040
San Luis Obispo	-0.44**	0.220	32,633
Ventura	-0.32***	0.093	135,945
Yolo	-0.45**	0.242	22,650

Notes: Regressions of log(VMT) on all covariates in (ii) in Table 5 run separately for selected counties. The coefficient on ln(average gasoline price) is shown. Many counties have relatively small sample sizes and thus are statistically insignificant when all of the controls are included. Robust standard errors are given in column (2), clustered on vehicle model. *** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level.

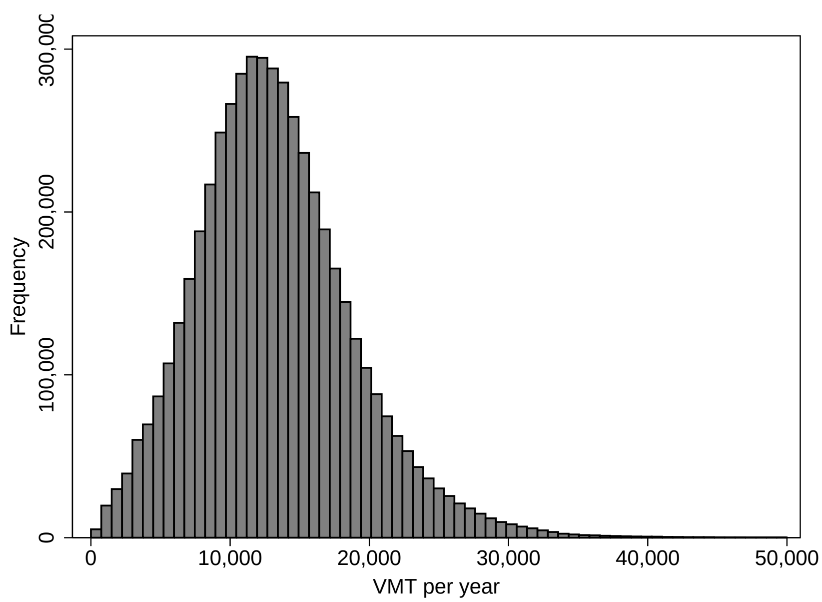

Figure 2: Driving per year by vehicles during their first six years of use in California has been remarkably single-peaked.

Table 11: Cluster Summary Statistics

Variable	Mean	Std Dev	Min	Max	N
Cluster 1: semi-rural, upper class					
zip density	3.28	3.19	0.00	22.04	498,620
zip income	115,572	9,565	104,313	139,257	498,620
Cluster 2: suburban, upper middle class					
zip density	4.49	3.81	0.00	31.52	1,399,177
zip income	76,231	3,876	67,828	84,575	1,399,177
Cluster 3: suburban, upper class					
zip density	4.01	3.53	0.00	32.13	881,663
zip income	93,039	6,062	84,774	104,029	881,663
Cluster 4: urban, low income					
zip density	6.91	7.50	0.00	51.60	1,255,451
zip income	40,236	7,040	0.00	49,790	1,255,451
Cluster 5: rural, wealthy					
zip density	1.65	1.19	0.48	6.73	100,809
zip income	165,487	21,939	141,614	375,000	100,809
Cluster 6: suburban, middle class					
zip density	5.51	4.77	0.00	52.18	1,731,055
zip income	59,349	5,416	49,865	67,709	1,731,055

Notes: Table presents summary statistics for each cluster based on a k-means clustering on zip code density and income

Table 12: Responsiveness by cluster

	cluster 1 **(i)**	cluster 2 **(ii)**	cluster 3 **(iii)**	cluster 4 **(iv)**	cluster 5 **(v)**	cluster 6 **(vi)**
log(avg gas price)	-0.12*	-0.17***	-0.18***	-0.26***	-0.30*	-0.19***
	(0.06)	(0.04)	(0.04)	(0.05)	(0.12)	(0.04)
firm	0.14***	0.16***	0.13***	0.17***	0.13***	0.14***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
rental	0.18***	0.12***	0.17***	0.13***	0.23	0.16***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.16)	(0.01)
government	-0.10	-0.11	-0.12	0.04	-0.20	-0.05
	(0.10)	(0.08)	(0.09)	(0.05)	(0.16)	(0.07)
lease	0.05***	0.05***	0.03***	0.04***	0.07***	0.05***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
constant	7.63***	6.97***	7.06***	7.16***	8.00***	6.71***
	(0.24)	(0.19)	(0.20)	(0.12)	(0.50)	(0.14)
Vehicle segment indicators	X	X	X	X	X	X
Vehicle characteristics	X	X	X	X	X	X
Purchase month indicators	X	X	X	X	X	X
Zip code % age brackets	X	X	X	X	X	X
Zip code % race brackets	X	X	X	X	X	X
County indicators	X	X	X	X	X	X
Model fixed effects	X	X	X	X	X	X
Economic conditions	X	X	X	X	X	X
% summer months	X	X	X	X	X	X
Counts of months covered	X	X	X	X	X	X
Time to test indicators	X	X	X	X	X	X
R-squared	0.100	0.071	0.079	0.051	0.156	0.062
N	401,253	1,123,723	785,472	1,091,046	92,084	1,543,398

Notes: Regressions of log(VMT) on all covariates used in (ii) in Table 5 run separately for each cluster. Robust standard errors are in parentheses, clustered on vehicle model. *** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level.

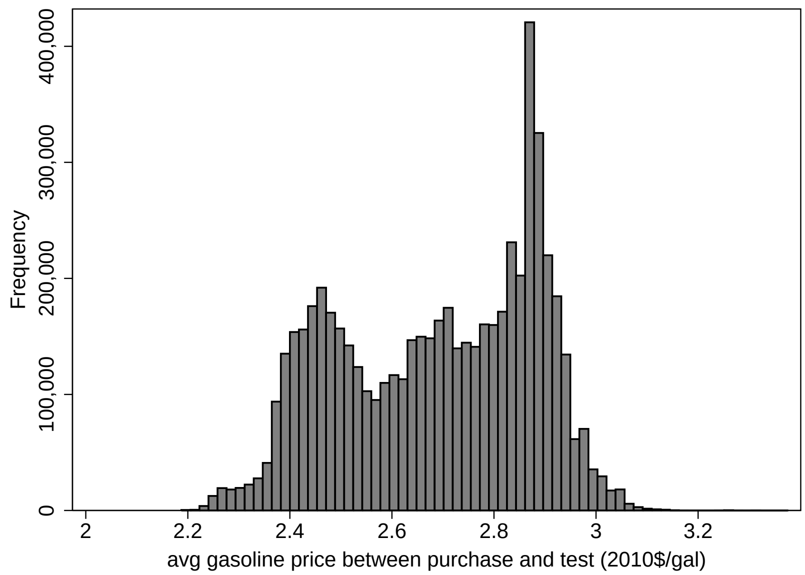

Figure 3: Average retail gasoline prices over the time frame to the first smog check show considerable variation.

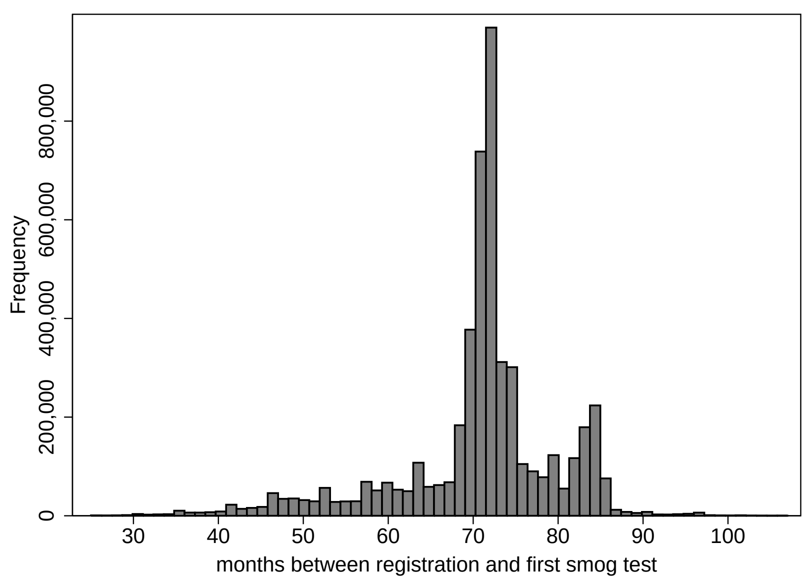

Figure 4: Most vehicles are tested within a few months of six years, but there is a mass at seven years as well.

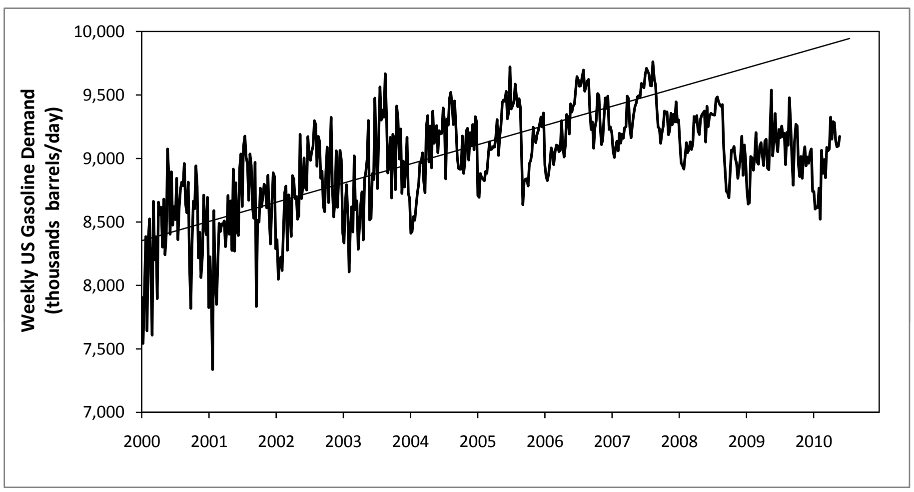

Figure 5: Gasoline demand in the US was increasing at a steady pace until the higher prices of 2007 and 2008 made an impact. Source: US Energy Information Administration.

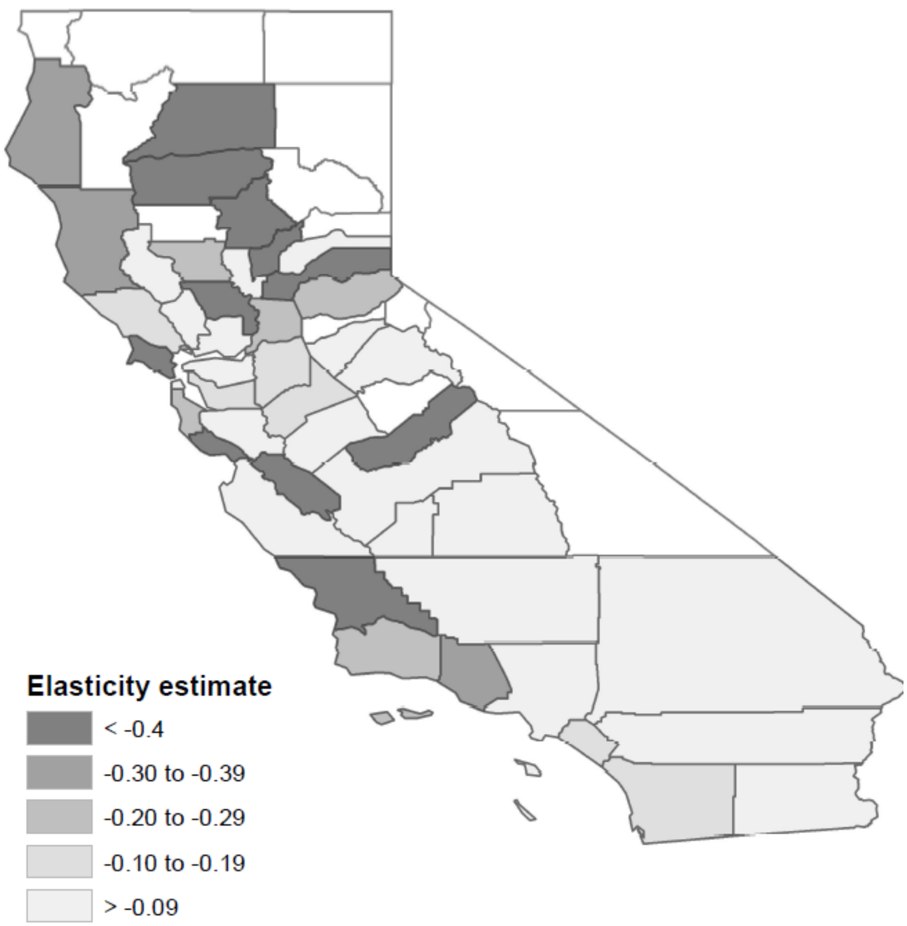

Figure 6: There is surprising heterogeneity in responsiveness across counties in California. Counties in white either are not subject to the smog check or do not have enough observations.