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

Kenneth Gillingham†

Submission for the Dennis J. O'Brian USAEE/IAEE Best Student Paper Award

July 8, 2010

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 vehicles registered in California in 2001-2003 and then subsequently given a smog check in 2005-2008. This covers a period of steady economic growth but rapidly increasing gasoline prices after 2005. The primary empirical result is a medium-term estimate of the utilization elasticity of driving – the elasticity of vehicle-miles-traveled with respect to gasoline price – in the range of -0.15 to -0.20. This estimate appears to be fairly robust to a variety of specification checks, and has important implications for the effectiveness of price policies, such as increased gasoline taxes or a carbon policy, in reducing greenhouse gases from the transportation sector. Further analysis indicates that there is considerable heterogeneity in this elasticity, both geographically and across income and vehicle types. This heterogeneity underscores differing distributional and local air pollution benefits of policies that increase the price of gasoline.

---

*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, and Larry Goulder for the funding that made this research possible. I would also like to thank Jim Sweeney, Tim Bresnahan, Jon Levin, and the participants at the UC-Berkeley Energy Institute at Haas Summer Workshop 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.

†Department of Management Science & Engineering and Department of Economics, Stanford University, P.O. Box 16336, Stanford, CA 94309, Phone: (650) 353-6578; Email: `kgilling@stanford.edu`.

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 either changing driving behavior or purchasing a more fuel-efficient vehicle. Quantifying exactly how consumers respond has been a research topic of great interest for energy economists for decades, yet it remains just a relevant as ever for policy analysis of a variety 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 gasoline prices. To identify this elasticity, this study brings 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 extremely rich dataset, along with a careful choice of time period, helps to overcome key identification challenges, while at the same time allowing for a closer look into the heterogeneity in consumer response at the income, vehicle class, 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 a rich 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 jointly estimate a utilization elasticity along with vehicle choice, often using the framework developed in Dubin and McFadden (1984).

More recent evidence suggests that the utilization elasticity may be declining 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, and find a utilization elasticity of -0.02 to -0.03 in the short run and -0.11 to -0.15 in the longer run. Small and Van Dender (2007) attribute their lower estimates to the growth of income and lower fuel prices over the time frame of their study, and Hymel, Small, and Van Dender (2010) use the same methodology with more recent data to reach similar conclusions. These results indicating a declining utilization elasticity 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 elasticity may be declining over time. Specifically, Hughes, Knittel, and Sperling (2008) find that the short run

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

[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.

price elasticity has decreased from -0.21 to -0.34 for 1975-1980 and is in the range of -0.03 and -0.08 for 2001-2006. Small and Van Dender (2007) also calculate a gasoline price elasticity and find slightly higher estimates of -0.05 to -0.09 in the short run and -0.43 in the long run. Older estimates of the long run elasticity from the 1970s and 1980s are often much higher (e.g., see the survey by Graham and Glaister (2002)).

There is remarkably little empirical evidence quantifying the heterogeneity in consumer responsiveness across any dimension and for either the gasoline price or utilization elasticity. West (2004) is a notable exception, which 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.

This paper uses the observed vehicle-level VMT to provide new evidence on the utilization elasticity of driving and explore the heterogeneity in this elasticity. A primary result is an estimate of the medium run utilization elasticity for vehicles in their first several years of use in the range of -0.15 to -0.20, suggesting that consumer responsiveness to the gasoline price shock of recent years was substantial. Moreover, the results from a k-means clustering technique suggest considerable heterogeneity in responsiveness across income, vehicle class, and geography.

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 novel 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 in a vehicle, identified by the 16 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, drive type, fuel, doors, body type, engine size, 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 and the transaction type (personal, firm, rental or government). 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.

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 all 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.

[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.

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 keep VIN matches where the make and model also match. Approximately 0.1 to 0.2 million of the vehicles each year are dropped due to miscoding, exemption from the smog check, were scrapped, violated the law, or were in a county that 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 drop VIN matches where the county in which the vehicle was registered when it was new does not match 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.

The primary question of this study is how consumers respond to gasoline price shocks, 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. 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 also some cross-county variation.[4] 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.

Next, I merge in zip code-level demographics from the American Factfinder (i.e., Census). 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

[4]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.

also include county-level unemployment over the time frame of the study from the Bureau of Labor Statistics to control for changing economic conditions.

Finally, one of the key characteristics of vehicles missing from the R.L. Polk and smog data is the fuel economy. This study is not attempting to identify an elasticity of VMT with respect to fuel economy, but the fuel economy of a vehicle may be an important control variable correlated with some of the regressors of interest. The EPA rates each vehicle by make, model, model year, trim, drive type, doors, transmission, fuel, body type, engine size, cylinders, and the presence of a turbo- or super-charger. I match the post-2008 vehicle fuel economy ratings to my vehicle-level dataset through an iterative process progressively matching based on fewer characteristics until all vehicles are matched to a fuel economy.[5]

2.2. Construction of sample

The sample chosen for this study is motivated by the structure of the smog check program and the economic downturn in the United States and California in late 2008. From 2001 until mid-2008, the economy was growing steadily, but in the third quarter of 2008, with the collapse of Lehman Brothers, the economy in both the US and California began taking a plunge. By 2009, the recession was fully underway, greatly changing consumer decision-making. The price of gasoline also dropped dramatically in late 2008. Thus, in order to avoid confounding the effect of changing gasoline prices with the economic downturn, I drop all observations with vehicles that had a smog test after September 2008.[6]

[5]Over 90% match on the make, model, engine size, and cylinders grouping.

[6]I perform a robustness analysis to show that the exact month chosen to truncate the sample has no appreciable impact on the results.

The sample is created as follows. Starting with roughly 1.5 million matches each year for vehicles registered in 2001 to 2003, the dataset drops to 2.3 million observations by dropping all test dates after September 2008. Next, I drop vehicles where the odometer reading six years later was coded to zero (2,521 vehicles) and vehicles for which the average miles driven per month is greater than 5,000 (1,730 vehicles), for both of these cases are very likely to be miscoded. Finally, I drop vehicles that do not run on gasoline (918 vehicles) or are commercial trucks (1,228 vehicles). This yields a dataset of 2,242,921 vehicles, 52% of which are registered in 2001, 38% 2002, and the remainder in 2003. Several thousand of these vehicles are in zip codes that do not have demographic variables, so the final sample used for estimation consists of 2,180,228 observations.

2.3. **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 $2.51/gal and the median is $2.54/gal. There are two prominent peaks corresponding to tests done in 2007 and 2008, occurring due to the seasonality of vehicle sales.

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 summers in the time frame between the new vehicle registration and the odometer reading also face a higher average price (from the raised summer prices), but respond differently because it is summer. To control for this possibility, I create a variable for the percentage of months during the odometer reading time period that are summer months. The mean of this variable is 24.9%, with a standard deviation of 1.1 percentage points.

Table 1 provides summary statistics for all of the non-demographic variables used in this study, including VMT per day and the average gasoline price. Most of these variables are indicator variables representing the 24 vehicle classes. The largest vehicle class (16% of the vehicles) is "Upper Middle," which includes several of the most popular cars such as the Honda Accord, Toyota Camry, Nissan Altima, and Ford Taurus. This class is followed by "Basic Economy" (13%), "Sport Utility" (12%), and "Fullsize Pickup" (11%). Also included in Table 1 are indicator control variables for whether the vehicle had the smog check due to a transfer of title (e.g., all vehicles getting a smog check much earlier than 6 years after initial registration) or whether the vehicle was late in getting a smog check.

Table 2 gives summary statistics for the zip code-level demographic variables, taken over observations (vehicles) in the dataset. One of the most important demographic variables for this study is the household income variable (Table 3). The subsample of my final dataset that includes income is 363,783 with 25% in 2001, 43% in 2002, and 33% in 2003. The distribution of income into the income brackets appears quite reasonable for those who purchased new vehicles and financed them.[7]

3. Estimation Strategy

3.1. Model

The demand for driving 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, the geography and demographics of the area they are driving in D_i, and the consumer's income I_i:

$$VMT_{it} = f(P_{it}, X_i, D_i, I_i).$$

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

$$VMT_{it} = (P_{it})^{\gamma} \exp(\alpha + \beta_X X_i + \beta_D D_i + \beta_I I_i) \exp(\varepsilon_i), \tag{1}$$

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

$$\log(VMT_{it}) = \alpha + \gamma \log(P_{it}) + \beta_X X_i + \beta_D D_i + \beta_I I_i + \varepsilon_i, \tag{2}$$

[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.

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). Heterogeneity in γ can be modeled in this reduced form by interacting price with other variables to see how the responsiveness to gasoline price shocks varies. Estimation is performed by ordinary least squares, fixed effects estimation, and quantile regression.

3.2. Identification

My research design and specification are developed to properly identify γ, which can be interpreted as the medium run elasticity of VMT with respect to the retail price of gasoline. Importantly, I observe consumers who purchased vehicles in 2001-2003 and then faced a gasoline price shock several years later, beginning in 2005 (recall Figure 1). Thus, by conditioning on the characteristics of the vehicle chosen and assuming imperfect foresight about the upcoming gasoline price shock at the time of purchase, I can focus on the VMT decision and do not need to simultaneously model VMT and vehicle choice decisions.

Another important aspect of my research design is that I am careful to avoid confounding my estimates with the downturn in the business cycle in late 2008. Up until 2008, the California

[8] Of course, since I only observe the average amount of driving over the entire time period of the odometer reading, the concave curvature of the log function may imply that my results would be biased due to Jensen's inequality. I explore this possibility in the robustness section and find it not to be a major consideration.

[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.

economy had experienced relatively steady growth since 2001. In late 2008 and into 2009, the gasoline price dramatically declined along with the economy. Since driving can be expected to follow economic activity, driving could be expected to decrease in 2009 along with gasoline prices. I avoid this issue by focusing on the responsiveness of driving before the economic downturn is fully in force. In addition, I control for more local macroeconomic forces by including the monthly unemployment rate in the county of the vehicle and the 2000 to 2007 rate of population growth in the zip code the vehicle is registered in.

My dataset allows me to attach perhaps the best estimate possible of the gasoline price faced by the drivers of a particular vehicle, but of course there is always some heterogeneity of gasoline prices within counties. In addition, some drivers may have registered in a county, moved for most of the life of the vehicle, and then registered or tested in that original county again. This measurement error would lead to an attenuation bias, possibly biasing my coefficient estimates downwards. I anticipate that this will be a relatively minor concern.

The dataset also allows for an analysis of individual-level decisions, so simultaneity of gasoline supply and demand is less likely to be a concern. At the individual level, consumers are not likely to take into account their influence on the gasoline price, so it seems reasonable that there is not an appreciable endogeneity concern here.

4. Results

4.1. Utilization elasticity

The primary results from estimating equation (2) on my sample of 2.1 million are given in Table 4. Column (1) includes vehicle class and zip code demographic controls, but does not include test year indicator variables. Column (2) includes these test year indicator variables to control for unobserved factors that vary over time. The results of these two columns suggests that controlling for the test year makes an important difference in identifying the utilization elasticity. The results in column (2) show a highly statistically significant medium run elasticity of VMT estimate of -0.17.

Column (3) includes vehicle model fixed effects to control for unobserved heterogeneity in driving behavior across vehicles that may be correlated with the average gasoline price. Column (4) includes zip code fixed effects to control for non-time-varying unobserved heterogeneity across zip codes that may be correlated with the average gasoline price.[10] The estimated elasticity values in these regressions, -0.2 and -0.18, are only very slightly larger than the result from column (2), largely corroborating the result in column (2).

Columns (5) through (7) present quantile estimation results to provide an overview of the heterogeneity in this elasticity. The 0.75 and 0.25 quantiles for the utilization elasticity are highly statistically significant and are -0.28 and -0.03 respectively. This suggests that there is considerable heterogeneity in the responsiveness across consumers, providing motivation for a further exploration of this heterogeneity. The median (0.5 quantile) elasticity estimation result in column (6) is a useful check on the results in columns (2) through (4) since the 0.5 quantile is more robust to outliers than the previous regressions. The result of a -0.15 elasticity is within the range of the results in columns (2) through (4), providing further evidence that the elasticity is well-identified.

To put these results into context, Figure 4 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

[10]Zip code demographics would be linearly dependent in the zip code fixed effects estimation and are thus not included.

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, Knittel, and Sperling (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 at all that a medium run elasticity for new vehicles in their first six years of life is in the range of -0.15 to -0.2.

The other coefficients in this estimation are best thought of as controls, and one should be cautious in interpreting them directly. For example, while it may be useful to condition on attributes of the vehicle, such as vehicle class and fuel economy, the effects of these variables on driving would best be estimated jointly along with the purchase decision.[11] Future work is underway to directly empirically estimate the rebound effect by simultaneously modeling these two decisions.

4.2. **Heterogeneity in responsiveness**

The richness of my dataset allows for further exploration into how people differ in consumer responsiveness. I examine heterogeneity in two ways: (1) perform a k-means cluster analysis on my income subsample and interact the clusters with the log of the gasoline price, and (2) interact the

[11]Interestingly, not including fuel economy or vehicle class as a covariate does not change the utilization elasticity very much at all, for neither are highly correlated with the gasoline price.

log of the gasoline price with county fixed effects. k-means clustering is a common approach used in statistics to partition observations into groups of similar observations. More specifically, the k-means clustering technique partitions N observations into K disjoint sets $S = \{S_1, S_2, ..., S_K\}$ in order to minimize the within-cluster sum of squares. In other words, S^* is given by

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

where x_j is an observation (i.e., vehicle) and μ_i is the mean of all of the observations in S_i. Just as in principal components analysis, the attributes of resulting clusters can then be examined and given a meaning in order to interpret the cluster. For the purposes of my analysis, I estimate clusters based on income groups, vehicle classes, and zip code density to give a sense of the heterogeneity across income, vehicle class, and degree of urbanization. I choose 15 clusters for this analysis.

I then perform the same regression as in column (2) of Table 4, only I include cluster indicator variables and interactions of the cluster indicator variables with the log of the gasoline price. The calculated values of the elasticities for each cluster are given in Table 5. I test if each elasticity is statistically different than zero with a Wald test and report the p-value of the test.[12] Nearly all of the clusters are statistically significant. The values of the cluster elasticities have several interesting interpretations, with a combination of income, need to drive, and the fuel economy or class of vehicle all interacting to produce substantial heterogeneity.

For urban dwellers, higher fuel economy cars have a lower elasticity than SUVs and pickups (compare clusters 2 and 3), suggesting that there may be some degree of switching vehicles within

[12] I use the delta method to get the variance-covariance matrix of the linear combinations of coefficients.

a household in response to a gasoline price shock. Similarly, for high income suburban dwellers, there appears to be a switch from low fuel economy pickups and SUVs to somewhat higher fuel economy luxury vehicles (compare clusters 7 and 8). High income suburban dwellers who own low fuel economy roadsters appear to either make a substantial switch to a more fuel-efficient vehicle – or alternatively, may cut out some unnecessary joy rides.[13] This behavior seems somewhat surprising, but perhaps makes sense in light of just how salient gasoline prices were during the gasoline price shocks of 2006-2008. In general, it appears that cars driven by higher income drivers are some of the least elastic.[14] Interestingly, rural low income dwellers driving pickups and SUVs appear to have a low elasticity. This may reflect the necessity of using these larger vehicles for work purposes.

To more closely examine the heterogeneity of consumer responsiveness across geography, I again use the same specification as in column (2) of Table 4, only this time I include county indicators and interactions between the county indictors and the log of the gasoline price. Most of the county-level elasticities are statistically significant to the 1% level and all but a few are significant at the 10% level. Figure 5 graphically shows the substantial degree of heterogeneity in responsiveness across California. The coefficients on the counties around the top and to the far right are identified off of very few observations, so it is more useful to closely examine the counties in the Bay Area (middle left), Central Valley (middle) and Southern California (bottom along the coast). The wealthier San Mateo and Santa Clara counties in the Bay Area appear to be less responsive. San Francisco appears to be slightly more responsive, perhaps due to better public transportation opportunities. Several

[13]Roadsters include vehicles such as the Mazda Miata, BMW M Roadster, Cadillac XLR, and Nissan 350Z.

[14]A similar cluster analysis that included a cluster for high income cars showed a very low elasticity – an element that may be manifesting itself here through the low elasticities in clusters 7 and 15.

of the less wealthy rural counties in the Central Valley appear to be more responsive, while the somewhat wealthier counties around Davis and Sacramento appear to be less responsive. Drivers in southern California appear to be somewhat responsive as well, but perhaps not as responsive as other counties.

These results underscore the substantial heterogeneity in responsiveness. Not all of the results make obvious sense a priori. For example, Marin County just north of San Francisco is very wealthy, but appears more responsive. However, upon further examination, it appears that the vehicle stock in Marin is considerably less fuel efficient on average than in many other counties in California (except for a large number of Priuses), which may contribute to the greater responsiveness.

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.

One important check is to examine the sensitivity of the results to changing the month of truncation of the sample. The fall of Lehman Brothers was in September 2008, and thus my primary sample uses this cut-off as the final month of the sample. The results change very little by varying this cut-off month by several months. For example, moving the cut-off month three months earlier changes elasticity estimate from the specification in column (2) of Table 4 to -0.16 and moving the cut-off month three months later changes the same elasticity estimate to -0.18. However, if I include large stretches of 2009, my estimated elasticity becomes much closer to zero. In the extreme case, I use the subset vehicles that did a test in 2007 and 2009 and estimate a small *positive* utilization elasticity. This result can easily be explained by the fact that in late 2008 and 2009, both the economy and gasoline prices were heading downward. I interpret this finding as important evidence that a careful research design is warranted in any attempt to estimate a utilization elasticity.

Another robustness check is to explicitly control for the vehicle purchaser's household income, rather than simply controlling for 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 makes nearly no quantitative change to the estimated elasticity value.

Finally, since the primary specification uses the log of the average gasoline price, I feel it is prudent to check whether the log of the averages would lead to a substantially different result due to Jensen's inequality. I create a variable for the average of the log gasoline price over the years of each vehicles odometer reading, and find nearly identical results to the primary specification results in Table 4. Future work will consider this issue more carefully, perhaps in a structural model of driving behavior.

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 consumers to the gasoline price shock of 2005-2008. I find evidence for a medium run elasticity of driving with respect to the gasoline price in the range of -0.15 to -0.20 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 energy policy.

The most striking implication is that policies that increase the price of gasoline may substantially 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, Knittel, and Sperling (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 – does not appear to be quite so inelastic. 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 of driving to their newer vehicle, implying that the newer vehicle would appear *less* elastic. Another explanation is that any gasoline price variations up 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. Using a k-means cluster analysis to divide up vehicles into logical groupings, I find noticeable shifts in driving behavior that differ along several margins. The evidence is suggestive of consumers switching to driving to more fuel-efficient vehicles, and perhaps cutting back on joy-riding. Urban and suburban drivers also appear to be affected somewhat differently than rural drivers, suggesting a geographic component to the heterogeneity. This geographic component comes through clearly in the differences in county-level elasticities. 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 will explore the implications of these results 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).

References

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

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

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., and S. Glaister (2002): "The Demand for Automobile Fuel: A Survey of Elasticities," *Journal of Transport Economics and Policy*, 36, 1–26.

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.

Hymel, K., K. Small, and K. Van Dender (2010): "Induced Demand and Rebound Effects in Road Transport," *Transportation Research B*, Forthcoming.

National Research Council (2002): *Effectiveness and Impact of Corporate Average Fuel Economy (CAFE) Standards*. National Academy Pres.

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

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

Appendix

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 counties. 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.

Table 1. Summary Statistics

Variable	Mean	Std Dev	Min	Max	N
VMT per month	1,129.7	480.9	0.01	4,993.6	2,242,921
gas price (real 2010$)	2.54	0.14	2.15	3.25	2,242,921
fuel economy (2008 ratings)	19.55	4.38	8.56	53	2,242,765
Entry Level	0.01	0.12	0	1	2,242,921
Basic Economy	0.13	0.34	0	1	2,242,921
Lower Middle	0.03	0.16	0	1	2,242,921
Upper Middle	0.16	0.37	0	1	2,242,921
Upper Middle Specialty	0	0.07	0	1	2,242,921
Traditional Large	0.01	0.09	0	1	2,242,921
Basic Sporty	0.04	0.2	0	1	2,242,921
Midsize Sporty	0	0.06	0	1	2,242,921
Prestige Sporty	0.01	0.08	0	1	2,242,921
Basic Luxury	0.05	0.21	0	1	2,242,921
Midsize Luxury	0.04	0.2	0	1	2,242,921
Prestige Luxury	0.02	0.14	0	1	2,242,921
Compact Pickup	0.06	0.23	0	1	2,242,921
Midsize Pickup	0.01	0.08	0	1	2,242,921
Fullsize Pickup	0.11	0.31	0	1	2,242,921
Mini Sport Utility	0.06	0.23	0	1	2,242,921
Sport Utility	0.12	0.32	0	1	2,242,921
Fullsize Utility	0.08	0.26	0	1	2,242,921
Minivan (Passenger)	0.06	0.24	0	1	2,242,921
Minivan (Cargo)	0.00	0.04	0	1	2,242,921
Window Van (Passenger)	0.00	0.05	0	1	2,242,921
Fullsize Van (Cargo)	0.01	0.09	0	1	2,242,921
Roadster	0.01	0.07	0	1	2,242,921
lease	0.16	0.37	0	1	2,242,921
firm	0.03	0.18	0	1	2,242,921
rental	0.07	0.26	0	1	2,242,921
govt	0.01	0.08	0	1	2,242,921
early smog check	0.33	0.47	0	1	2,242,921
late smog check	0.07	0.26	0	1	2,242,921
% months that are summer months	24.92	1.1	20	29.41	2,242,921
county unemployment rate	5.78	1.4	3.78	16.97	2,242,921
smog test in 2006	0.16	0.37	0	1	2,242,921
smog test in 2007	0.5	0.5	0	1	2,242,921
smog test in 2008	0.34	0.47	0	1	2,242,921

TABLE 2. Demographic Summary Statistics

Variable	**Mean**	**Std Dev**	**Min**	**Max**	**N**
zip density (thousand people/mi^2)	5.22	5.39	0	52.18	2,181,746
zip businesses 2000	1,605.9	1037.4	1	6,521	2,241,114
zip population 2007	41,820.4	19,803.1	1	109,549	2,180,909
zip pop growth rate 00-07	1.51	2.43	-32.5	199.2	2,180,909
zip median hh income 2007	71,063.96	26,975.02	5,952	375,000	2,180,389
zip % pop age 65+	11.19	5.11	0	100	2,180,909
zip % pop under 18	25.48	6	0	41.3	2,180,909
zip % pop white 2007	58.85	18.35	4.4	100	2,180,909
zip % pop black 2007	5.5	7.63	0	86.60	2,180,909
zip % pop hispanic 2007	31.44	21.13	0	97.8	2,180,909

TABLE 3. Tabulations of income

Income Category	Observations	Percent
<$15,000	26,896	0.07
$15,000 - $19,999	9,998	0.03
$20,000 - $29,999	29,705	0.08
$30,000 - $39,999	33,821	0.09
$40,000 - $49,999	36,951	0.10
$50,000 - $74,999	90,887	0.25
$75,000 - $99,999	60,159	0.17
$100,000 - $124,999	30,107	0.08
>$125,000	45,259	0.12
Total	363,783	1.00

Table 4. Primary results

	OLS		Fixed Effects			Quantiles	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	(Base)	(Time Controls)	(Model FE)	(Zip FE)	(0.25 Quant)	(0.50 Quant)	(0.75 Quant)
log(gas price)	-0.294***	-0.173***	-0.201***	-0.182***	-0.280***	-0.148***	-0.034***
	(0.006)	(0.010)	(0.025)	(0.019)	(0.014)	(0.010)	(0.010)
lease	0.060***	0.060***	0.062***	0.062***	0.090***	0.031***	-0.008***
	(0.001)	(0.001)	(0.005)	(0.002)	(0.001)	(0.001)	(0.001)
firm	0.226***	0.226***	0.232***	0.240***	0.215***	0.223***	0.231***
	(0.002)	(0.002)	(0.018)	(0.007)	(0.003)	(0.002)	(0.002)
rental	0.216***	0.216***	0.229***	0.225***	0.254***	0.178***	0.125***
	(0.001)	(0.001)	(0.008)	(0.011)	(0.002)	(0.002)	(0.002)
govt	0.007	0.008	-0.096	-0.054	-0.145***	0.062***	0.145***
	(0.006)	(0.006)	(0.092)	(0.032)	(0.006)	(0.004)	(0.004)
log(fuel economy)	0.193***	0.193***	0.196***	0.192***	0.180***	0.149***	0.146***
	(0.004)	(0.004)	(0.044)	(0.006)	(0.005)	(0.004)	(0.003)
% summer months	-0.000	-0.001**	-0.001	-0.001*	-0.000	-0.001*	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
early smog check	0.082***	0.076***	0.078***	0.077***	0.087***	0.064***	0.052***
	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
late smog check	0.031***	0.035***	0.038***	0.031***	0.045***	0.035***	0.024***
	(0.001)	(0.001)	(0.005)	(0.003)	(0.002)	(0.001)	(0.001)
unemployment rate	-0.005***	-0.005***	-0.004***	0.007	-0.007***	-0.006***	-0.004***
	(0.000)	(0.000)	(0.001)	(0.004)	(0.000)	(0.000)	(0.000)
smog test in 2006		0.131*	0.125*	0.136*	0.141	0.233**	0.158
		(0.059)	(0.063)	(0.067)	(0.111)	(0.085)	(0.081)
smog test in 2007		0.126*	0.121	0.135*	0.143	0.227**	0.143
		(0.059)	(0.063)	(0.067)	(0.111)	(0.085)	(0.081)
smog test in 2008		0.108	0.107	0.119	0.129	0.210*	0.123
		(0.059)	(0.063)	(0.066)	(0.111)	(0.085)	(0.081)
constant	7.106***	6.889***	6.908***	6.216***	6.686***	6.985***	7.272***
	(0.023)	(0.063)	(0.149)	(0.078)	(0.115)	(0.088)	(0.084)
Vehicle class controls	Y	Y	Y	Y	Y	Y	Y
Zip code demographics	Y	Y	Y	N	Y	Y	Y
Model FE	N	N	Y	N	N	N	N
Zip code FE	N	N	N	Y	N	N	N
Observations	2,180,228	2,180,228	2,180,228	2,242,765	2,180,228	2,180,228	2,180,228

*** indicates significant at 1% level, ** significant at 5% level, * significant at 10% level
Heteroskedasticity/Cluster-robust standard errors in parentheses
Dependent variable: log(VMT)

TABLE 5. Elasticities by cluster

cluster	elasticity	p-value (Wald test)	Observations
1 urban, medium income, all	-0.228	0.17	3,751
2 urban, medium income, cars	-0.158***	0.01	38,309
3 urban, medium income, pickups & SUVs	-0.258***	0.00	28,043
4 urban, medium income, luxury & SUVs	-0.267***	0.00	13,388
5 suburban, low income, pickups & SUVs	-0.305***	0.00	24,434
6 suburban, medium income, cars	-0.361***	0.00	77,248
7 suburban, high income, luxury	-0.154**	0.01	31,957
8 suburban, medium/high income, pickups & SUVs	-0.373***	0.00	53,024
9 suburban/rural, high income, roadsters	-1.189***	0.00	6,612
10 rural, low income, pickups & SUVs	-0.173***	0.01	27,071
11 rural, all incomes, cars	-0.147***	0.01	46,167
12 rural, medium income, SUVs	-0.286***	0.00	43,990
13 rural, medium income, vans & pickups	-0.435***	0.00	40,268
14 rural, high income, pickups & SUVs	-0.315***	0.00	46,331
15 rural, high income, luxury	-0.220***	0.00	40,317

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

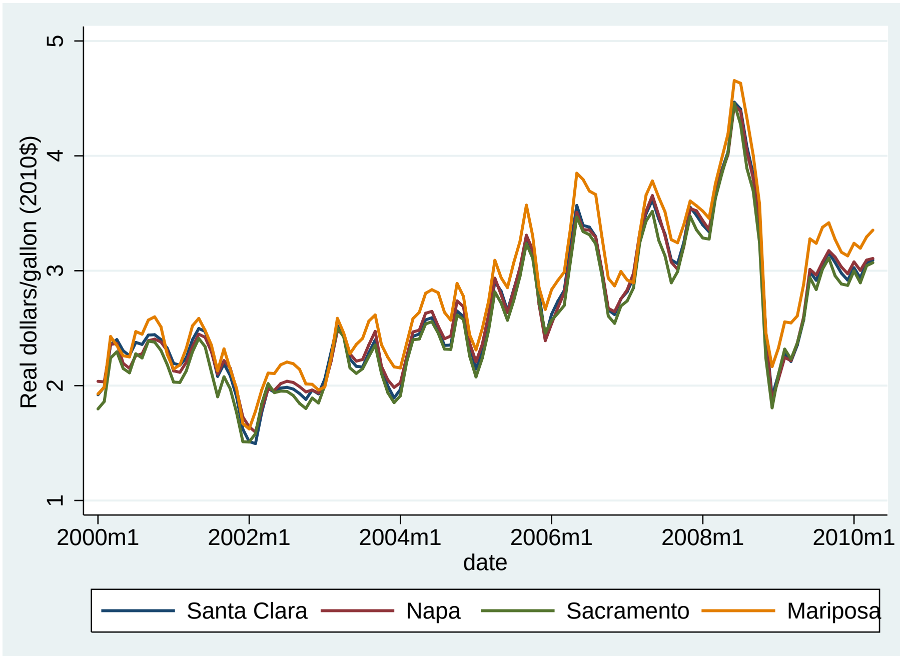

Figure 1. Retail gasoline prices in California were relatively flat and then rose substantially until 2008, providing substantial time series variation and some cross-county variation. Four representative counties are shown here.

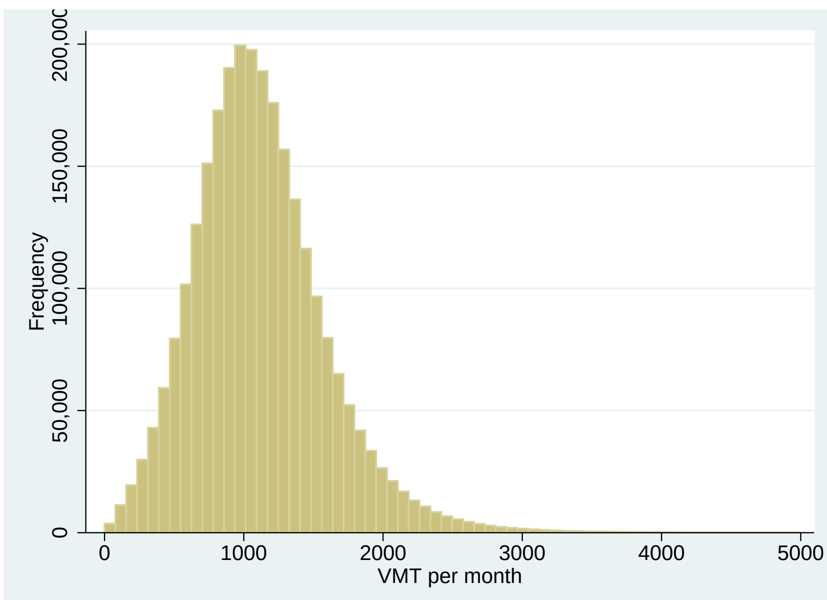

Figure 2. Driving per month by vehicles during their first six years of use in California has been remarkably sharp-peaked.

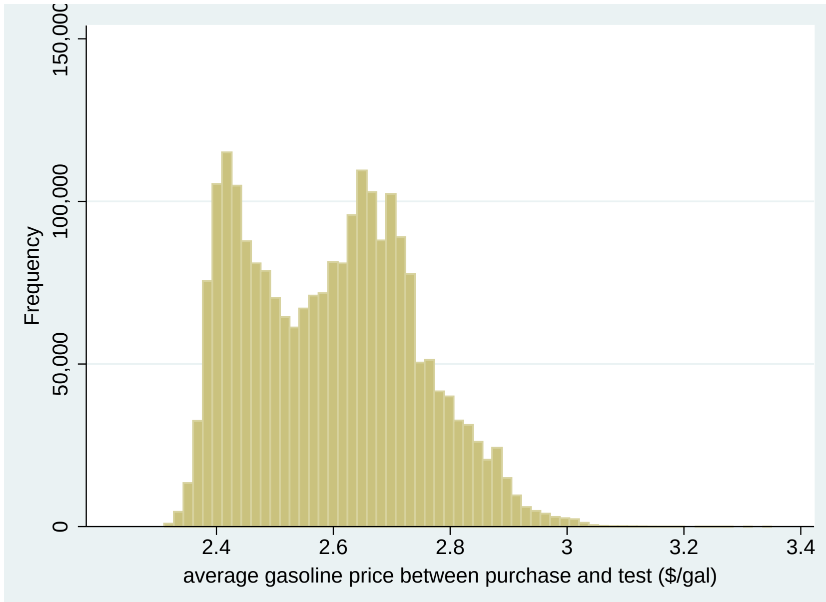

FIGURE 3. Average retail gasoline prices over the time frame to the first smog check show considerable variation.

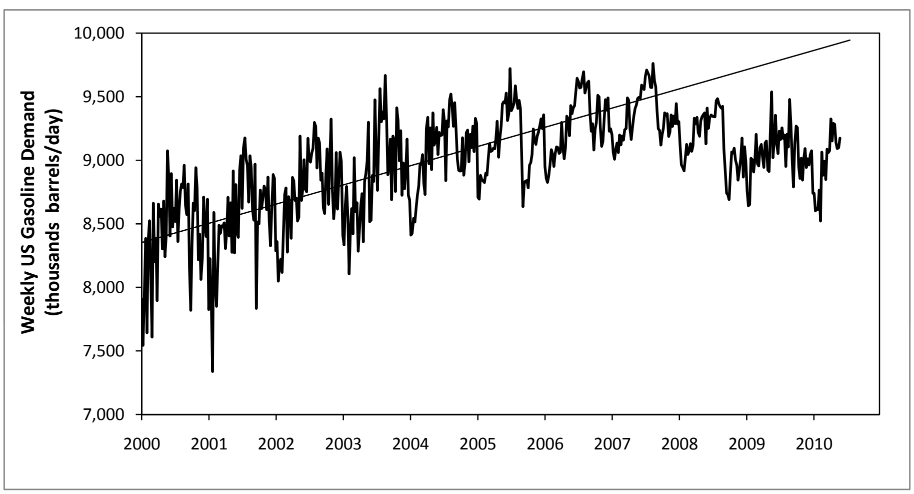

FIGURE 4. Gasoline demand in the US was increasing at a steady pace until the higher prices of 2007 and 2008 made an impact. Source: US Energy Information Administration.

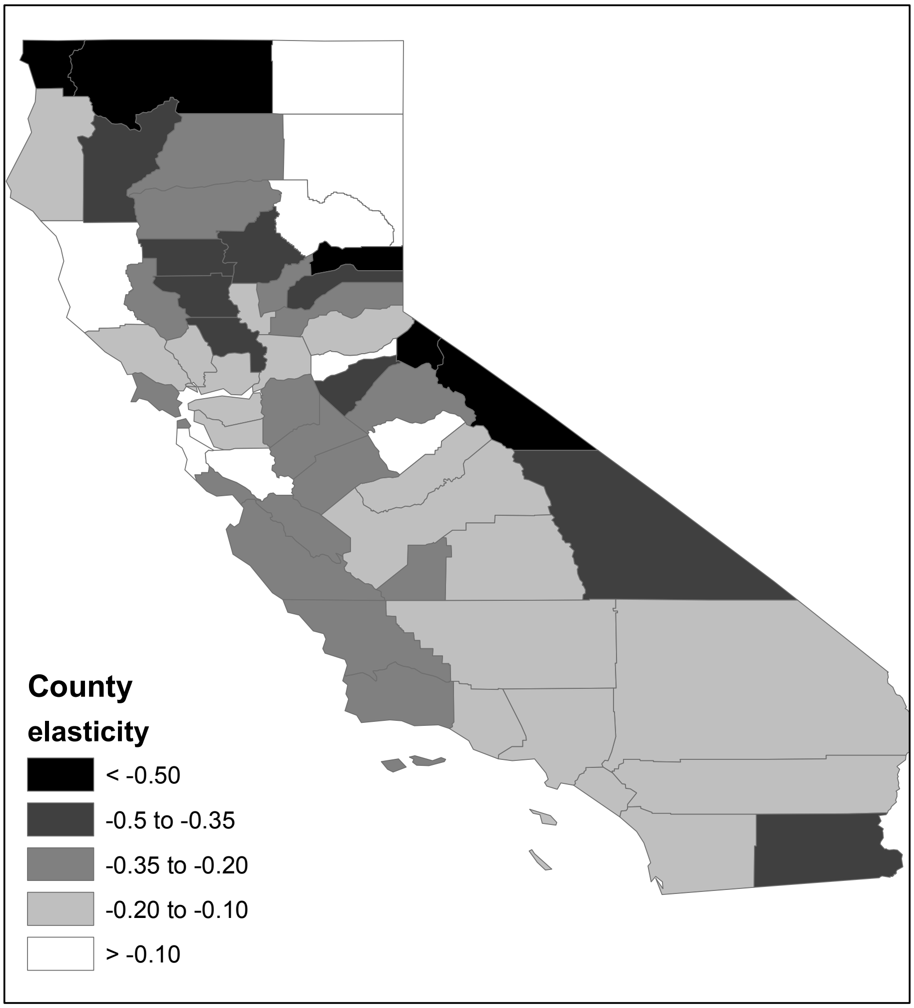

FIGURE 5. The map of California shows substantial heterogeneity across counties in California likely based largely on the local economy and income of the county.