Does Energy Efficiency Affect Demand: An Empirical Study on the Energy Paradox Within Automobile Industry*†

Travis Cao‡

October 10, 2018

Abstract

This paper studies "the energy paradox" within automotive market, where the slow adoption of vehicles with better fuel economy is observed. We obtain empirical survey data from California Energy Commission in attempt to reconstruct automotive market demand, investigating whether mpg (miles per gallon) has a significant contribution. We adopt the random coefficient logit framework developed by Berry, Levinsohn, and Pakes (1995), and create an empirical willingness to pay (WTP) distribution for one additional mpg to illustrate our finding. Our result shows that only about 30% of the empirical WTP density has WTP more than $600 (which is calculated as a reasonable cutoff for one additional mpg). Given that the majority (70%) are unwilling to pay for a reasonable amount in trading of one additional mpg, this suggests that consumer undervaluates mpg at the market level, illustrating the paradox.

Keywords: Energy Paradox, Demand Estimation, Fuel Economy

*Funded by the University of Wisconsin-Madison L&S Honors Program through Summer Senior Thesis Research Grant.

†Special thanks to Professor Matthew Wiswall at University of Wisconsin-Madison, Department of Economics, who served as advisor for this research project.

‡Legal name: Shengyang Cao

1 Introduction

"The energy paradox," an environmental economics term describing the slow adoption of cost-effective energy-efficient technologies, has been observed in a number of energy markets since its introduction in 1994 by Jaffe and Stavins (Jaffe and Stavins 1994). Most prominently, the automobile industry has been the classic example to illustrate the paradox, with its gradual increase of average fuel economy (or vehicle energy efficiency). Despite the technological advances made to enable vehicles with more than 100 miles per gallon (mpg) or equivalent unit measure for electricity charge, the mean fuel economy of vehicles owned by American households have increased merely by 5.4 mpg in the span from 2004 to 2016, reaching an average of 24.7 mpg for all types of vehicles and 28.5 mpg for cars only (EPA 2018).

To investigate the causes behind the slow fuel efficiency improvement in U.S. residential households, we adopt the random coefficients logit model developed by Berry, Levinsohn, and Pakes (Berry, Levinsohn, and Pakes 1995; henceforth BLP) where we perform a demand estimation on American automotive market. Due to the availability of data set, we decide to limit our focus to the state of California, studying whether fuel efficiency contributes to the market demand of home-use vehicles, and if so, whether the amount of contribution is significant enough.

The decision on restricting our geographical focus is not arbitrary. The gasoline or electricity charges incurred for refueling vehicles vary across states. At the same time, states might impose different policies and subsidy programs, which creates different impact on consumer preferences. By focusing on one state, we are able to control the aforementioned variations, enabling us to make meaningful interpretations from our result. In terms of which state to choose, California rises to the top due to its easily accessible public data. California Energy Commission (CEC) has been conducting transportation surveys on state residential and commercial sectors, providing detailed micro-data that contains vehicle types, makes, model year, acquired year, refueling mechanisms, mpg, and so on. With such detailed information, we are able to recreate market level vehicle share data used by BLP's random coefficient model, and then further recover the demand of each specified market.

During the course of our study, we have found evidence suggesting consumer undervaluing fuel economy. By constructing a willingness to pay distribution (WTP), we found that only roughly 30% of the sample population have reasonable WTP for one additional mpg. This leaves the majority of consumers with lower-than-reasonable WTP for an extra mpg, illustrating the energy paradox.

In terms of our methodology, we choose to deploy BLP's random coefficient logit model as our demand estimating technique. This model is designed to apply on aggregate-level data set with available sales, prices, and product characteristics, but it has also been successfully adapted to micro-level survey data with richer consumer demographic information (Berry, Levinsohn, and Pakes 2004). In the field of industrial organization, BLP's model has been widely applied due to two major advantages. One, BLP's random coefficient allows for price endogeneity. For a given product, its market demand usually depends, not surprisingly, on its price and own characteristics. While intuitively we can recover the demand of a product by simply including both price and product characteristics as regressors, one problem in doing so is that price itself is often determined by product characteristics. In other words, price is an endogenous variable since it is determined by our model. BLP address price endogeneity by using appropriate instrumental variables. Two, BLP's model creates reasonable substitution patterns. BLP allows for interaction between individual and product characteristics, creating substitution patterns that depend on consumer preferences. This prevents cross-price elasticities to be solely dependent on product market share; that is, two drastically different vehicles with same market share will have the same own-price demand derivatives. We will expand more on BLP's approach in section 2.

The study of "Energy Paradox" in the automotive market has had a long history. Among the many theories proposed to explain the paradox, one has been at researcher's focus: how fuel economy contributes to automotive market demand. More specifically, does consumer undervalue (or put no value on) fuel economy. If this turns out to be the case, then it becomes less surprising to see gradual fuel economy increase despite technological advances. Goldberg collected micro data from Consumer Expenditure Survey (CES), and adopted a discrete choice model for estimating automotive market demand (Goldberg 1995). In this model, buying a vehicle is considered as a nested logit sequence, in which consumer is presented with options of buying a car or selecting the outside good. Then, total market demand is weighted from CES's population data. Combined with an oligopoly supply with differentiated products, market equilibrium is reached. Goldberg found that fuel efficiency contributes positively to luxury and sports car market, but negatively to small and big car market. While Goldberg did not make explicit interpretation on these results, we can see that consumer valuation on fuel efficiency varies across markets. This confirms the need of considering heterogeneous preferences in different consumer markets.

Brownstone, Bunch, and Train looked into automotive market forecasting problems plagued by using either stated preferences (consumer response in experiment environment) or revealed preferences (what consumer actually purchase) data, and proposed using a joint model to harness each type of data's benefits (Brownstone, Bunch, and Train 2000). They conducted two rounds of interviews: stated preferences were collected in the first round, the revealed preferences in the second; the second round of interview is conducted 15 months after the first. To combine two types of data, Brownstone, Bunch, and Train developed simple mixed logit specifications, which incorporate unobserved correlations and scaling differences. Estimation from joint mixed logit suggests that having higher range between charges or refuels increases predicted market demand; that is, consumer places positive value on fuel economy. However, when there exists a large number of choices, Brownstone, Bunch, and Train's discrete choice approach may not be suitable for revealed choice modelling. More troubling, Brownstone, Bunch, and Train assumes away from the price endogeneity issue, which, as pointed out by BLP, may lead to overfitting on aggregate data and produces bias too large to be explained by sampling error. These give more weight to BLP's model, where market share is used to help with the large number of revealed choices, and price endogeneity problem is considered.

Allcott and Wozny applied a static discrete choice model for a hypothesis testing: to investigate whether consumer values fuel economy, Allcott and Wozny proposed null hypothesis that consumers are willing to pay $1 extra today to purchase a vehicle with $1 less forecasted future fuel cost (Allcott and Wozny 2014). If the null hypothesis is rejected with a more than $1 willingness to pay, then consumers are said to overvalue fuel economy; if willingness to pay is less than $1, consumers are said to undervalue fuel economy. Allcott and Wozny's study found that consumers are willing to pay only $0.76 for a reduction of $1 future discounted fuel cost, implying undervaluation. While the result is shown to be insensitive to market share endogeneity and preferences change over time, Allcott and Wozny mentioned that a higher discount rate and the inclusion of 2008 recession might move the willingness to pay measure upwards, casting doubts on the reliability of using such static model.

As seen by the aforementioned research, it is hard to draw a clear-cut conclusion on how consumer values fuel economy. In fact, a report issued by Environmental Protection Agency (EPA) found 42.9% research conducted between 1994 and 2010 (inclusive) conclude that consumer undervalues fuel economy, while 17.9% found overvaluation, and the remaining 39.2% state same level of valuation as other characteristics or otherwise (*How Consumers Value Fuel Economy: A Literature Review* 2010). Within the sample, only about half have considered price endogeneity, and about half have considered consumer heterogeneous taste; it is unclear from the report how many addressed these two issue at once.

What our research attempts to accomplish is to perform automotive market demand estimation that accounts for both price endogeneity and heterogeneous taste, and then evaluates how consumer WTP on mpg responds to different consumer demographics. These two criteria produce more accurate effect size of price, and generate more intuitive substitution pattern. Hence, it is crucial to consider both factors for demand estimation.

The rest of the paper is laid out as the following. Section 2 expands on the BLP random coefficients logit model that we will adopt, and explain how this model is able to account for both price endogeneity and heterogeneous taste. Section 3 explains the data we use, and how we generate data set that is suitable for BLP analysis. Section 4 reports our result and provides interpretation by creating empirical WTP distribution. Section 5 discusses the validity of our result, and the potential limitation of our interpretation. Section 6 concludes the paper, restates our findings, and suggests path for future research.

2 Methodology

We deploy BLP's random coefficient logit model, which is a mixed logit model under the discrete choice family. The model specifies that consumer derives utility from purchasing one good from the choice set, where consumer is allowed to choose buying nothing. The utility level depends on a vector of consumer characteristics ζ, and a vector of product characteristics (p, x, ξ). Here, p represents the product price, x contains characteristics that are observed by the researcher, and ξ are characteristics unobserved by the researcher. We can write our general utility function as the following

$$u(\zeta, p, x, \xi; \theta) \tag{1}$$

in which $u(\cdot)$ represents the utility function, and θ are parameters to be estimated.

By this specification, consumer i within market t can choose good $j \in \{0, 1, 2, \ldots, J\}$, where $j = 0$ implies consumer buying nothing. When good j is chosen, consumer i derives utility

$$u(\zeta_{it}, p_{jt}, x_{jt}, \xi_{jt}; \theta) \geq u(\zeta_{it}, p_{rt}, x_{rt}, \xi_{rt}; \theta), \text{ for } r \in \{0, 1, 2, \ldots, J\} \text{ and } r \neq j \tag{2}$$

Equation 2 denotes how one individual i in market t chooses. However, what we are more interested in is the market share of good j in market t (market share is defined as quantity sold in market t divided by total consumers in market t, including consumers buying no good). This can be solved by integrating the distribution of consumer characteristics over the domain where good j is chosen in market t. Such domain is defined by

$$A_{jt} = \{\zeta : u(\zeta, p_{jt}, x_{jt}, \xi_{jt}; \theta) \geq u(\zeta, p_{rt}, x_{rt}, \xi_{rt}; \theta),\ \text{for } r \in \{0, 1, 2, \ldots, J\} \text{ and } r \neq j\} \tag{3}$$

With equation 3, we should be able to recover the share function $s(\cdot)$ through integration

$$s_{jt}(p_{jt}, x_{jt}, \xi_{jt}; \theta) = \int_{\zeta \in A_{jt}} d\zeta \tag{4}$$

Up until here, we have been using the generic utility function $u(\cdot)$ and ignoring the practicality of solving for such integration. Now, let us define the specific form of $u(\cdot)$, and expand on how the parameters are solved.

For consumer i choosing good j in market t, we define its indirect utility received as the following

$$u_{ijt} = \boldsymbol{X_{jt}}\boldsymbol{\beta_{it}} + \alpha_{it} p_{jt} + \xi_{jt} + \epsilon_{ijt} \tag{5}$$

where $\boldsymbol{X_{jt}}$ is a $1 \times K$ matrix of K numbers of product characteristics; p_{jt} represents the inflation-adjusted transacting price (units: \$1,000); ξ_{jt} gives unobserved product characteristics; ϵ_{ijt} retains all other error terms.

$(\boldsymbol{\beta_{it}}, \alpha_{it})$ are the θ parameter for consumer demographics. They are defined as the following

$$\begin{pmatrix} \boldsymbol{\beta_{it}} \\ \alpha_{it} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta} \\ \alpha \end{pmatrix} + \boldsymbol{\Pi} \boldsymbol{D_{it}} + \boldsymbol{L} \boldsymbol{\nu_{it}} \tag{6}$$

in equation 6, $\binom{\boldsymbol{\beta}}{\alpha}$ is of dimension $(K+1) \times 1$, representing the coefficient of a characteristic for a representative consumer in the market; $\boldsymbol{D_{it}}$ is a $d \times 1$ matrix containing d numbers of consumer demographic variables; $\boldsymbol{\nu_{it}}$ denotes the unobserved consumer demographics component in a $(K + 1) \times 1$ matrix; $(\boldsymbol{\Pi}, \boldsymbol{L})$ are coefficients resulted from consumer heterogeneity, where $\boldsymbol{\Pi}$ is of dimension $(K + 1) \times d$, and $\boldsymbol{L}$ is of dimension $(K + 1) \times (K + 1)$.

Now, if we want to similarly recover the predicted market share, we can start with one individual i's probability of choosing good j among their choice set. Define

$$\tilde{A}_{ijt} = \{\epsilon_{it} : u_{ijt} \geq u_{irt},\ \text{for } r \in \{0, 1, 2, \ldots, J\} \text{ and } r \neq j\} \tag{7}$$

With the help of this notion, individual i's probability of choosing j becomes

$$Pr_{ijt} = \int_{\tilde{A}_{ijt}} dF(\epsilon_{ijt} | \boldsymbol{D_{it}}, \boldsymbol{\nu_{it}}) \tag{8}$$

Here F denotes the distribution of error term ϵ_{ijt}.

In logit, the error term ϵ_{ijt} is usually assumed to have Type-I extreme value distribution. After imposing this assumption, we can explicitly write out the integration in equation 8 as the following

$$Pr_{ijt} = \frac{\exp(\boldsymbol{X_{jt}}\boldsymbol{\beta_{it}} + \alpha_{it}p_{jt} + \boldsymbol{\xi_{jt}})}{1 + \sum_{m=1}^{J} \boldsymbol{X_{mt}}\boldsymbol{\beta_{it}} + \alpha_{it}p_{mt} + \boldsymbol{\xi_{mt}}} \tag{9}$$

Equation 9 yields one individual i's probability of choosing good j. If we integrate it across the entire sample, we will recover the probability that good j is sold over other goods in market t, which is the sample market share of good j in market t.

$$s(jt) = Pr_{jt} = \int_{\boldsymbol{D_{it}}} \int_{\boldsymbol{\nu_{it}}} Pr_{ijt} dF(\boldsymbol{D_{it}}|\boldsymbol{\nu_{it}}) dF(\boldsymbol{\nu_{it}}) \tag{10}$$

while equation 10 is hard to evaluate analytically, we can randomly draw R samples of $\boldsymbol{D_{it}}$ and $\boldsymbol{\nu_{it}}$ (where $\boldsymbol{D_{it}}$ comes from consumer demographic sample, $\boldsymbol{\nu_{it}}$ is of normal distribution $N(0, \boldsymbol{I_{K+1}})$), and then use Monte Carlo integration to approximate equation 10.

The reason behind calculating the predicted market share of good j in market t is to equate it with our observed market share, which helps us recover the θ coefficients that we are interested from our data. Specifically, BLP random coefficient logit is run in the following steps:

1. Draw R number of individuals from the consumer demographic sample.

2. Solve for an initial market mean utility coefficients, by simulating the integral for each good j's market share in market t and set it equal to the observed market share.

3. Use instruments $\boldsymbol{Z_t} = \begin{pmatrix} z_{11} & z_{21} & \dots & z_{J1} \\ \vdots & \vdots & \ddots & \vdots \\ z_{1l} & z_{2l} & \dots & z_{Jl} \end{pmatrix}$ and $\boldsymbol{\xi_t} = \begin{pmatrix} \xi_{1t} \\ \vdots \\ \xi_{Jt} \end{pmatrix}$ (where z_{jl} are the lth instrument for good j in market t), form sample-moment conditions $T^{-1}\sum_{t=1}^{T} \boldsymbol{Z_t}\boldsymbol{\xi_t}$, and further form GMM (Generalized Method of Moments) objective function based on that.

4. Search for coefficients $\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\Pi}, \boldsymbol{L}$ that minimizes the GMM objective function.

Vincent (2015) records more detailed operation on step 3 and 4. We will take a moment here to discuss which instruments are the most suitable for addressing price endogeneity. Because price correlates to demand-side product characteristics, and price itself also affects market demand, price is hence considered as an endogenous variable. And if the price endogeneity problem is left unaddressed, the coefficient on price variable is surely to be biased, affecting the interpretation of our result.

To address this issue, BLP suggested using instrumental variables. The fundamental of using instruments z is established by the following:

$$cov(z_{jt}, p_{jt}) \neq 0 \tag{11}$$

$$cov(z_{jt}, \xi_{jt}) = 0 \tag{12}$$

Essentially, to address the bias of price variable resulted from the endogeneity problem, we would want to use instruments to perform two-stage estimation in reducing bias. From a demand estimation perspective, we would choose variables that directly affect the pricing of a vehicle, without impacting other vehicle characteristics. Two classical sets of instruments have been proposed over the years that are deemed as satisfying equation 11 and 12. One is direct cost shifters that relate to the supply side of the market; this includes factors such as shipping cost, labor cost, and input component costs. The other is vehicle characteristics in the competing markets: that is, the sum of characteristics for the same type of vehicles produced in other firms, or the characteristics of other vehicles produced at the same company. These competing market characteristics can be viewed as cross market cost shifters.

Since our data solely reflects the demand side of the market, we adopted the second set of instruments. For each vehicle characteristics, we included the own and competing firm's sum of characteristics up to order 2, and their interactions in between. That is, if we have one vehicle characteristic x_{jt}, our set of instruments will be the following:

$$\begin{aligned} Z_{jt} = \Bigg\{ & \sum_{r \neq j, r \in \mathcal{F}_{jt}} x_{rt}, \sum_{r \neq j, r \notin \mathcal{F}_{jt}} x_{rt}, \\ & \Bigg(\sum_{r \neq j, r \in \mathcal{F}_{jt}} x_{rt} \Bigg)^2, \Bigg(\sum_{r \neq j, r \notin \mathcal{F}_{jt}} x_{rt} \Bigg)^2, \\ & \Bigg(\sum_{r \neq j, r \in \mathcal{F}_{jt}} x_{rt} \Bigg) \cdot \Bigg(\sum_{r \neq j, r \notin \mathcal{F}_{jt}} x_{rt} \Bigg) \Bigg\} \end{aligned} \tag{13}$$

where $\mathcal{F}_{jt}$ denotes the parent firm of good x_{jt}.

In terms of the more flexible and realistic substitution patterns, BLP achieves this by considering consumer heterogeneous taste. If we take away the random coefficients component, effectively making $\boldsymbol{\beta_{it}} = \boldsymbol{\beta}$ and $\alpha_{it} = \alpha$, then the own- and cross- product price elasticities of demand (changes of own or other product market share, versus changes in price) are

$$e_{jkt} = \begin{cases} -\alpha p_{jt}(1 - s_{jt}) & \text{if } j = k \\ \alpha p_{kt} s_{kt} & \text{if } j \neq k \end{cases} \tag{14}$$

Here, s_{jt} represents the market share of good j in market t, and e_{jkt} denotes the price elasticities of good j in market t to either itself (where $j = k$), or to another product (where $j \neq k$ and k is from market t).

Given that market share is often small, own-price elasticity is roughly linear in product's own price. This suggests that cheaper products will be more price inelastic, which is a rather unrealistic assumption to make. What is more problematic is the cross-product price elasticity. Consider a sports car, an SUV, and another sports car in market t, all have the same price and same market share. Suppose we increase the price of the second sports car, then equation 14 suggests that the market share of the first sports car and the SUV will rise to the same degree. In reality, we are more likely to observe consumers substitute for the same type of vehicle, so it is inappropriate to have cross-product elasticities held same for both the first sports car and the SUV in our example.

In BLP, the introduction of consumer heterogeneity addresses the above problems. The own- and cross-price elasticities in BLP random coefficient logit is the following:

$$e_{jkt} = \begin{cases} -\frac{p_{jt}}{s_{jt}} \int \alpha_i Pr_{ijt}(1 - Pr_{ijt}) dF(\boldsymbol{D_{it}}, \boldsymbol{\nu_{it}}) & \text{if } j = k \\ \frac{p_{kt}}{s_{jt}} \int \alpha_i Pr_{ijt} Pr_{ikt} dF(\boldsymbol{D_{it}}, \boldsymbol{\nu_{it}}) & \text{if } j \neq k \end{cases} \tag{15}$$

Because of the interaction between consumer heterogeneity, price elasticities generated by BLP's model allow consumers with similar demographics to have similar substitution patterns, which is more likely to be observed in data.

Now that we established the foundation of BLP random coefficient logit, we write out the specific variables that we decided to include. In terms of vehicle characteristics in equation 5, we include the following one at a time and reported the change of coefficients as a panel in our results section:

$$\begin{aligned} \boldsymbol{X_{jt}} = (&mpg_{jt}, hpoweight_{jt}, wheelbase_{jt}, fueltank_{jt}, \\ &cityrange_{jt}, rpm_{jt}, size_{jt}, typesports_{jt}, european_{jt}) \end{aligned} \tag{16}$$

Here jt represents model j in market t. All the variables themselves are defined as the following. mpg stands for miles per gallon (or equivalent measure for electric and hybrid vehicle), which is recorded as EPA combined mpg (55% city value, 45% highway value) and in the units of miles per 10 gallon. $hpoweight$ is short for horsepower over weight (units: 10 hp per lbs of curb weight). $wheelbase$ is the distance between the front and rear axles of a vehicle (units: inch / 100). $fueltank$ denotes the amount of fuel (or electricity charge) that a vehicle can hold (units: gal / 10). $cityrange$ is the range of distance a vehicle can be driven in the city (units: mi / 100). rpm stands for revolution per minute, which records how fast a vehicle's crankshaft is spinning (units: rpm / 1000). $size$ is calculated by vehicle's length times width (units: sq in / 10000). $typesports$ is a binary variable, recording whether the vehicle is of sports type. $european$ is also a binary variable, denoting whether the vehicle's parent company is in Europe.

By using these vehicle characteristics, we write out the explicit form of equation 5 as the following:

$$\begin{aligned} u_{ijt} =& \beta_{mpg_{it}} mpg_{jt} + \beta_{hpoweight_{it}} hpoweight_{jt} + \beta_{wheelbase_{it}} wheelbase_{jt} + \\ & \beta_{fueltank_{it}} fueltank_{jt} + \beta_{cityrange_{it}} cityrange_{jt} + \beta_{rpm_{it}} rpm_{jt} + \\ & \beta_{size_{it}} size_{jt} + \beta_{typesports_{it}} typesports_{jt} + \beta_{european_{it}} european_{jt} + \\ & \alpha_{it} p_{jt} + \xi_{jt} + \epsilon_{ijt} \end{aligned} \tag{17}$$

In terms of consumer demographic variables in equation 6, we include the following:

$$\boldsymbol{D'_{it}} = (income_{it}, education_{it}, gender_{it}) \tag{18}$$

In equation 18, it stands for consumer i in market t. $income$ reports the inflation-adjusted real annual income (units: \$). $education$ is an individual's total years of education (units: years). $gender$ is a binary variable recording an individual's gender (1 if male, 0 if female). We took the step to demean and standardize the demographic variables within each market, which helps us with the convergence of GMM objective function and reach result in our BLP random coefficient logit.

After defining all components in consumer demographic D_{it}, for any of the random coefficient we had in equation 17 (denote this random coefficient as a generic β), we can write out the explicit form of equation 6 for it:

$$\beta_{it} = \beta + \pi_{income_{it}} income_{it} + \pi_{education_{it}} education_{it} + \pi_{gender_{it}} gender_{it} + \tilde{\nu}_{it} \tag{19}$$

where $\tilde{\nu}_{it} = \boldsymbol{L}_r \boldsymbol{\nu}$ the product of matrix multiplication, and $\boldsymbol{L}_r$ denotes the rth row of $\boldsymbol{L}$.

We now turn to the data section on how such variables are being extracted.

3 Data

As briefly mentioned in introduction, our data comes from CEC's 2015-2017 California Vehicle survey on residential and commercial light-duty vehicle sectors in California. The main purpose of this survey is to help forecast energy demand as vehicle technologies advance over years. However, given its recording on vehicle characteristics, we manage to adapt this data set to our analysis purpose.

The CEC has published two versions of this data: Version 1 was made public in 2017; Version 2 in 2018. We ended up using the first version, since, for some unknown reason, the second version removes the reported combined mpg value of vehicles owned by consumers. Version 1 and 2 mainly differs on the level of data cleanliness, given that Version 1 was first published as raw survey data. Additionally, while the survey includes samples drawn from both residential and commercial sector, we only used the residential sector since it aligns with our study focus.

The survey samples were selected through a market research panel, where both the physical and online address were used to stratify samples into six regions: San Francisco, Sacramento, Central Valley, Los Angeles, San Diego, and the Rest of California (*2015-2017 California Vehicle Survey Consultant Report* 2018). Random sampling technique was applied within each strata. Additional questions were asked on respondent who owns a plug-in electric vehicle.

Selected samples received a postcard from CEC, on which a unique access password was printed for recipient to access the online survey. In case of no internet access, telephone line and mail-in address were provided as alternative ways to complete the survey, though only two respondents successfully utilized these methods. As pointed out by CEC's report, residential respondents on average completed the survey in more than 30 minutes. This suggests that the final sample might be skewed towards households with fewer vehicles, as less time is required for them to complete all survey questions.

Given our goal to account for consumer heterogeneous taste, we first need to define the markets in which vehicles were acquired. This step is needed since consumer demographic distributions may vary across markets. We followed two criteria to define markets in our sample: the year when the vehicle is acquired, and the region where the respondent resides. This division is necessary since consumer demographic distribution varies given different time and regional markets, and hence the distribution difference may result in variation in heterogeneous taste. The time market was acquired straight from survey's "year_acquired" variable, in which we selected market year 2005 to 2016 (inclusive) due to the ample amount of samples in this period.

The regional market was defined as whether consumer resides in an urban area. We were able to create the urban variable by matching respondent's zip code with its corresponding city, and then selected a population cutoff to divide urban and non-urban markets. Specifically, we retrieved 2010 U.S. Census data, and defined urban region as cities with population larger than .5 million (*2010 Census* 2010). Combining the year and regional division yields 24 markets in total.

While CEC data set includes many vehicle characteristics that help us identify the vehicles, it suffers from one major issue: CEC did not ask respondents on the transacting price of the vehicle acquired, nor did it provide a manufacturer suggested retail price (MSRP). Undoubtedly, price plays a major factor in such an oligopoly industry, and so to recover the price variable, we matched the CEC data set with a third-party vehicle-price-and-specs data, based on the vehicle makes and specs that we observe from CEC's data. The third-party data was purchased from Teoalida, INC., whose data was scraped off from Edmunds.com, a car dealership website with comprehensive data on vehicles sold in the United States. Although Edmunds also sells an application program interface (API) that allows access to its database, their API is not suitable for generating a table of specs and price for all vehicles sold. To ease our matching process, we fell back to Teoalida's solution and acquired the data table from them.

The matching between CEC and Teoalida data was done by identifying vehicle make, model, type (i.e. whether the vehicle is a normal car, an SUV, a truck, or a van), refueling mechanism, and EPA combined mpg measure (allowing plus or minus 5 difference on the mpg level). This allowed us to match 61.8% of the CEC data. While this yield rate is not ideally high, we choose to settle for a lower yield rate in gaining a more precise matching result.

In turning our matched CEC data into BLP's required format, we took the following necessary steps. One, we defined several vehicles as being the same model in a similar fashion as BLP: if vehicles are of the same make, and their horsepower, length, width, height, and wheelbase are all within plus or minus 5% of each other, then they are counted as one model, and such vehicle's market share is calculated by the quantities of such model in each market divided by the total amount of vehicles sold in the respective market. Two, when we are combining vehicles or similar traits into one model, we select the characteristics of the cheapest vehicle. This is in line with BLP's approach, where we treat the cheapest vehicle as the baseline model. Three, because all recorded entries in CEC data set has purchased a vehicle, and our BLP Random Coefficient model actually allows for buying good 0 (i.e. not purchasing any vehicle), so to correctly adjust for such factor, we obtained statistics on city-level percentage of households with at least one vehicle (*Car Ownership* *in U.S. Cities Data and Map* 2016), and multiplied our previously calculated vehicle share data with this percentage measure. Notice that one assumption we made here is that the percentage of vehicle ownership is static over time. This is an simplification, but given the difficulty in acquiring historical data and reconstructing percentage ownership for all 24 markets, we settled with this measure for our analysis.

Some descriptive statistics on vehicle characteristics can be found in **table 1**. It is worth mentioning that due to the nature of matching data, we have substituted the vehicle purchasing price with the MSRP. While we would like to know the exact transacting cost that took place when consumer bought the vehicle, the MSRP is the best we can achieve at this point. Additionally, notice that the average EPA combined mpg measure has increased over time for both urban and non-urban markets, but the amount of change is small over the years. Non-urban respondents in general report higher mpg than urban respondents, which is expected given that non-urban consumers tend to travel at a further distance.

In terms of the consumer demographic data, we turned to Current Population Survey (CPS; Flood et al. 2018), and used corresponding responding time and geographic variables to classify respondents into each of the 24 markets. While our CEC data set also provided several consumer demographic variables associated with vehicle characteristics, we could not utilize on this demographic information since CEC's survey was only conducted in between 2015 to 2016; that is, the consumer demographic recorded in CEC data was only appropriate for markets of year 2015 and 2016. Although some consumer demographics, such as gender and race, do not change over time, other more important characteristics like income and education level do. Out of this reason, we resorted to CPS data for consumer heterogeneous taste. Some descriptive statistics on CPS data can be found in **table 2**. The data looks reasonable, as it exhibits the income decrease throughout the Great Recession, and the average level in recent years is not far off from an expected middle-class family.

The matching and cleaning of all data was completed in R, while the running of BLP Random Coefficient is done in Stata. We report our findings in the following section.

4 Results

The utility coefficients for homogeneous consumer body along with coefficients on consumer demographics is reported in **table 3**. The table is constructed by adding one additional vehicle characteristics at a time, with the exception of model (2) that adds vehicle horsepower over weight, wheelbase, and fuel tank capacity variables all at once.

Such characteristics are included in the order of perceived relevance to mpg, our main variable of interest. According to Allcott and Wozny (2014), vehicle's fuel economy is mechanically correlated with vehicle's weight and horsepower, so we created horsepower over weight variable to capture this correlation. Wheelbase captures the vehicle's weight distribution, and approximates the vehicle type. Fuel tank capacity directly relates to how much fuel (or for electric and hybrid vehicles, how much energy) the vehicle can hold. Hence, these three variables are considered as the fundamental characteristics to include in our analysis.

The other included variables are mainly aiming at reducing the amount of bias in our mpg coefficient. As part of our model's assumption, we take that cov(mpg, $\epsilon_{ijt}) = 0$. Should this be violated, the mean coefficient on mpg would be biased. The p-value change on mpg mean coefficient illustrates the existence of bias: when only adding horsepower over weight, wheelbase, and fuel tank capacity variables in model (2), the p-value on mpg is only 0.903; after including more vehicle characteristics, we managed to reduce mpg's p-value down to around 0.5. This suggests that mpg was correlated with unknowns in ϵ_{ijt}, that other factors have cast noise on mpg. Hence, if such other variables were not included in our model, then mpg is recognized as not significantly contributing to the rationalized utility $V(\cdot)$:

$$\begin{aligned}&\Pr(U_{ijt} \geq U_{ikt}, \forall k \neq j)\\&= \Pr(V_{ijt} + \epsilon_{ijt} \geq V_{ikt} + \epsilon_{ikt}, \forall k \neq j)\\&= \Pr(\epsilon_{ijt} - \epsilon_{ikt} \geq V_{ikt} - V_{ijt}, \forall k \neq j)\end{aligned}$$

($U(\cdot) = V(\cdot) + \epsilon(\cdot)$, $U(\cdot)$ being the utility level, $V(\cdot)$ is the utility explained by attributes included, and $\epsilon(\cdot)$ is the unobserved error). In this case, some characters that could be rationalized by $V(\cdot)$ are left out from our model, so the error term would be pulling too much weight on explaining choices made, which leaves us with imprecise results.

Some trends in **table 3** are worth pointing out. One, for a representative consumer in a market, the marginal utility for mpg, horsepower over weight, fuel tank capacity, and vehicle size are positive across the models, suggesting that the demand elasticities of these characteristics are positive. Similarly, demand elasticities are negative for vehicle MSRP (for 4 out of 7 models, vehicle MSRP are statistically significant at 5% significance level), wheelbase, mileage range to drive in a city, and whether the vehicle is a sports car.

Two, in terms of the demographic variables that capture consumer heterogeneous taste, we see that while their coefficient value varies across models, the variation is restricted to a fairly small range (mostly within .5 or 1), and each coefficient's sign has stabilized as more vehicle characteristics are identified. Consumer income and gender dummy both have negative coefficients, whereas education provides positive coefficient. Such observation is understandable: consumer with higher education level tends to be more forward looking, so extra education boosts consumer's marginal utility on mpg; meanwhile, consumer with higher income is less sensitive to gas price, and male consumers may be more likely to purchase sports, truck, or van vehicles that are typically low on mpg, so both measures negatively affect marginal utility.

Going back to the representative consumer, we see that demand elasticities generate mostly reasonable results. For instance, while mpg and horsepower over weight both yield positive marginal utility, mpg is of a much smaller scale. Since the marginal utility of mpg fluctuates at .2, whereas horsepower over weight at 1, after putting these variables back to its original scale, we have

$$0.05 = \frac{1/10}{.2 \times 10} \approx \frac{\frac{\partial u}{\partial hpoweight}}{\frac{\partial u}{\partial mpg}} = \frac{\partial mpg}{\partial hpoweight} = \frac{\frac{\partial mpg}{\partial price}}{\frac{\partial hpoweight}{\partial price}} \tag{20}$$

we see that for the same change in price, consumer demands much less change in mpg compared with horsepower over weight (in this specific numeric example, .05 times of changes made to horsepower over weight), suggesting that the marginal demand for mpg is much lower. This implies that for the best bang out of their bucks, consumer is more concerned with vehicle horsepower and weight ratio (which is indicative of vehicle overall performance). Such observation is in line with consumer stated preferences, and is partially the reason why the energy paradox has existed in the first place.

An analytic approach like equation 20 already indicates the limited role mpg has played in automotive market demand. A better way to illustrate this is through a WTP measure. We decide to construct the WTP density distribution using both model (6) and (7), since they are included with the most amount of vehicle characteristics that we can identify from our data. Model (6) is appealing in its lower p-value on mpg for the representative consumer, and its statistically significant coefficient on vehicle price. Model (7), while having slightly higher p-value for both mpg and vehicle price for the representative consumer, is included with all identifiable variables in our data, so the coefficient on mpg should have the smallest correlation with unobserved error ϵ_{ijt}. Hence, we decided to construct WTP measure for both.

We define the WTP measure as consumer willingness to pay for one additional mpg, holding consumer expected utility constant. Accounting for variable's scale adjustment (mpg is of units miles per 10 gallon, or mpg / 10), we can write the expected utility for vehicle with same mpg and with one additional mpg from equation 17:

$$E[u_{ijt}|mpg_{jt}, \boldsymbol{X^*_{jt}}, p_{jt}] = \hat{\beta}_{mpg_{it}} mpg_{jt} + \boldsymbol{\hat{\beta}^*_{it} X^*_{jt}} + \hat{\alpha}_{it} p_{jt} \tag{21}$$

$$E[u_{ijt}|mpg_{jt} = mpg_{jt} + .1, \boldsymbol{X^*_{jt}}, \tilde{p}_{jt}] = \hat{\beta}_{mpg_{it}} (mpg_{jt} + .1) + \boldsymbol{\hat{\beta}^*_{it} X^*_{jt}} + \hat{\alpha}_{it} \tilde{p}_{jt} \tag{22}$$

Here, the star denotes separating mpg and its coefficients from the rest; that is, $\boldsymbol{\beta_{it}} = (\beta_{mpg_{it}}, \boldsymbol{\beta^*_{it}})$, and $\boldsymbol{X'_{jt}} = (mpg_{jt}, \boldsymbol{X'^*_{jt}})'$. And $\tilde{p}_{jt}$ denotes price not necessarily the same; that is, $\tilde{p}_{jt}$ does not necessarily have to equal to p_{jt}.

Now, equation 21 represents the expected utility achieved by consumer i buying vehicle j in market t, equation 22 represents expected utility of the same consumer buying the (almost) identical vehicle at a different price $\tilde{p}_{jt}$, only that the vehicle now has one additional mpg than before. To hold the expected utility constant, we equate equation 21 with 22:

$$\begin{aligned} E[u_{ijt}|mpg_{jt}, \boldsymbol{X^*_{jt}}, p_{jt}] &= E[u_{ijt}|mpg_{jt} = mpg_{jt} + .1, \boldsymbol{X^*_{jt}}, \tilde{p}_{jt}] \\ \hat{\beta}_{mpg_{it}} mpg_{jt} + \boldsymbol{\hat{\beta}^*_{it} X^*_{jt}} + \hat{\alpha}_{it} p_{jt} &= \hat{\beta}_{mpg_{it}} (mpg_{jt} + .1) + \boldsymbol{\hat{\beta}^*_{it} X^*_{jt}} + \hat{\alpha}_{it} \tilde{p}_{jt} \\ \hat{\alpha}_{it}(\tilde{p}_{jt} - p_{jt}) &= -.1\hat{\beta}_{mpg_{jt}} \end{aligned} \tag{23}$$

Since WTP is the difference between price paid before and after the increase of mpg, we arrive at

$$WTP_{it} = \tilde{p}_{jt} - p_{jt} = \frac{-.1\hat{\beta}_{mpg_{it}}}{\hat{\alpha}_{it}} \tag{24}$$

Equation 24 tells us that, in order to calculate the WTP of consumer i to purchase one additional mpg in market t, we only need to know the estimated random coefficients for mpg and price. By equation 19, we can derive the explicit form of these random coefficients as the following:

$$\hat{\beta}_{mpg_{it}} = \hat{\beta}_{mpg} + \hat{\pi}_1 income_{it} + \hat{\pi}_2 education_{it} + \hat{\pi}_3 gender_{it} \tag{25}$$

$$\hat{\alpha}_{it} = \hat{\alpha} + \hat{\pi}_4 income_{it} + \hat{\pi}_5 education_{it} + \hat{\pi}_6 gender_{it} \tag{26}$$

Hence, after estimating the coefficients for each variable in equation 25 and 26 by running the BLP random coefficient logit model, we can randomly draw from the same consumer demographic data used for BLP estimation, and calculate $\hat{\beta}_{it}$ and $\hat{\alpha}_{it}$ for each sampled consumer i in market t. Then, using the formula in equation 24, we obtain the WTP for each drawn individual. If we draw a large size of sample from our consumer demographic data, we can thus create the empirical WTP distribution for one additional mpg.

Before diving into the resulted WTP density, we want to take a step back and consider appropriate cutoffs for evaluating our WTP distribution. To do this, we calculate the energy savings generated by one additional mpg, for an average U.S. consumer.

The U.S. Energy Information Administration reported that 391 million gallons of gasoline was used every day to power motor vehicles in 2017 (*Gasoline Is the Main U.S. Transportation Fuel* 2018). With 263.6 million of vehicles on the road in 2015, per capita gasoline consumption is about 1.48 gallon per day. For a vehicle with 28 mpg, it can travel about 41.44 miles per day, or 15,125.6 miles per year. Say the consumer travels the same distance over the year with a vehicle of one higher mpg, then such individual only need 1.42 gallon of gasoline per day. The reduction is small on a daily scale, but over the year with average gas price as about \$2.5 in 2017 (*How Do Recent Gasoline Prices Compare With Historical Prices?* 2018), this will save such average consumer \$54.75 per year. Given that an average consumer keeps using their vehicle for about 11.6 years, the savings amount to \$635.1 over the lifetime. If we take into account that the \$2.5 per gallon gas cost is at a relatively cheap point in the past 10 years, then with a higher gas per gallon price (about \$3.5 per gallon in early 2010s), average consumer's lifetime savings would amount to \$889.14.

Based on these empirical statistics about the U.S. population, we round the figures and use \$600 and \$850 as two cutoff values (in consideration of consumer discounting over time). That is, after drawing the WTP density distribution from our model, we will evaluate the proportion of consumers that have WTP higher than \$600 and \$850. If the proportion is low, then we can conclude that fuel economy does not place strong influence on automotive market demand, and hence one way to explain the energy paradox is consumer's undervaluation on fuel efficiency.

The resulting WTP density graphs can be found in the figures section. **Figure 1** is generated using model (6), and **figure 2** for model (7). The WTP density from both models paint a somewhat similar picture. In model (6), about 29.2% of the sample population are willing to pay more than \$600 for one additional mpg; 22.8% are willing to pay more than \$850. In model (7), the figures are 29.2% for more than \$600, and 26.2% for more than \$850. There also exist a significant amount of consumers with negative WTP in both models: 40.8% in model (6), and 61.7% in model (7). We will expand more on the negative WTP in section 5.3.

Our result shows that only roughly 30% of the sample population are willing to pay a reasonable amount (\$600) or higher for one additional mpg. This leaves the majority of our sample (about 70%) unwilling to invest in an additional mpg that should generate savings in the future, suggesting that the majority of Californian residents who purchase home-use vehicles are undervaluing fuel economy. Hence, we conclude that energy efficiency has little positive effects on automotive market demand in the state of California.

5 Discussions

Our result shows some evidence regarding consumer undervaluation on the energy efficiency factor, but several steps taken along the path might sway our conclusion.

5.1 Data Cleaning and Merging

Back in section 3, we mentioned that only 61.3% of the CEC data was able to be matched with vehicle characteristics. We adopted a strict matching scheme, meaning that not only do we consider vehicle make, type, model year, and refueling mechanism, but we allow for plus or minus 5 reported mpg error in order to at best effort correctly identify each vehicle in our data.

This, however, does omit a good number of vehicles that could not be matched up with the combined mpg value in our reference data set. When the mpg matching is ignored, 75.2% of the original CEC data set was able to be matched. This creates a gap of 850 vehicles which, while were able to be identified through basic vehicle characteristics, have incorrect mpg values reported by consumers. The remaining 24.8% of the data have other matching problem. For instance, if consumer did not report vehicle make, type, model year, or refueling mechanism, we cannot use these characteristics to identify vehicles in our matching data set. A small number of respondents (40) also reported illogical vehicle model characteristic. If a vehicle is purchased in 2000, its model year is usually of 2001 or lower, at best 2003 (companies like Audi tend to grant their advanced vehicles with a future model year); however, such 40 respondents reported vehicle model year much higher than the year of purchase, which cannot be identified in our matching data set. Such respondents might have the purchase year and model year mixed up, or they recorded the numbers in their vehicle name as model year. Either way, incorrect survey responses led us to a relatively low matching rate, which could have skewed the composition of vehicles in our automotive market.

Another issue arose when we grouped vehicles with similar traits as one model for the purpose of BLP analysis. This is the step where we specified the instruments used for each vehicle model by summing up vehicle characteristics of competitor's models and its own manufacturer's other models. Due to the nature of sampling survey and micro data set, about 4.8% (104 in number) out of the grouped vehicle models does not have either same type of vehicles made by competing companies, or any other vehicles made by the same manufacturer. And among the two scenarios, the latter is more likely to occur. While the number of models we omitted in this step is not huge, but compounding with the amount we omitted during the characteristics matching step, this could make our sample less representative of the actual Californian automotive market.

In terms of generalizing our result to the rest of the U.S., we should also consider the representative of Californian automotive market. California is one of the few states in the U.S. that have stricter state-wide mpg standards than EPA's guideline (Evarts 2018), so it is possible that the result we see, while still demonstrating the undervaluation of fuel economy, is leaning on the better side of the scale.

5.2 Modeling

Among the seven models presented, model (6) and (7) should report the least biased coefficients on mpg due to the number of vehicle characteristics included. However, both have shortcomings as almost none of the vehicle characteristics reached 5 or 10 percent statistical significance on their mean utility level. This contradicts with numbers of research done in the field, where statistical significance is usually found on the average level.

As discussed in subsection 5.1, the various problems resulted from data matching have altered the shape of markets (omitting unmatchable responses and responses without appropriate instruments), which could have made our vehicle characteristics not significant for the market representative consumer.

Another theory is that some other vehicle characteristics are correlated with existing attributes, but we failed to include these due to the limitation of our referencing data set. This could leave the error term with too much weight in explaining the discrete choice made, and thus distorted the mean coefficients on vehicle characteristic terms. For instance, the original BLP paper has included an "air" dummy to record whether the vehicle has an air conditioning system. As the operation of air conditioning depends on engine power, which further depends on the amount of fuel (or electricity for electric vehicles) holds, having air conditioning could affect how much fuel can be used on the road. Although in recent years, air conditioning have become more as a standard feature of vehicles, other in-car systems – including high-tech navigation, internet connection, and entertainment systems – all rely on engine power. With the rise of smart devices and the increasing attention given to in-car systems, such characteristics could affect vehicle fuel efficiency and might have played a more crucial role in automotive market demand. Hence, our inability to include these traits might explain the statistically insignificant results we got.

A more prominent modelling problem is regarding the consumer demographic variables. We have chosen income, years of education, and gender as accounting for consumer heterogeneous taste, but more demographics could be included to see the effect on our result. The reason why we used these three is due to their more prominent role in affecting any demand market, and a strict restriction imposed by the amount of instruments available. For one additional consumer demographic we include, we need a minimum of $K+1$ instruments to keep our model identified. This could be challenging if we have no access to direct cost shifting instruments, or when we have more vehicle characteristics to include. And the more factors on consumer heterogeneity we consider, the longer it takes the BLP random coefficient logit to converge in its GMM objective function. If researchers can obtain more characteristics or cost shifters, the validity of including only three consumer demographics should be evaluated.

5.3 WTP Distribution

The empirical WTP distributions we generated in figure 1 and 2 have two interesting features. One, while figure 1 seems to have its distribution abruptly cut off at around $700, there actually exists draws beyond the $15,000 x-axis limit we set up. Such draws have extremely high WTP for one additional WTP, up to $70,000. This distribution is also in contrast with figure 2, which has WTP values more spread out. This could be the fact that model (6), while having more precise estimates, still suffers from bias resulted from omitting variables. Combined with the limitations mentioned in section 5.2, future research should attempt to recreate WTP distribution after further reducing bias on coefficients.

Two, the WTP distribution we created allows for negative WTP values. For such people, having one more mpg is so unpleasant that they have to be compensated for it. As we see in some electric or hybrid vehicles, these highly fuel-efficient cars tend to have strange designs. There does exist subsidy programs that incentivize people for buying these peculiarly designed vehicles, so the negative WTP can be rationalized in this way. But more conventionally, we would expect WTP to be a non-negative measure, or at least the negative values would not be as large as the -$10,000 we get. One potential explanation of our extremely negative WTP could be the biased coefficients we estimated, so future research should see if the distribution changes when reducing the omitted variable bias. Another way to address this issue is by forcing WTP into non-negative values: that is, if the WTP is calculated to be negative, we reassign its value to 0. This can also be attempted in future research to see how much the restriction changes the shape of WTP distribution.

6 Conclusion

In this paper, we attempted to investigate whether fuel efficiency has significantly affected Californian automotive market in the span of 2005 to 2016 for the purpose of studying the energy paradox. Through constructing an empirical distribution on consumer WTP for one additional mpg, we found that the majority of consumers (roughly 70%) are unwilling to pay an reasonable amount, suggesting the paradox. Discussions are given on issues related to data cleaning, merging, and modelling decisions, which all could impact the validity of our result. Some approaches on improving validity and reducing estimation bias have been introduced in section 5. We can also verify our result by using aggregate-level market data or other sources of data in the same region from the same period, and then comparing discrepancy in the WTP distribution created. There are obviously many other research that can still be done in this field, and hopefully our suggestions can lead to a better understanding on the energy paradox.

References

2010 Census (2010). United States Census Bureau.

2015-2017 California Vehicle Survey Consultant Report (2018). California Energy Commission.

Allcott, Hunt and Nathan Wozny (2014). "Gasoline Prices, Fuel Economy, and the Energy Paradox". In: *Review of Economics and Statistics* 96.5, pp. 779–795.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995). "Automobile Prices in Market Equilibrium". In: *Econometrica: Journal of the Econometric Society*, pp. 841–890.

— (2004). "Differentiated Products Demand Systems From a Combination of Micro and Macro Data: The New Car Market". In: *Journal of political Economy* 112.1, pp. 68–105.

Brownstone, David, David S Bunch, and Kenneth Train (2000). "Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-Fuel Vehicles". In: *Transportation Research Part B: Methodological* 34.5, pp. 315–338.

Car Ownership in U.S. Cities Data and Map (2016). Tech. rep. Governing.com. URL: `http://www.governing.com/gov-data/car-ownership-numbers-of-vehicles-by-city-map.html`.

EPA (2018). *Light-Duty Automotive Technology, Carbon Dioxide Emissions, and Fuel Economy Trends: 1975 Through 2017*. Tech. rep. United States Environmental Protection Agency. URL: `https://nepis.epa.gov/Exe/ZyPDF.cgi?Dockey=P100TGLC.pdf`.

Evarts, Eric (2018). *California Warns It Won't Follow Lower EPA Fuel Economy, Emissions Rules*. Tech. rep. Green Car Reports. URL: `https://www.greencarreports.com/news/1118136_california-warns-it-wont-follow-lower-epa-fuel-economy-emissions-rules`.

Flood, Sarah et al. (2018). *Integrated Public Use Microdata Series, Current Population Survey: Version 6.0 [Dataset]*. Tech. rep. DOI: `https://doi.org/10.18128/D030.V6.0`.

Gasoline Is the Main U.S. Transportation Fuel (2018). Tech. rep. U.S. Energy Information Administration. URL: `https://www.eia.gov/energyexplained/index.php?page=gasoline_use`.

Goldberg, Pinelopi Koujianou (1995). "Product Differentiation and Oligopoly in International Markets: The Case of the US Automobile Industry". In: *Econometrica: Journal of the Econometric Society*, pp. 891–951.

How Consumers Value Fuel Economy: A Literature Review (2010). Research rep. United States Environmental Protection Agency.

How Do Recent Gasoline Prices Compare With Historical Prices? (2018). Tech. rep. U.S. Energy Information Administration. URL: https://www.eia.gov/energyexplained/index.php?page=gasoline_prices.

Jaffe, Adam B and Robert N Stavins (1994). "The Energy Paradox and the Diffusion of Conservation Technology". In: *Resource and Energy Economics* 16.2, pp. 91–122.

Vincent, David (2015). "The Berry-Levinsohn-Pakes Estimator of the Random-Coefficients Logit Demand Model". In: *The Stata Journal* 15.3, pp. 854–880.

Tables

Year	In Urban?	Number of Models	Total Number of Vehicles Sold	Avg. Price (In 2018 $, Per $1000)	Avg. mpg (miles / 10 gallon)	Avg. Horsepower Over Curb Weight (10 hp / lb)	Avg. Wheelbase (in / 100)	Avg. Fuel Tank Capacity (gal / 10)	Avg. Range to Drive in City (mi / 100)	Avg. rpm (rpm / 1000)	Avg. Vehicle Size (sq in / 10000)	Proportion of Sports Vehicle	Proportion of European Vehicle
2000	0	4	4	25.02	2.03	0.54	1.08	2.05	3.17	0.03	1.34	0.00	0.00
2000	1	2	2	34.05	1.90	0.59	1.05	2.05	3.16	0.03	1.31	0.00	0.00
2001	0	21	25	18.77	2.22	0.52	1.04	1.67	3.24	0.03	1.25	0.04	0.04
2001	1	20	25	18.99	2.00	0.54	1.05	1.72	3.08	0.03	1.26	0.00	0.04
2002	0	32	44	19.51	2.20	0.54	1.06	1.71	3.20	0.03	1.28	0.02	0.05
2002	1	29	41	21.27	2.16	0.53	1.06	1.70	3.18	0.03	1.29	0.10	0.07
2003	0	37	61	19.45	2.23	0.53	1.04	1.62	3.15	0.03	1.24	0.03	0.08
2003	1	28	45	20.86	2.25	0.55	1.06	1.67	3.27	0.03	1.28	0.04	0.09
2004	0	47	62	19.65	2.16	0.52	1.07	1.67	3.17	0.03	1.28	0.03	0.08
2004	1	37	64	21.40	2.30	0.52	1.07	1.68	3.28	0.03	1.29	0.05	0.08
2005	0	50	82	22.65	2.37	0.54	1.07	1.72	3.34	0.03	1.30	0.00	0.12
2005	1	51	85	20.95	2.34	0.53	1.07	1.71	3.38	0.03	1.28	0.02	0.09
2006	0	55	90	22.23	2.43	0.51	1.07	1.68	3.45	0.03	1.29	0.01	0.04
2006	1	54	87	20.94	2.47	0.54	1.05	1.62	3.46	0.03	1.25	0.01	0.03
2007	0	59	111	21.35	2.57	0.54	1.06	1.58	3.48	0.03	1.26	0.03	0.09
2007	1	55	95	20.69	2.51	0.54	1.05	1.52	3.23	0.03	1.23	0.00	0.12
2008	0	48	86	20.81	2.52	0.54	1.06	1.58	3.37	0.03	1.26	0.01	0.06
2008	1	52	83	20.39	2.43	0.53	1.05	1.57	3.29	0.03	1.25	0.01	0.07
2009	0	57	100	21.52	2.67	0.53	1.07	1.59	3.51	0.04	1.26	0.01	0.07
2009	1	47	87	20.65	2.58	0.55	1.05	1.55	3.37	0.03	1.23	0.00	0.09
2010	0	71	131	22.95	2.60	0.54	1.07	1.64	3.75	0.04	1.27	0.00	0.12
2010	1	62	115	21.81	2.60	0.55	1.07	1.66	3.67	0.04	1.28	0.00	0.04
2011	0	67	106	24.03	2.57	0.56	1.07	1.63	3.55	0.04	1.29	0.01	0.08
2011	1	58	91	23.10	2.50	0.57	1.06	1.56	3.46	0.03	1.26	0.02	0.11
2012	0	97	174	22.76	2.55	0.56	1.06	1.59	3.52	0.04	1.27	0.00	0.09
2012	1	90	159	24.06	2.49	0.57	1.06	1.62	3.48	0.03	1.28	0.01	0.12
2013	0	101	214	23.75	2.68	0.55	1.06	1.60	3.62	0.04	1.29	0.01	0.07
2013	1	84	173	23.91	2.71	0.58	1.06	1.55	3.50	0.03	1.27	0.01	0.10
2014	0	109	244	24.92	3.17	0.56	1.07	1.56	3.52	0.04	1.28	0.01	0.09
2014	1	116	238	25.60	2.86	0.57	1.06	1.55	3.77	0.04	1.28	0.01	0.10
2015	0	127	275	27.25	2.87	0.59	1.08	1.60	3.53	0.04	1.31	0.02	0.12
2015	1	136	298	27.37	2.79	0.58	1.06	1.55	3.56	0.04	1.28	0.01	0.13
2016	0	125	234	25.95	2.59	0.57	1.08	1.65	3.63	0.04	1.32	0.06	0.12
2016	1	122	231	26.01	2.86	0.57	1.08	1.54	3.79	0.04	1.30	0.02	0.12

Table 1: Descriptive statistics of vehicle characteristics

Year	In Urban?	Sample Size	Real Annual Income (in 2018 $, per $)				Years of Education (years)				Gender (0 = female, 1 = male)			
			Mean	Median	Min	Max	Mean	Median	Min	Max	Mean	Median	Min	Max
2005	0	6529	43229.89	30815.90	1037.88	280229.92	12.78	13	2.5	20	0.51	1	0	1
2005	1	2445	39404.82	26870.23	1109.37	292751.04	12.62	13	2.5	20	0.52	1	0	1
2006	0	7638	43214.43	29892.80	1004.77	298994.73	12.78	13	2.5	20	0.51	1	0	1
2006	1	2622	40524.02	28009.60	1039.33	291995.90	12.68	13	2.5	20	0.52	1	0	1
2007	0	7627	44039.44	31306.42	1008.73	283088.82	12.82	13	2.5	20	0.51	1	0	1
2007	1	2539	41299.94	28986.35	1042.35	297162.27	12.61	13	2.5	20	0.52	1	0	1
2008	0	7675	42979.50	31892.68	1003.62	296002.00	12.88	13	2.5	20	0.51	1	0	1
2008	1	2714	40446.55	27878.22	1003.62	266090.91	12.85	13	2.5	20	0.51	1	0	1
2009	0	7960	42701.55	31363.64	1008.12	291158.72	12.93	13	2.5	20	0.51	1	0	1
2009	1	2638	40272.30	27057.86	1008.12	262558.45	12.76	13	2.5	20	0.51	1	0	1
2010	0	7907	41728.00	30164.37	1025.01	293785.29	12.99	13	2.5	20	0.51	1	0	1
2010	1	2793	38368.71	26731.25	1095.09	290516.43	12.93	13	2.5	20	0.50	1	0	1
2011	0	7828	40888.33	28812.86	1013.79	299621.77	12.95	13	2.5	20	0.51	1	0	1
2011	1	2666	38323.97	26635.89	1067.14	298800.07	12.90	13	2.5	20	0.51	1	0	1
2012	0	7854	39923.14	27047.13	1013.23	299662.40	12.95	13	2.5	20	0.51	1	0	1
2012	1	2634	37891.63	26006.85	1015.31	296910.88	13.12	13	2.5	20	0.51	1	0	1
2013	0	8157	40693.69	27667.21	1024.71	297284.13	13.08	13	2.5	20	0.51	1	0	1
2013	1	2811	40810.72	25617.78	1004.22	298526.08	13.36	13	2.5	20	0.50	1	0	1
2014	0	8525	41309.17	28446.92	1008.76	294211.72	13.17	13	2.5	20	0.51	1	0	1
2014	1	2436	39737.00	25218.90	1008.76	296069.85	13.38	13	2.5	20	0.49	0	0	1
2015	0	9698	41945.34	28898.33	1007.75	285200.50	13.11	13	2.5	20	0.51	1	0	1
2015	1	566	48358.69	36315.21	1008.76	248355.70	14.16	14	2.5	20	0.50	1	0	1
2016	0	7839	43223.14	30000.00	1001.00	291122.00	13.24	13	2.5	20	0.52	1	0	1
2016	1	856	52395.62	37000.00	1200.00	291308.00	13.95	14	2.5	20	0.49	0	0	1

Table 2: Descriptive statistics of consumer demographics, obtained from IPUMS-CPS

model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
mpg	.059 (.058)	.036 (.291)	.148 (.681)	.158 (.285)	.225 (.333)	.187 (.249)	.325 (.585)
realprice	-.015*** (.005)	-.045 (.031)	-.057*** (.027)	-.056*** (.019)	-.052 (.036)	-.057*** (.028)	-.039 (.044)
hpoweight		1.751 (2.112)	1.793 (1.598)	1.509** (.812)	1.227 (2.681)	1.239 (1.443)	.998 (1.489)
wheelbase		.091 (1.626)	-.061 (1.419)	-.057 (1.117)	-1.151 (3.906)	-1.262 (2.460)	-2.060 (6.589)
fueltank		.201 (.459)	.503 (1.299)	.558 (.451)	.210 (1.122)	.338 (.729)	.287 (.800)
cityrange			-.190 (.641)	-.202 (.214)	-.186 (.181)	-.211 (.272)	-.193 (.443)
rpm				.094** (.054)	.010 (.220)	.069 (.255)	-.063 (.392)
size					1.168 (2.600)	.837 (1.477)	1.083 (3.147)
typesports						-5.681 (43.469)	-1.726 (11.596)
european							-4.625 (3.540)
mpg p-value	0.304	0.903	0.828	0.579	0.500	0.453	0.578
realprice p-value	0.003	0.151	0.038	0.004	0.153	0.046	0.365
income	106080.3*** (3.877)	.206 (2.382)	.104 (2.717)	.079 (1.759)	-.458 (2.621)	-.560 (2.116)	-1.266 (3.235)
education	82698*** (2.451)	-.463 (1.890)	.179 (4.178)	.209 (2.317)	.853 (3.410)	1.170 (.888)	1.783 (2.758)
gender	83480.79*** (7.112)	.204 (3.443)	-.026 (11.892)	-.057 (5.446)	-.470 (4.652)	-.032 (1.453)	-.671 (5.616)

Table 3: Results from running BLP Random Coefficient Logit.

Notes:

1. Standard errors are reported in the parentheses.
2. *** indicates statistical significance at 5% significance level.
3. ** indicates statistical significance at 10% significance level.
4. The first section of the table reports mean coefficients ($\boldsymbol{\beta}$, α) on mpg, $realprice$, $hpoweight$, $wheelbase$, $fueltank$, $cityrange$, rpm, $size$, $typesports$, $european$.
5. The second section of the table reports p-values on β_{mpg} and α (i.e. on the representative consumer).
6. The third section of the table reports coefficients for $income$, $education$, $gender$ consumer demographics for $\beta_{mpg_{jt}}$.

Figures

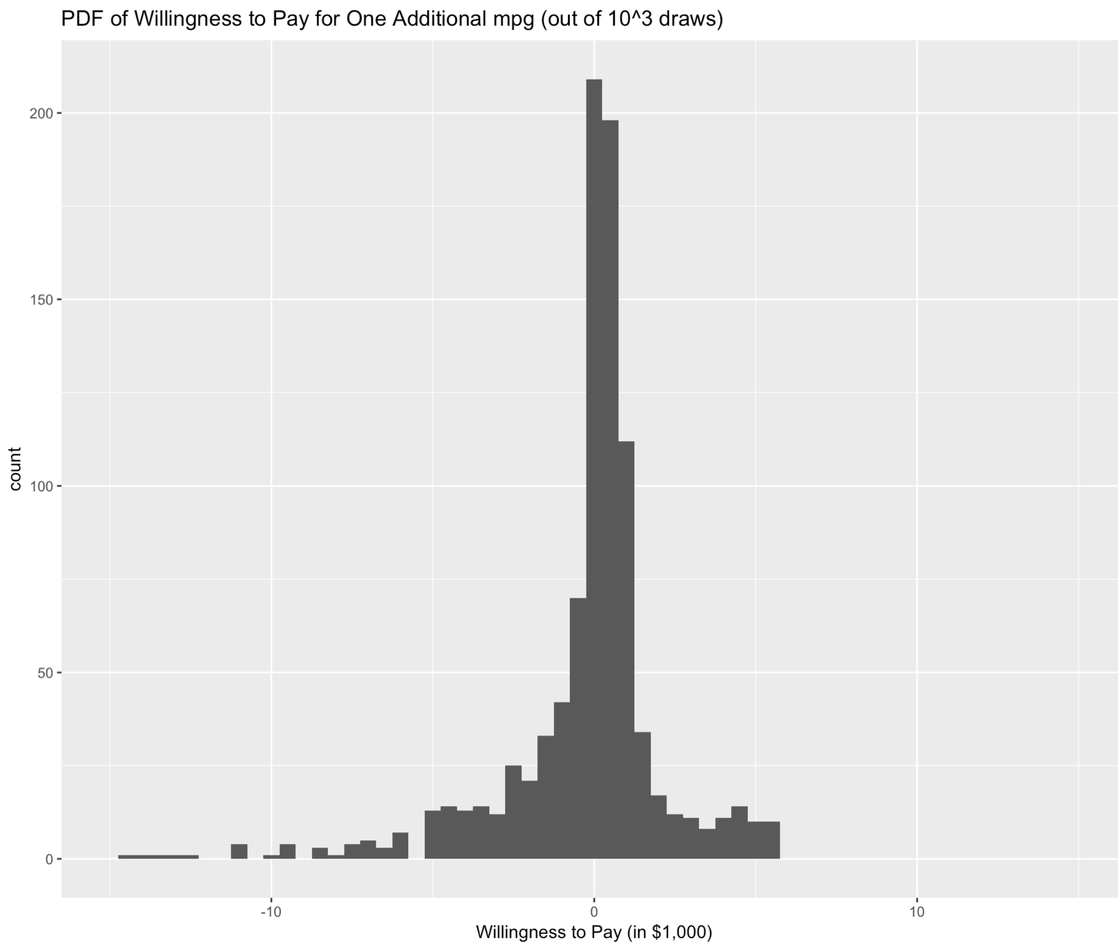

Figure 1: PDF of WTP for one additional mpg, constructed using model (6) (sample draws refer to the CPS data)

Notes:

1. The graph restricts x axis to be between -15 to 15.

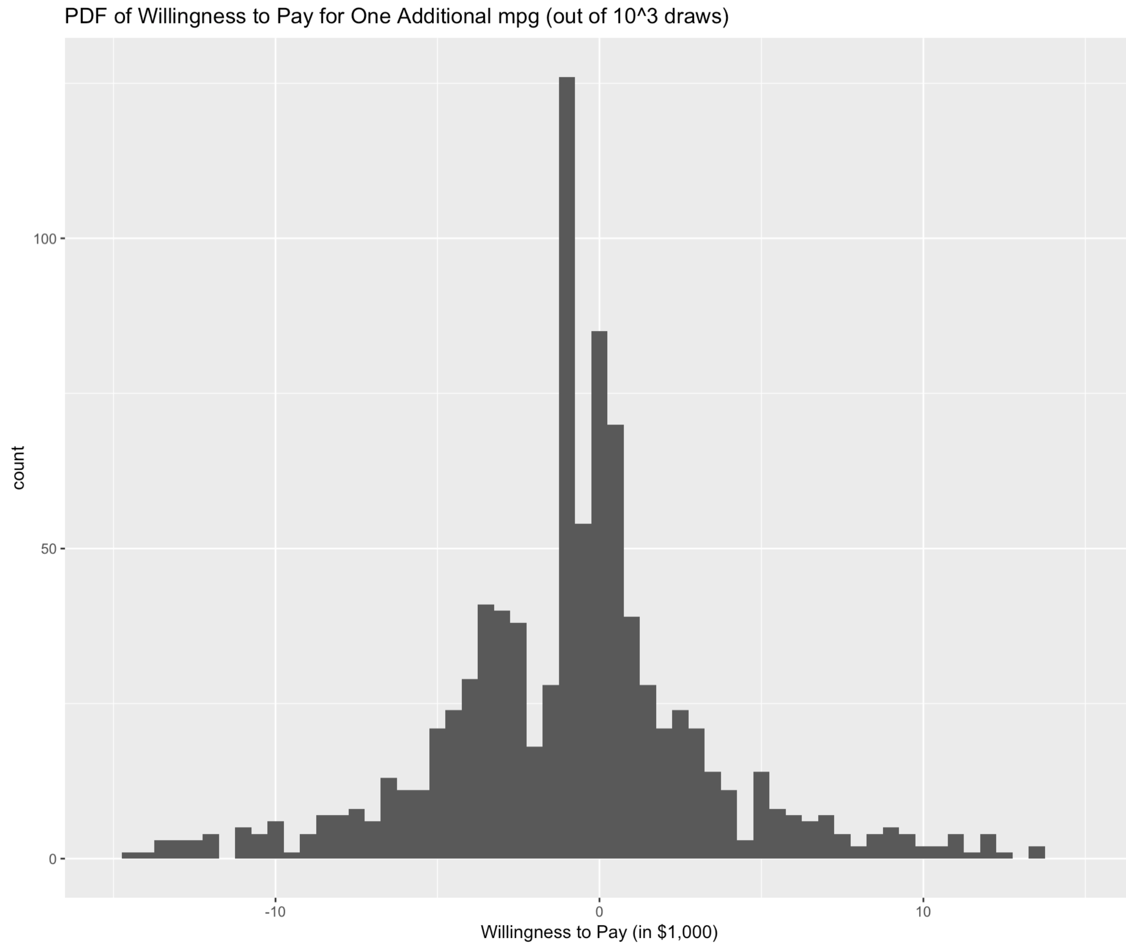

Figure 2: PDF of WTP for one additional mpg, constructed using model (7) (sample draws refer to the CPS data)

Notes:

1. The graph restricts x axis to be between -15 to 15.