How Gasoline Prices Impact Household Driving and Auto Purchasing Decisions: A Revealed Preference Approach

Elisheba Spiller
Duke University
PhD Candidate

October 2010

This research is supported by Resources for the Future’s Joseph L. Fisher Doctoral Dissertation Fellowship 2010-2011
Abstract

My paper seeks to identify how consumers respond to changes in gasoline prices and thus to evaluate the impact and effectiveness of different environmental policies and models. Understanding how gasoline prices affect consumer behavior—that is, correctly identifying the elasticity of the demand for gasoline—is fundamental to choosing optimal environmental policies, and has important policy implications.

While many prior micro-economic studies have found relatively inelastic gasoline demands, this does not seem to be the case, at least anecdotally. Over the last two years, as gasoline prices began to sharply fluctuate, we have witnessed a large shift in consumer behavior with regards to vehicle purchase and driving decisions. By improving the assumptions and methods used, I have demonstrated how to minimize the downward bias on elasticity of demand for gasoline.

Most prior research has aggregated the choice set in order to make estimation more feasible, yet this creates a downward bias in the calculated elasticity, since more subtle changes along the extensive margin are ignored by this aggregation. In order to maintain a large choice set, I use a revealed preference approach where a household chooses the vehicles in its garage only if that vehicle bundle gives the household higher utility than other bundles. The estimation technique uses two steps of binomial probits where each household’s utility is evaluated at random unilateral deviations from their observed vehicle purchases.

I also allow for vehicle fixed effects in order to capture unobserved heterogeneity in vehicle purchase decisions. Through the use of revealed preference, I am able to implement a swapping estimator that allows me to difference out the fixed effects in the first stage and recover them later, in order to facilitate estimation.

Finally, I model purchase and vehicle miles travelled (VMT) decisions of each vehicle in the household’s garage as dependant on the full stock of vehicles owned by the household. The decision for use is inseparable from the household’s bundle of vehicles- how relatively efficient each of the vehicles in their bundle is compels how much the household chooses to drive each. Modeling these decisions as independent biases down the elasticity of demand, as short-run shifts in VMT within a household’s garage are not captured.

My research intends to gain a more accurate understanding of how households make vehicle bundle purchase and driving decisions, and how they respond to changes in gasoline prices. By using a discrete-continuous household model, not artificially restricting the choice set, and by modeling dependent household decisions at the bundle level, I have minimized the bias on the elasticity of demand for gasoline.
1. Introduction
Externalities such as air pollution, climate change, and national security risks are all associated with the consumption of gasoline above a socially optimal level. Current US policies aimed at minimizing gasoline consumption have focused on mandates for emissions, such as CAFE (Corporate Average Fuel Efficiency) standards, or gasoline refinement guidelines. Other possible solutions, such as emissions taxes or gasoline taxes, have not been implemented widely because of insufficient technologies (no dashboard technology exists to measure each car’s emissions output) and lack of political will to apply taxes. Yet this does not mean that command and control mandates are more effective than their market-based alternatives in achieving this goal. In order to be able to identify the most effective policy, it is important to understand exactly how consumer demand responds to different operating cost changes.

Correctly identifying the elasticity of demand for gasoline is essential for these policy decisions. Depending on how elastic consumer demand is, one set of policies may make more sense than another: if the demand for gasoline is elastic, higher gas taxes will reduce gasoline consumption without imposing extra costs on the production of automobiles. But if the demand is inelastic, higher gasoline taxes will not affect demand, and thus a different policy should be devised. On the other hand, if the goal of the gasoline tax is to balance net costs and benefits to society (as opposed to diminishing gasoline use), then a higher elasticity creates greater distortions and results in lower revenues to the government. There are other policy issues that depend on the correct measurement of the elasticity of demand for gasoline. For example, the elasticity of demand is used as a critical input in certain climate policy models. Thus, if it is measured incorrectly, it will lead to unintended consequences that were mis-forecast. Furthermore, many policy makers place importance on the economic burden or tax incidence of certain carbon or gasoline taxes- if the demand is elastic, then the burden falls more on the producer than on the consumer. In sum, the elasticity of demand for gasoline must be correctly measured in order to

---

1 Policy makers tend to shy away from gasoline taxes due to lack of public support for increased taxes, as well as the idea that these types of taxes can be regressive.
create optimal environmental policies that impact the price of gasoline. Through my research, I approach this issue by assessing the impact of gasoline prices on consumer behavior.

Prior micro-economic studies have found a wide range of gasoline demand elasticities, ranging from -0.02 to -1.59\(^3\), although newer studies (post 1990s) have found less elastic demands. However, two things call that finding into question. One, anecdotal evidence suggests that consumers actually do change their car purchasing behavior as gasoline prices rise. Even before gasoline prices spiked between 2007 and 2008, nearly 50% of car buyers claimed that gasoline prices were affecting their purchasing decisions (Kelley Blue Book 2004). Furthermore, McManus (2007) and Woodyard (2009) demonstrate that the price of fuel-inefficient vehicles was negatively correlated with the price of gasoline during the period of high gasoline prices. Vehicle miles travelled was also affected by higher gasoline prices, as demonstrated in Figure 1 – when gasoline prices spike in 2008, there is a clear sudden drop in VMT. While this picture does not demonstrate directly the elasticity of demand for gasoline, it does imply that individuals are more responsive to gasoline prices than most studies would suggest.

\(^3\) See Espey (1996), Dahl and Sterner (1991) for surveys on elasticity measurements.
Second, and even more important, the economic studies that have measured the elasticity of demand for gasoline have approached this problem in a misspecified way that may bias their results. Some of these problematic methods include not taking into account households’ fleet of vehicles, not modeling the car choice jointly with the vehicle use decision, and aggregating the choice set to make estimation easier. My model and methods better reflect the household’s vehicle driving and purchase decisions and thus more accurately capture the elasticity of demand for gasoline.

First, my model takes into account not only how much people drive (the intensive margin) but also what kinds of automobiles they decide to purchase (the extensive margin). These two margins are intimately related: how many miles we need to drive influences the type of car we buy, and the type of car we own will also affect how much we drive it. If these two decisions are not modeled jointly, the elasticity of demand for gasoline may be underestimated. For example, a household with a long commute time may choose a larger, more comfortable vehicle. Yet, because the vehicle is comfortable, it may induce the household to drive it more frequently. If
these two margins are modeled separately, then this household would appear to be less sensitive to gasoline price than is the case. Even if this does not bias the elasticity, modeling these margins separately does not allow the researcher to capture exactly how the household makes decisions given a change in gasoline price. My model allows the household to maximize its utility by choosing the number of miles to drive each car, given the type (or types) of car that it owns, and allows for a simultaneous estimation of both these margins.

Second, I allow the household to have a “fleet” of vehicles (i.e. all the vehicles in the garage) from which it optimally allocates vehicle miles travelled (VMT) to each vehicle. The relative efficiency of the vehicles in a household’s fleet affects the household’s allocation of VMT. For example, if a household owns an SUV and a car, an increase in gas prices would presumably result in a shift of VMT from the SUV to the car. However, if these two vehicles were treated as independent, then the researcher may interpret the increase in the car’s VMT as a household that is insensitive to gasoline price changes. Thus, modeling the VMT choices within a fleet as independent decisions may create a downward bias on the elasticity of demand for gasoline-while the household may not be able to change its purchasing decision in the short run, it can easily shift its driving behavior within its fleet in the medium-run. That is, by allowing households to substitute between the vehicles in their garage as relative operating costs change, I am allowing for movement along another margin that many researchers disregard.

Third, I include in my specification vehicle specific fixed effects. Adding these alternative specific constants allows me to capture attributes of the cars that are unobserved by the econometrician, but are observable to both the producer and the consumer. There may be factors (such as brand style) that increase the price of the car, which, if ignored in estimation, make it seem that consumers value high prices or are less price sensitive. Berry et al. (1995) (from here out BLP) stress that not including product-specific fixed effects can cause a downward bias on price:

If producers know the values of \( \xi \) (the unobserved car attribute)...then prices are likely correlated with them. (There exists) nonzero disturbance associated with unobserved determinants of demand that are correlated across consumers in a market. If these disturbances are known to the
Thus, these constants capture unobserved attributes that impact the vehicle purchasing decision, thereby minimizing the bias on the coefficient of price.

Finally, my model expands the choice set to as large as the number of vehicle model-year combinations that are present in my dataset. Most researchers choose to aggregate the choice set into broad categories such as SUV, van, truck, or car in order to facilitate estimation. However, this does not allow the researcher to register a change within a category (such as from a less fuel efficient larger car to a more efficient smaller car) as an actual change in consumer behavior. This means that the changes along the extensive margin must be much larger in order to capture adjustments in the vehicle purchasing decision, creating a downward bias in the elasticity of demand for gasoline. Thus, by expanding the choice set to all model-year combinations within my dataset, I am able to capture more subtle changes in consumer behavior along the extensive margin.

However, expanding the choice set generates a significant problem in estimation. If households can choose up to 3 cars, and my dataset includes 3,751 distinct car-year choices, this means that the household has \( \binom{n}{k} \) choices = 8,789,061,875 possible different car combinations that they can have in their garage. Any conventional discrete choice model would be very difficult to estimate with such a large choice set. Moreover, since I am incorporating alternative specific constants, this would imply that I would have to estimate 2,418 extra parameters, further increasing the difficulty of the problem. Other techniques, such as the contraction mapping used in BLP, would not work very well under these huge choice sets, as many of the bundle choices would have practically zero market shares. Therefore, I use a method that simultaneously allows me to do estimation over a choice set of any size, as well as eliminate the need to estimate the fixed effects through differencing.

The remainder of this paper is organized as follows. Section 2 provides the literature review; Section 3 describes the model; Section 4 demonstrates the estimation procedure; Section 5
presents the data; Section 6 discusses the results; Section 7 discusses some policy outcomes from the bias in the elasticity estimate; and Section 8 concludes.

2. Literature Review

Many prior studies have estimated separately either the demand for vehicles, or the demand for gasoline (indirectly) through vehicle miles travelled (VMT). Among them are Schmalensee and Stoker (1999), West and Williams (2004), Petrin (2002), and West (2007). None of these papers incorporate the information that is present along both margins. As West (2007) writes, “In order to accurately estimate vehicle choice probabilities, one must estimate the joint choice of vehicle and miles traveled (p. 12).” My research, on the other hand, models both these margins jointly, thus more accurately capturing household behavior when faced with changing gasoline prices.

More immediately relevant to my proposed study are the studies by West (2004), Mannering and Winston (1985), Goldberg (1998), and Berkowitz et al. (1990). As does my study, these four papers jointly model gasoline demand and car choice. However, these studies adapt a method found in Dubin and McFadden (1984), which introduces a joint demand system for energy use and the demand for durable goods. Their method is a two-step, sequential method that first estimates the purchase decision (intensive margin) in a nested logit model, and then uses these parameters to model the energy use decision (extensive margin). While this approach imposes correlation between the extensive and intensive margin, all these researchers have aggregated the choice set into very broad categories. Furthermore, only Berkowitz et al. allow for interdependence among the vehicles in the household’s garage. In any case, they estimate an

\[ \text{4 Schmalensee and Stoker estimate an elasticity range of } -0.29 \text{ to } -1.13; \text{ West and Williams find a range of } -0.18 \text{ to } -0.738 \text{ depending on income: upper-income quintiles are less responsive to gasoline price changes than lower-income quintiles; Petrin measures the impact of the introduction of the minivan on the overall demand for vehicles, though gasoline price elasticity does not factor into his measurement; and West (2007) does not measure the elasticity of demand for gasoline, but does find that higher gasoline prices lead to a shift in consumer purchases towards more fuel-efficient vehicles.} \]
inelastic demand for gasoline, which may be due in part to their restrictive choice set assumptions\(^5\).

There have been some papers that use a simultaneous, one-step approach to estimating both VMT and car purchase decisions, yet they also find that demand for gasoline is relatively inelastic. Feng, Fullerton and Gan (2005) estimate VMT and a nested logit purchase probability in a simultaneous maximization step. Furthermore, they also look at the household fleet as arising from a jointly determined decision process. However, they only incorporate the probabilities of choosing either an SUV or a car, which likely leads to a very inelastic demand for gasoline between -0.024 and -0.07. Bento et al. (2008) also employ a utility maximization framework through the use of a random-parameter Bayesian estimation technique. However, they assume that each car is independent from all other cars in the household’s fleet—an assumption that also may bias their result, which is an inelastic demand for gasoline of -0.35. In my study, I do not assume that the cars in a household’s garage were chosen independently of one another, and I also allow for a large and much more detailed choice set. By incorporating these measures, I am able to capture a more realistic elasticity of demand.

While most papers do not take into account the household’s fleet and thus model car choices as independent, Green and Hu (1985) estimate a fleet model, where they look at how households substitute between their cars when gasoline prices change. However, they only allow households to choose between trucks, large cars, and small cars. Given these broadly defined choices—choices that are much more narrowly defined in my study—it is not surprising that they find a negligible impact on elasticity due to shifting VMT between cars. My choice set includes all model-year combinations (e.g. 2010 Ford Taurus) that are present in the dataset, and thus will not artificially restrict the choice set.

Finally, the majority of the papers reviewed above have not taken advantage of the large changes in gasoline prices over the past two years. Klier and Linn (2008), on the other hand, utilize new gasoline data to estimate how this affects shares of new vehicle sales, yet are forced

\(^5\) West (2004) estimates a VMT elasticity of -0.87 to -1.03; Mannering and Winston estimate an operating cost elasticity of around -0.25; Goldberg finds a fuel cost elasticity of -0.5; Berkowitz et. al. find a fuel price elasticity of -0.24.
to use aggregate data. As Hensher (1985) writes, “[disaggregate] data alone is inadequate as a basis for identifying influences on the demand for fuel. The two critical inputs into the energy consumption equation are energy efficiency of the durable technology and the level of utilization of the technology (p. 303).” Thus, by looking only at disaggregate macroeconomic data on gasoline use, these researchers disregard two important margins that are in play at the household level: the relative efficiency of the vehicles chosen, and the decision of how much to use these vehicles given a change in gasoline price. In my research, the use of new 2008 disaggregate data helps me to better model household level decisions, and thus estimate more accurately both the extensive and intensive margins.

While many researchers have estimated a relatively inelastic demand for gasoline, not all have yielded the same results. In fact, as detailed above, there has been a wide range of values, which depend not only on the data used by the researcher, but also the methods and models employed. Molly Espey (1996) conducts a meta-analysis on research estimating the elasticity of demand for gasoline. She finds that the estimates range between -0.02 to -1.59, averaging at -0.53. She writes that these wide ranges of differences arise because of the data used, the assumptions employed, the demand function specification, and the econometric technique. Thus, the downward bias that can be found in assuming independence of vehicles and choice set aggregation techniques has to be compared within a single model with the same dataset and functional forms. I will demonstrate the bias by running my model with my specification and comparing these results to my model with different underlying assumptions.

Summarizing, my method and model allow me to improve upon prior studies through the use of new disaggregate data and an unlimited choice set, and through an integrated study of household decisions at the fleet level.

3. Model
My analysis is conducted at the household level- I assume each household optimally chooses how many vehicles to own, what mix of models these vehicles should be, and how much to drive each of these. These decisions depend on both vehicle and household characteristics.
Since optimization is at the household fleet level, there can be substitution between vehicles within the household.

The per-car utility is expressed in the following way:

\[ u_{ij} = \alpha_j VMT_{ij}^\rho + \gamma_i X_j + \xi_j + \epsilon_{ij} \]

where \( VMT_{ij} \) is the number of miles household \( i \) drives car \( j \) per year; \( 0 \leq \rho \leq 1 \) (in order to satisfy diminishing marginal utility of driving) is the constant elasticity of substitution parameter on driving and consumption; \( X_j \) is a \((m \times 1)\) vector of car characteristics; \( \xi_j \) is vehicle \( j \)'s attributes unobservable to the econometrician; and \( \alpha_{ij} \), \( i \)'s marginal utility of driving car \( j \), is parameterized in the following manner: \( \alpha_{ij} = \exp(\bar{Z}_i A \bar{X}_j) / n_i \). \( \bar{Z}_i \) is a \((1 \times k)\) vector subset of the \((m \times 1)\) vector of household characteristics \( Z_i \), \( A \) is the diagonalized \((k \times 1)\) vector of marginal utility coefficients \( \alpha \), \( \bar{X}_j \) is a \((k \times 1)\) vector subset of \( X_j \), and \( n_i \) is the number of vehicles the household owns\(^6\). Finally, \( \epsilon_{ij} \sim N(0, \sigma^2_{\epsilon}) \) and \( \gamma_i = g_0 + g_1 Z_i \). This implies that the marginal utility of a car characteristic for a given household will have a baseline level that is similar across households \((g_0 X_j)\) and a variable utility depending on certain household characteristics \((Z_i g_1 X_j)\) - where \( g_1 \) is the diagonalized \((m \times 1)\) vector of interaction coefficients. This creates the fixed effect for vehicle \( j \): \( \theta_j = g_0 X_j + \xi_j \). Thus, the utility function is additively separable in two parts, one which depends only on car characteristics, and the other which depends on household-car characteristic interactions. One downfall to this specification is that it implies that all households have the same utility from certain vehicle characteristics, such as horsepower, although it is not difficult to imagine that different households value these characteristics in very different ways. However, this assumption is utilized frequently in the relevant literature.

Each household maximizes its utility subject to its budget constraint:

\[ \sum_{j=1}^{n_i} VMT_{ij} \leq B_i \]

\(^6\) Dividing the marginal utility of driving by the number of vehicles a household owns implies that if a household has few vehicles, then the household receives greater marginal utility from each of the vehicles in the garage.
\[
\max_{VMT,c} U_i = \sum_j u_{ij} + c_i^\rho
\]
\[
s.t. \quad y_i = \sum_j P^u_j + \sum_j P^d_{ij} VMT_{ij} + P^c c_i
\]

where \(y_i\) is household income, \(P^u_j\) is the used price per car, \(P^d_{ij}\) is the price of driving (i.e. $/mile), and \(P^c\) is the price of consumption, which is normalized to 1. The story underlying this utility maximization problem is that each household wakes up in the morning and, given the price of gasoline, goes out and buys their fleet of vehicles in the used vehicle market\(^8\). Thus, the household is assumed to be at an optimal point with respect to the vehicles they own and how much they use them\(^9\).

Utility maximization provides the optimal level of consumption and VMT:

\[
c_i^* = \frac{y_i - \sum_k P^u_k}{1 + \sum_k \left( P^d_{ik} \left( \frac{\rho}{\rho - 1} \right) \left( \frac{1}{\alpha_{ik}} \right)^{\frac{1}{\rho - 1}} \right)}
\]

\[
VMT_{ij}^* = \left( \frac{P^d_{ij}}{\alpha_{ij}} \right)^{\frac{1}{\rho - 1}} \frac{y_i - \sum_k P^u_k}{1 + \sum_k \left( P^d_{ik} \left( \frac{\rho}{\rho - 1} \right) \left( \frac{1}{\alpha_{ik}} \right)^{\frac{1}{\rho - 1}} \right)}
\]

Once the optimal consumption and VMT choices have been identified, the indirect utility function, \(V_i = U_i^\ast(c_i^*, VMT_{ij}^\ast)\), is additively separable in the \(\theta_j\), but not in the sub-utilities \(u_{ij}\), which are correlated through \(VMT_{ij}^\ast\). Since the sub-utilities enter into the overall utility in the

---

\(^7\) The usage cost of the vehicle also depends on maintenance costs, which are implicitly captured through the vehicle fixed effect: since the fixed effect is at the model-vintage level, the higher maintenance costs associated with older vehicles is thus captured as an unobserved vehicle attribute.

\(^8\) This makes the implicit assumption that there are no transactions cost associated with the purchase of vehicles.

\(^9\) This specification thus does not allow households to purchase any of their vehicles conditional on the vehicles that are already in their garage (i.e., it does not allow for dynamics in the purchase decision). However, I need to assume this because I do not have data on the purchase dates of the vehicles for each household. Furthermore, this is a customary assumption employed by many in the relevant literature.
same manner (because there is no coefficient multiplying the sub-utilities or in any way differentiating the sub-utilities within the overall utility function), this implies that marginal utility of ownership is constant in the number of vehicles. Furthermore, the marginal rate of substitution is constant and equal to 1. Thus, a household with 5 vehicles has the same marginal utility of owning each vehicle as a household with only 1 vehicle. However, diminishing marginal utility of use is incorporated into the model through the $\alpha_0$ parameter, since each vehicle’s marginal utility of use diminishes with the number of vehicles in the garage. Therefore, while the household will enjoy driving each vehicle more if there are less of them, the household does not have diminishing returns from owning a large vehicle fleet. I choose to employ this assumption because I do not have the data in order to estimate the marginal returns from owning a “first” or “second” vehicle- while I know the vintage of the vehicle, I do not know in which order the vehicles were purchased. Identification of these potential coefficients on the sub-utilities is therefore not possible with the data I have. Furthermore, assuming a constant marginal rate of substitution equal to one will not affect my short-run elasticity estimate. For example, assume household $i$ owns two vehicles, A and B (and I know in which order they were purchased). If the coefficient on vehicle A’s sub-utility is larger than the coefficient on vehicle B’s sub-utility, then by not estimating these coefficients, I will be under-estimating the marginal utility of driving A (and thus, $VMT_{iA}^*$), while over-estimating the marginal utility of driving B (and similarly, $VMT_{iB}^*$). However, when the elasticity of demand for gasoline is calculated, the change in driving under each scenario (new and old gasoline prices) will not be affected by having a higher (or lower) baseline $VMT$, since the counterfactual $VMT$ will also be higher (or lower). Therefore, this assumption does not affect my counterfactual scenarios when calculating a short-run elasticity.

One downside to this model is the fact that any two households who look alike and who own the same fleet will be forced to have the same driving patterns. However, different households may have very different driving patterns, based on certain household unobservables such as length of commute to work. For example, a household with a fixed commute time will be less able to adjust its overall driving patterns than my model would predict. Thus, it is important to account for unobserved consumer heterogeneity in order to improve the estimation of the
elasticity of demand for gasoline, and in order to minimize an upward bias on the elasticity of demand. Therefore, I extend my model to incorporate unobserved consumer heterogeneity in Section 4.4, and find that the elasticity estimates drop.

4.1 Estimation Overview
In order to estimate the model, I rely on a revealed preference approach. This estimation technique is based on Manski (1975)’s semiparametric maximum score estimation for discrete choice models, which rank ordered choice probabilities. This revealed preference method was employed by Bajari and Fox (2005), Fox (2007) and Ellickson, Houghton, and Timmins (2008) to solve high dimensionality problems in a wide variety of industrial organization applications. The basis behind this approach is that strategic players maximize their utility by choosing the outcome that is observed in the data. The researcher recovers the parameters that make the observed outcome provide the player/consumer with more utility than any other alternative. The main econometric difference between Manski’s approach and my own is that I assume a distribution on the error term, while Manski does not10.

In order to use this revealed preference approach, I need to assume that the household’s car purchase and driving choices provide it with the highest indirect utility possible. Thus, any deviation from the observed outcome will decrease household i’s indirect utility. Through this assumption, I propose unilateral deviations from the household’s optimizing behavior by comparing deviations across different households, in order to difference away all the fixed effects that make typical estimation techniques infeasible.

Once I identify the parameters that drive the vehicle use decision, I use a similar optimality assumption in order to recover the fixed effects. This allows me to identify parameters that affect the car purchasing decision, so that I can predict how consumers would adjust their long-run behavior given a change in policy.

10 I choose this more parametric route because it improves the second stage estimation technique. In my method, I first difference away the vehicle fixed effects and recapture them in the second stage. If I did not impose a structure on the error term in the first stage, I would only be able to set-identify the vehicle fixed effects in the second stage. By imposing a distribution on the error term in the first stage, I am able to point identify the vehicle fixed effects, which allows me to have more policy-relevant results that can be transferred to other settings.
4.2a Estimation: First Stage

In the first stage of the estimation procedure, I model the purchase decision in a probit framework. This identifies the model parameters $\alpha$ (the marginal utility of driving), $\rho$ (the CES parameter), and $g_i$ (how different types of households value different car characteristics). The revealed preference approach assumes that households are observed in a long-run equilibrium. That is, although their cars may have been bought at different times, the fleet they own now is as optimal as if they had just bought all their cars today. The identifying assumption is thus the following:

Assumption 1: Each household is at an equilibrium with respect to the number and types of cars in its fleet, how many miles to travel in each car, and how much consumption to engage in.

This assumption implies that the household has chosen its fleet of cars so as to maximize indirect utility. Thus:

$$V_{iF_i^*} \geq V_{iF_i} \forall F_i \neq F_i^*$$  \hspace{1cm} (1)

where $F_i^*$ is the fleet of cars chosen by household $i$. Since any change in the household’s observed car holdings diminishes its indirect utility, this also holds if I only change one car in the household’s fleet. This allows me to compare the utility of different variations from the observed outcome. The fortunate impact of doing unilateral deviations is that all the fixed effects ($\theta_j$) for the cars that are being held fixed drop out from both sides of the equation, since they enter linearly into the indirect utility function. This leaves only two fixed effects to deal with: the $\theta_j$ from the car that is being swapped out of the fleet on the left hand side of equation (1), and the $\theta_j$ from the car that is being added to the fleet on the right hand side of equation (1). My main goal is to try to eliminate all of the fixed effects in this stage. Therefore, I propose a “swap” which would eliminate these final two $\theta_j$ from the estimation. The idea is to take two households who have different vehicles, and swap each of their vehicles with one from the other household. For example, assume we have two households: 1 and 2, that each has respective vehicles A and B. Equation (1) can thus be rewritten in the following manner:
\[
\tilde{V}_{1A} + \theta_A \geq \tilde{V}_{1B} + \theta_B \\
\tilde{V}_{2B} + \theta_B \geq \tilde{V}_{2A} + \theta_A
\]

where \( \tilde{V} \) is the part of the indirect utility that does not include the fixed effect. By adding across these two inequalities, \( \theta_A \) and \( \theta_B \) cancel out from each side of the equation, leaving only the \( \tilde{V} \):

\[
\tilde{V}_{1A} - \tilde{V}_{1B} + \tilde{V}_{2B} - \tilde{V}_{2A} \geq 0
\]

The probability that a random car swap between two households makes each household worse off is the probability that equation (2) holds. Rearranging equation (2), and since the composite error term \( \tilde{e} \) is distributed normally, this gives:

\[
\text{Pr}(\tilde{V}_{1A} - \tilde{V}_{1B} + \tilde{V}_{2B} - \tilde{V}_{2A} \geq 0) = \text{Pr}(\tilde{e}_{12,AB} \leq \tilde{V}_{1A} - \tilde{V}_{1B} + \tilde{V}_{2B} - \tilde{V}_{2A}) = \Phi\left(\frac{\tilde{V}_{1A} - \tilde{V}_{1B} + \tilde{V}_{2B} - \tilde{V}_{2A}}{\sigma_{\tilde{e}}}\right)
\]

Where \( \tilde{V}_{ij} \) is the deterministic part of the indirect utility that does not include \( \theta_j \). I can set up many of these inequalities- at least one for each vehicle in every household’s fleet- and create a log likelihood that any random vehicle swap between two different households makes each household worse off:

\[
LL(\beta) = \sum_{\text{swap}(A_{i2}, B_{j2})} \log \Phi\left(\frac{\tilde{V}_{i_{h1}, j_{h2}} - \tilde{V}_{i_{h2}, j_{h2}} + \tilde{V}_{i_{j1}, j_{j2}} - \tilde{V}_{i_{j2}, j_{j2}}}{\sigma_{\tilde{e}}}\right)
\]

The normalization in this step is through the fixing of the consumption parameter to 1. While this allows me to eliminate the fixed effects and estimate the first stage parameter vector, I am imposing some correlation between the composite error terms. That is, suppose household A swaps with household B, and household B swaps with household C. Both composite error terms in these swaps include \( \tilde{e}_{Bj} \). Thus, the very nature of the cross-household swaps imposes
correlation through the idiosyncratic household-vehicle error term included in the composite error term. While this implies that some of the composite error terms are not i.i.d., Han (1987) proves that as long as the cross-sectional error terms (i.e. the $\varepsilon_{ij}$) are i.i.d., there is uniform convergence and strong consistency of pair-wise estimation. Since I am controlling for both unobserved and observed car characteristics, all $\varepsilon_{ij}$ are i.i.d., even within households. Thus, I am able to consistently estimate the first stage parameter vector with this pair-wise estimation technique.

There does exist another technique that also allows for household bundling of goods under a high dimensioned choice set. McFadden et al. (1978) demonstrate that if the I.I.A. property holds under a logit error, then the parameter estimates obtained from using a subsample of all the alternatives are statistically equal to those obtained from the full sample. This implies that it is possible to use only a randomly selected fraction of all the bundles available to the households and thus minimize greatly the choice set. However, this technique presents two distinct problematic issues that the swapping technique I employ avoids. First, by construction, this method demonstrates the I.I.A. property that has an unfortunate outcome, as demonstrated by Chipman (1960) and Debreu (1960). This outcome, known as the “red bus-blue bus” scenario, allows an alternative that is practically identical to one of the choices (i.e. a red bus and a blue bus in a transportation method decision) to change the probabilities of choosing the other choices. Thus, if we were to add to the choice set the ability to take a blue and a red bus, instead of just a blue bus, then the probabilities of choosing a personal vehicle instead of a public bus decreases. This is an unrealistic outcome that is inherent to the logit method, as well as the subsampling of alternatives technique, which is unfortunate under a vehicle choice model. Second, in order to use the subsampling of alternatives technique with bundles, the researcher would have to assume bundle-specific error terms. Given the way in which they arise, these error terms should be correlated across the bundles that share any vehicles between them. Thus, the error terms are no longer i.i.d. and violate the assumptions of the logit model. My technique, as described above, does not exhibit correlated error terms until

---

11 I.I.A. is the “independence of irrelevant alternatives” outcome that emerges from logit probabilities.
after the swapping has occurred, thus maintaining my i.i.d. assumption throughout the estimation.

By eliminating the fixed effects in the first stage, I am able to estimate the parameter vector \( \beta = (\alpha, \rho, g_1, \sigma_\epsilon) \) through maximum likelihood without having to artificially restrict the choice set available to households, and by avoiding certain implications that are present in other techniques that allow for a large choice set.

### 4.2b Estimation: First Stage Starting Values and Optimization

In order to incorporate more information on observed household driving behavior, and also to ensure that this non-linear maximization technique does not get stuck at a local maximum, I implement a three-stage procedure. First, I use information on observed driving behavior in order to minimize the distance between optimal and observed annual miles travelled for each household. Once I have found the parameters \( \alpha \), I hold these fixed and search for the \( \beta \) parameters that solve the likelihood function (4). Once these two parameter sets have been found, I use these values as starting values for my first stage. This ensures that I am searching closer towards the correct parameter space, which is important with such a non-linear high variable function as the one I am implementing.

I use a simulated annealing optimizer in order to approach the maximum of my likelihood. Simulated annealing is helpful in that it searches over a large parameter space and does not tend to get trapped in a local maximum. Once the function has reached sufficient convergence, I implement a simplex technique in order to find the peak of the curve.

### 4.3 Estimation: Second Stage

Once I have estimated the utility parameters in the first stage, I can recover the fixed effects in order to find how individuals value car characteristics. In order to identify the crucial parameter \( g_0 \), I employ the following two assumptions:

**Assumption 2:** The household purchases vehicles as long as the net sub-utility from each vehicle is greater than or equal to zero.
**Assumption 3:** The household purchases vehicles to maximize overall utility, so any extra vehicle purchased would result in a decrease in total utility.

Under Assumption 2, the same household 1 as before that owns car A has the following indirect utilities:

\[ V_{1A} \geq V_1^0 \] 

(4)

where the indirect utility of the outside option \( V_1^0 \equiv \) not owning a car) is normalized to zero.

Under Assumption 3, household 2 would have the following indirect utility:

\[ V_{2B} \geq V_{2B,A} \] 

(5)

That is, adding an extra vehicle to household 2’s fleet decreases household 2’s overall indirect utility. In order to recover the parameter \( g_0 \), I can rewrite the above equations in the following way:

\[ V_{1A} + \theta_A + \epsilon_{1A} \geq 0 \] 

(4.1)

\[ V_{2B} + \theta_B + \epsilon_{2B} \geq V_{2B,A} + \theta_A + \theta_B + \epsilon_{2B} + \epsilon_{2A} \] 

(5.1)

In equation (5.1), both the fixed effects and the idiosyncratic error term for car B drop out, leaving only car A’s fixed effects and errors for both households. Thus, the probability that car A provides household 1 (and all other households who own car A) with positive sub-utility is:

\[ P(\epsilon_{1A} \leq \Delta V_i + \theta_A) \]

where \( \Delta V_i \) is the change in indirect utility from removing car A from the household’s fleet. For household 2 (and all others who do not own car A), the similar expression would be:

\[ P(\epsilon_{1A} \leq \Delta V_i - \theta_A) \]

---

12 By doing unilateral deviations in the second stage, I do not impose correlation between the error terms, so they are i.i.d. for the second round of probits.
Therefore, the likelihood that the sub-utility of any car $j$ is positive for the households who own it (and negative for those who don’t) is:

$$LL(\theta_j) = \sum_i y_{ij} \ln \Phi \left( \frac{\Delta V_{ij} + \theta_j}{\sigma_{\varepsilon_j}} \right) + (1 - y_{ij}) \ln \left( 1 - \Phi \left( \frac{\Delta V_{ij} - \theta_j}{\sigma_{\varepsilon_j}} \right) \right)$$

I can thus recover the set of $\theta_j$ and $\sigma_{\varepsilon_j}$ through $J$ binomial probits, and by treating the $\xi_j$ as random effects instead of as unobserved car attributes, I can run the following regression through OLS and estimate $g_0$:

$$\theta_j = g_0 X_j + \xi_j$$

Therefore, through two rounds of binomial probits, I can first identify the parameters that impact the utility of driving and car ownership, and then estimate the parameters that determine how households choose their fleet.

**4.4 Unobserved Consumer Heterogeneity**

In my estimation, I incorporate actual driving behavior into my estimation through the starting values procedure, as described in 4.2b. However, structurally, my model does not allow two households that have the same vehicle and are identical in observable behavior to drive different amounts. Yet there may be characteristics that are unobservable to the econometrician that impact how much individuals drive— for example, long commutes to work. Furthermore, these differences in households needs for driving impact more than just the parameters that are in the marginal utility of driving. Because optimal driving, $VMT_{ij}^*$, includes the marginal utility of driving, the overall indirect utility is also affected by $\alpha_{ij}$. This implies that the other parameters ($\rho$, $g$, and $\sigma$) will also be affected by each household’s particular driving needs. For example, a household with a long commute not only will choose a more comfortable car in order to increase the marginal utility of that drive, but will also probably steer towards a more fuel efficient car (even though fuel efficiency does not affect the marginal utility of driving). Thus, not allowing for unobserved consumer heterogeneity into the overall estimation technique may bias the results. For example, the estimate for optimal driving for a household
with an unobserved long commute (or any fixed commute length that would not change with gasoline price) would be biased downwards - the household would appear to be more responsive to gasoline prices than they truly are, since the drivers have less capability of adjusting their driving amount. Since not allowing for unobserved consumer heterogeneity will likely bias my estimate of the elasticity of demand for gasoline up, I incorporate information on how much households actually drive each vehicle and I include an error term that varies between households. In order to do this, I re-parameterize $\alpha_{ij}$ in the following way:

$$\alpha_{ij} = \exp(\bar{Z}_i, \bar{X}_j)/n_j + \mu_i$$

where $\mu_i$ is an idiosyncratic household specific unobservable that affects driving behavior. This implies that each household’s optimal driving pattern will now depend on some unobserved household characteristic, so that two similar households may drive different amounts:

$$VMT_{ij}^* = \left( \frac{P^d_{ij}}{\alpha_{ij} + \mu_i} \right)^{\frac{1}{\rho-1}} \left( \frac{y_i - \sum_j P^u_j}{1 + \sum_j \left( \frac{P^d_{ij}}{\rho} \left( \frac{1}{\alpha_{ij} + \mu_i} \right)^{\frac{1}{\rho-1}} \right) \right)$$  \hspace{1cm} (1)

While this does account for unobserved consumer heterogeneity, it greatly increases the difficulty of the estimation procedure. Integrating this function over the distribution of $\mu_i$ is out of the scope of this paper. Instead, I will implement an extra step at each iteration where I intend to solve the following equation:

$$\left( \frac{1}{j} \sum_j VMT_{ij}^* - \frac{1}{j} \sum_j VMT_{ij} \right)^2 = 0$$  \hspace{1cm} (2)

where $VMT_{ij}^*$ is the optimal driving as defined in equation (1), and $\overline{VMT}_{ij}$ is the observed driving behavior. Equation (2) implies that $\mu_i$ causes the actual driving behavior to equal the optimally
calculated VMT through an average weighting mechanism. Thus, the average VMT of all the vehicles in the household’s garage is shifted up or down based on a single household idiosyncratic error\textsuperscript{13}. Furthermore, this unobserved error term affects all parameters, in that it is an integral part of the indirect utility function, and its value depends on the entire parameter vector $\beta = (\alpha, \rho, g_i, \sigma_z)$. In order to incorporate this assumption into the optimal driving mechanism, I implement the following iterative procedure:

1) Guess at the parameter vector $\beta = (\alpha, \rho, g_i, \rho)$.
2) Find $\mu_i$ that makes equation (2) hold for each household.
3) Given current $\beta'$ and $\mu_i$, calculate $VMT_i^*$ as described in equation (1).
4) Proceed with estimation as outlined in Section 5 – find new parameter vector $\beta'$ that maximizes objective function.
5) Repeat steps 2-4 until convergence.

By including this unobserved heterogeneity, I am able to more accurately capture household behavior by incorporating more information that is currently available, allow for differences in household driving behavior, and minimize the bias on the estimate of elasticity.

5. Data
The main dataset that I am using in order to analyze household car use and purchase decisions is the National Household Transportation Survey (NHTS) from 2001 and 2009. These data provide information on household characteristics, including what types of vehicles each household owns, and how much they drive each. I am using the national sample of the 2001 dataset, which is comprised of 26,038 households, and 53,275 car-household observations, and the complete 2009 dataset, which is comprised of 143,084 households, and 309,163 car-household observations. The complete 2001 dataset has certain inconsistencies in the vehicle observations, and I was recommended by the NHTS data center to focus instead only on the national sample. However, the 2009 dataset did not have these issues, so I was able to start

\textsuperscript{13} While this weights each of the vehicles in a household’s garage equally in terms of affecting total miles travelled, there is no data particular to the dataset I am using that would favor any other weighting mechanism.
with the entire dataset. In order to not bias the results towards the 2001 observations, I randomly cut households (after dropping households with missing values) from the 2009 data until the sample size resembled the 2001 data sample size. Table 1 shows some summary statistics of the households. Because of missing data in household income, I impute income based on household size, race, number of vehicles in fleet, home ownership, and urban residence.

Detailed characteristics of the vehicles in the NHTS dataset come from 2001 and 2009 Polk Data, which I have matched to NHTS. However, not every car in the NHTS dataset was reflected in the Polk Data, and vice versa, making it necessary to drop 6,733 observations from 2001 and 151,294 observations from 2009 due to no match being found between datasets. For cars in NHTS that matched more than one car in Polk, if one of these vehicles was purchased at a much larger level than the other vehicles in the set, then this vehicle was assigned to the household. However, if the popularity of the vehicle models in the set was similar, then randomization was used to choose the final vehicle assigned to the household.
Table 1: Summary Statistics NHTS Dataset

<table>
<thead>
<tr>
<th></th>
<th>2001 National Sample</th>
<th>2001 Final 11% Sample</th>
<th>2009 Full Dataset</th>
<th>2009 Final Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Observations</td>
<td>53,275</td>
<td>18,166</td>
<td>309,163</td>
<td>18,259</td>
</tr>
<tr>
<td>Number of Households</td>
<td>26,038</td>
<td>11,354</td>
<td>143,084</td>
<td>11,366</td>
</tr>
<tr>
<td>% White</td>
<td>87 %</td>
<td>85 %</td>
<td>87 %</td>
<td>89 %</td>
</tr>
<tr>
<td>% Urban</td>
<td>70 %</td>
<td>79 %</td>
<td>77 %</td>
<td>72 %</td>
</tr>
<tr>
<td>Average Family Income</td>
<td>$55,832</td>
<td>$56,338</td>
<td>$58,411</td>
<td>$64,618</td>
</tr>
<tr>
<td>Average Household Size</td>
<td>2.82</td>
<td>2.66</td>
<td>2.62</td>
<td>2.46</td>
</tr>
<tr>
<td>Average Number of Vehicles in Household</td>
<td>2.69</td>
<td>2.04</td>
<td>2.71</td>
<td>1.61</td>
</tr>
<tr>
<td>Average MPG</td>
<td>25.72</td>
<td>25.87</td>
<td>21.13</td>
<td>21.64</td>
</tr>
<tr>
<td>Average Vehicle Age</td>
<td>8.49</td>
<td>7.21</td>
<td>9.24</td>
<td>7.96</td>
</tr>
<tr>
<td>Average Yearly VMT</td>
<td>10,995</td>
<td>11,594</td>
<td>9,729</td>
<td>10,401</td>
</tr>
</tbody>
</table>

The Polk dataset includes characteristics such as wheelbase, length, mpg, horsepower, and MSRP. An OLS regression of price on car characteristics yields a high $R^2$ and highly significant coefficients. Table 2 has the results of the regression: ln(price) is regressed on car characteristics including dummies for origin and air conditioning, as well as size, strength and efficiency parameters. European cars fetch the highest price, as do more fuel efficient cars and larger cars.

Table 2: Polk Data OLS Regression of Car Characteristics

<table>
<thead>
<tr>
<th>Regressor: ln(Price)</th>
<th>Coefficient (Std. Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>8.258*** (0.0589)</td>
</tr>
<tr>
<td>Domestic</td>
<td>0.1212*** (0.0124)</td>
</tr>
<tr>
<td>European</td>
<td>0.4044*** (0.0137)</td>
</tr>
<tr>
<td>Japanese</td>
<td>0.2202*** (0.0124)</td>
</tr>
<tr>
<td>Number of Doors</td>
<td>0.0236*** (0.0016)</td>
</tr>
<tr>
<td>Air Conditioning</td>
<td>0.2247*** (0.0041)</td>
</tr>
<tr>
<td>Length</td>
<td>0.0011*** (0.0004)</td>
</tr>
<tr>
<td>Width</td>
<td>-0.0078*** (0.0012)</td>
</tr>
<tr>
<td>Wheelbase</td>
<td>0.0038*** (0.0007)</td>
</tr>
<tr>
<td>Horsepower</td>
<td>0.0065*** (0.0001)</td>
</tr>
<tr>
<td>MPG</td>
<td>0.0013*** (0.0005)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.7812</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>18,273</td>
</tr>
</tbody>
</table>

(***: statistically significant at 99%)
One drawback of the NHTS dataset is that it does not give the transaction prices of the vehicles, nor does the dataset contain information on when the vehicles were purchased. Instead, I use the used vehicle prices from NADA 2001 and 2009 datasets, which give me national average used vehicle prices for most vehicles in 2001 and 2009. For 2001, I have used vehicle prices for vintages 1982-2002, while for 2009, I have used vehicle prices for vintages 1990-2010\textsuperscript{14}.

Finally, I use ACCRA data to find the gasoline prices in the household’s MSA. The ACCRA data provide me yearly information on gasoline prices at the city level, which I have aggregated up to the MSA level in order to match with the NHTS dataset. Figure 2 shows the distribution of the gasoline prices for the two years over the MSAs in both datasets, with the vertical lines demonstrating the median gasoline price in the corresponding year. As the NHTS 2001 has fewer MSAs, the number of points in the 2001 distribution is lower than that of the 2009 distribution. The 2001 NHTS also does not provide me with a complete location - if a household is located outside of an MSA, I only know the state they are located in, as long as the state has a population of more than 2 million. Thus, for all observations outside of an MSA, and within a designated state, I use an average of all the gas prices within the state. Since NHTS only records states if the state has a population of more than 2 million, households who live in 15 states do not have the state recorded. This causes me to drop observations from these states, since I cannot assign these households a gasoline price. The final sample is thus 18,166, approximately 34\% of the national sample. The 2009 dataset does not have this problem, as all states are recorded. However, the 2009 ACCRA dataset does not give me information on gas prices in South Dakota, so I need to drop households that live in this state. Furthermore, I need to drop households that have missing crucial information on vehicles, so the final sample is thus 149,867. I then proceed to take a 12\% random sample of the dataset, resulting in 18,259 observations for 2009.

\textsuperscript{14} 2010 values in 2009 were for vehicles that came out at the end of 2009 and have a 2010 vintage according to the manufacturer (likewise for 2002 values in 2001).
6.1 First Stage Results

In the first stage of the estimation, I calculate the household type-specific parameters contained in the utility of driving. Since the fixed effects are differenced away in this step, I do not estimate the mean utility parameters of ownership and use until the second stage. Once I have estimated the utility of driving parameters, I can use the optimal $\nu^*MT$ from the utility maximization in order to estimate the short-run and medium-run elasticity of demand for gasoline.

Results from the first stage of estimation are presented in Table 3. The marginal utility of driving parameters includes three different interaction terms: urban residence and acceleration (horsepower/weight); urban residence and wheelbase; vehicle size (length*width) and household size; and household size divided by the number of vehicles in the household. The
final interaction term implies that the discomfort from having to share a vehicle with multiple family members does not vary with the vehicle choice.

The signs of some of these parameters are intuitive: urban households prefer vehicles that have greater acceleration power, and larger households prefer having more vehicles (although this parameter value is not statistically different from zero). However, the coefficient on household size and vehicle size is negative, which is counterintuitive, and it does not appear that urban households care about the width of their vehicles. The other interaction terms are a set of interactions between household location (one of 9 census divisions - which can be seen in Figure 3) and urban/rural) and income with vehicle characteristics such as origin, size, age, and “greenness” (defined as MPG greater than 30). Many of these coefficients are intuitive. For example, green and small vehicles are preferred in the Pacific census division (although the green*Pacific coefficient is surprisingly not significant), while larger vehicles are preferred in the Mountainous regions. Domestic vehicles are preferred by households in the Midwest, while richer households prefer European and Japanese cars over domestic vehicles.
Given these parameter values, I calculate how a one percent increase in gas prices will affect optimal driving, and thus estimate the elasticity of demand for gasoline. When gasoline prices change, households can react in the following way: in the immediate short run, they can drive each of their cars less. In the medium run, they can begin to reallocate driving between the vehicles in their fleet. Finally, in the long run, they can choose to buy different vehicles and re-optimize their driving decision over their fleet. In order to estimate the long-run elasticity of demand for gasoline I would need to integrate out over the distribution of the household-vehicle specific error terms in order to find which bundle of vehicles was preferred for the household (given optimal driving and consumption under each bundle). Measuring this elasticity is something I am leaving for future research, and I expect to find a higher elasticity estimate with the long-run elasticity, as has been demonstrated in the literature.
Table 3: Full Model, First Stage Parameter Values

<table>
<thead>
<tr>
<th>Interaction Term</th>
<th>Parameter Value (Std. Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marginal Utility of Driving Parameters</td>
<td></td>
</tr>
<tr>
<td>(HP/weight) * urban</td>
<td>2.341*** (0.055)</td>
</tr>
<tr>
<td>Vehicle Size * household size</td>
<td>-0.252* (0.147)</td>
</tr>
<tr>
<td>Wheelbase* urban</td>
<td>0.051 (1.531)</td>
</tr>
<tr>
<td>Household size/# vehicles</td>
<td>-0.166 (0.353)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Indirect Utility of Driving Parameters</td>
<td></td>
</tr>
<tr>
<td>Green* income</td>
<td>-0.063 (0.268)</td>
</tr>
<tr>
<td>Green* Pacific</td>
<td>0.124 (1.147)</td>
</tr>
<tr>
<td>Green*New England</td>
<td>0.731** (0.347)</td>
</tr>
<tr>
<td>Size* Income</td>
<td>0.547*** (0.146)</td>
</tr>
<tr>
<td>Size* Pacific</td>
<td>-1.703*** (0.220)</td>
</tr>
<tr>
<td>Size* Mountain</td>
<td>2.085*** (0.204)</td>
</tr>
<tr>
<td>Domestic* Income</td>
<td>0.101 (0.439)</td>
</tr>
<tr>
<td>Domestic*Midwest (West North Central)</td>
<td>2.147*** (0.142)</td>
</tr>
<tr>
<td>Domestic*Midwest (East North Central)</td>
<td>1.118*** (0.168)</td>
</tr>
<tr>
<td>European* Income</td>
<td>0.989*** (0.137)</td>
</tr>
<tr>
<td>Japanese* Income</td>
<td>0.495*** (0.149)</td>
</tr>
<tr>
<td>Japanese* Pacific</td>
<td>1.170*** (0.161)</td>
</tr>
<tr>
<td>Vehicle Age* Income</td>
<td>-0.037 (0.144)</td>
</tr>
<tr>
<td>Size* Urban</td>
<td>-2.021*** (0.154)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.176 (0.817)</td>
</tr>
</tbody>
</table>

Std. Dev. of Error Term $\sigma_{\tilde{\varepsilon}}$ 1.539*** (0.084)

Elasticity of Demand for Gasoline, $\Delta$Price = 1%

Mean (Standard Deviation)

(-***: statistically significant at 99%, **: statistically significant at 95%, *: statistically significant at 90%)

The coefficient results are similar when I account for unobserved heterogeneity, as most are the same sign and size. I find that my elasticity becomes more inelastic by about 15%, as can be seen in Table 4. This is as expected- by not allowing households to have fixed levels of driving needs, I am overestimating the ability of households to respond to changing gasoline prices. Thus, this presents an upward bias in the elasticity estimate when I do not account for these differences.
Table 4: Unobserved Consumer Heterogeneity Model, First Stage Parameter Values

<table>
<thead>
<tr>
<th>Interaction Term</th>
<th>Parameter Value (Std. Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(HP/weight) * urban</td>
<td>2.391*** (0.015)</td>
</tr>
<tr>
<td>Vehicle Size * household size</td>
<td>-0.357*** (0.033)</td>
</tr>
<tr>
<td>Wheelbase* urban</td>
<td>1.128*** (0.019)</td>
</tr>
<tr>
<td>Household size/# vehicles</td>
<td>0.169* (0.096)</td>
</tr>
<tr>
<td>Marginal Utility of Driving Parameters</td>
<td></td>
</tr>
<tr>
<td>Green* income</td>
<td>-0.045 (0.498)</td>
</tr>
<tr>
<td>Green* Pacific</td>
<td>-0.155 (1.004)</td>
</tr>
<tr>
<td>Green*New England</td>
<td>2.441*** (0.164)</td>
</tr>
<tr>
<td>Size* Income</td>
<td>2.310*** (0.077)</td>
</tr>
<tr>
<td>Size* Pacific</td>
<td>-7.388*** (0.083)</td>
</tr>
<tr>
<td>Size* Mountain</td>
<td>-1.231*** (0.413)</td>
</tr>
<tr>
<td>Domestic* Income</td>
<td>-0.195 (0.359)</td>
</tr>
<tr>
<td>Domestic*Midwest (West North Central)</td>
<td>5.103*** (0.082)</td>
</tr>
<tr>
<td>Domestic*Midwest (East North Central)</td>
<td>1.377*** (0.154)</td>
</tr>
<tr>
<td>European* Income</td>
<td>1.290*** (0.094)</td>
</tr>
<tr>
<td>Japanese* Income</td>
<td>0.106 (0.621)</td>
</tr>
<tr>
<td>Japanese* Pacific</td>
<td>2.449*** (0.091)</td>
</tr>
<tr>
<td>Vehicle Age* Income</td>
<td>-0.038 (0.100)</td>
</tr>
<tr>
<td>Size* Urban</td>
<td>-2.281*** (0.174)</td>
</tr>
<tr>
<td>ρ</td>
<td>-0.119 (0.361)</td>
</tr>
<tr>
<td>Std. Dev. of Error Term</td>
<td>2.000*** (0.036)</td>
</tr>
<tr>
<td>Elasticity of Demand for Gasoline, ΔPrice = 1%</td>
<td>-1.108*** (0.193)</td>
</tr>
<tr>
<td>Mean (Standard Deviation)</td>
<td></td>
</tr>
</tbody>
</table>

(***: statistically significant at 99%, **: statistically significant at 95%, *: statistically significant at 90%)

6.2 Measuring the Bias

In order to quantify the bias present in aggregating the choice set and assuming independence between vehicles in the household’s garage, I estimate my model under these different assumptions. First, I quantify the bias present in assuming independence. In order to do this, I restrict households to be unable to reallocate VMT between the vehicles in the fleet. This is
comparable to assuming independence: the household chooses how much to drive each car without taking into consideration the rest of the vehicles in the garage. In order to adapt the model so as to reflect this decision making process, I assume that each vehicle in the dataset is a different household, and implement the utility maximization procedure as before. Thus, any household who owns \( n \) vehicles will be counted as \( n \) households each owning one vehicle.\textsuperscript{15} Following Bento \textit{et al.} (2008), I divide household income by a measure of household size (number of adults) in order to scale down the level of consumption to a more realistic measure. The parameter estimates are given in Table 5, and demonstrate a downward bias in elasticity on the order of 12\%.

In the second step of measuring bias, I aggregate the choice set. In order to do this, I follow Bento \textit{et al.} and aggregate the choice set into 233\textsuperscript{16} different choices, based on the following stratification: nine classes of vehicles\textsuperscript{17} (compact, midsize, fullsize, luxury, small truck, large truck, small SUV, large SUV/van, and minivan), vehicle age (1 year old, 2-3 years old, 4-7 years old, 8-12 years old, and 13-18 years old), and seven make categories (Ford, Chrysler Daimler, GM, Honda, Toyota, other East Asian, and European). The parameter estimates are given in Table 6, and demonstrate that not allowing for a more detailed choice set biases the elasticity towards zero by 21.4\%.

In the third step of measuring the bias, I not only aggregate the choice set, but I also assume independence of the vehicles in the household’s garage. The results are presented in Table 7, where it is shown that the elasticity furthers decreases to -0.992, a 22.3\% downward bias.

Since allowing for inter-household substitution of VMT increases the medium run elasticity of demand to -1.277, this demonstrates that there is indeed a downward bias in the elasticity of

\textsuperscript{15} The only difference in the swapping technique is that I have to be more careful about the swaps I make. For example, say household 1 owns cars 2 and 3. If I choose household 1 to swap car 2 with household 1’s car 3, then the inequality is awash, and it adds nothing to the maximum likelihood estimation.

\textsuperscript{16} While the total number of possible choices is much larger under this specification, there are only 233 choices with observed household outcomes.

\textsuperscript{17} The complete set of choices is slightly different from Bento et.al., because of differences in our vehicle characteristic datasets- they have ten class categories: non-luxury compact, nonluxury midsize, non-luxury fullsize, luxury compact, luxury midsize/fullsize, small truck, large truck, small SUV, large SUV/van, and minivan. However, my dataset does not differentiate luxury compact with luxury midsize/fullsize, so I have to group these into one category.
demand under the assumptions of independence and aggregation, although the main part of the bias is due to the aggregation of the choice set. Thus, households become more elastic if they are able to adapt along more margins to changing gasoline prices.

This bias implies that policy makers should take into account whether the researcher has aggregated the choice set or has assumed independence in order to facilitate estimation, since these assumptions may be biasing their results.

Table 5: Independent Model, First Stage Parameter Values

<table>
<thead>
<tr>
<th>Interaction Term</th>
<th>Parameter Value (Std. Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marginal Utility of Driving Parameters</td>
<td></td>
</tr>
<tr>
<td>(HP/weight) * urban</td>
<td>3.231*** (0.134)</td>
</tr>
<tr>
<td>Vehicle Size * household size</td>
<td>1.369*** (0.147)</td>
</tr>
<tr>
<td>Wheelbase* urban</td>
<td>-1.313*** (0.194)</td>
</tr>
<tr>
<td>Household size/# vehicles</td>
<td>-2.391*** (0.144)</td>
</tr>
<tr>
<td>Indirect Utility of Driving Parameters</td>
<td></td>
</tr>
<tr>
<td>Green* income</td>
<td>-0.019 (0.607)</td>
</tr>
<tr>
<td>Green* Pacific</td>
<td>0.012 (8.404)</td>
</tr>
<tr>
<td>Green* New England</td>
<td>0.322 (0.579)</td>
</tr>
<tr>
<td>Size* Income</td>
<td>0.273 (0.311)</td>
</tr>
<tr>
<td>Size* Pacific</td>
<td>-1.111*** (0.348)</td>
</tr>
<tr>
<td>Size* Mountain</td>
<td>0.968*** (0.399)</td>
</tr>
<tr>
<td>Domestic* Income</td>
<td>-0.061 (0.578)</td>
</tr>
<tr>
<td>Domestic* Midwest (West North Central)</td>
<td>1.450*** (0.296)</td>
</tr>
<tr>
<td>Domestic* Midwest (East North Central)</td>
<td>0.556* (0.329)</td>
</tr>
<tr>
<td>European* Income</td>
<td>0.528* (0.296)</td>
</tr>
<tr>
<td>Japanese* Income</td>
<td>0.221 (0.318)</td>
</tr>
<tr>
<td>Japanese* Pacific</td>
<td>0.833** (0.305)</td>
</tr>
<tr>
<td>Age* Income</td>
<td>-0.023 (0.292)</td>
</tr>
<tr>
<td>Size* Urban</td>
<td>-2.274*** (0.357)</td>
</tr>
<tr>
<td>Std. Dev. of Error Term</td>
<td>(\sigma_e)</td>
</tr>
<tr>
<td></td>
<td>1.147*** (0.253)</td>
</tr>
<tr>
<td>Elasticity of Demand for Gasoline, (\Delta)Price = 1%</td>
<td>-1.123*** (0.321)</td>
</tr>
</tbody>
</table>

**Mean (Standard Deviation)**

(***: statistically significant at 99%, **: statistically significant at 95%, *: statistically significant at 90%)
Table 6: Aggregation Model, First Stage Parameter Values

<table>
<thead>
<tr>
<th>Interaction Term</th>
<th>Parameter Value (Std. Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marginal Utility of Driving Parameters</td>
<td></td>
</tr>
<tr>
<td>(HP/weight) * urban</td>
<td>0.123 (0.224)</td>
</tr>
<tr>
<td>Vehicle Size * household size</td>
<td>0.396*** (0.128)</td>
</tr>
<tr>
<td>Wheelbase* urban</td>
<td>-1.458*** (0.304)</td>
</tr>
<tr>
<td>Household size/# vehicles</td>
<td>-1.043*** (0.252)</td>
</tr>
<tr>
<td>Indirect Utility of Driving Parameters</td>
<td></td>
</tr>
<tr>
<td>Green* income</td>
<td>-0.165 (0.579)</td>
</tr>
<tr>
<td>Green* Pacific</td>
<td>1.383** (0.564)</td>
</tr>
<tr>
<td>Green*New England</td>
<td>1.748*** (0.598)</td>
</tr>
<tr>
<td>Size* Income</td>
<td>1.881*** (0.535)</td>
</tr>
<tr>
<td>Size* Pacific</td>
<td>-2.338*** (0.590)</td>
</tr>
<tr>
<td>Size* Mountain</td>
<td>-1.366* (0.740)</td>
</tr>
<tr>
<td>Domestic* Income</td>
<td>-4.561*** (0.533)</td>
</tr>
<tr>
<td>Domestic*Midwest (West North Central)</td>
<td>5.515*** (0.528)</td>
</tr>
<tr>
<td>Domestic*Midwest (East North Central)</td>
<td>1.911*** (0.539)</td>
</tr>
<tr>
<td>European* Income</td>
<td>-2.698*** (0.545)</td>
</tr>
<tr>
<td>Japanese* Income</td>
<td>-3.610*** (0.536)</td>
</tr>
<tr>
<td>Japanese* Pacific</td>
<td>2.682*** (0.532)</td>
</tr>
<tr>
<td>Age* Income</td>
<td>-0.110 (0.527)</td>
</tr>
<tr>
<td>Size* Urban</td>
<td>3.189*** (0.443)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.019 (0.014)</td>
</tr>
<tr>
<td>Std. Dev. of Error Term</td>
<td>$\sigma_{\varepsilon}$</td>
</tr>
<tr>
<td>Elasticity of Demand for Gasoline, $\Delta$Price = 1%</td>
<td>-1.003*** (0.005)</td>
</tr>
</tbody>
</table>

*Mean (Standard Deviation)*

(***: statistically significant at 99%, **: statistically significant at 95%, *: statistically significant at 90%)
Table 7: Independence and Aggregation Model, First Stage Parameter Values

<table>
<thead>
<tr>
<th>Interaction Term</th>
<th>Parameter Value (Std. Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Marginal Utility of Driving Parameters</strong></td>
<td></td>
</tr>
<tr>
<td>(HP/weight) * urban</td>
<td>-0.421 (1.123)</td>
</tr>
<tr>
<td>Vehicle Size * household size</td>
<td>1.675*** (0.082)</td>
</tr>
<tr>
<td>Wheelbase* urban</td>
<td>-1.692*** (0.144)</td>
</tr>
<tr>
<td>Household size/# vehicles</td>
<td>-3.328*** (0.069)</td>
</tr>
</tbody>
</table>

| **Indirect Utility of Driving Parameters**   |                             |
| Green* income                                | -0.006 (0.169)              |
| Green* Pacific                               | 0.023 (0.287)               |
| Green*New England                            | 0.049 (0.248)               |
| Size* Income                                 | -0.013 (0.139)              |
| Size* Pacific                                | -0.059 (0.209)              |
| Size* Mountain                               | -0.066 (0.232)              |
| Domestic* Income                             | -2.411*** (0.582)          |
| **Domestic*Midwest (West North Central)**    | 0.123* (0.067)              |
| Domestic*Midwest (East North Central)        | 0.042 (0.146)               |
| European* Income                             | -2.371*** (0.591)          |
| Japanese* Income                             | -2.394*** (0.586)          |
| Japanese* Pacific                            | 0.061 (0.090)               |
| Age* Income                                  | -0.003 (0.055)              |
| Size* Urban                                  | 0.339*** (0.065)            |
| **Std. Dev. of Error Term**                  | -1.489*** (0.031)           |
| Elasticity of Demand for Gasoline, ΔPrice = 1%| -0.992*** (0.003)          |

Mean (Standard Deviation)

(***: statistically significant at 99%, **: statistically significant at 95%, *: statistically significant at 90%)

6.3 Second Stage Results

In the second stage estimation, I recapture the fixed effects through the imposition of an optimality condition in order to estimate the mean utility of certain vehicle characteristics. Since \( y = g_0 + g_1Z_i \), \( g_0 \) must contain the same parameter vector as the interaction terms in \( g_1Z_iX_j \). Thus, once I have estimated the \( \theta_j \) through \( J \) separate probit maximizations, I regress them on the parameter vector \( g_0 = (\text{green}, \text{size}, \text{Domestic}, \text{European}, \text{Japanese}, \text{age}) \) and a vector of ones. Results from this regression are given in Table 8. These results are intuitive-
individuals value “green” vehicles, larger vehicles, and newer vehicles. Japanese vehicles impart the highest utility relative to domestic and European vehicles.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Parameter Value (Std. Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>16.086*** (1.231)</td>
</tr>
<tr>
<td>Green</td>
<td>0.869** (0.321)</td>
</tr>
<tr>
<td>Size</td>
<td>1.081*** (0.435)</td>
</tr>
<tr>
<td>Domestic</td>
<td>2.941*** (1.012)</td>
</tr>
<tr>
<td>European</td>
<td>1.181* (0.876)</td>
</tr>
<tr>
<td>Japanese</td>
<td>7.777*** (1.452)</td>
</tr>
<tr>
<td>Vehicle Age</td>
<td>-0.512*** (0.035)</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.675</td>
</tr>
</tbody>
</table>

With this information, it is possible to estimate a long-run elasticity, where households are allowed to alter their vehicle bundle when gasoline prices change. In order to estimate the long-run elasticity, it is necessary to integrate out over the distribution of \( \epsilon_{ij} \) in order to rank the different bundle choices. Because of the high dimensionality that occurs if I allow households to have more than 2 vehicles, it would be necessary to restrict the sample to households with only 2 vehicles, which is the main portion of the NHTS dataset. Once I have the utility levels of all different bundles, I could estimate how a change in gasoline prices would lead individuals to change the vehicles they have in their garage.

7. Policy Impact
As I stated in the introduction, correctly identifying the elasticity of demand for gasoline is crucial to many policy outcomes. Policies, such as gasoline taxes or carbon taxes, depend on the elasticity estimate for correct outcome forecasts. I have demonstrated that there exists a substantial bias in the elasticity estimates when one does not account for bundling or high dimension of a choice set. Thus, since the large part of the literature on the subject has either aggregated the choice set or restricted the household’s decisions to one vehicle at a time it is possible that their estimates are biased down to zero by 20%. Parry and Small (2005) analytically calculate an optimal gasoline tax that maximizes social welfare and government
revenue, deals with congestion externalities, and minimizes pollution. The optimal tax is shown in the following equation (p. 1279):

$$t^*_f = \frac{MEC_f}{1 + MEB_f} + \frac{(1 - \eta_{ML})c^e_{LL}}{\eta_{FF}} t_L (q_F + t_F) + \frac{\beta M}{F} E^c \left\{ \epsilon_{LL} - (1 - \eta_{ML})c^e_{LL} \right\} \frac{t_L}{1 - t_L}$$

\textit{Adjusted Pigovian Tax} \hspace{2cm} \textit{Ramsey Tax} \hspace{2cm} \textit{Congestion Feedback}

The second portion of the optimal gasoline tax, the Ramsey tax, depends directly on the negative of the elasticity of demand for gasoline, $\eta_{FF}$. If this estimate is biased by 20%, (i.e. the true parameter $\eta_{FF} = 1$, but because of the bias, we use $\hat{\eta}_{FF} = 0.8$), then the misestimated Ramsey tax will be 1.25 times larger than it should. This is due to the fact that when we tax goods that are more elastic, there is a deadweight loss associated with taxation, and less government revenue is collected.

On the other hand, many policy makers view the purpose of the gasoline tax as a method with which to decrease gasoline demand, and rely on a low estimate of elasticity to argue against taxation. This would lead them to choose another method, such as CAFE standards that increase average fuel efficiency in the market. However, if the elasticity estimate is in fact elastic, then increasing the fuel efficiency may actually be counter-productive. As overall fuel efficiency increases, it becomes relatively cheaper to drive more, and if demand is elastic, then there will be a larger rebound effect, increasing VMT and gallons consumed.

Thus, not only can underestimating consumers’ response to changing gasoline prices lead policy makers to choose sub-optimal policies, the resulting consumer response could also be counterproductive to the policy.

8. Conclusion
This paper has tested the impact of a change in gasoline price on household driving behavior through a discrete-continuous household fleet model. Households make their driving decisions based on the fleet of vehicles they have in their garage- how relatively comfortable, strong,
large, and fuel efficient each of their cars is will affect how they choose to allocate vehicle miles to each car. I find that by allowing households to optimize over all their vehicles increases the elasticity of demand for gas, relative to assuming independence between the vehicles in the household’s garage. Thus, I have confirmed that the independence assumption causes an important downward bias in the calculated elasticity, although this bias is not as strong as the bias created by aggregation of the choice set. By widely increasing the choice set, I allow households to make larger changes along the extensive margin by substituting between the vehicles in their garage as operating costs change, demonstrating that households are more elastic than previously believed.

My research presents a more accurate understanding of how households make vehicle fleet purchase decisions, and how they respond to changes in gasoline prices. By using a discrete-continuous household model that allows for bundles and unobserved consumer heterogeneity, and a method that does not artificially restrict the choice set, I have found an elasticity of demand for gasoline of -1.108. Furthermore, I have demonstrated that by aggregating the choice set and not allowing for households to have bundles, many researchers may have up to a 22.3% downward bias on their estimate.
References


