Household Vehicle Bundle Choice and Gasoline Demand: A Discrete-Continuous Approach

Elisheba Spiller

Duke University, Resources for the Future

Abstract

Accurately estimating the price elasticity of demand for gasoline is fundamental in many different policy settings, in that how consumers respond to changing gas prices significantly affects the outcome of policy forecasts and debates. This paper simultaneously addresses four common modeling assumptions frequently used in gasoline demand estimation that cause the researcher to underestimate this elasticity. These are (i) ignoring the role of bundle effects in the driving decision; (ii) over-aggregation of the choice set; (iii) ignoring the inter-relatedness of the decisions about how much to drive and what vehicles to purchase; and (iv) failing to account for unobserved vehicle attributes. While methods typically found in the literature can only deal with a subset of these issues at a time, the empirical technique I employ (which is based on revealed preference inequalities) easily allows for an integrated analysis that deals with all of these concerns. Furthermore, this approach allows me to test the impact of each of these assumptions, both independently and jointly, on the elasticity estimate. By utilizing disaggregate household data from the National Household Transportation Survey in 2001 and 2009, I demonstrate that the researcher may underestimate the elasticity up to 66% if these aspects of the vehicle purchase and driving decisions are not taken into consideration.

1. Introduction

Negative 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 (e.g., no dashboard technology exists to measure each vehicle'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 changes in operating costs.

Correctly identifying the elasticity of demand for gasoline is essential for these policy decisions. Depending upon how elastic consumer demand is, one set of policies may make more sense than another: if the demand for gasoline is elastic, higher gasoline taxes will reduce gasoline consumption[1]. 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[2]. There are other policy issues that also depend on the correct measurement of the elasticity of demand for gasoline. For example, the elasticity of demand is used as a critical

[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.
[2] See Parry and Small (2005).

input in certain climate policy models. Thus, if it is measured incorrectly, it may lead to bad forecasts and suboptimal policies. 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 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 vehicle purchasing behavior as gasoline prices rise. Even before gasoline prices spiked between 2007 and 2008, nearly 50% of vehicle 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.

Second, and even more important, there has been a consistent misspecification by researchers measuring the elasticity of demand for gasoline to such an extent that their elasticity estimates may have been impacted. This misspecification includes not taking into account households' bundle of vehicles, not modeling the vehicle choice jointly with the vehicle use decision, and aggregating the choice set to facilitate estimation. My model and methods better

[3] See Espey (1996), Dahl and Sterner (1991) for surveys on elasticity measurements.

reflect the household's vehicle driving and purchase decisions and thus more accurately capture the elasticity of demand for gasoline.

1.1 Contribution to the Literature

My research intends to not only correct some of these misspecifications in the calculation of elasticity of demand for gasoline, but also to quantify the impact that these have on the elasticity estimate.

First, my model takes into account both how much people drive (the intensive margin) and what types of vehicles they decide to purchase (the extensive margin). These two margins are intimately related: how many miles we need to drive influences the type of vehicle we buy, and the type of vehicle we own will also affect how much we drive it. If these two decisions are not modeled jointly, the resulting underlying preference parameters may be misestimated, leading to a distortion in the elasticity of demand for gasoline. 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 prices than is the case. Even if this does not impact 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 vehicle, given the type (or types) of vehicle that it owns, and allows for a simultaneous estimation of both these margins.

Second, I allow the household to own a bundle of vehicles from which it optimally allocates vehicle miles travelled (VMT) to each vehicle. The relative efficiency of the vehicles in a household's bundle 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 would not capture the full response of the household, and would ignore the household's ability to adjust along this extra margin. Thus, modeling the VMT choices within a bundle as independent decisions may underestimate 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 bundle 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. Since an independence assumption may cause the researcher to underestimate the elasticity of demand for gasoline, I specifically model the household's choice of bundle and allow for the household to allocate VMT optimally across its entire vehicle holdings. Furthermore, given the flexibility of the method, I am able to identify the impact on the elasticity estimate from assuming independence between the vehicles in the household's garage.

Third, I include in my utility function specification vehicle specific fixed effects. Adding these alternative specific constants allows me to capture attributes of the vehicles 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 vehicle, which, if ignored in estimation, make it seem that consumers value high prices or are less price sensitive[4]. Berry *et al.* (1995) (from here out BLP) stress that not including product-specific fixed effects causes a bias on vehicle price:

> If producers know the values of ξ (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 producers and the consumers…, then equilibrium…prices will depend upon the disturbances. (p.9-10)

Thus, these constants capture unobserved attributes that impact the vehicle purchasing decision, thereby minimizing the bias on the vehicle price coefficient. The same sort of analysis holds for gasoline prices- if a less fuel efficient vehicle is chosen more frequently than a more fuel efficient vehicle (everything else equal), the researcher may attribute this higher market share to a negative preference for fuel efficiency instead of an unobserved attribute (such as comfort or style) that is actually driving the purchase decision. This would cause the researcher to underestimate[5] how sensitive individuals are to gasoline prices in the long-run.[6] In the short-run, the elasticity estimate may also be impacted, in that excluding vehicle specific fixed effects creates a distortion in the underlying preference parameters that influence this estimate. However, vehicle specific fixed effects have been largely absent from the relevant literature, which distinguishes my method from those previously used by allowing me to capture these

[4] The bias can also go the other way – if consumers purchase the less expensive vehicle in much greater quantities, then they may appear to be more price sensitive than they actually are if unobserved vehicle attributes are not taken into account.

[5] As in the vehicle price problem, the bias can also go the other way- if individuals "over" consume the hybrid relative to the non-hybrid, it may be that individuals are choosing the status of the hybrid or the environmentally friendly image it provides them rather than the fuel efficiency gains. Thus, the direction of the bias needs to be estimated with data.

[6] Long-run elasticities allow households to reoptimize over the choice of vehicle bundle, while short-run elasticities maintain fixed the vehicle holdings and allow only a reoptimization over the VMT decision.

important unobserved vehicle attributes. Furthermore, my method allows me to test the impact of the inclusion or exclusion of these fixed effects on the elasticity estimate.

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 SUV to a more efficient SUV) 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, causing the elasticity of demand for gasoline to be underestimated in the long-run. In essence, this aggregation misspecifies the choice set faced by the consumer at the time of purchasing the vehicle, which can cause the researcher to misestimate the underlying preference parameters that impact 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, while at the same time better representing the household's true decision making process. I also am able to quantify the impact of a certain level of aggregation previously used in the literature on the elasticity impact.

However, expanding the choice set generates a significant problem in estimation. If households can choose up to 3 vehicles, and my dataset includes 3,751 distinct vehicle-year choices, this means that the household has ${}_nC_k$ choices = 8,789,061,875 possible different vehicle 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 vehicle specific fixed effects, this would imply that I would have to estimate 3,751 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 estimate using 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 provides some policy analysis; 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)[7]. 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

[7] Schmalensee and Stoker estimate an elasticity range of -0.29 to -1.13; West and Williams find a range of -0.18 to -0.738 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.

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 vehicle 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 (extensive margin) in a nested logit model, and then uses these parameters to control for the endogeneity of vehicle choice in the energy use decision (intensive 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 inelastic demand for gasoline, which may be due in part to their restrictive choice set assumptions[8].

There have been some papers that use a simultaneous, one-step approach to estimating both VMT and vehicle 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 bundle 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

[8] 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.

maximization framework through the use of a random-parameter Bayesian estimation technique. However, they assume that each vehicle is independent from all other vehicles in the household's bundle—an assumption that also may impact their result, which is an inelastic demand for gasoline of -0.35. In my study, I do not assume that the vehicles 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 minimize the distortions on the estimation of the elasticity of demand for gasoline.

While most papers do not take into account the household's bundle and thus model vehicle choices as independent, Green and Hu (1985) estimate a bundle model, where they look at how households substitute between their vehicles 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 vehicles. 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 few 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 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 or vehicle purchases, 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 2009 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 differences arise because of the data used, the assumptions employed, the demand function specification, and the econometric technique. Thus, the downward effect 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 impact on elasticity by running my model with the same specification employed by Bento *et al.* (2008) and comparing these results to the same 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 bundle level.

3. Model

My analysis is conducted at the household level. I assume each household optimally chooses how many vehicles to own, what the mix of models should be, and how much to drive each vehicle. Furthermore, the household decision on what types of vehicles to own and how much to drive depend on household characteristics, vehicle characteristics, and the current price of gasoline. Thus, the household is assumed to be at an optimal point with respect to the vehicles they own and how much to drive[9]. Since optimization is at the household bundle level, there can be substitution between vehicles within the household.

I begin with a baseline model, adapted from Bento *et al.* (2008)[10], which is based on the following indirect utility function:

$$V_{ij} = \frac{-1}{\lambda_i}\exp\left(-\lambda_i\left(\frac{y_i}{T_i} - P_j^u\right)\right) - \frac{1}{\beta_{ij}}\exp\left(\alpha_{ij} + \beta_{ij}P_{ij}^d\right) + \tau_{ij} + \varepsilon_{ij}$$

$$\begin{aligned}
\alpha_{ij} &= \alpha z_{ij}^{\alpha} \\
\beta_{ij} &= -\exp\left(\beta z_{ij}^{\beta}\right) \\
\lambda_i &= \exp\left(\lambda z_i^{\lambda}\right) \\
\tau_{ij} &= \tau z_{ij}^{\tau} \\
\varepsilon_{ij} &\sim N\left(0, \sigma^2\right)
\end{aligned}$$

where *i* denotes household and *j* denotes vehicle; $\left(z_{ij}^{\alpha}, z_{ij}^{\beta}, z_{ij}^{\tau}\right)$ are alternative automobile characteristics interacted with household demographics; z_i^{λ} contains just household

[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 (of static models).

[10] The main differences between my model and Bento *et al.*'s model are that they include random coefficients and normalize by the Hicksian composite commodity price, while I do not.

characteristics; T_i is number of adults in the household; P_j^u is the vehicle's used price[11]; P_{ij}^d is the vehicle's operating cost (gasoline price divided by the vehicle's fuel efficiency); and y_i is the household's income. The interaction terms allow different households to value vehicle characteristics in dissimilar manners- for example, this specification lets households with more children prefer larger vehicles to smaller vehicles. Through Roy's Identity, this indirect utility function implies that the optimal VMT for each vehicle is as follows:

$$VMT_{ij}^* = \exp\left(\alpha_{ij} + \beta_{ij} P_{ij}^d + \lambda_i \left(\frac{y_i}{T_i} - P_j^u\right)\right)$$

While this indirect utility function is based on vehicle and household characteristics, it does not include fixed effects and the resulting optimal VMT does not allow for substitution between the household's bundle of vehicles. In order to expand this to a more flexible specification, I follow Feng, Fullerton and Gan (2005) and adapt the utility function in order to allow for cross-price effects between the vehicles in the garage, along with including an additively separable vehicle specific fixed effect. However, I make a slight change: instead of normalizing the second term in the indirect utility function by $\frac{1}{\beta_{i1}}$, I normalize by $\frac{1}{\sum_{k=1}^{J_i} \beta_{ik}}$, where J_i is the bundle owned by household *i*. [12] Thus, the resulting indirect utility function is:

$$V_i^* = -\frac{1}{\lambda_i} \exp\left(-\lambda_i \left(y_i - \sum_{k=1}^{J_i} P_k^u\right)\right) - \frac{1}{\sum_{k=1}^{J_i} \beta_{ik}} \exp\left(\sum_{k=1}^{J_i} \left(\alpha_{ik} + \beta_{ik} P_{ik}^d\right)\right) + \sum_{k=1}^{J_i} \left(\theta_k + \tau_{ik} + \varepsilon_{ik}\right)$$

[11] This can be thought of as the opportunity cost of not selling the vehicle- although the household might not purchase their vehicles in every period, they choose not to sell it on the market.

[12] β_{i1} is the β_{ij} that corresponds to an arbitrarily selected vehicle in the garage.

where vehicle specific fixed effects are denoted by θ_j and the other variables are defined as before. In this specification, $\theta_j = \gamma z_j^{\gamma} + \xi_j$, where z_j^{γ} are vehicle characteristics, and ξ_j are vehicle specific unobservable attributes. This indirect utility function implies the following VMT demands through Roy's identity:

$$VMT_{ij}^{*} = \frac{\beta_{ij}}{\sum_{k=1}^{J_i} \beta_{ik}} \exp\left\{ \sum_{k=1}^{J_i} \left(\alpha_{ik} + \beta_{ik} P_{ik}^{d} \right) + \lambda_i \left(y_i - \sum_{k=1}^{J_i} P_k^{u} \right) \right\}$$

This model therefore allows the driving decisions of each vehicle to depend not only on that vehicle's characteristics and operating costs, but also on that of the other vehicles in the household's garage. Furthermore, both the baseline model and this extension produce identical optimal driving patterns for households who own only one vehicle.

The indirect utility function has an error term that is the summation of all the household-vehicle specific error terms associated with each household's chosen bundle. However, this resulting error term is not bundle specific. That is, it is not an idiosyncratic error term associated with the particular bundle, since a different household owning the same bundle would not have the same bundle error term. Thus, I do not assume that a particular bundle of vehicles delivers a specific shock to the household who owns it. Instead, this grouping of error terms creates a correlation between the observed bundle and any unilateral deviation from that bundle (which will be important in the next section). Specifically, a household will be more likely to move from one bundle to another bundle that differs only by one vehicle. Thus, the correlation between error terms is explicitly incorporated into the model and is across households, and not across vehicles or vehicle bundles.

One downside to this model is the fact that any two households who look alike and who own the same bundle 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 avoid overestimating the elasticity of demand. Therefore, I extend my model to incorporate unobserved consumer heterogeneity in Section 4.5.

3.1 Model Benefits and Contributions

This discrete-continuous model of household behavior allows me to model not only the substitution patterns between vehicles in a household's garage, but also to explicitly measure the impact on the estimate of elasticity of demand for gasoline caused by different research assumptions. Since this model allows for a very detailed choice set and bundles, it also allows for more flexible specifications. I can relax the detailed choice set assumption by aggregating across vehicle types, as I will show later in the paper. In doing this, I can also recover aggregate type fixed effects instead of vehicle specific fixed effects. The assumption on bundles can also be relaxed by assuming that each vehicle observation is a separate household, thus imposing independence between the vehicles in the household's garage. Neither of these assumptions changes the underlying specification or method employed in this paper, thus allowing me to measure exactly how they affect the estimate of the elasticity of demand for gasoline.

Also, very few papers have estimated vehicle specific fixed effects in a bundle model. By allowing for fixed effects at the model-year level, I add significantly more variation along the extensive margin than has previously been accounted for. However, since the method is so flexible, I can also run the model without fixed effects to see if this affects the elasticity estimates.

Furthermore, I present a discrete-continuous model that improves upon previous methods. The Dubin-McFadden (1984) method, which has been used extensively in discrete-continuous choice applications, presents two fundamental problems in this application. First, allowing for bundles in the household would normally require the researcher to assume a bundle-specific error term[13]. Second, the Dubin-McFadden two-step sequential estimation does not account for the cross-equation restrictions that are present in a unified model of vehicle and VMT choice.

Similarly, the method I employ improves upon others which allow for household bundling of goods under a high dimensional choice set. McFadden *et. al.* (1978) demonstrate that if the I.I.A. property[14] 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 problems for the bundle model that my method avoids. First, by construction, this method imposes the I.I.A. property, which has an unfortunate outcome in

[13] As is also true in the subsampling of alternatives method presented by McFadden et. al. (1978). A more in depth discussion of this method and the problems of bundle-specific error terms can be found in Section 4.1.
[14] I.I.A. is the "independence of irrelevant alternatives" outcome that emerges from logit probabilities.

vehicle choice, 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. That is, 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, and can be problematic under a vehicle choice model. Second, in order to use the subsampling of alternatives technique with bundles, the researcher would have to assume independent bundle-specific error terms. However, assuming independence between the error terms of all bundles is a strong, and perhaps, unrealistic assumption. This assumption maintains that each bundle has a certain idiosyncratic shock associated with its purchase that is in no way related to a different bundle, even if the two bundles share the majority of the vehicles. If this is not the case, in that these error terms are in fact correlated across bundles that share vehicles, then the error terms are no longer i.i.d. and violate the assumptions of the logit model. While the I.I.A. assumption can be relaxed through nested logit, nested logit still would need to assume bundle-specific error terms in order to accommodate bundles. My technique, as I will describe in more detail in the next section, does not exhibit correlated error terms, thus maintaining my i.i.d. assumption throughout the estimation.

Finally, this method can be used in many other discrete-continuous applications, such as the fuel-use decisions of households in developing countries or recreation demand (the decisions being where to visit and how often to visit). This method has been applied to several discrete choice applications but, as of now, has still not been applied to discrete-continuous choices. Thus, I am improving upon this literature by presenting a method that not only improves upon previous techniques, but can be applied to many diverse settings.

4. Estimation

4.1 Overview of Estimation Procedure

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 or 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 not, and I allow for a continuous element in the decision process.

In order to use this revealed preference approach, I need to assume that the household's vehicle 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 the vehicle specific fixed effects have been differenced away, I am able to focus on estimating the underlying preference parameters that drive the vehicle choice and driving decisions. I then proceed to calculate the short-run elasticity- through Roy's identity, I identify the optimal VMT for each household's vehicle under the original gasoline price and a small increase in gasoline price, divide by MPG to calculate gallons demanded, and then find the mean elasticity of all households. This step allows me to focus on the short-run elasticity, as calculating a long-run elasticity is outside the scope of this paper.

4.2 Estimation

In the estimation procedure, I model the purchase decision in a probit framework. This identifies the model parameters $\Omega = (\alpha, \beta, \lambda, \tau)$, which describe how households value vehicle characteristics and consumption, and how household characteristics affect vehicle choice. The revealed preference approach assumes that households are observed in a long-run equilibrium. That is, although their vehicles may have been bought at different times, the bundle they own now is as optimal as if they had just bought all their vehicles today. The identifying assumption is thus the following:

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

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

$$V_{iJ_i^*} \geq V_{iJ_i} \quad \forall J_i \neq J_i^* \tag{1}$$

where J_i^* is the bundle of vehicles chosen by household *i*. Since any change in the household's observed vehicle holdings diminishes its indirect utility, this also holds if I only change one vehicle in the household's bundle. 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 (θ_j) for the vehicles 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 θ_j from the vehicle that is being swapped out of the bundle on the left hand side of equation (1), and the θ_j from the vehicle that is being added to the bundle 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 θ_j from the estimation. The idea is to take two households who have different vehicles, and swap one of their vehicles with one from the other household. For example, assume we have two households: 1 and 2, that each has a bundle with the respective vehicles A, B, and C, D. Equation (1) can thus be rewritten in the following manner:

$$\widetilde{V}_{1A,B} + \theta_A + \theta_B \geq \widetilde{V}_{1B,C} + \theta_B + \theta_c$$

$$\widetilde{V}_{2C,D} + \theta_D + \theta_C \geq \widetilde{V}_{2A,D} + \theta_A + \theta_D$$

where $\widetilde{V}$ is the part of the indirect utility that does not include the fixed effect. The first equation compares household 1's original bundle A and B with a proposed deviation to a bundle that includes B and C. Similarly, the second equation compares household 2's original bundle C and D with a proposed deviation: A and D. In essence, I am "swapping" vehicles A and C across households 1 and 2.[15] Once this swap is complete, I can add across these two inequalities, and all the fixed effects cancel out from each side of the equation, leaving only the $\widetilde{V}$:

$$\widetilde{V}_{1A,B} - \widetilde{V}_{1B,C} + \widetilde{V}_{2C,D} - \widetilde{V}_{2A,D} \geq 0 \tag{2}$$

Thus, the probability that a random vehicle swap between two households makes each household worse off is the probability that equation (2) holds. Under the baseline model with no vehicle specific fixed effects, this swap is unnecessary- instead, equation (1) would replace equation (2) in the likelihood function, where the right hand side is another vehicle in the dataset that has been randomly chosen. Rearranging equation (2), and since the composite error term $\widetilde{\varepsilon}$ is distributed normally[16], this gives:

$$\begin{aligned}\Pr\left(\widetilde{V}_{1A,B} - \widetilde{V}_{1C,B} + \widetilde{V}_{2C,D} - \widetilde{V}_{2A,D} \geq 0\right) &= \Pr\left(\widetilde{\varepsilon}_{12,ABCD} \leq \overline{V}_{1A,B} - \overline{V}_{1C,B} + \overline{V}_{2C,D} - \overline{V}_{2A,D}\right) \\ &= \Phi\left(\frac{\overline{V}_{1A,B} - \overline{V}_{1C,B} + \overline{V}_{2C,D} - \overline{V}_{2A,D}}{2\sigma}\right) \end{aligned} \tag{3}$$

Where $\overline{V}_{iJ_i}$ is the deterministic part of the indirect utility that does not include any θ_j . I can set up many of these inequalities- at least one for each vehicle in every household's bundle- and create a log likelihood that any random vehicle swap between two different households makes each household worse off[17]:

$$LL(\beta) = \sum_{swaps_{(i_1,i_2)(j_1,j_2)}} \log \Phi\left(\frac{\overline{V}_{i_1J_{11}} - \overline{V}_{i_1J_{12}} + \overline{V}_{i_2J_{21}} - \overline{V}_{i_2J_{12}}}{2\sigma}\right) \tag{4}$$

[15] This swap does not mean that each household negotiates with each other to exchange vehicles. Instead, it is tantamount to identifying an alternative reality where the households had chosen these alternative vehicles instead of the vehicles they actually chose in the data.

[16] Since each $\varepsilon_{ij} \sim N(0,\sigma^2)$, then the sum over the four error terms associated with this double inequality will be distributed normally with variance equal to $4^*\sigma^2$.

[17] The normalization in this case is on the fixed effects: I set $\theta_1 = 0$.

I use Matlab and a simplex technique to maximize this likelihood function with the use of parallel processing over 8 virtual machines. This allows me to eliminate the fixed effects and estimate the parameter vector.

The procedure is thus as follows:

1. Guess at the parameter vector $\Omega = (\alpha, \beta, \lambda, \tau)$
2. For each vehicle observation in the dataset:
   a. Randomly choose a vehicle from another household.
   b. Swap the chosen vehicles between the two households.
   c. Calculate the indirect utility under each scenario (observed and proposed) for each household- $\overline{V}_{i_1 J_{11}}, \overline{V}_{i_1 J_{12}}, \overline{V}_{i_2 J_{21}}, \overline{V}_{i_2 J_{12}}$.
   d. Difference the indirect utilities, and add across the two resulting preference inequalities.
3. Calculate the objective function (i.e. the summed log of differences).
4. Find the next parameter vector, Ω', that increases the objective function.
5. Repeat until a sufficient measure of convergence has been reached.

However, the swapping is done in a particular way in order to avoid creating correlation between the error terms in the household. Suppose household 1 swaps with household 2, and household 2 swaps (the same vehicle) with household 3. Both composite error terms in these swaps include ε_{2j} . Thus, the very nature of the cross-household swaps imposes correlation through the idiosyncratic household-vehicle error term included in the composite error term. I can avoid this by not using the same vehicle-household combination in more than one swap.

While this will decrease the number of observations in my final likelihood, it eliminates any correlation between the error terms, and I am thus able to consistently estimate the parameter vector with this pair-wise estimation technique.

By eliminating the fixed effects, I am able to estimate the parameter vector Ω 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.

In order to estimate the elasticity under an assumption of a utility function with no vehicle specific fixed effects, it is unnecessary to swap vehicles between households. Instead, I evaluate the preference inequality for each household separately, by randomly choosing alternative vehicles for each observation and comparing the indirect utility under the observed and the proposed deviation. In this setting, each preference inequality's resulting idiosyncratic error term is necessarily i.i.d. from all others, so it is not necessary to adjust the style of deviations in order to avoid correlations among error terms.

4.3 Parameters and Short-run Elasticity

The parameter vector Ω tells us how households respond to certain vehicle characteristics, and indirectly, how these affect the optimal driving decision. Thus, once this vector has been estimated, it is now possible to simulate a short-run elasticity. A short-run elasticity is defined in this paper as how individuals immediately respond to gasoline prices – they can adjust their gasoline consumption, either through VMT substitution between the vehicles in their garage, or by diminishing overall VMT. In the long-run, households can reoptimize to a new vehicle bundle, which will also change their gasoline consumption. However, in order to simulate a long-run elasticity it is necessary to calculate indirect utility under *all* possible bundles. Therefore, simulating a long-run elasticity has a methodological hurdle that is non-trivial to overcome. In order to choose the optimal new bundle of vehicles, the household's utility must be compared to the universe of bundles available to the household. While this simulation is outside the scope of the paper, it is nonetheless an important step to understanding how households react to changing gasoline prices.

In order to simulate the short-run elasticity, once the preference parameters have been recovered, it is necessary to calculate optimal VMT under two scenarios- the original gasoline price, and a 1% increase in gasoline prices – and divide these VMTs by the fuel efficiency of each vehicle in the bundle in order to calculate total gallons consumed. Then the calculation of total gallons demanded and elasticity is straightforward:

$$gallons_0 = VMT^*_{iJ_i^*}\left(gasoline\_price_0\right)/\,mpg_{J_i^*}$$

$$gallons_1 = VMT^*_{iJ_i^*}\left(gasoline\_price_0 * 1.01\right)/\,mpg_{J_i^*}$$

$$\varepsilon = \frac{1}{N}\sum_i^N \frac{gallons_1 - gallons_0}{gas\_price_1 - gas\_price_0} * \frac{gas\_price_0}{gallons_0}$$

where $VMT^*_{iJ_i^*}$ and $mpg_{J_i^*}$ are vectors of VMT and miles per gallon for all vehicles in a household's bundle. Thus, I am able to simulate a short-run elasticity for each household in the dataset, and my resulting elasticity estimate is an average elasticity over the entire sample.

4.4 Extension: Unobserved Consumer Heterogeneity

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 households drive- such as long commutes to work. A household with a long commute time not only will choose a more comfortable vehicle in order to increase the marginal utility of that drive, but will also probably steer towards a more fuel efficient vehicle (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 impact 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 away from zero- 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 overestimate the elasticity of demand for gasoline, I incorporate information on how much households actually drive each vehicle so that different households can vary in their preference for driving. In order to do this, I non-parametrically recover λ_i for each household. That is, instead of parameterizing λ_i as an interaction between a parameter and a household characteristic, a singular value will be recovered for each household. Thus, each household will care differently about consumption relative to driving, which can result in a higher or lower driving pattern which is idiosyncratic to each household. The optimal driving for each vehicle will therefore be the following:

$$VMT_{ij}^{*} = \frac{\beta_{ij}}{\sum_{k=1}^{J_i} \beta_{ik}} \exp\left\{ \sum_{k=1}^{J_i} \left(\alpha_{ik} + \beta_{ik} P_{ik}^{d} \right) + \tilde{\lambda}_i \left(y_i - \sum_{k=1}^{J_i} P_k^{u} \right) \right\} \tag{6}$$

where $\tilde{\lambda}_i$ is the recovered value that varies across each household.

In order to recover $\tilde{\lambda}_i$, I implement an extra step at each iteration of the maximum likelihood estimation routine where I find the $\tilde{\lambda}_i$ that solves the following equation:

$$\left(\frac{1}{J_i} \sum_{k=1}^{J_i} VMT_{ij}^{*} - \frac{1}{J_i} \sum_{k=1}^{J_i} \overline{VMT}_{ij} \right)^2 = 0 \tag{7}$$

where J_i is the number of vehicles household i owns, VMT_{ij}^{*} is the optimal driving as defined in equation (6), and $\overline{VMT}_{ij}$ is the observed driving behavior as detailed in the data. Equation (7) implies that $\tilde{\lambda}_i$ 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 idiosyncratic household value[18]. Furthermore, this idiosyncratic term affects all parameters, in that it is an integral part of the indirect utility function, and its value depends on a subset of the parameter vector: α and β. Thus, $\tilde{\lambda}_i$ can be described in the following way: $\tilde{\lambda}_i = f(\alpha, \beta, z_{ij})$. That is, $\tilde{\lambda}_i$ is a function of observable vehicle and household characteristics and a subset of the parameters which I am recovering in the maximum likelihood routine. Thus, it is not analogous to a fixed effect. Instead, it is more analogous to a correlated random effect and therefore it does not create an incidental parameters problem in identifying α and β.

[18] 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.

In order to nonparametrically recover $\widetilde{\lambda}_i$ and incorporate it into the estimation routine, I implement the following iterative procedure:

1) Guess at the parameter vector $\Omega_1 = (\alpha, \beta, \tau)$.
2) Find the value of $\widetilde{\lambda}_i$ that makes equation (7) hold for each household.
3) Given current Ω_1 and $\widetilde{\lambda}_i$, calculate the indirect utility function $V_i^*\left(\widetilde{\lambda}_i, \Omega_1, z_{ij}\right)$ as described in Section 3.
4) Proceed with estimation as outlined in Section 4.2 – find a new parameter vector $\Omega_1^{'}$ that increases the objective function.
5) Repeat steps 2-4 until a sufficient measure of convergence has been reached.

While this estimation routine allows for a quasi-"contraction mapping" between the heterogeneity parameters and non-heterogeneity parameters, it is only valid if $\widetilde{\lambda}_i$ is unique. If, on the other hand, $\widetilde{\lambda}_i$ has more than one solution, this iteration may converge around a sub-optimal value, thus distorting the non-heterogeneity parameter vector and subsequently incorrectly measuring the elasticity of demand for gasoline.

Fortunately, there is only one unique value of $\widetilde{\lambda}_i$ that solves equation 7. This value is:

$$\widetilde{\lambda}_i = \frac{1}{y_i - \sum_{k=1}^{J_i} P_k^u} \ln\left\{\left(\frac{\left(\frac{1}{J_i}\sum_{k=1}^{J_i}\overline{VMT}_{ij}\right)}{J_i\left(\sum_{g=1}^{J_i}\frac{\beta_{ig}}{\sum_{k=1}^{J_i}\beta_{ik}}\right)}\right)\right\} - \sum_{k=1}^{J_i}\left(\alpha_{ik} + \beta_{ik}P_{ik}^d\right)$$

Thus, since $\widetilde{\lambda}_i$ is unique and is a function of data and parameters, then the value of the unobserved heterogeneity parameter is consistent with the non-heterogeneity parameters estimated in the previous iteration.

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 thus more correctly estimate the elasticity of demand for gasoline.[19]

5. Data

The main dataset that I use in order to analyze household vehicle 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 vehicle-household observations, and the complete 2009 dataset, which is comprised of 143,084 households, and 309,163 vehicle-household observations. In 2001, the NHTS' national sample was not the complete sample, in that as many areas were missing from the dataset, they included a number of add-on areas per request of certain companies or municipalities. However, the complete 2001 dataset has certain inconsistencies across the national sample and add-on areas, and thus I was recommended by the NHTS data center to focus exclusively on the national sample. However, the 2009 dataset

[19] This procedure is not a typical maximum likelihood, since it imposes a constraint on the parameters, or a penalization on those who do not cause optimal VMT to equal observed VMT. Thus, the asymptotics of this type of estimator have still not been determined, and is part of future work.

did not have any add-on areas, making the national sample the full sample, so I was able to start with the entire dataset. Since gasoline prices were about a dollar higher in 2009 than in 2001, I wanted to make a more balanced panel between both years in order to not bias the results towards the 2009 observations. Thus, I randomly cut households (after dropping those with missing values and cleaning the sample) from the 2009 data until the sample size resembled the 2001 data sample size. Table 1.1 shows some summary statistics for the households. Because of missing data in household income, I impute income based on household size, race, number of vehicles in bundle, home ownership, and urban residence.

Detailed characteristics of the vehicles in the NHTS dataset were hand entered from the Ward's Automotive Yearbooks, which I then matched to NHTS. Since the Ward's dataset is more disaggregate than the NHTS dataset (i.e. each model is distinguished by trim level and fuel type), I had to average over the vehicles in Ward's that fit into each model. However, not every vehicle in the NHTS dataset was reflected in the Ward's Data, and vice versa, making it necessary to drop 6,733 observations from 2001 and 151,294 observations from 2009. For vehicles in NHTS that matched more than one vehicle in Ward's, if one of these vehicles had a much higher market share 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

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

The Ward's dataset includes characteristics such as wheelbase, length, mpg, horsepower, and MSRP. An OLS regression of price on vehicle characteristics yields a high R^2 and highly significant coefficients. Table 1.2 has the results of the regression: ln(price) is regressed on vehicle characteristics including dummies for origin and air conditioning, as well as size, strength and efficiency parameters. European vehicles fetch the highest price, as do more fuel efficient vehicles and larger vehicles.

Table 2: Ward's Data OLS Regression of Vehicle Characteristics

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

(***: 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[20]. The underlying assumption in this paper is that a household views the used vehicle prices in the vehicle market and the gasoline prices simultaneously, and then optimizes. Thus, the used vehicle prices reflect the level of gasoline prices at the moment of the decision- if gasoline prices are very high, this leads to low used vehicle prices for gas guzzlers, so the decision made at the household level is more optimal than if the decision had been made when gasoline prices were at a different level.

[20] 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).

Finally, I use the American Chamber of Commerce Researchers Association (ACCRA) data to find the gasoline prices in the household's MSA. The ACCRA data provide me quarterly information on gasoline prices at the city level, which I have aggregated up to the year and MSA level in order to match with the NHTS dataset. Figure 1.1 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 or household demographics, 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.

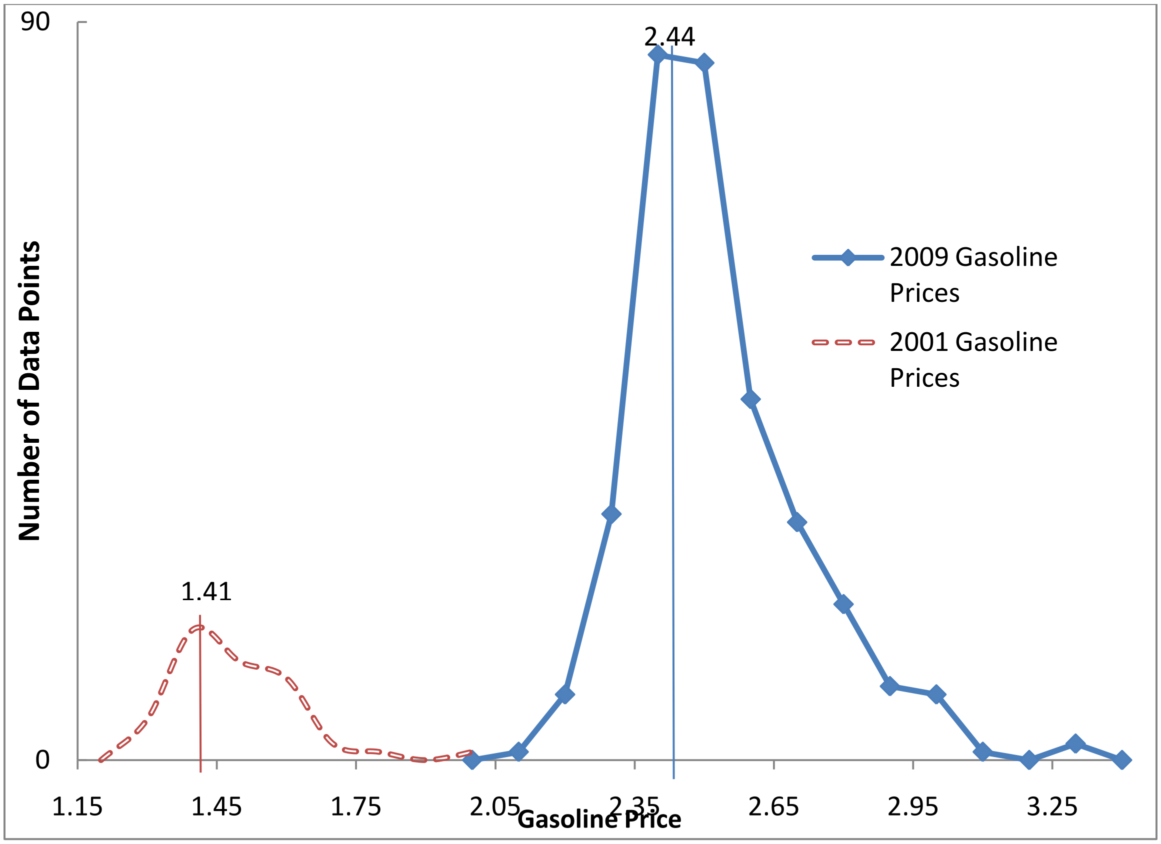

Figure 1: Gasoline Price Distributions

6. Results

6.1 Estimates from Swapping Method

In the estimation, I calculate the household type-specific parameters contained in the indirect utility of driving. Once I have estimated these utility of driving parameters, I can use the optimal VMT from the utility maximization in order to estimate the short-run elasticity of demand for gasoline.

The results for the full model are presented below in Table 1.3.

Table 3: Full Model Parameter Estimates

	Interaction Term	Parameter Value (Std. Err.)
Λ	Household size	0.113 (0.097)
	Income	-0.133***(0.050)
	Urban	-0.067 (0.431)
β	(Horsepower/weight)* urban	-0.018***(0.002)
	Wheelbase* Mountain	-0.020 (0.111)
	Green*Pacific	3.263***(0.026)
	Size* Income	-0.092***(0.000)
α	Domestic * Midwest	4.987***(0.000)
	European * Income	0.229***(0.000)
	Japanese * Pacific	0.236***(0.002)
τ	Vehicle Age * Income	-0.008 (0.026)
	Vehicle Size * Household Size	0.033* (0.019)
Std. Dev. of Error Term	σ_ε	0.211***(0.000)
Elasticity of Demand for Gasoline, ΔPrice = 1% *Mean (Std. Err.)*		-0.915***(0.262)

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

The signs of most parameters are intuitive: rural households prefer vehicles that have greater acceleration power, larger households prefer larger vehicles, and wealthier families prefer smaller, European vehicles. Furthermore, families living in the Pacific census block (one of 9 census divisions- which can be seen in Figure 1.2) prefer environmentally friendly (green) and Japanese vehicles, while Midwestern families prefer 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, which is -0.89. When gasoline prices change, households can react in the following way: in the immediate short-run, they can drive each of their vehicles less. In the medium run, they can begin to reallocate driving between the vehicles in their bundle. Finally, in the long-run, they can choose to buy different vehicles and re-optimize their driving decision over their bundle. 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 long-run 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.

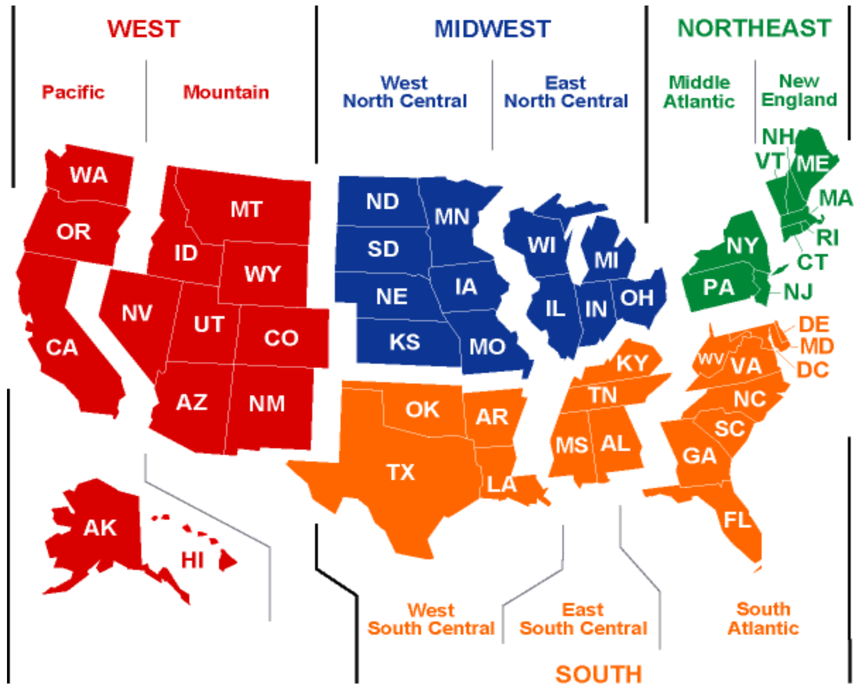

Figure 2: Census Divisions

6.2 Measuring the Impact on Elasticity Estimates

In order to quantify how aggregating the choice set and assuming independence between vehicles in the household's garage diminishes the elasticity of demand for gasoline, I estimate my model under these different assumptions. I start by estimating the baseline model, which reflects the indirect utility function utilized by Bento *et al.* In order to do this, I have to assume three things. First, I eliminate fixed effects[21]. Second, I do not allow for households to substitute between the vehicles in the garage. This is comparable to assuming independence: the household chooses how much to drive each vehicle without taking into consideration the rest of the vehicles that it owns. In order to adapt the model so as to reflect this decision making

[21] Not including fixed effects allows a one way swap instead of requiring a two way swap: for each observation in the dataset, I difference the indirect utility of the observed bundle with the unilateral deviation from that bundle. Thus, I don't have to look at two households at once, instead I evaluate each household separately

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.[22] Following Bento *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. Third, I aggregate the choice set (aggregation assumption), in the same way that Bento *et al.* aggregated their choice set. I thus aggregate the choice set into 233 different choices[23], based on the following stratification: nine classes of vehicles[24] (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). My parameter results for the baseline model estimate an elasticity of -0.30. This is very similar to the estimate by Bento *et al.*, who calculated an elasticity of -0.35.

In the second step, I calculate what happens to the elasticity when I relax either the aggregation assumption, or the elasticity assumption. The elasticity under independence but no aggregation is -0.32, which is a slight increase in elasticity. However, the elasticity under aggregation, but no independence significantly changes to -0.43, a 40% increase relative to the

---

[22] 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.

[23] While the total number of possible choices is much larger under this specification, there are only 233 choices with observed household outcomes.

[24] 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.

baseline. These results demonstrate that by not capturing more subtle behavior by individuals, be it through movements along the extensive margin or substitution between vehicles as operating costs change, we can drastically underestimate the elasticity of demand for gasoline.

In the third step, I move away from the non-fixed effects estimation. At first, I allow for independence and aggregation in order to test the impact of allowing for fixed effects on the elasticity estimate. I find that allowing for vehicle fixed effects increases the elasticity to -0.52, a 73% change in elasticity relative to the baseline. Next, I relax both the independence and aggregation assumption individually. Relaxing the aggregation assumption drops the elasticity to -0.62, which is a 106% increase in elasticity relative to the baseline model. The impacts of these different assumptions on the elasticity estimate are detailed below in Table 1.5.

Table 4: Elasticity Estimates Under Different Assumptions

	Baseline	**1**	**2**	**3**	**4**	**5**	**6**
No Fixed effects	Yes	Yes	Yes	No	No	No	No
Independence	Yes	No	Yes	Yes	Yes	No	No
Aggregation	Yes	Yes	No	Yes	No	No	No
Unobserved Cons. Het.	No	No	No	No	No	No	Yes
Elasticity estimate (Std. Err.)	-0.30 (0.08)	-0.32 (0.09)	-0.43 (0.12)	-0.52 (0.15)	-0.62 (0.17)	-0.89 (0.25)	-0.58 (0.17)
% Underestimated	66%	64%	51%	41%	29%	-	-

Since allowing for inter-household substitution of VMT increases the medium-run elasticity of demand to -0.89, this demonstrates that the assumptions of independence and aggregation cause the elasticity of demand to be severely underestimated (over 65%), although the main part of this mis-estimation 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.

However, this result changes when the household's heterogenous preferences are incorporated into the analysis. As hypothesized, allowing households to have different idiosyncratic preferences over driving and consumption decreases the elasticity of demand for gasoline. Households' inability to change their driving behavior due to fixed commute times make them less sensitive to gasoline prices, and if we don't control for their specific driving needs/preferences, then we will overestimate (by 34%) the elasticity measure.

These differences in elasticity demonstrate how important it is for the policy maker to critically analyze the paper from which they are taking their elasticity estimate. Whether the researcher has incorporated unobserved consumer heterogeneity and vehicle specific fixed effects, aggregated the choice set, or has assumed independence in order to facilitate estimation will impact the resulting elasticity estimate.

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 impact on 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 underestimated by almost 70%. 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):

$$\underbrace{t_f^* = \frac{MEC_f}{1+MEB_L}}_{\textit{Adjusted Pigovian Tax}} + \underbrace{\frac{(1-\eta_{MI})\varepsilon_{LL}^c}{\eta_{FF}} \frac{t_L(q_F+t_F)}{1-t_L}}_{\textbf{\textit{Ramsey Tax}}} + \underbrace{\frac{\beta M}{F} E^C \{\varepsilon_{LL} - (1-\eta_{MI})\varepsilon_{LL}^c\}}_{\textit{Congestion Feedback}} \frac{t_L}{1-t_L}$$

The second portion of the optimal gasoline tax, the Ramsey tax, depends directly on the negative of the elasticity of demand for gasoline, η_{FF}. If this estimate is underestimated by 66%, (i.e. the true parameter η_{FF} = 0.9, but because of the mis-estimation, we use $\hat{\eta}_{FF}$ = 0.3), then the misestimated Ramsey tax will be almost 3 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 prices on household driving behavior through a discrete-continuous household bundle model. Households make their driving decisions based on the bundle of vehicles they have in their garage- how relatively comfortable, strong, large, and fuel efficient each of their vehicles is will affect how they choose to allocate vehicle miles to each vehicle. I find that by allowing households to optimize over all their vehicles increases the elasticity of demand for gasoline, relative to assuming independence between the vehicles in the household's garage. Thus, I have confirmed that the independence and aggregation assumptions cause an important downward distortion on the calculated elasticity, although this effect is not as strong as that created by the exclusion of vehicle specific fixed effects. 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.

However, I also demonstrate that not including unobserved consumer heterogeneity can substantially overestimate the elasticity of demand for gasoline, and thus this should also be incorporated into the analysis.

The method that I have presented in this paper allows the researcher to test the impact of various assumptions on the elasticity estimate. Because of its flexibility with respect to functional form, it not only allows the researcher to test underlying assumptions, it also reconciles household bundling behavior within an otherwise impossibly large choice set. However, future research should include an estimation of the long-run elasticity in order to be able to analyze a policy such as CAFE standards, or any other that affects vehicle prices and thus will affect the types of bundles purchased.

My research thus presents a more accurate understanding of how households make vehicle bundle purchase decisions, and how they respond to changes in gasoline prices. By using a discrete-continuous household model that allows for bundles and fixed effects, and a method that does not artificially restrict the choice set, I have found an elasticity of demand for gasoline of -0.89. Furthermore, I have demonstrated that by aggregating the choice set and not allowing for households to have bundles, many researchers have underestimated the elasticity of demand for gasoline by up to 66%.

References

Bento, A., L.H. Goulder, M.R. Jacobsen, and R.H. von Haefen (2008). "Distributional and Efficiency Impacts of Increased U.S. Gasoline Taxes", *American Economic Review*, Vol. 99, No. 3, pp. 667-699.

Beresteanu, A. and F. Molinari (2008). "Asymptotic Properties for a Class of Partially Identified Models", *Econometrica*, Vol. 76, No. 4, pp. 763–814.

Berkowitz, M.K., N.T. Gallini, E.J. Miller, and R.A. Wolfe (1990). "Disaggregate Analysis of the Demand for Gasoline", *The Canadian Journal of Economics*, Vol. 23, No. 2, pp. 253-275.

Berry, S., J. Levinsohn and A. Pakes (1995). "Automobile Prices in Market Equilibrium", *Econometrica*, Vol. 63, No. 4, pp. 841-890.

Chipman, J. (1960). "The foundations of utility", *Econometrica,* Vol. 28, pp. 193-224.

Debreu, G. (1960). "Review of R.D. Luce individual choice behavior", *American Economic Review,* Vol. 50, pp. 186-188.

Dubin, J.A., and McFadden, D.L. (1984). "An econometric analysis of residential electric appliance holdings and consumption", *Econometrica*, Vol. 52, No. 2, pp. 345-362.

Espey, M. (1996). "Explaining the Variation in Elasticity Estimates of Gasoline Demand in the United States: A Meta-Analysis", *The Energy Journal*, Vol. 17, No. 3, pp. 49-60.

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

Fox, J.T. (2007). "Semiparametric Estimation of Multinomial Discrete-Choice Models Using a Subset of Choices", *RAND Journal of Economics*, Vol. 38, No. 4, pp. 1002-1019.

Goldberg, P.K. (1998). "The Effects of the Corporate Average Fuel Efficiency Standards in the US", *The Journal of Industrial Economics*, Vol. 46, No. 1, pp. 1-33.

Greene, D.L., and P.S. Hu (1985). "The influence of the price of gasoline on vehicle use in multi-vehicle households", *Transportation Research Record*, Vol. 988, pp. 19–24.

Han, A.K. (1987). "Non-Parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator", *Journal of Econometrics*, Vol. 35, pp.303-316.

Hensher, D.A. (1985). "An Econometric Model of Vehicle Use in the Household Sector", *Transportation Research Part B: Methodological*, Vol. 19, No. 4, pp.303-313.

Kelley Blue Book (2004). "Nearly 50% of Car Buyers Say Gas Prices Are Affecting Purchase Decision" http://mediaroom.kbb.com/index.php?s=43&item=34.

Klier, T. and J. Linn (2008). "The Price of Gasoline and the Demand for Fuel Efficiency: Evidence from Monthly New Vehicles Sales Data", Working Paper.

Mannering, F. and C. Winston (1985). "A Dynamic Empirical Analysis of Household Vehicle Ownership and Utilization", *The RAND Journal of Economics*, Vol. 16, No. 2, pp. 215-236.

Manski, C.F. (1975). "Maximum Score Estimation of the Stochastic Utility Model of Choice", *Journal of Econometrics*, Vol. 3, pp. 205-228.

McFadden, D., K. Train, and W. Tye (1978). "An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model", *Transportation Research Record*, Vol. 637, pp. 39-46.

McManus, W. (2007). "The link between gasoline prices and vehicle sales: economic theory trumps conventional Detroit wisdom". Munich Personal RePEc Archive, Paper No. 3463.

"Oil Prices and Vehicle Miles", calculatedriskblog.com, March 21, 2010.

Parry, I.W.H. and K.A. Small (2005). "Does Britain or the United States Have the Right Gasoline Tax?" *American Economic Review*, Vol. 95, No. 4, pp. 1276-1289.

Petrin, A. (2002). "Quantifying the Benefits of New Products: The Case of the Minivan", NBER Working Paper 8227.

Schmalensee, R. and T.M. Stoker (1999). "Household Gasoline Demand in the United States", *Econometrica*, Vol. 67, No. 3, pp. 645-662.

US Department of Energy (2009). "Motor Gasoline Retail Prices, U.S. City Average" http://www.eia.doe.gov/emeu/mer/pdf/pages/sec9_6.pdf

West, S.E. (2004). "Distributional effects of alternative vehicle pollution control policies", *Journal of Public Economics*, Vol. 88, pp. 735-757.

------------- (2007). "The Effect of Gasoline Prices on the Demand for Sport Utility Vehicles", Working Paper.

West, S.E. and R.C. Williams (2004). "Estimates from a consumer demand system: implications for the incidence of environmental taxes", *Journal of Environmental Economics and Management*, Vol. 47, pp. 535–558.

Woodyard, C. (2009) "Small used cars aren't big sellers as gas stays cheap." USA Today, Sept. 19.