Attribute Substitution in Household Vehicle Portfolios

James Archsmith	Kenneth Gillingham	Christopher R. Knittel	David Rapson*
UC Davis	Yale	MIT	UC Davis

September 12, 2017

Abstract

Household preferences for goods with a bundle of attributes may have complex substitution patterns when one attribute is changed. For example, a household faced with an exogenous increase in the size of one television may choose to decrease the size of other televisions within the home. This paper quantifies the extent of attribute substitution in the context of multi-vehicle households. We deploy a novel identification strategy to examine how an exogenous change in the fuel economy of a kept vehicle affects a household's choice of a second vehicle. We find strong evidence of attribute substitution in the household vehicle portfolio. This effect operates through car attributes that are correlated with fuel economy, including vehicle footprint and weight. Our findings suggest that attribute substitution exerts a strong force that may erode a substantial portion of the expected future gasoline savings from fuel economy standards, particularly those that are attribute-based. Elements of our identification strategy are relevant to a broad class of settings in which consumers make sequential purchases of durable portfolio goods.

*Archsmith: PhD Candidate in Economics, UC Davis, jesmith@ucdavis.edu. Gillingham: Associate Professor, Yale University and NBER, kenneth.gillingham@yale.edu. Knittel: George P. Shultz Professor of Applied Economics, Sloan School of Management, and Director of the Center for Energy and Environmental Policy Research, MIT, and NBER, knittel@mit.edu. Rapson: Associate Professor of Economics, UC Davis, dsrapson@ucdavis.edu.

This paper has benefited from conversations with seminar participants at Universidad Carlos III de Madrid, CESifo, ETH Zurich, University of Mannheim, European University Institute, London School of Economics, Toulouse School of Economics, Fondazione Eni Enrico Mattei, the Energy Institute @ Haas Summer Energy Camp, IIOC and NBER EEE Spring Meetings. We gratefully acknowledge financial support from the California Air Resources Board. The statements and conclusions presented are those of the authors and not necessarily those of the California Air Resources Board. The mention of commercial products, their source, or their use in connection with material reported herein is not to be construed as actual or implied endorsement of such products.

1 Introduction

The idea that *products* can act as complements or substitutes is commonplace in economics. What is less well understood is how households trade off particular *attributes* of one product with the attributes of another. For example, a household may choose to purchase a smaller television in one room if faced with an exogenous increase in the size of another television in the household. One immediate implication of this "attribute substitution" is that the demand for products within a household will not be independent. A second implication is that attribute substitution may influence the efficiency or cost effectiveness of policies that influence the attributes of products, such as energy efficiency standards.

Few papers have been able to empirically measure these forces. There is a literature, for example, that suggests households engage in attribute substitution when it comes to children. Ben-Porath and Welch (1976) and Angrist and Evans (1998) show that households that have had two children of the same gender are more likely to have a third child, compared to two-children households endowed with one boy and one girl. This suggests that the attributes of the first two children, namely gender, affect the utility from having a third child. There is also evidence that households trade off, or substitute, the characteristics of occupations across spouses. For example, households may avoid having two occupations in the same sector as a way to reduce risk.[1] While intuitive, identification of such a phenomena in the broader labor market is difficult because matching costs may be lower within an occupation or sector; physicians tend to meet other physicians and not economists.

Our setting, a household's demand for vehicles, allows us to overcome similar identification challenges. Attribute substitution may manifest itself in a number of dimensions. For example, households may prefer to have one SUV and one sedan, or one powerful vehicle and one fuel-efficient vehicle. Identification in this case is challenging because households may also fall into certain household "types." Just as Anderson et al. (2015) and Mannering and Winston (1985) show that some households have a preference for a certain brand of vehicle, some households may also have a preference for certain attributes. In such instances, if households have a preference for, say, horsepower or fuel economy, the attribute will be present in high or low levels across all goods in the portfolio.

In this paper, we overcome these identification challenges through the use of a rich data set that enables us to track households over a number of vehicle replacement and purchase decisions, and through the use of an instrumental variables strategy. We use panel data on the portfolio of vehicles within a household over time to estimate how a household's choice of vehicle depends on the attributes of the other vehicles owned by the household. Our empirical strategy focuses on two-vehicle households, which offer a clean experiment that we will discuss in detail. We estimate how the fuel economy of a newly-added vehicle depends on the fuel economy of the vehicle already

[1] See, for example, Udry et al. (1995) and Ellis (2000).

held by the household and then show the implications of this dependence.

While attribute substitution is relevant across a broad range of products and settings, understanding the patterns of demand for vehicles is particularly important for a number of policy issues, most notably the effect of fuel economy standards or gasoline taxes on vehicle choice. Empirical models used to analyze the costs and benefits of such policies often capture many key drivers of patterns of vehicle demand, but typically assume away interactions between preferences for multiple vehicles within a household. That is, these models of fuel economy standards assume that consumers choose only a single vehicle or, alternatively, that the choice of each vehicle in a household is independent of the others (Bento et al., 2009; Goulder, Jacobsen, and van Benthem, 2012; Jacobsen, 2013).

However, there are likely to be two sources of interdependence. The first is that households may have particularly strong preferences for certain vehicle attributes that put them into different household "types." This source of dependence is *implicitly* captured in empirical models that allow for variation in the willingness to pay for vehicle attributes. For example, the choice of fuel economy across vehicles within a household will be positively correlated for a household that particularly values horsepower. The second source of interdependence is that households may have complementarities and other portfolio considerations, as suggested by Gentzkow (2007) for newspapers and Wakamori (2011) for Japanese vehicle purchasers. Wakamori (2011) focuses on the combination of a small automobile and minivan, but the more policy-relevant consideration in the United States would be a case where a household endowed with a more fuel-efficient vehicle will compensate by reducing the fuel economy (perhaps by increasing the horsepower or size) of the second vehicle.

The presence of this second form of interdependence can alter the predictions from policy counterfactuals related to fuel economy standards and gasoline taxes. Suppose a policy were to increase the chosen fuel economy of the newest vehicle for a given household at time t. When the household subsequently replaces the other vehicle at a later date, attribute substitution could lead them to purchase a lower fuel economy vehicle than they might have otherwise. Because fuel economy is correlated with other attributes, when there are attribute-based standards, such as the footprint-based standards in the United States (Gillingham, 2013; Ito and Sallee, 2014; Kellogg, 2017), this effect could further drive a wedge between the economic efficiency of fuel economy standards and Pigouvian gasoline taxes. Pigouvian gasoline taxes would still be economically efficient in the presence of attribute-subsititution effects, but these substitution effects may affect the counterfactual level of emissions reductions under Pigouvian taxes.

The ideal experiment to answer our research question would randomly assign the "kept" vehicle to households in the market for a new or used vehicle and then observe the relationship between the fuel economy of this kept vehicle and the fuel economy of the newly-acquired vehicle. Since this ideal experiment is obviously not feasible, our identification strategy must overcome two potential sources of endogeneity stemming from the non-random assignment of the kept vehicle. The first is the choice of which vehicle to replace. Since the household preference for particular features of a multi-car portfolio will directly inform the decision of which car to keep or drop, the attributes of the kept car are endogeneous when using observational data. The second is related to the presence of unobserved household preferences for vehicle attributes. Household fixed effects allow us to focus on within-household variation and can address time-invariant unobserved preferences, but there would still be a concern if preferences change over time. Time-varying preferences may produce a correlation between the desired attributes of the kept and newly-acquired vehicle, again implying that the attributes of the kept vehicle are endogenous.

We employ two sets of instruments to account for these potential sources of bias. The first set of instruments is derived from the observation that changes in the relative price of cars in a portfolio systematically affect the probability that the lowest fuel economy car is dropped. We discuss and present three instruments that rely on this feature of the choice setting, with our preferred instrument based on deviations from the expected change in relative vehicle prices at the time when the kept car was initially purchased. To the best of our knowledge, this instrument is new to the literature. The second instrument is the gasoline price at the time of the purchase of the kept vehicle. A number of papers (Klier and Linn, 2010; Allcott and Wozny, 2014; Busse, Knittel, and Zettelmeyer, 2013; Gillingham, 2011) have shown that vehicle purchase behavior is influenced by contemporaneous gasoline prices. Given this literature and the finding in Anderson, Kellogg, and Sallee (2013) that consumers tend to use the current gasoline price in forming expectations of future prices, we would expect the fuel economy of the kept vehicle to be influenced by the gasoline price at the time of that purchase. We argue that this instrument for the kept vehicle fuel economy satisfies the exclusion restriction because after controlling for current gasoline price, past gasoline prices should not influence the choice of the new vehicle. This assumption rests on limited serial correlation in the residuals and assuming that consumers are using the contemporaneous gasoline price to form expectations of future gasoline prices.

We find evidence that households substitute attributes across the vehicle portfolio. Increasing the fuel economy of the kept car induces households to demand less fuel economy in the purchased car. We show this using a continuous measure of the newly-acquired vehicle fuel economy as the dependent variable and by estimating the probability a household purchases a vehicle in the upper and lower quartiles of the fuel economy distribution. Increases in the fuel economy of the kept car reduce the probability the household purchases a car in the lower quartile of gallons per mile (the highest fuel economy quartile), while such increases raise the probability the household buys a car in the upper quartile (the lowest fuel economy quartile). The effects we estimate are equilibrium effects on the market, explicitly allowing for the entire bundle of vehicle attributes to change. We focus on fuel economy as the measure of interest due to its high correlation with many other attributes (Knittel, 2011) and its particular policy relevance, as we readily admit that households may be substituting an attribute that is correlated with fuel economy, such as size or power, and not fuel economy itself.

We further find that changes in gasoline prices affect household preferences in intuitive ways. As gasoline prices increase, the effect of the kept vehicle fuel consumption and the probability of buying a car in the lower quartile of fuel consumption becomes even more positive. In contrast, as gasoline prices increase, the effect of the kept vehicle fuel consumption and the probability of buying a car in the upper quartile of fuel consumption becomes even more negative.

To gauge the importance of attribute substitution, we use our reduced form results to explore the net effects of an exogenous increase in the fuel economy of the kept vehicle. We are agnostic as to what is causing this increase in fuel economy. It could be, for example, from fuel economy standards, subsidizing fuel economy, or improved technology. In such a thought experiment, a number of adjustments by the household are put in motion. The first, and the focus of our paper, is that the attribute substitution effect implies a decrease in the fuel economy of the newly purchased vehicle. Our estimates suggest that attribute substitution will have a large countervailing effect on the fuel economy of the newly purchased vehicle. For example, in our preferred specification, *increasing* the fuel economy of the kept vehicle by 10 percent results in a 4.8 percent *decrease* in the fuel economy of the purchased vehicle.[2]

Understanding the full impact of our thought experiment requires estimating how an increase in the fuel economy of the kept car alters not only the fuel economy of the newly purchased vehicle, but also how the two vehicles are used. When the kept vehicle has higher fuel economy, the cost per mile of driving is lower, so the well-known rebound effect implies that the kept vehicle will be driven more miles. Two additional forces are also at play. First, a decrease in the fuel economy of the newly-purchased vehicle will imply fewer miles traveled by this vehicle. This is a negative rebound brought about by attribute substitution. Second, by changing the relative fuel economy of the two vehicles within the household, miles will naturally flow away from the now less efficient vehicle to the now more efficient vehicle. Using data on household miles driven by both vehicles, we measure the magnitudes of these effects. We find that these effects also reduce the savings from the exogenous increase in fuel economy from our thought experiment, but the main channel remains through the fuel economy of the newly-purchased vehicle.

Attribute substitution erodes over 60% of the fuel savings from the fuel economy increase of the kept vehicle on net after accounting for all of these factors. As a specific example, consider a 10 percent increase in fuel economy from the average vehicle in our sample. Given the average miles driven (688 per vehicle in our sample), this 10 percent fuel economy increase would directly lead to a 69 gallon decrease in annual fuel consumption. However, due to attribute substitution the next vehicle the household purchases will be less fuel efficient than it otherwise would have been. This decrease in fuel economy of the newly-purchased vehicle reduces the fuel savings from our thought experiment to 40 gallons, holding usage of the two vehicles constant. But we also find significant changes in usage patterns that further reduce the net fuel savings. Mileage of the kept

[2]We also show that the attribute substitution effect operates strongly through attributes that are correlated with fuel economy, including vehicle footprint and weight. We do not attempt to separately identify these channels.

car increases significantly. A large fraction of this increase is due to shifts in miles traveled from the now less fuel efficient purchased vehicle; however, we also find a net increase in overall mileage across the two-vehicle portfolio. Accounting for all of the changes, the net savings of the exogenous increase in fuel economy falls from the naive estimate of 68 gallons to 24 or 27 gallons, depending on whether the initial vehicle was the most fuel efficient vehicle in the household.

The remainder of the study proceeds as follows. The next section describes the household vehicle choice problem and outlining a simple theoretical model (Section 2). We then describe our datasets, the restrictions that determine the sample used for our empirical tests, our identification strategy and empirical approach (Section 3). We next present our results and their economic importance (Section 4). We conclude with a brief discussion of the implications for policymakers and empiricists (Section 6).

2 Context and Model

We begin by developing a simple economic framework of utility maximization in a setting where consumers enjoy a portfolio comprised of multiple units of a similar good. The model builds upon and shares elements from the frameworks presented in Gentzkow (2007) and Wakamori (2011). We consider the possibility that the optimal level of attributes of one good in the portfolio may be related to the level of those attributes in another. The framework maintains generality by referring to any product that may have portfolio characteristics, but the reader may prefer to hold in mind a multi-vehicle household (our empirical application). For simplicity, the household is considered an autonomous decision-maker (i.e., a single consumer) such that issues of within-household bargaining are not considered. When the consumer derives utility from the attributes of each good, the model will demonstrate how preferences may lead the consumer to jointly optimize the level of attributes across goods in their portfolio. This section will also highlight identification challenges that motivate the estimation approach that we later deploy.

Consider a consumer that possesses one unit of the good and is purchasing another. Consider further a standard discrete choice framework with a random utility model. Let the characteristics of the good be given by the vector θ_V, where $V \in \{A, B, ...\}$ denotes distinct bundles of attributes embodied by the product, which we will call "types." For example, in the case of vehicles product types may be defined broadly, such as the class of vehicle (e.g., SUV or small car), or at a finer level such as at the make-model level.

Suppose the consumer is endowed with a product of type A and is deciding which subsequent product to purchase for her portfolio.[3] The consumer receives utility based on the characteristics of each product type that she holds. She may also receive utility from having a portfolio of products with different characteristics. Let the contribution to utility from having a particular portfolio be

[3]We assume that the probability of not purchasing the subsequent product is not altered with a change in an attribute of product A. This allows us to simplify the model by ignoring the outside option.

given by Γ_{AB} such that Γ_{AB} is utility that is additional and derived from the fact that the consumer possesses a portfolio comprised of products A and B concurrently. The indirect utility derived from such a portfolio is given as:

$$u^{AB} = f(\theta_A) + f(\theta_B) + \Gamma_{AB} - \alpha(p_A + p_B), \tag{2.1}$$

where $f(.)$ is a function that maps product attributes into consumer utility, and p_V is the price of product V.

In the context of many household durable goods, a portfolio is likely to arise from sequential purchases. This is typically (though not necessarily) the case with multi-car households, financial asset portfolios, some household durables (e.g., electronics, art, decor), clothing fashion, media subscriptions, higher education choices, and more. While the utility formation described above does not change in the case of sequential purchases, the choice of the paired product will occur conditional on the attributes of the pre-existing item. We now shift focus to such a sequential purchase scenario.

2.1 Implications for Product Choice

We seek to understand how an exogenous change in the attributes of the already-owned good influences the choice of the second product. Conceptually, there exists a cross-attribute elasticity that functions much like a cross-price elasticity. Whereas the cross-*price* elasticity reflects the change in demand for one product as a function of the *price* of the other, the cross-*attribute* elasticity reflects how the change in an *attribute* influences the probability of choosing a particular follow-on product. Extend the above framework by assuming that the consumer may choose between options B and C for the second product. The household chooses portfolio AB rather than AC if $u^{AB} > u^{AC}$. Thus, AB is chosen if:

$$f(\theta_B) - f(\theta_C) + \Gamma_{AB} - \Gamma_{AC} - \alpha p_B + \alpha p_C > 0. \tag{2.2}$$

This simple inequality indicates that the consumer will choose B as the second product when the net utility of the product characteristics, cost, and portfolio utility from B dominate the net utility of the product characteristics, cost and portfolio utility from C.

A set of policy-relevant comparative statics emerge from this framework. Consider what the model implies for the equilibrium portfolio choice probabilities. Let $g(.)$ be the distribution of preferences in the population. Then conditional on purchasing a second product, for the simple choice between B and C the choice probabilities are given as follows:

$$Pr_{AB} = \int_{\mathbf{u}} I(u_{AB} > u_{AC}) dG(\mathbf{u}),$$
$$Pr_{AC} = \int_{\mathbf{u}} I(u_{AC} > u_{AB}) dG(\mathbf{u}).$$

Here $I(\cdot)$ is an indicator and $G(\cdot)$ is the distribution of utilities in the population.

The goal of this paper is to examine how a change in the attributes of a kept good, A, alters choice probabilities relating to a subsequent purchase. Thus, for a given consumer we are interested in the choice between portfolios AB and AC when θ_A changes. How the choice between AB and AC changes with a change in θ_A reflects consumer preferences for different portfolios of products, which is governed by the derivative of (2.2) with respect to θ_A: $\frac{\partial \Gamma_{AB}}{\partial \theta_A} - \frac{\partial \Gamma_{AC}}{\partial \theta_A}$. Many different preferences are possible.

Allow θ_V^m to denote the element in θ_V associated with the individual attribute $m \in M$. If the endowed product A experiences an increase in a desirable attribute (i.e., θ_V^m increases), this lessens the need for m in a subsequent purchase if there is attribute substitution and increases it if there is attribute complementarity. If households display a preference for attribute substitution for attribute m, then increasing θ_V^m for the initial good would decrease the willingness to pay for m in the subsequent purchase. Thus, *ceteris paribus*, the probability of purchasing a good abundant in attribute m would decrease. For precision, we can more formally define this preference for attribute substitution as follows:

Definition 1. Consider a change in the (initial) product A's level of attribute m, θ_A^m, and let B possess a lower concentration of m than type C such that $\theta_B^m < \theta_C^m$. A consumer exhibits a preference for *attribute substitution* in m when $\frac{\partial Pr_{AB}}{\partial \theta_A^m} > \frac{\partial Pr_{AC}}{\partial \theta_A^m}$. That is, an increase in θ_A^m increases the probability of the portfolio AB relative to the probability of portfolio AC.

For the remainder of this section and beyond, we will focus on attribute substitution rather than attribute complementarity. There is no a priori reason why one wouldn't observe attribute complementarity in any number of settings; however, to limit the number of new concepts, we focus on attribute substitution since it appears to be most relevant to the household vehicle portfolio. Furthermore, we now begin to shift our focus towards the empirical application which examines fuel economy of vehicles.

2.2 Considerations Relating to Household Vehicle Portfolios

It is important to note the differences between attribute substitution and a household's potential desire to *diversify* attributes. Attribute substitution implies that an exogenous increase in the fuel economy of the kept vehicle A leads to the purchase of a lower fuel economy vehicle B, regardless of whether vehicle A is the more fuel efficient or less fuel efficient vehicle in the household. Thus, attribute substitution may lead to a convergence or divergence in levels of fuel intensity in the portfolio. In contrast, when there is a desire for attribute diversification, the paired product choice in the portfolio depends on whether vehicle A is the more or less fuel efficient vehicle in the household. Consider, for example, the case where vehicle A is the less fuel efficient vehicle. If households desire diversification, increasing the fuel economy of vehicle A could lead the household to further *increase* the fuel economy of vehicle B because of the preference for diversity in the portfolio. This is the opposite of what would occur under attribute substitution. Now consider the case where vehicle A is the more fuel efficient vehicle. Under a preference for diversification, increasing the fuel economy of vehicle A would lead the household to decrease the fuel economy of vehicle B. Note that in this case diversification is observationally equivalent to attribute substitution.

This discussion alludes to the challenge when attempting to empirically distinguish between attribute substitution and a preference for diversification. Our empirical specifications will allow the change in vehicle B's fuel economy to depend on the relative fuel economies of the two vehicles in order to empirically test whether the data supports attribute substitution or diversification. Irrespective of whether the kept car is the high or low fuel economy vehicle, there is evidence of attribute substitution if households repeatedly choose replacement cars that move the average fuel economy towards a "target" average fuel economy for the vehicles in their portfolio.

3 Data and Identification

The cornerstone of our dataset is the universe of California vehicle registration records that occurred from 2001-2007.[4] The DMV dataset includes every vehicle registered under the residential designation code (e.g., not commercial or government). In California every vehicle must be registered annually. Each record includes the 17-digit vehicle identification number (VIN) that uniquely identifies the vehicle, that year's registration date, the date when the vehicle was last sold, and various other information. A confidential version of the data includes registrant surnames and premise address. This information allows us to construct a household-level panel dataset of vehicle ownership in partnership with the California Air Resources Board.

Basic vehicle attributes (e.g., horsepower, weight, etc.) are available via a VIN decoder that we purchased from DataOne Software. We augment the decoder to include vehicle fuel economy, which is available from the US Environmental Protection Agency. Vehicle-miles traveled (VMT) are available for each VIN whenever the vehicle is sold and upon receiving biannual Smog Check certification.[5] We thus have an average measure of miles traveled by each vehicle and, by extension, each household for each year in our sample. The coarseness of these data are not optimal for examining high-frequency effects of VMT-switching between vehicles in response to changes in gasoline prices. Nonetheless, gasoline prices are a variable of interest in this study, since they affect the household's optimal portfolio of vehicle fuel economy. Our gasoline price data are from the Oil

[4] We thank the California Department of Motor Vehicles (DMV) for making these data available for research.

[5] New vehicles are not required to have a smog check until six years after registration, subsequently it is a biennial check. Hybrid electric vehicles and electric vehicles are also exempt.

Price Information Service (OPIS) at the county-month level.

3.1 Describing the Sample

In each year households are characterized by the starting and ending number of vehicles in their portfolio. In year t a household's starting portfolio size (N^s) is the number of vehicles registered in that year. If the household also registers exactly N^s vehicles in year $t+1$ or $t+2$, then the ending portfolio size (N^e) in year t is N^s. If the number of vehicles registered in years $t+1$ and $t+2$ are identical to each other, but not equal to N^s then the ending portfolio size is the number of vehicles registered in the later years.[6]

[Table 1 about here]

Table 1 shows the distribution of household portfolio transitions. Rows indicate the number of cars in year t, and columns indicate the number of cars in $t+1$. The table represents all possible household transitions. The large mass on diagonals indicates that many households do not increase or decrease the number of cars that they register from year to year. A careful interpretation of "0" is necessary: a household with 0 cars is not in our dataset, so transitions from 0 occur when a Californian household without a car in t registers one in $t+1$, or with observationally-equivalence, a household moves to California from another state. Similarly, transitions to 0 occur either when a household sells all of its registered cars, if it exits the data via a move to another state or a dissolution of the household.

The key regressions that follow are estimated using a sample of two-car households that replace one of their cars, a sample which we call "2x2 replacement households".[7] While other transitions are certainly interesting, two-car replacement households provide the cleanest experiment. Households increasing the number of cars in their portfolio are likely to be experiencing an unobserved development that increases their demand for transportation (e.g., having a baby). Furthermore, it is unclear how to characterize the channels through which the consumers may have preferences for attribute substitution when there are multiple kept cars. Do these households substitute attributes based on the highest-VMT kept car, or the newest? Or is a higher dimensional analysis required?

Given that no clear answer exists to these questions, we choose the transparent path of focusing on the replacement decisions of two-car households, consistent with the simple theory model

[6] We examine one and two years in the future as a household that may register more cars in one year than they ever owned simultaneously. For example, consider a household that owns two cars in year t. In year $t+1$ they re-register both previously owned vehicles and the registrations expire. Then, toward the end of the year, they sell one vehicle and replace it with a new one (which requires registration of the new vehicle). This household has registered three unique vehicles in year $t+1$ but only ever owned two at any given time. In year $t+2$, barring the purchase of yet another new vehicle, the household would return to registering two vehicles.

[7] We define a household as replacing one vehicle if the starting (in year t) and ending (in year $t+1$ or $t+2$) portfolios differ by one vehicle. The household may conduct multiple vehicle transactions, as long as one of the two vehicles appears in both the starting and ending portfolios. We do not consider households where both vehicles in the two-vehicle portfolio change as the relative timing of each purchase becomes important for defining the portfolio at the time of each vehicle's purchase.

presented above. Moving forward with our 2x2 replacement sample is valid when considering small deviations from an interior consumer choice optimum. For such deviations, the probability of a different transition (e.g., not purchasing a second vehicle or purchasing a third vehicle) is not affected by a small change in an attribute of the kept vehicle.

Table 2 shows summary statistics for all 2x2 replacement households, including segmentation based on the fuel economy of the bought car. Households that purchase relatively fuel efficient vehicles (gallons per mile quartile 1) tend to keep relatively fuel efficient cars as well. The converse is true for households buying fuel inefficient vehicles, suggesting that households may have an overall preference for either high or low fuel economy cars. In the remainder of this paper we use fuel intensity in terms of gallons-per-mile (GPM), rather than fuel economy (miles-per-gallon), because the fuel intensity better captures the fuel savings from changing the fuel economy of the vehicle (Larrick and Soll, 2008).

[Table 2 about here]

Some of the analyses that follow use the quartile of fuel economy to describe bought and sold cars. The GPM cutoffs are presented in Table 3, along with their corresponding fuel economy analogs in miles-per-gallon (MPG) for reference.

[Table 3 about here]

3.2 Identification

The goal of this study is to understand attribute substitution: the effect of a change in fuel economy of the kept car on the fuel economy and other attributes of the bought car. This is a challenging object to identify due to the complex nature of consumer preferences over the bundles of vehicle attributes. Our empirical approach must disentangle household preferences over the *level* of attributes (e.g., some households demand low fuel intensity in both cars) from a preference for substituting attributes across vehicles in the portfolio. In terms of our simple model from section 2, we are aiming to see if $\frac{\partial \Gamma_{AB}}{\partial \theta_A} > \frac{\partial \Gamma_{AC}}{\partial \theta_A}$ when car B is more fuel efficient than car C.

The identification challenges in our setting can be most easily understood by considering the ideal experiment for answering our research question. Take every two-car household that is about to exchange a vehicle, randomly assign one car to be the "kept" car, exogenously perturb the GPM of this kept car (f^k), and see how this exogenous change affects the household's observed choice of f^b, the GPM of the car purchased. To see the first identification challenge, consider a case in which we observe a cross-section of household vehicle pairs, where one is a newly-purchased second vehicle. Using "between" variation, we may see that the purchased vehicle's fuel intensity is increasing in the kept vehicle's fuel intensity. There are two (non-conflicting) reasons why this may be the case. Households that already own a fuel efficient kept vehicle may prefer more efficient vehicles in general, and this may be true even if they also have a preference for attribute substitution. We ideally want to exogenously change the fuel economy of the kept vehicle A and observe its effect on the probability of choosing vehicle B versus C, but in the cross-section we cannot separate preferences over the level of attributes from the preferences for the mix of attributes.[8]

Panel data allow us to overcome this challenge by observing repeated replacement choices by the same household. For example, a household repeatedly responding to an exogenous improvement in the fuel intensity of the kept car by choosing a less fuel efficient second car allows us to identify preferences over level and gradient. That is, we would be able to see whether $\frac{\partial \Gamma_{AB}}{\partial \theta_A} > \frac{\partial \Gamma_{AC}}{\partial \theta_A}$ without contaminating the change in θ_A with a change in overall preferences for fuel intensity.

However, even when utilizing repeated choices, there remain two threats to identification. First, in observational data the attributes of the kept vehicle are not exogenously determined. That is, there may be time-varying household attributes that affect their fuel economy decisions (e.g., adding a household member). Second, households often have the ability to choose which vehicle to drop from their portfolio. For example, Jacobsen and van Benthem (2015) show that when gasoline prices increase, households choose to scrap less-efficient vehicles more often. Both of these features of household decision-making could bias an attempt to estimate the preference relationship between portfolio attributes because even with repeated choices the researcher would observe the jointly determined preference for attribute level and mix. We address them using two instrumental variables: one intended to provide exogenous variation in which car to keep and another to perturb the fuel intensity of the kept car. Instrumenting may also address a wide variety of other potential confounders such as unobserved car attributes or time-varying unobserved household attributes.

Instruments 1: Price Differentials

A contribution of this work is the proposal of a new instrumental variables strategy. In our context, we need to instrument for the choice of kept vehicle and the fuel economy of the kept vehicle. A valid instrument will provide exogenous variation in the process that determines which of the household vehicles is kept and which is replaced. The exclusion restriction requires that the instrument affects the household's choice of f^b only indirectly, through the choice of which car to keep. We assert that variation in the price differential between the kept and dropped car contains such identifying variation. There are three functions of the price differential that we explore. For exposition, let P_t^k and P_t^d be the average retail value of the kept and dropped cars, respectively, at the time when the car is dropped (t). The first candidate instrument is the price difference at time t: $\Delta P^{kt} = P_t^k - P_t^d$. This is highly predictive of which car is dropped (generally the less expensive one), but one might be concerned that attributes of the car that are correlated with both the choice of which car to drop and the price difference. This would violate the exclusion restriction.

The second potential instrument is the change in price differences between time t and time 0, when the kept car was purchased. That is $\Delta\Delta P^{kd} = (P_t^k - P_t^d) - (P_0^k - P_0^d)$. To the extent that market forces are exogenous to portfolio preferences, this instrument has promise. However, one

[8] A cross-sectional analysis is analogous to taking the derivative of (2.2) with respect to a composite variable of θ_A and an unobserved preference variable.

may be concerned that the change in relative prices was expected by the buyer in time t, and thus potentially correlated with preferences in time t as well.

The third candidate instrument addresses the above concerns by extracting only the portion variation in the price difference-in-difference that occurs after the time of purchase (i.e., deviates from expectations about the trend of relative prices). We assume that households form expectations using lagged 1-, 3-, or 5-year depreciation rates at the make-model-model year level, and project these into the future. Deviations from these projections are what we refer to as the "deviation from trend," and we use differences in these deviations as an instrument. We find it difficult to come up with a violation of the exclusion restriction in this case. Recall that any concern must posit a correlation between the portfolio preference exhibited in the initial purchase of the kept car and the instrument. Relying on an instrument using market-level changes in relative prices that arise only after the purchase of the kept car would be problematic only in the unlikely case that those broader market-level changes were correlated with individual household preferences over the vehicle portfolio.

[Figure 1 about here]

Figures 1 (a)-(c) display the reduced form relationship between these price differentials and the probability the sold vehicle is the least valuable in the portfolio, partialed of covariates. Each of the instruments appears to have power. It is clear the relationship between the potential instruments, and in particular the Price Deviation DiD, and the choice of the vehicle to drop from the portfolio is best approximated by a cubic polynomial of the instrument. Consequently, we deploy these instruments as third-order polynomials.

Instruments 2: Gasoline Prices at Time of Purchase

Our second instrument is the price of gasoline at the time of the kept car purchase, $p^{gas}_{it_k}$, which provides exogenous variation in the level of kept car fuel economy. Both theory and evidence (e.g., Busse, Knittel, and Zettelmeyer (2013)) demonstrate that households consider future operating costs of the vehicle in their purchase decision. Changes in California gasoline prices several years prior are exogenous with respect to today's household choice, vary extensively over the time period of our data, and alter the expected lifecycle cost of vehicles according to each vehicle's fuel efficiency. Based on this logic, when gasoline prices are high at the time of the kept car purchase, we would expect the household to purchase a more fuel efficient car than when gasoline prices are low (as also demonstrated in Busse, Knittel, and Zettelmeyer (2013)). The price of gasoline at the time of the kept car purchase thus provides exogenous variation in the potentially endogenous variable of interest, fuel economy of the kept vehicle f^k.

Recall that the relationship between f^k and f^b is theoretically ambiguous. If there is attribute substitution there will be a negative correlation, but complementarity between attributes is also possible and will lead to a positive correlation. Furthermore, the relationship between $p^{gas}_{it_k}$ and f^b may also be positive or negative.

The reduced form relationship between the the gasoline price instrument and our outcome variable of interest, f^b, is presented in Figures 2a and 2b. Many factors influence a consumer's choice of vehicle attributes, including f^b, so a plot of the raw data reveals little about the underlying relationship between our variables of interest. Instead, we present the variables after partialing out other covariates. The x-axis and y-axis are the residuals retrieved from regressing $p^{gas}_{it_k}$ and f^b, respectively, on covariates.

[Figures 2a and 2b about here]

A clear positive relationship emerges, indicating a positive correlation between the gasoline price at the time of purchase of the kept car and the fuel economy of the bought car. This figure provides some of the first suggestive evidence of attribute substitution, whereby households increase the fuel intensity of the bought car when gasoline prices are higher at the time of the purchase of the kept car (accordingly leading to lower fuel intensity of the kept car).

Our preferred regression specification includes an endogenous indicator for whether the kept vehicle was more fuel efficient than the dropped vehicle. In these specifications, we also include the gasoline price at the time the dropped vehicle was purchased ($p^{gas}_{it_d}$) as an additional instrument. This gasoline price is similarly correlated to the fuel economy of the dropped vehicle but exogeneous with respect to the household's current choice.

Final Sample

Identifying household fixed effects requires observing at least two transactions per household, which imposes a restriction on our viable sample. Figures 3a - 3b present histograms of the number of transactions per household under various sample restrictions. It reveals that, while many households must be excluded to estimate specifications with household fixed effects, we are still left with approximately 235,000 households in the instrumental variables specification that includes household fixed effects.

[Figures 3a to 3b about here]

3.3 Regression Specifications

The basic regression strategies examine the relationship that GPM of the kept car has on the chosen GPM of the bought car. The dependent variable is thus either GPM of the bought car itself (f^b_{it}), or quartile indicators of that variable. Regressors of interest include gasoline price at the time of purchase, GPM of kept car (f^k_{it}), and their interaction. The right-hand side includes an indicator variable for whether the high or low fuel economy vehicle is kept from the original portfolio. This allows the change in utility from a change in fuel economy of the kept vehicle to differ depending on whether the more efficient or less efficient vehicle is kept, providing additional flexibility in our model and allowing us to test for evidence of attribute substitution versus diversification. For notational simplicity, denote the chosen vehicle by the following indicators:

$$\mathbb{1}^{k>d} \equiv \mathbb{1}\{f^k > f^d\} \tag{3.1}$$
$$\mathbb{1}^{d\geq k} \equiv \mathbb{1}\{f^d \geq f^k\} = \left(1 - \mathbb{1}^{k>d}\right) \tag{3.2}$$

Many of the regression results that follow are retrieved from estimating a linear model of the probability of purchasing vehicles in a given GPM quartile (a nearly direct mapping from our simple model in section 2). For ease of exposition of the results, and to allow a focus on what happens in the top and bottom quartile, we combine vehicles in the 2nd and 3rd quartiles are into a single category, "med," as was shown in the summary statistics above. The baseline specification is:

$$\begin{aligned} Pr\left(q(f_{it}^b) = s\right) = \; & \beta_0 + \beta_g p_{it}^{gas} + \mathbb{1}^{k>d} + \mathbb{1}_{it}^{k>d} \times \left(\beta_{fk} \times f_{it}^k + \beta_{gfk} p_{it}^{gas} \times f_{it}^k\right) + \\ & \mathbb{1}_{it}^{d\geq k} \times \left(\beta_{fd} \times f_{it}^k + \beta_{gfd} \times p_{it}^{gas} \times f_{it}^k\right) + \alpha_X X_{it}^k + \varepsilon_{it}, \end{aligned} \tag{3.3}$$

where the dependent variable, $Pr\left(q(f_{it}^b) = s\right)$, equals one if f_{it}^b falls within the range of quartile $s \in \{1, med, 4\}$. We also estimate a continuous model where the dependent variable is f_{it}^b, keeping the rest of the specification as presented in (3.3). GPM of the vehicles bought (b) and kept (k) by household i in time t are denoted f_{it}^b and f_{it}^k; i's contemporaneous gas price in t is p_{it}^{gas}, whereas $P_{it}^{gas_k}$ is the price of gasoline *at the time household i purchased the kept car in time t*. Control variables, denoted X_{it}, include vehicle attributes (e.g., class, make, value, age), nonparametric time controls (year and month-of-year fixed effects) and household/demographic (household fixed effects and county-level unemployment).

Deploying such a specification accounting for the endogenous explanatory variables described above requires estimating five endogenous variables: an indicator for observations where households replace the relatively efficient vehicle in the portfolio ($\mathbb{1}^{k>d}$), this indicator interacted with the endogenous kept vehicle GPM variables (f^k and $p^{gas} \times f^k$), and corresponding terms interacted with an indicator for when households replace the relatively more fuel-intense vehicle in the portfolio, leading to the following system of endogenous variables:

$$\mathbf{Z}_{it} = \left[\begin{array}{ccccc} \mathbb{1}_{it}^{k>d} & \mathbb{1}_{it}^{k>d} \times f_{it}^k & \mathbb{1}_{it}^{k>d} \times p_{it}^{gas} \times f_{it}^k & \mathbb{1}_{it}^{d\geq k} \times f_{it}^k & \mathbb{1}_{it}^{d\geq k} \times p_{it}^{gas} \times f_{it}^k \end{array}\right]'.$$

The IV specifications deploy instruments for this vector of endogenous regressors. In each specification, we instrument using the gas price at the time the kept vehicle was purchased ($p_{it_k}^{gas_k}$) and the gas price at the time the dropped vehicle was purchased ($p_{it_d}^{gas_d}$). We augment this set of instruments with the instruments based on vehicle price differences that were briefly described in Section 3.2 and projections from the space of exogenous variables as explained below.

Before proceeding, we describe the vehicle price difference instruments precisely. In the "Price Difference" specification, we include the difference in the current resale value of the kept and sold vehicles ($\Delta P_{it}^{kd} = P_{it}^k - P_{it}^d$) as an additional instrument. The "Price Difference-in-Difference" specification uses the change in value for the kept and dropped vehicles between the point the vehicle was purchased and the current time period: $\Delta\Delta P_{it}^{kd} = (P_{it}^k - P_{i0}^k) - (P_{it}^d - P_{i0}^d)$.

The third instrument, which we call "Price Deviation from Trend Difference-in-Difference" (DfT), is constructed from the deviation of the difference between the kept and dropped vehicles relative to their expected deprecation rates at the time of the kept car purchase. For each of the kept and dropped vehicle we proxy the household's expectation of annual vehicle depreciation using an estimate of the depreciation of similar vehicles over the previous five years. Specifically, for vehicle make m and model year y, and value $V_{m,y,t}$ in year t, the expected depreciation is:[9]

$$\mathbf{E}[Dep_{m,y,t}] = \left(\prod_{s=1}^{5} \frac{V_{m,y-s+1,t-s+1} - V_{m,y-s,t-s}}{V_{m,y-s,t-s}} \right)^{\frac{1}{5}}. \tag{3.4}$$

We can then calculate the deviation from this expected depreciation rate for each car in the portfolio, and construct the DfT instrument. Assuming vehicle j has resale value $P_{j,t}$ in year t, this is:

$$\Delta\Delta V_{it}^{kd} = (P_{it}^k - \mathbf{E}[Dep_{it}^k] \cdot P_{i,t-1}^k) - (P_{it}^d - \mathbf{E}[Dep_{it}^d] \cdot P_{i,t-1}^d). \tag{3.5}$$

The set of three price difference instruments is $W = \{\Delta P_{it}^{kd}, \Delta\Delta P_{it}^{kd}, \Delta\Delta V_{it}^{kd}\}$. Following the relationship evident in Figure 1 and the functional form of the second stage model in Equation (3.3), we specify a vector of instruments (V_{it}^{wkd}) consisting of a cubic of a price difference instrument (one of $\{\Delta P_{it}^{kd}, \Delta\Delta P_{it}^{kd}, \Delta\Delta V_{it}^{kd}\}$), gas prices at the time the kept vehicle was purchased ($p_{it_d}^{gas_k}$), and the gasoline price at the time the dropped vehicle was purchased ($p_{it_d}^{gas_d}$):

$$V_{it}^{wkd} = \begin{bmatrix} w_{it} & (w_{it})^2 & (w_{it})^3 & p_{it_k}^{gas_k} & p_{it_d}^{gas_d} \end{bmatrix}. \tag{3.6}$$

The first stage thus consists of the following system of five equations for each of the instruments $w \in W$ where $\mathbf{\Xi}_{it}^w$ is a vector of idiosyncratic errors:

$$\mathbf{Z}_{it}^w = \mathbf{\Gamma}_0 + \mathbf{\Gamma}_V V_{it}^{wkd} + \mathbf{\Theta}\mathbf{X}_{it} + \mathbf{\Xi}_{it}^w. \tag{3.7}$$

Estimating Equation 3.7 using instruments V leads to very low first stage power. The endogenous regressors are a system of interactions with both an exogenous regressor (the gasoline price at the time of bought vehicle purchase) and an endogenous binary indicator leading to a system that is difficult to approximate using linear models. To more closely approximate the hypothesized relationship between the endogenous variables and instruments, one may consider forming addi-

[9]As a more concrete example, for a household in year $t = 2005$ owning a 2002 Honda Civic, the expected depreciation is the geometric mean annual depreciation rate of 2001 Hondas in 2004, 2000 Hondas in 2003, 1999 Hondas in 2002, etc.

tional instruments that follow the functional form of these relationships by interacting V with the exogenous current gas price or using pairwise interactions from the Kronecker product of instruments, $V \otimes V$. This however, can lead quickly to a proliferation of instruments.[10] Instead, we form a narrow set of instruments, approximating the functional form of the endogenous variables using interactions of projections from the space of exogenous variables.

We estimate the first-stage relationships for the uninteracted endogenous variables $\mathbb{1}^{k>d}$ and f_{it}^k and projections from the space of exogenous variables ($\widehat{\mathbb{1}^{k>d}}$ and $\widehat{f_{it}^k}$). We then compute four new instruments as interactions of these predictions:

$$\begin{aligned}
&\widehat{\mathbb{1}^{k>d} \times f_{it}^k} = \widehat{\mathbb{1}^{k>d}} \times \widehat{f_{it}^k} && \widehat{\mathbb{1}^{k>d} \times f_{it}^k \times p_{it}^{gas}} = \widehat{\mathbb{1}^{k>d}} \times \widehat{f_{it}^k} \times p_{it}^{gas} \\
&\widehat{\mathbb{1}^{d\geq k} \times f_{it}^k} = (1 - \widehat{\mathbb{1}^{k>d}}) \times \widehat{f_{it}^k} && \widehat{\mathbb{1}^{d\geq k} \times f_{it}^k \times p_{it}^{gas}} = (1 - \widehat{\mathbb{1}^{k>d}}) \times \widehat{f_{it}^k} \times p_{it}^{gas}.
\end{aligned} \tag{3.8}$$

We augment the vector of instruments (V_{it}^{wkd}) with these four additional instruments and estimate the full system using GMM.

4 Results

This section presents our main estimation results. We first demonstrate the importance of the instrumental variables approach and inclusion of household fixed effects, both of which qualitatively and quantitatively alter key coefficient estimates. We then present the marginal effects of kept car GPM on bought car GPM, which reveal household preferences for attribute substitution. Motivated by the correlation between GPM and other vehicle attributes, we also examine the relationship between kept car GPM and footprint, engine displacement, and weight of the bought car. These results provide context for the discussion of policy implications that follows.

4.1 Effect of Kept GPM on Bought GPM

Table 4 presents the baseline regression results from new and used car purchases. The effect of kept vehicle fuel intensity is allowed to vary depending on whether households make the (endogenous) decision to keep the more ($\mathbb{1}^{k>d}$) or less ($\mathbb{1}^{d\geq k}$) fuel-intense vehicle in the portfolio. Column 1 estimates the parameters using OLS, ignoring potential unobserved household-level heterogeneity and endogeneity. Column 2 instruments using gas price at the time of kept vehicle purchase, gas price at the time of dropped vehicle purchase, the "Price deviations from trend" instruments, and projections from the space of exogenous variables described in Section 3.3 (this suite of IVs is referred to as "GP+DfT+I"). Column 3 assumes all variables are exogenous but controls for household-level unobserved heterogeneity using household fixed effects. Finally, Column 4–our preferred specification–estimates parameters using the GP+DfT+I instruments and household fixed

[10] Our preferred specification deploys 5 instruments. Simply forming all pairwise interactions and the interactions with gasoline prices would lead to 50 instruments, with the potential to greatly exacerbate any IV finite sample bias.

effects.

It is clear from this progression of specifications that it is important to account for both endogeneity and unobserved household heterogeneity. Specifications without household fixed effects primarily rely on variation across households and do not reflect the thought experiment described earlier, which uses within-household variation. In many cases, the inclusion of household fixed effects flips the sign of the estimated coefficient, indicating that source of variation (within versus across) may lead to important differences in interpretation. Deploying instruments has the overall effect of magnifying coefficient estimates. This not surprising in a setting in which many factors enter into the household vehicle purchase decision, including those outside of our channels of interest. The extent of bias associated with OLS underscores the importance of valid instruments. We expose each IV model to the Cragg-Donald minimum eigenvalue test for weak identification. The statistics associated with each of our baseline regressions in Table 4 offer reassurance that the instruments are indeed strong.

[Table 4 about here]

When the regression is identified using across-household variation (columns 1 and 2), the evidence indicates that households will tend to replace their dropped car with one that is qualitatively similar in GPM to the kept car. For example, the coefficient in the first row of column 1 indicates that when the household drops the more fuel-intense vehicle, an increase in the fuel intensity of the kept vehicle is correlated with an increase in the fuel intensity of the bought vehicle. The same result also holds in row 2 of column 1, where increasing the fuel intensity of the kept vehicle is correlated with an increase in the fuel intensity of the bought vehicle. Both of these results suggest that when looking across households, we see higher fuel intensity in the kept car (regardless of which one is dropped) being correlated with higher fuel intensity in the bought car. This is consistent with different households having different "types," such as a preference for fuel sippers or gas guzzlers (presumably due to correlated attributes including power, comfort, safety, etc).

It is only when we explore within-household variation that attribute substitution can be seen. Our preferred results in column 4 for new cars show such an effect. Both the first and second rows of column 4 indicate that an increase in fuel intensity of the kept car *decreases* the fuel intensity of the bought car. For households that dropped the more fuel-intense car (kept the more fuel-efficient car), this implies that the household responds to an exogenous increase in fuel economy of the already more-efficient kept car by acquiring a less-efficient car. This is consistent with substitution across attributes: if the kept car is made more efficient, the household prefers to substitute fuel economy for other attributes in the bought car, reducing the fuel economy of the bought car.[11]

[11] In the case of the kept vehicle being more-efficient, this result is also consistent with households diversifying their portfolio, for it suggests that if the more-efficient kept car has an even further increase in fuel economy, households would respond by decreasing the fuel economy of the less-efficient bought car. Note this study is focused on attribute substitution, rather than a preference for diversification.

For households that kept the more fuel-intense car, the results in column 4 imply that the household responds to an exogenous increase in the fuel economy of the less-efficient kept car by acquiring an slightly less-efficient bought car. This does *not* correspond to the natural definition of diversification, but rather, it again implies a preference for attribute substitution: if the kept car is made more efficient, the household again prefers to substitute fuel economy for other attributes in the bought car.

Next we present marginal effects of f_{it}^k on f_{it}^b, which reveal the presence and extent of a portfolio effect that arises in vehicle fuel economy. Table 5 shows results from specifications using f_{it}^b (continuous) as the dependent variable as estimated at different gasoline prices. We separately compute marginal effects for cases where households choose (endogenously) to drop the more ($f^d \geq f^k$) or less ($f^k > f^d$) vehicle and report these effects in alternating columns. Figures 4a and 4b offer a graphical representation of the population average marginal effect in the highest and lowest GPM quartiles.

[Table 5 about here]

[Figures 4a - 4b about here]

We focus on results from the preferred specification in columns 7 and 8 in Table 5. Negative coefficients reflect a household's revealed desire to buy a car with more (less) GPM as the GPM of its kept car decreases (increases). All of the marginal effects have a negative sign, and all are statistically significant.

The overall story is clear: households incorporate portfolio considerations in their vehicle purchase decisions and have a preference for substitution across attributes. That is, if we were to increase the fuel intensity of the kept car, households would buy a second car that has attributes associated with lower fuel intensity.

For households that keep their more fuel-efficient car, the magnitude of the attribute substitution effect increases with the gasoline price. This may be due to the relative importance of fuel economy (or attributes correlated with fuel economy) in household decisions when gasoline prices are higher. Thus far, all of the qualitative results hold similarly for both new and used car purchase instances.

Figures 4a and 4b display the marginal effects of f_{it}^k on the probability of buying a car in the highest (red and least fuel efficient) or lowest (blue and most fuel efficient) GPM quartile. The qualitative story remains the same. Over most of the gasoline price range, increasing the GPM of the kept car increases (decreases) the probability of purchasing a car in the most (least) fuel efficient quartile. This finding is qualitatively similar across new and used cars.

4.2 The Role of Non-GPM Attributes

Our empirical specification intentionally omits many kept-car vehicle attributes from the set of control variables. This allows us to interpret the portfolio effect in what we believe is the most policy-relevant way: allowing other vehicle attributes to change along with GPM. We now directly explore the effect of changing kept car GPM on three vehicle attributes: footprint, engine displacement, and weight. We chose these attributes because they appear with good coverage in our dataset and because of their economic relevance. *Ceteris paribus*, increases in weight, power, and size decrease fuel efficiency, and thus it is likely that the results we have discussed thus far are (at least in part) operating via these attribute channels.

Table 7 displays marginal effects from specifications that are analogous to columns 7 and 8 of Table 5, except with the alternative attribute of the bought car as the dependent variable.[12] For each attribute, a negative estimate can be interpreted as households demanding less of that bought car attribute as f_{it}^k increases. All point estimates are negative, however some are statistically indistinguishable from zero, such as when the outcome is engine displacement in used vehicles. This provides evidence in favor of the hypothesis that the GPM portfolio effect that we observe is operating at least in part through portfolio preferences over other attributes.

We can compare the magnitude of the portfolio effect between households keeping their fuel efficient car with those that keep their fuel inefficient car. These results are also consistent with the hypothesis that households keeping their more efficient car exhibit a stronger portfolio effect in attributes that are correlated with fuel economy. The effect on the gradients in gasoline price also conform to our baseline results from Table 5.

The importance of the effect on vehicle footprint has direct policy relevance. Fuel economy standards in many countries worldwide are vehicle attribute-based (e.g., footprint in the US and weight in several countries), whereby larger or heavier cars receive a less stringent fuel economy requirement. To the extent that the portfolio effect manifests through preferences for vehicle size, there will be consequences for the realized effectiveness of fuel economy standards relative to expectations. We now turn to an exercise designed to illustrate the implications for fuel economy standard policy.

5 Gauging the Strength of the Attribute-Substitution Effects

In this section we quantify the strength of the forces we uncover in Section 4. We do this through two thought experiments. First, we measure the net effect of an increase in the fuel economy of the kept vehicle allowing for the fuel economy of the bought vehicle to change in ways consistent with the results in Section 4. The second thought experiment uses empirical estimates on the welfare costs associated with Corporate Average Fuel Economy (CAFE) standards to measure the added welfare costs associated that result from attribute substitution.

To be clear, we do not claim to provide a complete counterfactual analysis. For example, we do not allow for firms (or potential used car suppliers) to react in terms of the vehicles they offer or the pricing of those vehicles. The calculations do, however, provide us with evidence as to the

[12] We also condition on the same attribute of the kept vehicle, which we assume to be exogenous.

power of the attribute substitution forces at work and their relevance for counterfactual analysis of policies designed to increase fuel economy.

Our first thought experiment investigates the net effect on gasoline consumption of increasing the fuel economy of a household's initial ("kept") vehicle by 10%. We are agnostic as to what leads to the increase in fuel economy of the initially purchased vehicle. One could imagine, for example, a one-time increase in fuel economy standards or some other vehicle-level incentive that operates only at the time of the purchase of the kept vehicle, such as the Cash for Clunkers program. Because incentives might exist for used-car purchases, we also include used vehicles in the thought experiment. We report the results for used cars in the appendix.

Such a thought exercise will put in motion a number of forces. Our estimates in Section 4 imply that, given an increase in the fuel efficiency of the kept vehicle, the next vehicle purchased by the household will be less fuel efficient. We show below that this has a dramatic effect on the net fuel savings. The exogenous increase in the fuel economy of the kept vehicle may also lead to changes in usage patterns across the household's two vehicles. For one, we might expect to see a rebound effect: increasing the fuel efficiency of the kept vehicle reduces the marginal cost of driving, leading to more miles traveled within the household.[13] We might also expect to see the usage across vehicles in the household change given that the *relative* fuel economies of the two vehicles has changed. Furthermore, this shifting of mileage will be exacerbated by the fact that the newly purchased vehicle becomes even less fuel efficient due to attribute-substitution.

To implement this thought experiment, we augment our empirical results on attribute substitution with estimates on how changes in fuel economy affect a household's total vehicle miles traveled, as well as how these miles are divided across the two vehicles within the household. The details of this empirical exercise are provided in Appendix A. We exploit two sources of variation in vehicle operating costs: variation in gasoline prices over time while holding the vehicle portfolio fixed and changes in operating costs resulting from changes in the fuel intensity of vehicles in the portfolio. For each vehicle $i \in \{1, 2\}$ in a two-vehicle portfolio, we compute the fuel cost in dollars per mile DPM_i as the price of gasoline, in dollars per gallon, times the fuel intensity, in gallons per mile. As these are two-vehicle portfolios, attributes of the other vehicle included in the regression are subscripted j.

We construct a yearly panel of two-vehicle households. For each vehicle i in year t, we compute the mean annual VMT (VMT_{it}) as miles driven between the closest preceding (at time $\underline{t}$) and upcoming (at time $\overline{t}$) odometer measurements for that vehicle.[14] We estimate the impact of operating costs on VMT using the following specification:

$$\log(VMT_{h,i,t}) = \beta_i DPM_{h,i,t} + \beta_j DPM_{h,j,t} + \Xi_h + \Theta_{h,t} + \Psi(\underline{t}, \overline{t}) + A^i(i) + A^j(j) + \varepsilon_{h,i,t}, \quad (5.1)$$

[13]See Borenstein (2015) and Gillingham, Rapson, and Wagner (2016) for more on the rebound effect.

[14]We obtain odometer readings through DMV records each time a vehicle is transacted and at the time of biannual smog checks for vehicles six years and older.

where Ξ_h are household fixed effects, $\Theta_{h,t}$ are fixed effects for the county of residence of household h in year t, $\Psi_{h,\bar{t}}$ are fixed effects controlling for seasonality in driving,[15] $A()$ are controls for vehicle attributes,[16] and $\varepsilon_{h,i,t}$ is an idiosyncratic error which may have arbitrary correlation within households.

Estimates from Equation (5.1) are shown in Table 8. The top panel shows the impact of vehicle operating costs on VMT of the more fuel-intense vehicle and the second set show the impacts for the relatively fuel-efficient vehicle. In either case, an increase in the cost of driving (DPM) of one vehicle introduces an incentive to shift VMT from that vehicle to the other vehicle in the household portfolio. We focus here on Column (5) which includes operating costs, household fixed effects, county of residence fixed effects, seasonality fixed effects (captured by the quarter the odometer was read), and controls for vehicle age, attributes, and leases. We discuss each column as well as robustness in the appendix.

The results on usage shifting are intuitive. Increasing the cost per mile of a given vehicle in the household reduces the number of miles that particular vehicle is driven, but increases the mileage of the other vehicle. For example, for the average vehicle in our households (i.e., a gallons per mile of 0.052) and at a gasoline price of \$3 per gallon, the estimates in Panel 1 of Column (5) imply that increasing the dollars per mile of vehicle 1, the more fuel intensive vehicle in the household, by 10% (i.e., a change of $0.0052 \cdot 3$) decreases the number of miles driven by vehicle 1 by 5.76% ($-3.697 \cdot 0.0052 \cdot 3$). The estimates in Panel 2 suggest that a large portion of these miles will be shifted to vehicle 2. In particular, that same 10% change in vehicle 1's cost per mile *increases* vehicle 2's miles driven by 3.35% ($2.152 \cdot 0.0052 \cdot 3$).

[Table 8 about here]

With the estimates on usage in hand, we can calculate the full impact of our thought experiment: increasing the kept vehicle's fuel economy by 10%. Table 9 breaks down the effect into the components discussed above for the average vehicle in our data and at a gasoline price of \$3. Because our estimates differentiate by whether the kept vehicle is the most or least fuel-efficient vehicle in the household there are two panels. The first row in each panel reports the baseline annual gasoline consumption for the kept (initial) and purchased vehicles (follow-on). Given technical progress and the positive trend in fuel prices in our data, the fuel consumption of the bought vehicle is lower than for the initial vehicle.

[15] Demand for VMT follows seasonal patterns and odometer readings do not necessarily occur at the same time each year for a given vehicle. In fact, one may be concerned that the timing of odometer readings may be correlated with demand for VMT. We deploy two sets of controls to account for seasonality in the VMT measurement. First, in a simpler specification, we include fixed effects for the quarter-of-year of the upcoming odometer reading. In our preferred specification, we interact these fixed effects with counts of each quarter-of-year elapsed since the previous odometer reading.

[16] All specifications include nonparametric controls for the age of both vehicles in the portfolio and indicators for leased vehicles. Additional attribute controls in our preferred specification include indicators for vehicle class and continuous measures of vehicle curb weight, wheelbase, vehicle width, and engine displacement.

The second row reports the first effect of our thought experiment. By construction the fuel consumption of the initial vehicle falls by 10%. Given the attribute substitution effect, this will *increase* the fuel consumption of the follow-on bought vehicle by 29 gallons; this is over 42% of the fuel savings from the 10% improvement of the initial vehicle. The next two rows report the impact from changes in usage. The first effect comes directly from the 10% increase in the fuel efficiency of the initial vehicle. This shifts mileage from the follow-on vehicle to the initial vehicle, holding constant the fuel efficiency of the follow-on vehicle. For the two average vehicle types in our sample, this increases fuel savings because miles traveled are shifting to the now more-efficient vehicle. The net effect is theoretically ambiguous because the 10% increase in fuel efficiency of the initial vehicle also leads to a net increase in miles driven.[17] The savings from this shift in miles traveled to the initial vehicle is necessarily larger if the initial vehicle is more efficient (panel b), but given the closeness in the fuel efficiencies of the average initial and follow-on vehicles, the two panels are similar. The next row then calculates the impact of the additional vehicle-usage shifting that comes from the fact that the follow-on vehicle's fuel economy will be changed through attribute substitution. We refer to this as the indirect effect on VMT. This shifts additional miles from the follow-on bought vehicle to the initial vehicle and reduces the fuel savings from our thought experiment.

Once all of the forces are considered, the fuel savings from our thought experiment falls from the naive estimate of 68.8 gallons to either 27.3 (panel a) or 24.4 (panel b) gallons. Therefore accounting for all of the effects reduces the fuel savings by over a 60% in both cases. These results are quite startling and may have unfortunate implications for the effectiveness of policies that explicitly regulate fuel economy, especially if such policies are short term in nature. Specifically, these calculations are most relevant to a fuel economy standard that increases, but then plateaus. These forces would also be especially relevant for such policies as Cash for Clunkers, which provided a one-time subsidy with the aim of improving new vehicle fuel economy. Our estimates would suggest that over 60% of the initial fuel economy savings from Cash for Clunkers would have been eroded from attribute substitution and rebound.

One could argue that these estimates are conservative. The used car market, which is not covered by fuel economy standards, is another channel through which attribute substitution may manifest. Increases in the fuel economy of initial kept vehicles due to an increased standard will increased demand for used fuel-inefficient vehicles. The increase in demand will lead used gas guzzlers to be more valuable, and thus more slowly retire from the fleet (similar to the effect

[17] In households where the kept vehicle is much less fuel efficient than the follow-on bought vehicle, this direct effect could decrease the fuel savings from our thought experiment. The reader might wonder whether our VMT specification should include the possibility of a knife-edge around the point where the kept vehicle's fuel efficiency equals the follow-on vehicle's fuel efficiency. A complete analysis of VMT shifting is beyond the scope of this paper but the topic of current research. An argument against such a knife-edge is that comfort and fuel economy tend to be inversely related. Therefore even if the kept vehicle is less efficient compared to the follow-on bought vehicle, a marginal increase in the fuel efficiency of the kept vehicle will lead to more miles traveled because it is, on average, the more comfortable vehicle within the household.

documented in Jacobsen and van Benthem (2015)).

Our second thought experiment focuses on the welfare costs of CAFE. It is simpler in nature and relies on the empirical estimates of the welfare costs of CAFE in Jacobsen (2013). The thought experiment is a sustained change in the average fuel economy required under CAFE of 1 MPG. Therefore, the fuel economy of both the initial (kept) and follow-on (bought) vehicles are forced to increase by 1 MPG. Jacobsen (2013) calculates the equivalent variation from such a 1 MPG change in CAFE, but his calculations do not include the additional welfare costs that would operate through attribute substitution. The welfare costs accounting for attribute substitution will be larger because the *desired* fuel economy of the follow-on vehicle will now be lower. Therefore forcing the follow-on vehicle's fuel economy to increase by 1 MPG will have a larger welfare consequence.[18]

We can put numbers to this effect. Table 8 in Jacobsen (2013) implies that the average equivalent variation across all households from a 1 MPG increase in CAFE standards is \$264 (in year 10). Our results suggest that the welfare costs for the bought vehicle will be larger because the desired fuel economy for the average bought vehicle in our data *decreases* by 0.41 MPG due to the increase in the initial vehicle's fuel economy. Therefore, a sustained 1 MPG increase in fleet fuel economy will be a 1.41 MPG increase in the average "desired" fuel economy of follow-on vehicles, again noting that these are not true counterfactual estimates. This would in turn increase the welfare costs by 41% to \$372.

6 Conclusions

Much in the same way that *products* can be complements or substitutes, households may view *attributes* of one product as substitutable with or complementary to attributes of another. This observation is potentially relevant to understanding consumer decisions relating to a broad set of goods ranging from financial asset portfolios, household durables and clothing fashion to media subscriptions, higher education, and more. We develop an identification strategy to separate household preferences for level effects from attribute substitution within the household portfolio. Aspects of the methodology may be generally applied to many of the aforementioned household goods. This paper focuses on the vehicle market due to both the particular suitability of the choice setting for identifying attribute substitution and the potential implications for transportation market regulations.

The effects of a number of polices applied to the vehicle market depend crucially on consumer choice patterns. Empirical estimates of vehicle choice typically assume that the vehicle choices within a household are made independently. We provide evidence that this assumption does not hold. Using panel data on the portfolio of vehicles within a household and an instrumental variables

[18]To observe this effect, one would likely have to have a continual increase in the standard, rather a one-time increase, due to technological progress. For example, Knittel (2011) estimates that technological progress increases fuel economy by roughly 2 percent per year if other attributes are held constant. Based on this estimate, the standard would thus have to increase by 2 percent per year for the follow-on bought vehicle's fuel economy to be diminished by the full amount we estimate.

approach, we find evidence that households exhibit a preference for attribute substitution. Exogenous increases in the fuel economy of the kept car lower the fuel economy of the purchased car. We show this using both a continuous measure of fuel economy, as well as by estimating the probability a household purchases a vehicle in the upper and lower quartiles of the fuel economy distribution. An increase in the fuel economy of the kept car reduces the probability the household purchases a car in the lower quartile of gallons per mile, while such an increase reduces the probability the household buys a car in the upper quartile.

We also find that gasoline prices affect the preference for attribute substitution in intuitive ways. As gasoline prices increase, the effect of the fuel intensity of the kept vehicle on the probability of buying a car in the lower quartile of fuel intensity becomes even more positive. In contrast, as gasoline prices decrease, the effect of fuel intensity of the kept vehicle on the probability of buying a car in the upper quartile of fuel intensity becomes even more negative. These effects manifest through substitution across vehicle attributes beyond fuel economy. We find evidence households view attributes such as vehicle size (footprint), weight, and horsepower as substitutes across their portfolio of vehicles.

These results have substantial economic importance for the understanding of major policies to improve the fuel economy of the light duty vehicle fleet, such as fuel economy standards and subsidies for more fuel-efficient vehicles (e.g., Cash-for-Clunkers). We use our results to estimate the net effect of a one-time exogenous increase in fuel economy of the kept vehicle and find that the attribute substitution effect can erode as much as 60% of the fuel savings from the increase in fuel economy. Moreover, our results suggest that this erosion of the savings is likely to be especially problematic under attribute-based standards, such as the current footprint-based standard in the United States and other countries in the world.

While this research setting considers household purchases in the vehicle market, these results highlight the challenges in design or evaluation of any policy intending to alter consumer choices over a portfolio of goods. When households view the attributes of those goods as substitutes, any shift in the attributes of one good will shift the unconstrained optimal choice for attributes of other goods in the portfolio in the opposite direction. This will lead to subsequent purchases which potentially erode or magnify effects of the policy or, if the policy also prevents the unconstrained optimal purchase, increases the true welfare costs of the policy.

References

Allcott, Hunt and Nathan Wozny. 2014. "Gasoline Prices, Fuel Economy, and the Energy Paradox." *Review of Economics and Statistics* 96 (5):779–795.

Anderson, Soren T., Ryan Kellogg, Ashley Langer, and James M. Sallee. 2015. "The Intergenerational Transmission of Automobile Brand Preferences." *The Journal of Industrial Economics* 63 (4):763–793. URL http://dx.doi.org/10.1111/joie.12092.

Anderson, Soren T., Ryan Kellogg, and James M. Sallee. 2013. "What Do Consumers Believe About Future Gasoline Prices?" *Journal of Environmental Economics and Management* 66 (3):383–403.

Angrist, Joshua D. and William N. Evans. 1998. "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." *The American Economic Review* 88 (3):450–477. URL http://www.jstor.org/stable/116844.

Ben-Porath, Yoram and Finis Welch. 1976. "Do Sex Preferences Really Matter?" *Quarterly Journal of Economics* 90 (2):285–307.

Bento, Antonio M., Lawrence H. Goulder, Mark R. Jacobsen, and Roger H. von Haefen. 2009. "Distributional and Efficiency Impacts of Increased US Gasoline Taxes." *American Economic Review* 99 (3):667–699.

Borenstein, Severin. 2015. "A Microeconomic Framework for Evaluating Energy Efficiency Rebound and Some Implications." *Energy Journal* 36:1–21.

Busse, Meghan R., Christopher R. Knittel, and Florian Zettelmeyer. 2013. "Are Consumers Myopic? Evidence from New and Used Car Purchases." *American Economic Review* 103 (1):220–256.

Ellis, Frank. 2000. "The Determinants of Rural Livelihood Diversification in Developing Countries." *Journal of Agricultural Economics* 51 (2):289–302.

Gentzkow, Matthew. 2007. "Valuing New Goods in a Model with Complementarity: Online Newspapers." *American Economic Review* 97:713–744.

Gillingham, Kenneth. 2011. "How Do Consumers Respond to Gasoline Price Shocks? Heterogeneity in Vehicle Choice and Driving Behavior." Manuscript: Yale University.

———. 2013. "The Economics of Fuel Economy Standards versus Feebates." Manuscript: Yale University.

Gillingham, Kenneth, David Rapson, and Gernot Wagner. 2016. "The Rebound Effect and Energy Efficiency Policy." *Review of Environmental Economics and Policy* 10:68–88.

Goulder, Lawrence H., Mark R. Jacobsen, and Arthur A. van Benthem. 2012. "Unintended Consequences from Nested State and Federal Regulations: The Case of Pavley Greenhouse-Gas-per-Mile Limits." *Journal of Environmental Economics and Management* 63:187–207.

Ito, Koichiro and James Sallee. 2014. "The Economics of Attribute-Based Regulation: Theory and Evidence from Fuel-Economy Standards." *NBER Working Paper 20500* .

Jacobsen, Mark R. 2013. "Evaluating U.S. Fuel Economy Standards In a Model with Producer and Household Heterogeneity." *American Economic Journal: Economic Policy* 5 (2):148–187.

Jacobsen, Mark R. and Arthur A. van Benthem. 2015. "Vehicle Scrappage and Gasoline Policy." *American Economic Review* 105 (3):1312–1338.

Kellogg, Ryan. 2017. "Gasoline Price Uncertainty and the Design of Fuel Economy Standards." *NBER Working Paper 23024* .

Klier, Thomas and Joshua Linn. 2010. "New Vehicle Characteristics and the Cost of the Corporate Average Fuel Economy Standard." Resources for the Future Discussion Paper 10-50.

Knittel, Christopher R. 2011. "Automobiles on Steroids: Product Attribute Trade-Offs and Technological Progress in the Automobile Sector." *American Economic Review* 101 (7):3368–3399.

Larrick, Richard and Jack Soll. 2008. "The MPG Illusion." *Science* 320 (5883):1593–1594.

Mannering, Fred and Clifford Winston. 1985. "A Dynamic Empirical Analysis of Household Vehicle Ownership and Utilization." *The RAND Journal of Economics* 16 (2):215–236. URL http://www.jstor.org/stable/2555411.

Udry, Christopher, John Hoddinott, Harold Alderman, and Lawrence Haddad. 1995. "Gender Differentials in Farm Productivity: Implications for Household Efficiency and Agricultural Policy." *Food Policy* 20 (5):407–423.

Wakamori, Naoki. 2011. "Portfolio Considerations in Differentiated Product Purchases: An Application to the Japanese Automobile Market." *Bank of Canada Working Paper 2011-27* .

Table 1: Number of Unique Households by Portfolio Size

Start	End Portfolio Size			
Portfolio Size	1	2	3	4+
1	7,262,111	1,360,594	187,558	75,150
2	1,172,278	4,632,425	839,546	259,098
3	168,745	849,703	2,169,948	675,040
4+	35,810	141,618	381,226	1,489,926

Each cell represents the count of unique households from 2001 to 2007 observed to have the starting portfolio size shown in each row and the ending portfolio size shown in the column. These counts provide a measure of the number of households providing identifying variation in each portfolio cell. A single household may appear in multiple cells if their portfolio changes over time but is counted at most once in each cell. For example, two-car household that replaces one car every year would add one to the count of the (2,2) cell. If instead, that household adds a third vehicle in 2004 and returns to a two-car portfolio in 2006 it would add one to the count of the (2,2) cell, one to the count of the (2,3) cell, one to the (3,3) cell, and one to the count of the (3,2) cell. Each household may have zero, one, or multiple vehicle transactions during this time period.

Table 2: Summary Statistics for Continuous Variables - 2x2 Replacement Households

	All Households	Bought GPM Qtile 1	Bought GPM Qtile 2 or 3	Bought GPM Qtile 4
Kept Vehicle GPM	0.0522	0.0507	0.0523	0.0533
	(0.0106)	(0.0106)	(0.0105)	(0.0108)
Bought Vehicle GPM	0.0516	0.0388	0.0503	0.0664
	(0.0108)	(0.0037)	(0.0036)	(0.0063)
Dropped Vehicle GPM	0.0511	0.0478	0.0507	0.0549
	(0.0103)	(0.0098)	(0.0094)	(0.0112)
Gasoline Price at Bought Purchase (US$)	2.380	2.434	2.377	2.335
	(0.747)	(0.763)	(0.745)	(0.733)
Gas Price at Kept Vehicle Purchase (US$)	2.064	2.105	2.056	2.041
	(0.539)	(0.558)	(0.534)	(0.527)
(Kept - Sold) Value DfT (US$)	4.483	22.629	2.567	-9.056
	(848.059)	(731.634)	(850.055)	(942.603)
Kept Vehicle Age (yr)	7.320	7.466	7.450	6.919
	(5.924)	(5.962)	(5.944)	(5.828)
Dropped Vehicle Age (yr)	9.948	10.651	9.990	9.187
	(5.899)	(5.849)	(5.847)	(5.957)
Kept vehicle value (US$)	9,905	9,082	9,953	10,626
	(8,352)	(7,295)	(8,457)	(9,024)
Bought Vehicle Value (US$)	11,283	7,468	11,819	13,911
	(9,162)	(5,267)	(9,999)	(9,273)
Dropped Vehicle Value (US$)	7,794	6,149	7,957	9,072
	(7,871)	(5,970)	(8,126)	(8,671)
N Transactions	2,004,312	491,010	1,003,044	510,258
N Households	1,452,896	392,168	768,517	413,367

Summary statistics of continuous variables for 2x2 replacement households. Standard deviations shown in parentheses.

Table 3: Distribution of observed fuel economy

Percentile	Gallons per Mile (GPM)	Miles Per Gallon (MPG)
25th Percetile	0.045	22.0
Median	0.052	19.3
75th Percentile	0.059	17.0

Table 4: Regression Estimates

	OLS (1) No IV/FE	IV (2) No FE	HHFE (3) No IV	HHFEIV (4) FE+IV
New				
$\mathbb{1}^{d\geq k} \times GPM^k$	0.4170	-0.0376	0.0881	-0.6440
	(0.0097)***	(0.0788)	(0.0390)**	(0.2879)**
$\mathbb{1}^{k>d} \times GPM^k$	0.2205	-0.0691	-0.0821	-0.5686
	(0.0076)***	(0.0489)	(0.0311)***	(0.1775)***
$\mathbb{1}^{d\geq k} \times GPM^k \times p^{gas}$	-0.0325	-0.0933	-0.2562	-0.3121
	(0.0041)***	(0.0232)***	(0.0154)***	(0.0526)***
$\mathbb{1}^{k>d} \times GPM^k \times p^{gas}$	-0.0110	-0.0422	-0.1905	-0.1907
	(0.0032)***	(0.0120)***	(0.0126)***	(0.0444)***
p^{gas}	0.0004	0.0027	0.0117	0.0131
	(0.0002)**	(0.0009)***	(0.0008)***	(0.0025)***
N Non-singleton	384,692	384,692	140,209	140,209
Cragg-Donald Stat		58.544		159.57
Used				
$\mathbb{1}^{d\geq k} \times GPM^k$	0.2561	0.3495	-0.0473	0.1660
	(0.0104)***	(0.0726)***	(0.0407)	(0.2342)
$\mathbb{1}^{k>d} \times GPM^k$	0.1584	0.1709	-0.1852	-0.1456
	(0.0079)***	(0.0499)***	(0.0327)***	(0.1402)
$\mathbb{1}^{d\geq k} \times GPM^k \times p^{gas}$	-0.0176	0.0153	-0.2582	-0.4104
	(0.0043)***	(0.0327)	(0.0162)***	(0.0327)***
$\mathbb{1}^{k>d} \times GPM^k \times p^{gas}$	-0.0101	0.0068	-0.1907	-0.2612
	(0.0034)***	(0.0161)	(0.0130)***	(0.0268)***
p^{gas}	0.0005	-0.0008	0.0112	0.0169
	(0.0002)**	(0.0012)	(0.0008)***	(0.0015)***
N Non-singleton	395,754	395,754	140,256	140,256
Cragg-Donald Stat		61.194		140.38
Instrumental Vars	N/A	GP+DfT+I	N/A	GP+DfT+I
Fixed Effects	None	None	HH	HH

Regression of the continuous bought vehicle GPM on covariates. Standard errors robust to arbitrary heteroskedasticity clustered by household shown in parentheses. *,**,*** denote significance at the 10%, 5%, and 1% levels, respectively. $\mathbb{1}^{d\geq k}$ ($\mathbb{1}^{k>d}$) is an indicator denoting the dropped vehicle was the most (least) fuel intense vehicle in the portfolio prior to the purchase. “GP+DfT+I” instrument deploys gas price at the time of kept vehicle purchase, gas price at the time of dropped vehicle purchase, the “Price deviations from trend” instruments, and projections from the space of exogenous variables described in Section 3.3 as instruments for endogenous regressors.

Table 5: Marginal Effect of Kept Vehicle GPM on Bought Vehicle GPM

	OLS (1) $f^d \geq f^k$	OLS (2) $f^k > f^d$	IV (3) $f^d \geq f^k$	IV (4) $f^k > f^d$	HHFE (5) $f^d \geq f^k$	HHFE (6) $f^k > f^d$	HHFEIV (7) $f^d \geq f^k$	HHFEIV (8) $f^k > f^d$
New								
$p^{gas} = \$2.00$	0.3519	0.1986	-0.2241	-0.1534	-0.4242	-0.4630	-1.2681	-0.9500
	(0.0046)***	(0.0035)***	(0.0985)**	(0.0592)***	(0.0218)***	(0.0180)***	(0.2731)***	(0.1744)***
$p^{gas} = \$3.00$	0.3194	0.1876	-0.3174	-0.1956	-0.6804	-0.6535	-1.5802	-1.1407
	(0.0059)***	(0.0046)***	(0.1143)***	(0.0671)***	(0.0259)***	(0.0219)***	(0.2807)***	(0.1891)***
$p^{gas} = \$4.00$	0.2869	0.1766	-0.4106	-0.2377	-0.9366	-0.8440	-1.8923	-1.3314
	(0.0090)***	(0.0071)***	(0.1322)***	(0.0760)***	(0.0366)***	(0.0309)***	(0.2975)***	(0.2123)***
Used								
$p^{gas} = \$2.00$	0.2208	0.1382	0.3802	0.1845	-0.5637	-0.5665	-0.6549	-0.6679
	(0.0049)***	(0.0036)***	(0.0716)***	(0.0566)***	(0.0211)***	(0.0172)***	(0.2177)***	(0.1269)***
$p^{gas} = \$3.00$	0.2032	0.1281	0.3955	0.1913	-0.8219	-0.7572	-1.0653	-0.9291
	(0.0061)***	(0.0047)***	(0.0909)***	(0.0659)***	(0.0250)***	(0.0204)***	(0.2165)***	(0.1283)***
$p^{gas} = \$4.00$	0.1856	0.1180	0.4108	0.1981	-1.0801	-0.9479	-1.4757	-1.1903
	(0.0093)***	(0.0073)***	(0.1163)***	(0.0774)**	(0.0364)***	(0.0297)***	(0.2201)***	(0.1352)***

Marginal effects of kept vehicle GPM from a regression of the continuous bought vehicle GPM on covariates. Standard errors robust to arbitrary heteroskedasticity clustered by household shown in parentheses. *,**,*** denote significance at the 10%, 5%, and 1% levels, respectively. $f^d \geq f^k$ ($f^k > f^d$) show marginal effects when the dropped vehicle was the most (least) fuel intense vehicle in the portfolio prior to the purchase. "GP+DfT+I" instrument deploys gas price at the time of kept vehicle purchase, gas price at the time of dropped vehicle purchase, the "Price deviations from trend" instruments, and projections from the space of exogenous variables described in Section 3.3 as instruments for endogenous regressors.

Table 6: Bought Vehicle Attributes

	Footprint (1)	Displacement (2)	Curb Weight (3)
New			
$\mathbb{1}^{d\geq k} \times GPM^k$	-622.10	-151.36	8.6899
	(248.38)**	(87.16)*	(8.7104)
$\mathbb{1}^{k>d} \times GPM^k$	-649.78	-119.58	-2.8888
	(172.78)***	(74.96)	(5.3321)
$\mathbb{1}^{d\geq k} \times GPM^k \times p^{gas}$	-175.43	-21.369	-10.513
	(41.40)***	(3.933)***	(1.311)***
$\mathbb{1}^{k>d} \times GPM^k \times p^{gas}$	-96.082	-10.170	-7.4679
	(34.592)***	(3.207)***	(1.2278)***
p^{gas}	7.2982	0.8328	0.4730
	(1.9291)***	(0.1886)***	(0.0654)***
Kept Vehicle Attribute	-0.0423	0.2289	-0.0566
	(0.1219)	(0.5109)	(0.1237)
N Non-singleton	142,402	143,460	142,418
Cragg-Donald Stat	148.55	135.05	240.42
Used			
$\mathbb{1}^{d\geq k} \times GPM^k$	25.28	-55.09	-1.6599
	(194.05)	(127.48)	(8.2521)
$\mathbb{1}^{k>d} \times GPM^k$	41.12	-67.42	-1.1678
	(154.82)	(120.72)	(6.4835)
$\mathbb{1}^{d\geq k} \times GPM^k \times p^{gas}$	-147.14	-40.276	-9.1561
	(35.88)***	(7.687)***	(1.7653)***
$\mathbb{1}^{k>d} \times GPM^k \times p^{gas}$	-83.254	-24.895	-5.6494
	(28.691)***	(4.791)***	(1.4737)***
p^{gas}	5.2106	1.6169	0.3580
	(1.5968)***	(0.3208)***	(0.0803)***
Kept Vehicle Attribute	-0.5259	0.2393	-0.3978
	(0.1348)***	(0.9707)	(0.2420)
N Non-singleton	138,163	140,410	138,384
Cragg-Donald Stat	129.50	127.98	154.13
Outcome	Footprint	Engine Disp.	Curb wt.
Unit	ft^2	L	tons

Regression of the continuous bought vehicle GPM on covariates. Standard errors robust to arbitrary heteroskedasticity clustered by household shown in parentheses. *,**,*** denote significance at the 10%, 5%, and 1% levels, respectively. $\mathbb{1}^{d\geq k}$ ($\mathbb{1}^{k>d}$) is an indicator denoting the dropped vehicle was the most (least) fuel intense vehicle in the portfolio prior to the purchase. All specifications deploy the preferred GP+DfT+I instrumental variables.

Table 7: Bought Vehicle Attributes - Kept GPM Marginal Effects

	Footprint (1) $f^d \geq f^k$	Footprint (2) $f^k > f^d$	Curb wt. (3) $f^d \geq f^k$	Curb wt. (4) $f^k > f^d$	Displacement (5) $f^d \geq f^k$	Displacement (6) $f^k > f^d$
New						
$p^{gas} = \$2.00$	-972.96	-841.94	-12.337	-17.824	-194.09	-139.92
	(235.81)***	(183.91)***	(7.621)	(5.002)***	(89.82)**	(75.63)*
$p^{gas} = \$3.00$	-1,148.4	-938.0	-22.850	-25.292	-215.46	-150.08
	(240.2)***	(198.5)***	(7.371)***	(5.276)***	(91.38)**	(76.16)**
$p^{gas} = \$4.00$	-1,323.8	-1,034.1	-33.364	-32.760	-236.83	-160.25
	(251.4)***	(217.6)***	(7.350)***	(5.802)***	(93.07)**	(76.83)**
Used						
$p^{gas} = \$2.00$	-269.00	-125.38	-19.972	-12.467	-135.64	-117.21
	(188.86)	(177.48)	(8.536)**	(7.909)	(141.16)	(128.78)
$p^{gas} = \$3.00$	-416.15	-208.64	-29.128	-18.116	-175.92	-142.10
	(196.31)**	(194.25)	(9.198)***	(8.907)**	(148.13)	(132.88)
$p^{gas} = \$4.00$	-563.29	-291.89	-38.284	-23.765	-216.20	-167.00
	(209.72)***	(213.57)	(10.127)***	(10.023)**	(155.17)	(137.03)

Marginal effects of kept vehicle GPM from a regression of the continuous bought vehicle GPM on covariates. Standard errors robust to arbitrary heteroskedasticity clustered by household shown in parentheses. *,**,*** denote significance at the 10%, 5%, and 1% levels, respectively. $f^d \geq f^k$ ($f^k > f^d$) show marginal effects when the dropped vehicle was the most (least) fuel intense vehicle in the portfolio prior to the purchase. All specifications deploy the preferred GP+DfT+I instrumental variables.

Table 8: Regression of Log VMT on Fuel Cost Per Mile

Outcome		(1)	(2)	(3)	(4)	(5)
$\log(VMT_1)$	DPM_1	-0.557	-3.705	-3.709	-3.711	-3.697
		(0.057)***	(0.113)***	(0.113)***	(0.114)***	(0.114)***
	DPM_2	-0.806	2.433	2.435	2.449	2.435
		(0.065)***	(0.139)***	(0.139)***	(0.140)***	(0.141)***
$\log(VMT_2)$	DPM_2	-0.420	-3.926	-3.925	-3.968	-3.953
		(0.071)***	(0.145)***	(0.145)***	(0.145)***	(0.146)***
	DPM_1	-0.363	2.105	2.100	2.153	2.152
		(0.058)***	(0.116)***	(0.116)***	(0.116)***	(0.117)***
N		2,942,024	2,942,024	2,942,024	2,905,962	2,903,315
N Households		854,299	854,299	854,299	845,665	845,121
Household FE		Y	Y	Y	Y	Y
County FE		N	N	Y	N	Y
Seasonality FE		N	N	N	Y	Y
Attribute Controls		N	Y	Y	Y	Y

Regression of vehicle log VMT on covariates. Variables subscripted with 1 denote the more fuel intense vehicle and 2 denotes the less fuel intense vehicle. Vehicle cost per mile (DPM) instrumented using gasoline price at the time the vehicle was purchased and its interaction with current gasoline prices. All regressions include household fixed effects, nonparametric controls for the age of both vehicles in the portfolio, county-level unemployment, and indicators for leased vehicles. Standard errors clustered by household shown in parentheses. Seasonality fixed effects account for seasonal patterns in driving behavior and consist of the quarter of year of the most recent VMT measurement interacted with counts of each quarter type since the previous VMT measurement. Attribute controls include indicators for vehicle class and continuous measures of vehicle curb weight, wheelbase, vehicle width, and engine displacement.

Table 9: Effect of a 10% Decrease in Fuel Intensity of Portfolio Fuel Consumption, New Vehicles

(a) Initial Vehicle is Least Efficient

	Initial Vehicle	Follow-on Vehicle	Portfolio Total
Base Fuel Consumption (gal/yr)	688.18	610.24	1,298.42
Direct Effect (gal/yr)	-68.82	29.05	-39.77
	[-10.00%]	[4.76%]	[-3.06%]
Direct AS Effect on VMT (gal/yr)	40.07	-16.39	23.68
	[5.82%]	[-2.69%]	[1.82%]
Indirect AS Effect on VMT (gal/yr)	12.17	-23.40	-11.23
	[1.77%]	[-3.83%]	[-0.86%]
Total Effect (gal/yr)	-16.57	-10.75	-27.32
	[-2.41%]	[-1.76%]	[-2.10%]

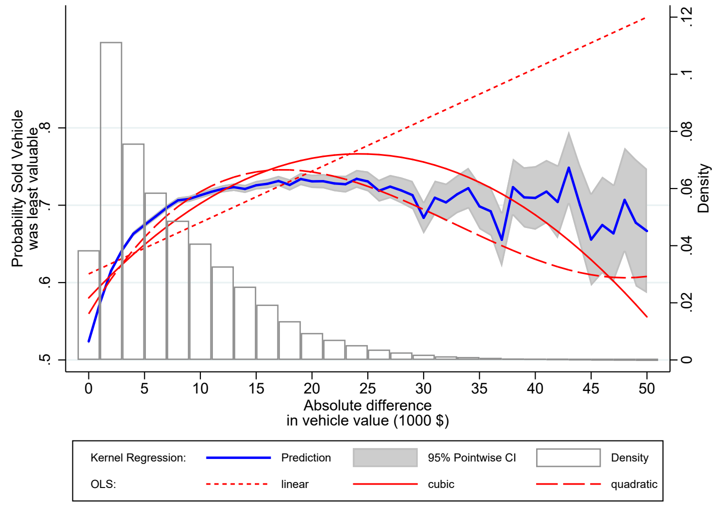
(b) Initial Vehicle is More Efficient

	Initial Vehicle	Follow-on Vehicle	Portfolio Total
Base Fuel Consumption (gal/yr)	688.18	610.24	1,298.42
Direct Effect (gal/yr)	-68.82	29.05	-39.77
	[-10.00%]	[4.76%]	[-3.06%]
Direct AS Effect on VMT (gal/yr)	42.85	-17.52	25.32
	[6.23%]	[-2.87%]	[1.95%]
Indirect AS Effect on VMT (gal/yr)	10.76	-20.69	-9.93
	[1.56%]	[-3.39%]	[-0.76%]
Total Effect (gal/yr)	-15.21	-9.16	-24.38
	[-2.21%]	[-1.50%]	[-1.88%]

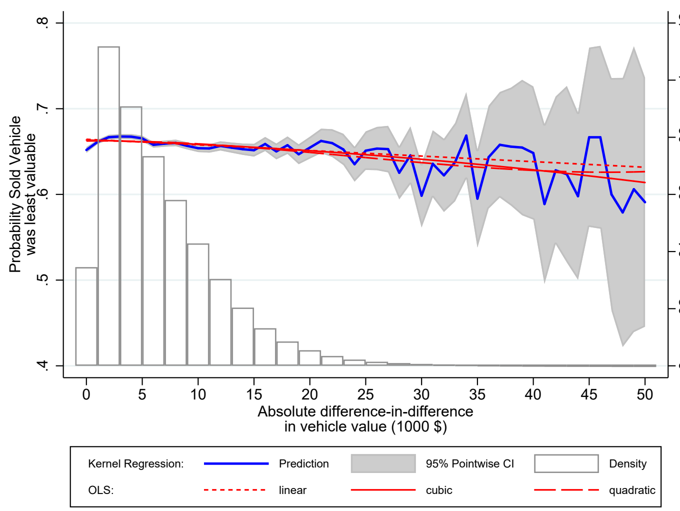
Effect if a 10% decrease in fuel intensity of a vehicle through the purchase of the next vehicle. Direct Effect is the effect from the exogeneous shock to the fuel intensity of the kept vehicle and the resulting change in fuel intensity of the follow-on purchase. Direct AS Effect is the own-vehicle effect in fuel consumption due to the change in operating costs. Indirect AS Effect is the cross-vehicle effect. Base fuel consumption and vehicle VMT are the sample mean for two car households. VMT effect assume a gasoline price of $3 per gallon.

Figure 1: Instrumental Variables Reduced Form Relationships

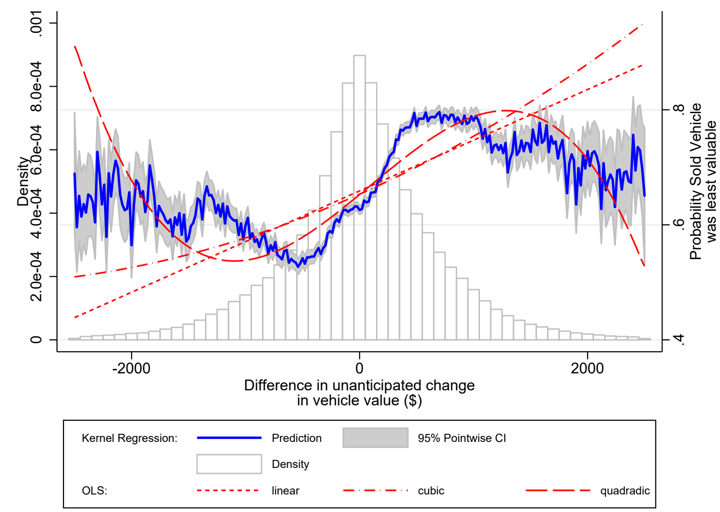
(a) Price Difference IV



(b) Price DiD IV



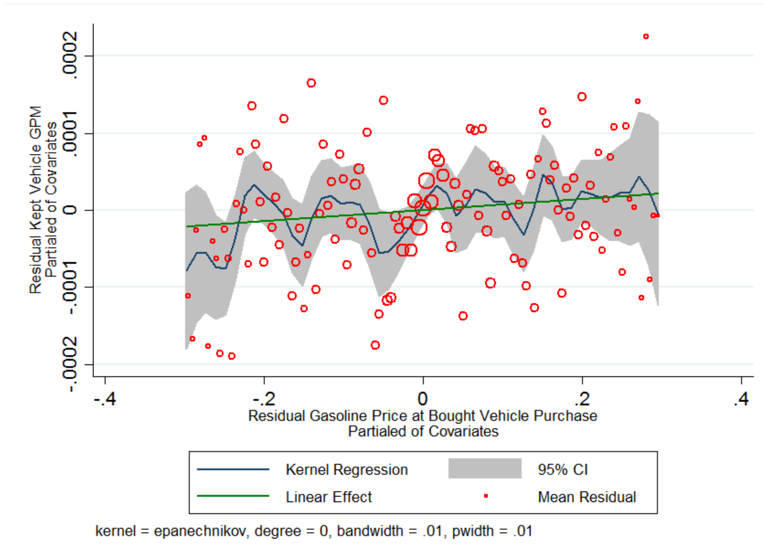
(c) Price Deviation from Trend DiD IV



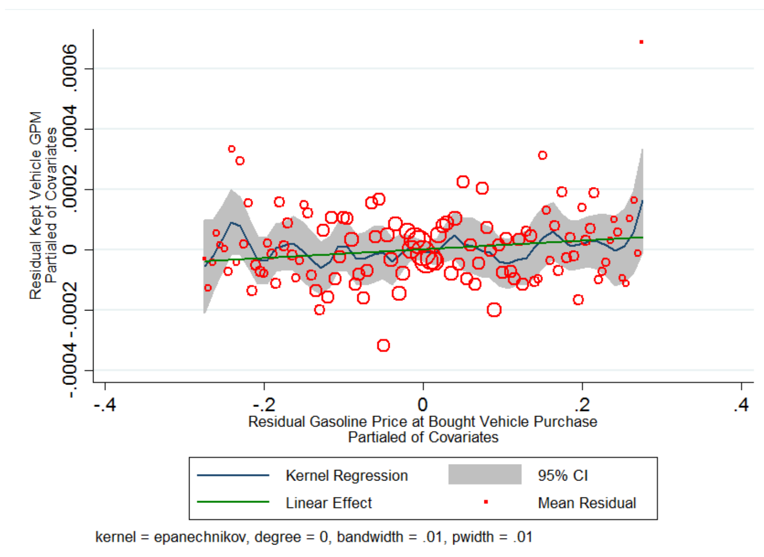
All 2x2 households. Probabilities conditional on a vehicle purchase (new or used) estimated within $1,000 bins. Binomial 95% confidence intervals shown in dashed lines. Values of the instruments in the Price Difference IV and Price DiD IV less than or greater than zero perfectly predict the least valuable vehicle in the portfolio and graphs are shown for the absolute value of these variables.

Figure 2: Reduced form relationship: Gas price at time of kept car purchase

(a) New Vehicle Purchases



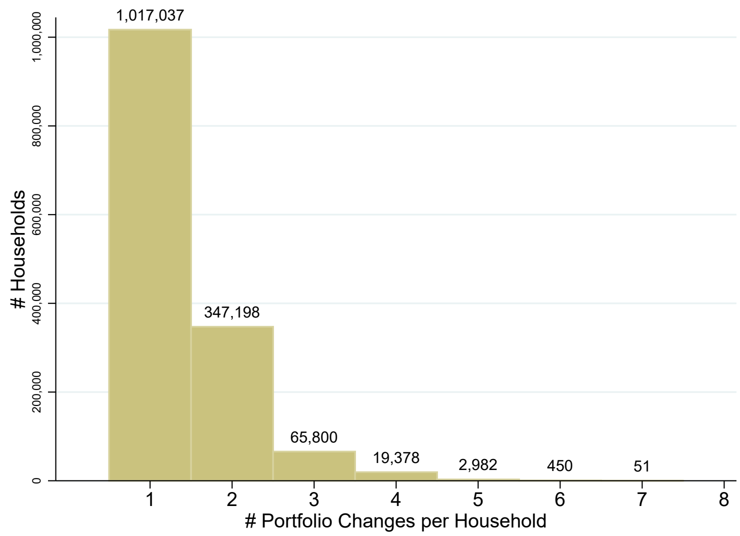
(b) Used Vehicle Purchases



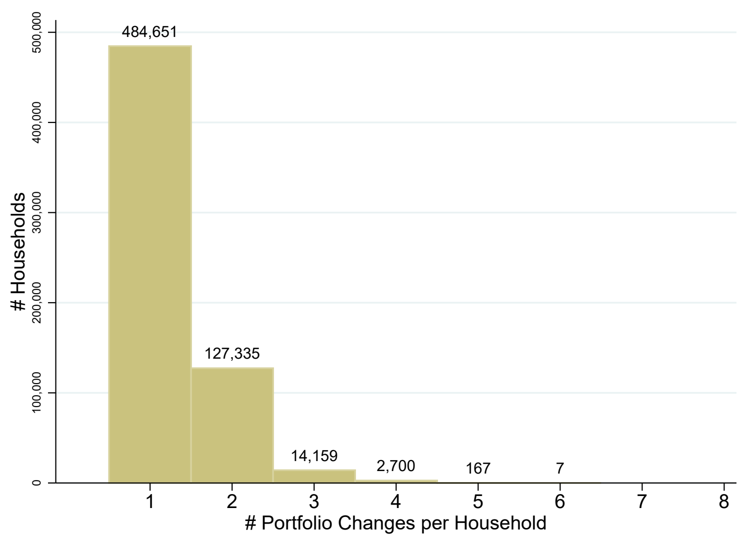
Plot of the reduced-form relationship between gasoline price at the time of kept vehicle purchase and the fuel economy (in GPM) of the purchased vehicle. Both variables are partialed of all other regression covariates. Graphs are limited to the 1st through 99th percentiles of residual kept vehicle gasoline price. Excludes observations where the household fixed effect perfectly predicts the outcome. Blue line is a kernel regression with Epanechnikov kernel and bandwidth 0.1. The gray band is the 95% confidence interval using the same kernel and bandwidth. The green line is the linear relationship estimated using OLS. Red circles are mean residuals for each 0.005 in kept vehicle GPM. The size of each circle is proportional to the number of observations used to compute the mean residual.

Figure 3: Number of Transactions per 2x2 Replacement Household

(a) Full Sample



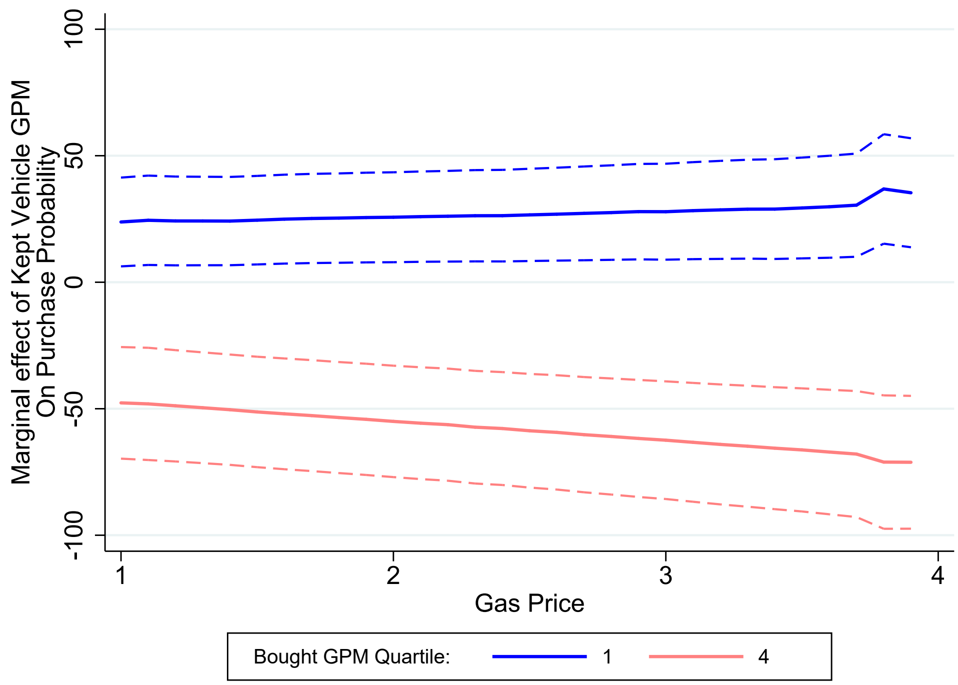
(b) IV Sample



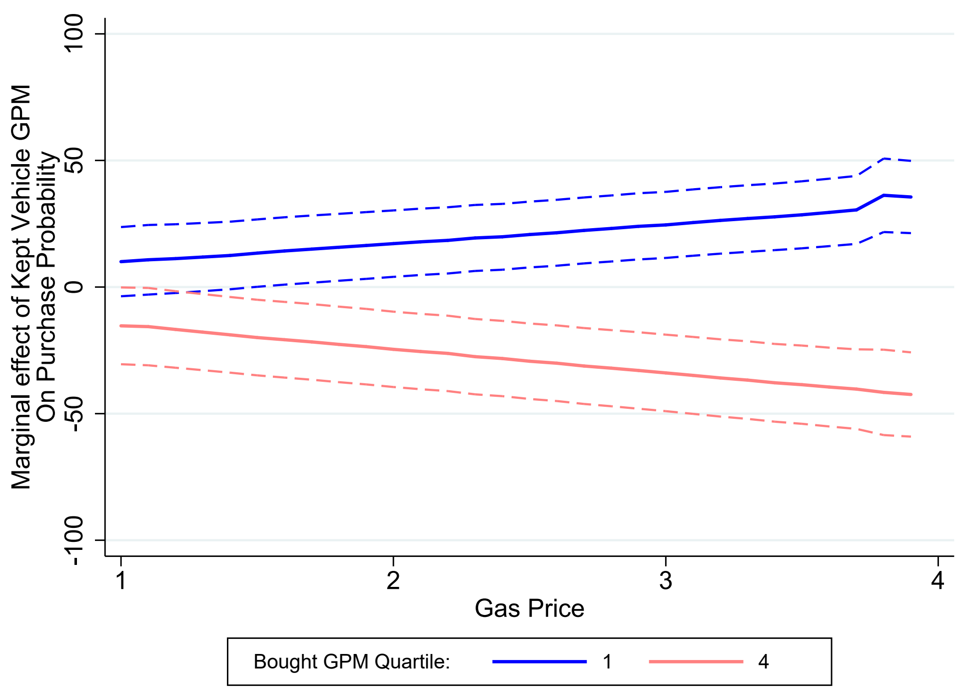
Distribution of the total number of observed vehicle transactions for each household from 2001 to 2007 for (a) the full sample of 2x2 replacement households and (b) households for which the data support deploying our IVs. In specifications including household fixed effects the fixed effect perfectly predicts the decision of a household if it only engages in one transaction. Other model parameters are identified by households engaging in multiple transactions from 2001 to 2007.

Figure 4: Marginal Effect of Kept Vehicle GPM on Bought Vehicle GPM

(a) New Vehicle Purchases



(b) Used Vehicle Purchases



Population average marginal of the kept vehicle GPM on the probability model a household purchases a vehicle in the 1st (blue) or 4th (red) quartile of the GPM distribution for used vehicle purchases. Estimated as a linear probability model using the specification shown in Equation (3.3) and the GP+DfT+I instruments. 95% confidence intervals robust to heteroskedastiticy and clustered by household shown in dashed lines.

A Substituting Vehicle-Miles Traveled

When the cost of transportation services changes differentially across vehicles in the portfolio, households may adjust their usage on the intensive margin by substituting towards the less expensive vehicle. We investigate the relationship between vehicle operating cost per mile and the household's allocation of VMT across the vehicles in its portfolio. We exploit two sources of variation in vehicle operating costs: variation in gasoline prices over time while holding the vehicle portfolio fixed and changes in operating costs resulting from changes in the fuel intensity of vehicles in the portfolio. For each vehicle $i \in \{1, 2\}$ in a two-vehicle portfolio, we compute the fuel cost in dollars per mile DPM_i as the price of gasoline, in dollars per gallon, times the fuel intensity, in gallons per mile. As these are two-vehicle portfolios, attributes of the other vehicle included in the regression are subscripted j.

We construct a yearly panel of two-vehicle households. For each vehicle i in year t, we compute the mean annual VMT (VMT_{it}) as miles driven between the closest preceding (at time $\underline{t}$) and upcoming (at time $\bar{t}$) odometer measurements for that vehicle.[19] We estimate the impact of operating costs on VMT using Equation A.1,

$$VMT_{h,i,t} = \beta_i DPM_{h,i,t} + \beta_j DPM_{h,j,t} + \Xi_h + \Theta_{h,t} + \Psi(\underline{t}, \bar{t}) + A^i(i) + A^j(j) + \varepsilon_{h,i,t} \tag{A.1}$$

where Ξ_h are household fixed effects, $\Theta_{h,t}$ are fixed effects for the county of residence of household h in year t, $\Psi_{h,\bar{t}}$ are fixed effects controlling for seasonality in driving,[20] $A()$ are controls for vehicle attributes,[21] and $\varepsilon_{h,i,t}$ is an idiosyncratic error which may have arbitrary correlation within households.

Estimating the causal impact of operating costs on VMT consumption using a household panel presents challenges to identification. A household's vehicle operating costs are the product of the current gasoline price, which we assume to be exogenous, and the fuel intensity of the household's vehicles. Over time the change in a household's preference for vehicles with low fuel intensity, or other attributes correlated with low fuel intensity, may be correlated with changes in demand for VMT over time. For example, a positive household income shock may increase demand for vehicle horsepower (increasing fuel intensity) but also increase demand for driving in general. Similar to previous regressions, we instrument for operating cost per mile using the gasoline price at the time

[19] We obtain odometer readings through DMV records each time a vehicle is transacted and at the time of biannual smog checks for vehicles six years and older.

[20] Demand for VMT follows seasonal patterns and odometer readings do not necessarily occur at the same time each year for a given vehicle. In fact, one may be concerned that the timing of odometer readings may be correlated with demand for VMT. We deploy two sets of controls to account for seasonality in the VMT measurement. First, in a simpler specification, we include fixed effects for the quarter-of-year of the upcoming odometer reading. In our preferred specification, we interact these fixed effects with counts of each quarter-of-year elapsed since the previous odometer reading.

[21] All specifications include nonparametric controls for the age of both vehicles in the portfolio and indicators for leased vehicles. Additional attribute controls in our preferred specification include indicators for vehicle class and continuous measures of vehicle curb weight, wheelbase, vehicle width, and engine displacement.

the vehicle was purchased and its interaction with current gasoline prices. Finally, as households may shift VMT between the more and less fuel-intense vehicles in response to changes in operating costs, we estimate separate regressions for the more (subscript 1) and less (subscript 2) fuel intense vehicle in the portfolio.

Estimates from Equation (A.1) are shown in Table 8. The top two sets of coefficients show the impact of vehicle operating costs on VMT of the more fuel-intense vehicle and the second set show the impacts for the relatively fuel-efficient vehicle. In either case, an increase in the cost of driving (DPM) of one vehicle introduces an incentive to shift VMT from that vehicle to the other vehicle in the household portfolio.

The specification in column (1) includes only operating costs, household fixed effects, and controls for vehicle age and leases. Here an increase in the operating cost of one vehicle appears to decrease the VMT of both vehicles in the portfolio. This may be reasonable in the presence of large income effects, but we suspect that it may be driven by omitted vehicle attributes. Column (2) adds controls for vehicle attributes to the previous specification. Here, all estimated parameters are of larger magnitude and the sign on the cross-effect for both vehicles becomes positive. That is, an increase in the operating cost of vehicle A reduces VMT of vehicle A but increases the VMT of vehicle B. These effects are similar in magnitude across the more and less fuel-intense vehicles in the portfolio, but the difference between the direct and indirect effects are larger for the relatively fuel-efficient vehicle, implying that households shift VMT demand from less to more fuel efficient vehicles in the face of increasing operating costs.

These results are robust to inclusion of alternative controls. The next three columns include additional fixed effects to account for various forms of unobserved heterogeneity. Column (3) adds fixed effects for the county of residence of household h in year t. Column (4) adds indicators for the quarter-of-year of the upcoming odometer reading to the specification in column (2). Column (5) is our preferred specification and includes county fixed effects from column (3) and provides robust controls for seasonality in driving patterns. It does so by using fixed effects for quarter-of-year of the upcoming odometer reading interacted with counts of each quarter-of-year elapsed since the previous odometer reading. This will flexibly control for determinants of VMT such as the number of summer seasons (high VMT) that elapsed between odometer readings. Coefficient estimates are quite similar cross Columns (2) to (5), showing our results are robust to the specific set of included fixed effects.