Household Diversification: The Vehicle Portfolio Effect

James Archsmith
UC Davis

Kenneth Gillingham
Yale

Christopher Knittel
MIT

David Rapson*
UC Davis

March 16, 2017

PRELIMINARY AND INCOMPLETE.
Please do not cite or circulate.

Abstract

Households value diversity in many settings, including financial assets, gender of children, and occupations. This paper quantifies the extent to which multi-car households exhibit preferences for a diversified vehicle portfolio. 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 purchased and find strong preferences for a diverse portfolio in fuel economy. We further find that this effect operates via car attributes that are correlated with fuel economy, including vehicle footprint and weight. This new evidence suggests that the portfolio effect 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.

*Archsmith: PhD Candidate in Economics, UC Davis, jesmith@ucdavis.edu. Gillingham: Assistant 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 the Energy Institute @ Haas Summer Energy Camp. 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

Households exhibit a taste for diversity across a number of settings. The value of diversity can arise because preferences are convex across products or because of risk aversion. Households routinely diversify financial asset holdings due to risk aversion, although evidence suggests that consumers may undervalue diversification with respect to international asset holdings, leading to what is known as the "diversification puzzle."[1] The taste for diversification in other settings is even more clear. Ben-Porath and Welch (1976) and Angrist and Evans (1998) show that households value diversity in the gender makeup of their children. This preference for diversity leads two-children households with two kids of the same gender to be much more likely to have a third child, compared to two-children households endowed with one boy and one girl.

Households may also value diversity in other settings where identification of these preferences is more difficult. For example, households may value diversity in occupations to reduce the risk of over-exposure to a single sector, a phenomenon likely to be particularly important for agricultural households in developing nations (Udry et al., 1995; Ellis, 2000). 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.

Similar identification challenges exist in estimating the preference for diversity in a household's vehicle portfolio. Diversity could 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, such as horsepower or fuel economy.

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 the use of a novel 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 other vehicles owned by the household. Our empirical strategy focuses on two-vehicle households and estimates how the fuel economy of a newly-added vehicle depends on the fuel economy of the vehicle already held by the household. Identification relies both on the richness of the panel, which allows us to control for household-level fixed effects, as well as a novel instrumental variables approach to control for the endogeneity of the fuel economy of the existing vehicle.

Understanding the patterns of demand for vehicles is 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

[1] This issue is not without controversy. Two papers (and titles) that underscore this debate are: Baxter and Jermann (1997) and Heathcote and Perri (2013).

multiple vehicles within a household. That is, these models of fuel economy standards assume that consumers choose only only 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 a taste for diversification, as suggested by Wakamori (2011) for Japanese households. In this case, a household endowed with a high horsepower vehicle will favor a more fuel-efficient 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. For example, 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, a strong preference for diversification would 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 diversification effects, but the taste for diversification 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 possible, 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 can address time-invariant unobserved preferences, but there would still be a concern if preferences change over time. Time-varying preferences may imply 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 are 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 rests on adequately addressing 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 value diversification in the vehicle portfolio. Increasing 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.

Changes in gasoline prices affect the preference for diversification in intuitive ways. As gasoline prices increase, the effect of the fuel consumption of kept vehicle 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 fuel consumption of kept vehicle and the probability of buying a car in the upper quartile of fuel consumption becomes even more negative.

To gauge the importance of the portfolio effect, we use our results to estimate the net effect of an exogenous increase in the fuel economy of the kept vehicle. We calculate the decrease in the fuel economy of the newly purchased vehicle when we increase the fuel economy of the kept vehicle by 10 percent. These calculations suggest that the portfolio effect can have large consequences on the net effect a one-time increase in fuel economy. The magnitude of the portfolio effect implies that 30-50 percent of fuel savings from increasing the fuel economy of the kept vehicle are eroded from the resulting decrease in fuel economy of the newly purchased vehicle, assuming an unconstrained consumer choice set. We find that the portfolio effect operates strongly through attributes that are correlated with fuel economy, including vehicle footprint and weight. This implies that fuel economy standards based on these attributes will be more exposed to the erosion of benefits due to this effect.

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 then present our results and their economic importance (Section 4). We conclude with a brief discussion of the implications for policymakers and empiricists (Section 5).

2 Context and Model

We begin by developing a simple economic framework inspired by Gentzkow (2007) to fix ideas and motivate our empirical work. Consider a household that has a vehicle and is considering adding a second vehicle. They may have just sold their second car, or they may be adding a new car to the household's vehicle portfolio. For simplicity, ignore the outside option of not purchasing a second vehicle. Consider a standard discrete choice framework with a random utility model. The household is the decision-maker in this framework; we abstract from issues of within-household bargaining.

Let the characteristics of a vehicle be given by the vector θ_V, where $V \in \{A, B, ...\}$ is the vehicle type. Vehicle 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 household has a vehicle of type A to start.

The household receives utility based on the characteristics of each type of vehicle, and may also receive utility from having a diversity of vehicles, which allows them to optimize their use of the vehicles (e.g., use the larger one for hauling goods and the more efficient one for the longer commutes). Let the contribution to utility from the diversity of the portfolio be given by Γ_{V_1,V_2}, where V_1 and V_2 are the vehicles types for the first and second car respectively.

The indirect utility for household i starting with a vehicle of type A and purchasing a vehicle of type B is thus given as:

$$u_i^{AB} = f(\theta_A) + f(\theta_B) + \Gamma_{AB} - \alpha(p_{A1} + p_{B2}) + \epsilon_i,$$

where $f(.)$ is a function that maps characteristics into consumer utility, and p_{Vj} is the remaining "present value lifetime ownership cost" for a vehicle of type V and order in the household j, where $j = 1$ refers to the vehicle the household already holds and $j = 2$ refers to the new vehicle. To understand the present value lifetime ownership cost, note that for p_{B2} the ownership cost includes the purchase price in addition to the future costs of fuel and maintenance, while for p_{A1} the ownership cost is just the future fuel and maintenance costs. These can be thought of as expectations based on the future expected driving of each of the vehicles, which need not necessarily be the same across vehicle types. α is the marginal utility of money.

In making the choice of which vehicle type to buy for the second vehicle, the household will compare u_i^{AB} to the utility from all other options. For example, suppose the household with vehicle of type A already is considering the option of buying another vehicle of type A. The utility u_i^{AA} would then be given by:

$$u_i^{AA} = f(\theta_A) + f(\theta_A) + \Gamma_{AA} - \alpha(p_{A1} + p_{A2}) + \varepsilon_i.$$

This equation allows for Γ_{AA}, although this might plausibly be assumed to be 0, for there may be no added utility from having a diverse portfolio if a household buys two of the same type of vehicle. One story for why $\Gamma_{AA} > 0$ would be that by having two of the same type of vehicle, the household receives additional utility by showing peers that they identify with a particular type (e.g., two hybrids showing the household is eco-friendly or two pickups showing the household is "tough").

2.1 Implications for vehicle choice

This paper is about how consumers choose their portfolio of vehicles. In other words, in this simple setting, it is about how the kept vehicle (i.e., vehicle 1 in this setting) influences the choice of the second vehicle.

Continuing the thought experiment where the consumer can only choose options A and B for the second vehicle, we can consider the conditions under which A or B is chosen. Specifically, the household chooses portfolio AB rather than AA if $u_i^{AB} > u_i^{AA}$, which is equivalent to (assuming $\Gamma_{AA} = 0$)

$$f(\theta_B) + \Gamma_{AB} - \alpha p_{B2} > f(\theta_A) - \alpha p_{A2}. \tag{2.1}$$

This simple inequality indicates that the household will choose B when the net utility of the vehicle characteristics, lifetime ownership cost, and portfolio effect from B dominate the net utility of the vehicle characteristics and lifetime ownership cost from A. Rewritten differently, we have

$$\Gamma_{AB} > f(\theta_A) - f(\theta_B) + \alpha(p_{B2} - p_{A2})$$

This states that if the added utility from having a portfolio is greater than the difference in utility from the characteristics of the two types of vehicles plus the difference in the lifetime ownership cost of purchasing the vehicles of the two types (converted to utility terms), then the household will choose a vehicle of type B. In other words, B will be chosen if a positive effect from portfolio diversity is larger than the household "type" effect due to the household valuing the characteristics and lifetime ownership cost. This can be rephrased as an empirically testable prediction:

The portfolio effect will dominate if we observe the household choosing vehicle B when the first *vehicle is A (as in the setting described so far), and the "type" effect will dominate if we observe the household choosing vehicle A when the first vehicle is also A.*

2.2 Changes in choice with gasoline price or fuel economy standards

This simple model also lends itself to a set of policy-relevant comparative statics. Consider what the model implies for the equilibrium household portfolio choice probabilities. Let $g(.)$ be the distribution of utilities in the population. Then for the simple choice between A and B for the new vehicle, the choice probabilities are given as follows:

$$Pr_{AB} = \int_{\mathbf{u}} I(u_{AB} > 0) I(u_{AB} > u_{AA}) dg(\mathbf{u}),$$
$$Pr_{AA} = \int_{\mathbf{u}} I(u_{AA} > 0) I(u_{AA} > u_{AB}) dg(\mathbf{u}).$$

Changes in Gasoline Prices

Consider a permanent increase in gasoline prices, with perfect foresight of this change. This will imply that p_{B2} and p_{A2} will change based on the relative fuel economy of the two vehicle types, which would directly influence Pr_{AB} and Pr_{AA}. But, the vehicle market would also re-equilibrate, with the relative prices of vehicles changing, which would indirectly influence Pr_{AB} and Pr_{AA}. These two effects can be summarized as:

1. Direct Effect: The probability of choosing the higher fuel economy vehicle will increase.
2. Indirect Effect: The relative prices of new vehicles in equilibrium will change, so that higher fuel economy vehicles will increase in price relative to others.

The indirect effect will work in the opposite direction as the direct effect, and depending on the marginal utility of income α, may even dominate. We will see the net of these two effects in the empirics.

Changes in Fuel Economy Standards

Fuel economy standards are substantially more complicated due to the complex nature of the attribute-based standards and the variety of possible automaker responses to meet the standard, such as changing attributes or changing prices. But one insight emerges clearly from the theory: if there is a strong portfolio effect, it can tilt the balance towards diversification (i.e., with a sufficiently large Γ_{AB}, $u_i^{AB} > u_i^{AA}$). This would mean that if fuel economy standards lead to a higher fuel economy new vehicle within the household, a strong portfolio effect would imply that the next vehicle would have lower fuel economy. Of course, if fuel economy standards are tightening, the fuel economy of all newer vehicles would be higher. However, fuel economy is highly correlated with other attributes, so if there is an attribute-based standard and a preference for diversification, the fleet-wide average fuel economy in future time period could be lower when there is a strong portfolio effect. Moreover, in the long-run, this portfolio effect may encourage automakers to increase the footprint of vehicles in order to meet the increased demand for vehicles with attributes that are negatively correlated with fuel economy.

3 Data and Identification

The cornerstone of our dataset is the universe of California vehicle registration records that occurred from 2001-2007.[2] The DMV dataset includes every vehicle registered under the residential designation code. In California every vehicle must be registered annually. Each record includes the registrant's US Census block group identifier, 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 are available for each VIN whenever the vehicle is sold and upon receiving biannual Smog Check certification. 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 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 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, but not equal to N^s then the ending portfolio size is the number of vehicles registered in the later years.[3]

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

[3]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.

[Table 1 about here]

Table 1 shows the distribution of household portfolio transitions. Specifically, 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.

Many of the 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."[4] 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 portfolio of households with more than two cars affects replacement decisions. Does a portfolio effect for those households operate via the highest-VMT kept car, or the newest? Or must the portfolio effect be defined in a higher dimension? Given that no clear answer exists to these questions, we choose the transparent path of focusing on the two-car replacement households.

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 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]

Many 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]

[4]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.

3.2 Identification

To understand the challenges associated with identifying the portfolio effect, we consider a thought exercise. For a given two-car household that replaces one of its vehicles, we would like to know the effect that randomly dropping one of the cars and exogenously perturbing the fuel economy of the "kept" car has on the choice of fuel economy of the "bought" car. That is, we would like to randomly assign one car to be the "kept" car, and to randomly assign it a GPM (f^k) to see how changes in f^k affect the household's observed choice of f^b, the GPM of the car purchased. This is what we mean when we refer to the "portfolio effect". There are two identification challenges to operationalizing this thought experiment to retrieve an estimate of the portfolio effect in our observational dataset. We propose instrumental variables to address each.

Identification Challenge 1: Which Vehicle to Keep? As described earlier, our sample isolates two-car households that replace one of their cars with another. In general, the choice of which car to keep is endogenous and many potential stories could be told about preferences and conditions that would lead to one or the other of the cars being kept. Such a choice is inconsistent with our thought experiment of randomly assigning the household its f^k. However, our data offer several appealing instrumental variables.

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 use. 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: $\Delta P^{kt} = P_t^k - P_t^d$. 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, which would violate the exclusion restriction.

The second 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 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". We find it difficult to come up with a violation of the exclusion restriction for this instrument. Recall that the concern is that a correlation exists between 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 if those 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 in their respective first stages.

Identification Challenge 2: Omitted Variables. The household's choice of f^k may be influenced by many factors that are unobservable to the researcher. These may include unobserved car attributes that are correlated with GPM (e.g. safety via weight) or unobserved household attributes (e.g. features of commutes). Of particular interest in our setting is fuel economy, and the confounding effect that unobservables may have on f^k. To address this identification challenge, we follow an instrumental variable approach and control for time-invariant household preferences via household fixed-effects.

Our preferred instrument for addressing omitted variables is the price of gasoline at the time of the kept car purchase, $p^{gas}_{it_k}$. 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 are exogenous with respect to the 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, f^k.

Recall that the relationship between f^k and f^b is theoretically ambiguous. A preference for diversification in the household portfolio will lead to a negative correlation, but complementarity between attributes associated with fuel economy may lead to a positive correlation. By extension, the relationship between $p^{gas}_{it_k}$ and f^b may also appear to 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 relationship emerges. highlighting the complementarity between vehicles in the portfolio. If gasoline prices were high when a household purchased the vehicle it is keeping, that vehicle will tend to be more fuel efficient. The household, valuing complementarity in its portfolio of vehicles, will tend to purchase a less efficient (higher GPM) vehicle for the complementary vehicle in the portfolio. This relationship is born out in observed household behavior. Low gasoline prices at the time the kept vehicle is purchased (a low value of $p^{gas}_{it_k}$) are correlated with purchases of more fuel efficient vehicles (lower f^b) to replace the other vehicle in the household's portfolio.

Further Consideration: Within-Household Variation. We further refine our identification strategy to address potential concerns that first-order household preferences for cars with high (or low) fuel economy may overwhelm our ability to identify the potentially second-order preference for a diversified fuel economy portfolio. The richness of our panel dataset allow us to deploy household fixed effects to control for time-invariant unobservable preferences such as this. The importance of these controls can be seen via a simple example. Suppose that there are two types of households. They both prefer a diverse vehicle portfolio, but one (say household A) has an overall preference for gas guzzlers and the other (household B) for fuel efficiency. Examining each household's portfolio may reveal that household A holds cars that are both in the highest GPM quartile, whereas household B holds cars that are both in the lowest GPM quartile. Were we to randomly remove one of the cars from each portfolio, they would each be left with a car in the GPM quartile of their preference. They would also be likely to purchase a new car that is also in that GPM quartile. On the surface, it would appear as though the households have a low preference for a diversified vehicle portfolio. However, that may be a false conclusion. Were we to examine GPM *within* the preferred quartile, we may discover that the household prefers diversification within that range. Using household fixed effects as controls allows our empirical approach to reveal the true impact of an exogenous marginal change in the fuel economy of the kept vehicle on the (marginal) fuel economy of the vehicle purchased.

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 IV 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_{it}^b), or quartile indicators of that variable. Regressors of interest include gasoline price at the time of purchase, GPM of kept car (f_{it}^k), and their interaction.

If households value diversity in their portfolio of vehicles, it is a reasonable prior belief that the effect of kept vehicle GPM on the vehicle purchase decision may differ if the household is replacing the relatively fuel-efficient vehicle in the portfolio as opposed to the fuel-intense vehicle. Our base specification allows the kept vehicle GPM effect and its interaction with current gasoline prices to vary across decisions where households replace either the most or least fuel-intense vehicle. For notational simplicity, let these these divisions of the sample be denoted 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. 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". The baseline specification is

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

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 equation 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 car that it keeps 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.

First, we describe the vehicle price difference instruments precisely. In "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 estimate the households expectation of annual vehicle depreciation using 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[5]

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

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 ($p_{it_d}^{gas_d}$).

[5] 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.

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

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 additional 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.[6] 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 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:

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

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

4 Results

The objective of this section is to present and justify our empirical approach, illuminate the effect of key variables on the choice of bought car GPM, and present a simple counterfactual analysis that demonstrate the policy-relevance of our findings. The section is comprised of two main parts.

First, we show results from various regression specifications. This allows us to demonstrate the importance of our instrumental variables approach and the inclusion of household fixed effects, both of which qualitatively alter key coefficient estimates. We then present marginal effects of

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

kept car GPM on bought car GPM, which reveals household preferences for a diversified portfolio. Motivated by the correlation between GPM and other vehicle attributes, we also examine the effect of kept car GPM on footprint, engine displacement, and weight of the bought car. These results provide context for interpreting the counterfactual analysis that follows.

In the second subsection, we present results from a counterfactual analyses in which we exogenously perturb the fuel economy of the kept vehicle, which is roughly what a successful fuel economy standard would do.[7] We observe the extent to which potential gasoline consumption reductions are either magnified or eroded due to portfolio considerations.

4.1 Regressions and Marginal Effects

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. Columns 1 and 2 estimate parameters using OLS, ignoring household-level heterogeneity and potentially endogenous variables. Columns 3 and 4 instrument for endogenous variables using the GP+DfT+I instrument. Columns 5 and 6 assume all variables are exogenous but control for household-level heterogeneity using household fixed effects. Finally, Columns 7 and 8 estimate parameters using the GP+DfT+I instrument and household fixed effects.

It is clear from this progression of specifications that accounting for endogeneity and unobserved household heterogeneity are both important. Specifications without household fixed effects compare decisions across households and do not reflect the thought experiment described earlier, which looks within households. 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 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, the evidence indicates that households will tend towards replacing their dropped car with one that is qualitatively similar in GPM to the kept car. This supports the hypothesis of a strong household "type" effect. In other words, some households simply prefer fuel efficiency and others that prefer gas guzzlers (presumably

[7] A fuel economy standard will cause new car buyers to purchase more fuel efficient cars, on average. The question of interest in this counterfactual is whether and to what extent this effects the desired choice of fuel economy in the next vehicle purchase.

due to power, comfort, safety, etc). It is only when we look within the household that the portfolio effect of interest is seen. Despite the possibility that some households prefer low- or high-GPM cars in general, there appears to be a preference for diversity in GPM within that band.

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 4 to 5 offer a graphical representation of the population average marginal effect in the highest and lowest GPM quartiles.

[Table 5 about here]

[Figures 4 - 5 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 it's kept car decreases (increases). All of the marginal effects have a negative sign, and all are statistically significant with 99 percent confidence.

The overall story is clear: households incorporate portfolio considerations in their vehicle purchase decisions. That is, if we were to increase the fuel efficiency of the kept car, households would buy a second car that has attributes associated with lower fuel efficiency. The results also exhibit other interesting patterns. Households keeping the more efficient vehicle in their initial portfolio (column 7) exhibit a stronger portfolio effect than those that keep the less efficient vehicle (column 8). This result is consistent with a narrative that when households keep their more efficient car, fuel economy is a more important factor in their decision than other attributes.

For households that keep their fuel efficient car, the magnitude of the portfolio effect increases with the gasoline price. This may once again 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 4(a) to 5(b) display 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.

4.1.1 Attribute Regressions and Marginal Effects

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 attributes.

Table 7 displays marginal effects from specifications that are analogous to columns 7 and 8 of Table 5, except with the alternative car attribute as the dependent variable.[8] 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, particularly 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 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 a counterfactual exercise that is designed to reveal the size of the portfolio effect as it relates to fuel economy standards.

4.2 Counterfactuals

We view our results as particularly informative about the long-run impacts of fuel economy standards, such as Corporate Average Fuel Economy (CAFE) standards in the United States. CAFE affects the suite of cars available for purchase, and their purchase prices. Once these cars become part of a household's vehicle portfolio, the change in attributes (relative to no CAFE standard) may influence the subsequent choice of vehicle purchased. In particular, if households exhibit a preference for diversification in their portfolio, increasing the fuel efficiency of their kept car will lead to a less fuel efficient second-car purchase.

The counterfactuals that we describe here examine the net effect of an exogenous decrease of 10% in the kept car GPM (i.e. a fuel efficiency increase) on predicted gasoline consumption. The net effect includes changes in gasoline consumption relating to use of the (now more efficient) kept

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

car, but also changes in gasoline consumption relating to the use of the bought car whose GPM is influenced via the portfolio effect that we estimate above. Throughout the exercise, we assign cars the vehicle-miles traveled (VMT) that are observed in our dataset. We do not adjust these to account for a rebound effect, although in reality, one may exist.[9] Notice that this implies the only changes in gasoline consumption that we estimate will occur due to changes in the (extensive margin) choice of bought car fuel economy that are induced via the portfolio effect. It may well be the case that there is a first-order effect of gas prices on bought car GPM, but we seek to isolate only the portfolio effect in this exercise.

Underpinning the validity of our counterfactual exercise is an assumption about the equilibrium in characteristic space. We are assuming that our baseline regression estimates emerge from a hedonic equilibrium in which consumers are trading off optimally between various vehicle attributes, producers are providing the level of those attributes optimally, and the marginal rate of substitution equals the marginal rate of transformation. Our counterfactual illustrates the consequence of increasing one attribute along thus envelope.

[Tables 8a and 8b about here]

Turning to Table 8a, first note the observed gas consumption of the kept and bought vehicles. Each is estimated to consume approximately 530-570 gallons of gasoline per year (the product of VMT and vehicle GPM). Changes in f_{it}^k do have a large effect; a 10% decrease in f_{it}^k mathematically reduces kept car gas consumption by 10%, as can be seen in the top row of the panel. The increases in bought car gas consumption reflect the portfolio effect. It is immediately apparent that these effects are quite large, and that the majority of gasoline conservation enjoyed by a kept car GPM improvement is eroded by the household's response via the choice of lower bought car fuel economy. Indeed, our estimates predict that the portfolio effect offsets about 48% of the fuel savings from increasing the kept vehicle's fuel economy.

The story is qualitatively similar when the bought car is used instead of new. Table 8b reveals that households purchasing used cars have, on average, lower gasoline consumption associated with the vehicle added to the portfolio, but higher gasoline consumption with the kept vehicle. The exogenous GPM changes associated with the kept car are 31% offset by the portfolio effect.

These results are quite startling and have unfortunate implications for the effectiveness of fuel economy standards as a way to reduce gasoline consumption. The magnitude of the portfolio response implies that strong forces will be at work, particularly where standards are attribute-based. The used car market, which is not covered by CAFE, is another channel through which the portfolio effect preferences may be manifested. Increased demand for used, fuel inefficient cars will occur as a result of increased efficiency from CAFE. The increase in demand will lead used gas

[9] As fuel economy changes due to policy, the cost per mile traveled also changes. Consumers faced with this change in relative prices may choose a different VMT. See Borenstein (2015) and Gillingham, Rapson, and Wagner (2016) for more on the rebound effect.

guzzlers to be more valuable, and thus more slow to be retired from the fleet (similar to the effect documented in Jacobsen and van Benthem (2015)).

5 Conclusions

The effects of a number of polices applied the light duty vehicle market depend crucially on vehicle choice patterns. Typically empirical estimates of vehicle choice 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 a novel instrumental variables approach, we find evidence that households value diversification. Exogenous increases in the fuel economy of the kept car lowers 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. Increases 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 increases reduce the probability the household buys a car in the upper quartile.

We also find that gasoline prices affect the preferences for diversification in intuitive ways. As gasoline prices increase, the effect of the fuel consumption of kept vehicle 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 fuel consumption of kept vehicle and the probability of buying a car in the upper quartile of fuel consumption becomes even more negative.

To understand the economic importance of this taste for diversification, we use our results to estimate the net effect of an exogenous increase in the fuel economy of the kept vehicle. These calculations suggest that the portfolio effect can have large consequences of the net affect a one-time increase in fuel economy, particularly when the standard is based on attributes that are correlated with fuel economy.

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.

Baxter, Marianne and Urban J Jermann. 1997. "The International Diversification Puzzle Is Worse Than You Think." *American Economic Review* 87 (1):170–180. URL https://ideas.repec.org/a/aea/aecrev/v87y1997i1p170-80.html.

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, F. 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.

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.

Heathcote, Jonathan and Fabrizio Perri. 2013. "The international diversification puzzle is not as bad as you think." *Journal of Political Economy* 121 (6):1108–1159.

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 Portfolio Size	End 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	OLS	IV	IV	HHFE	HHFE	HHFEIV	HHFEIV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$f^d \geq f^k$	$f^k > f^d$	$f^d \geq f^k$	$f^k > f^d$	$f^d \geq f^k$	$f^k > f^d$	$f^d \geq f^k$	$f^k > f^d$
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: Net Effect of Kept Vehicle GPM Changes on Gasoline Consumption

(a) Households Purchasing New Vehicles

Vehicle	Observed Gasoline Consumption (gal/yr)	Change in Gasoline Consumption New Vehicles
Kept	537.64	-10.00
Bought	555.34	4.76
Total	1,092.98	-5.24

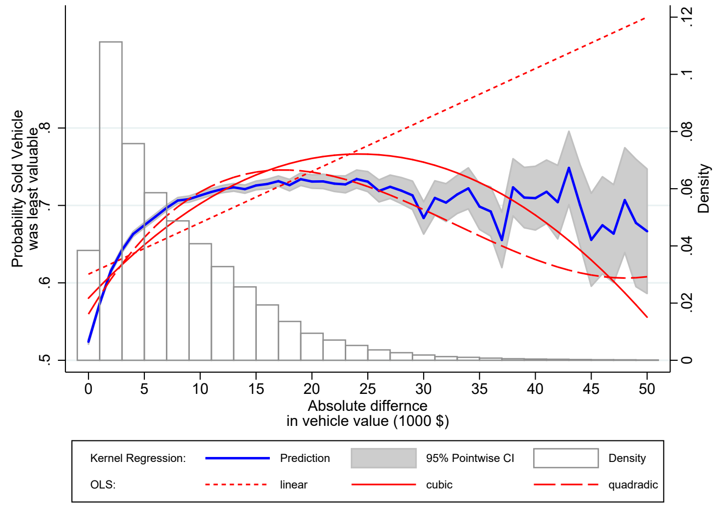
(b) Households Purchasing Used Vehicles

Vehicle	Observed Gasoline Consumption (gal/yr)	Change in Gasoline Consumption Used Vehicles
Kept	569.12	-10.00
Bought	537.29	3.14
Total	1,106.41	-6.86

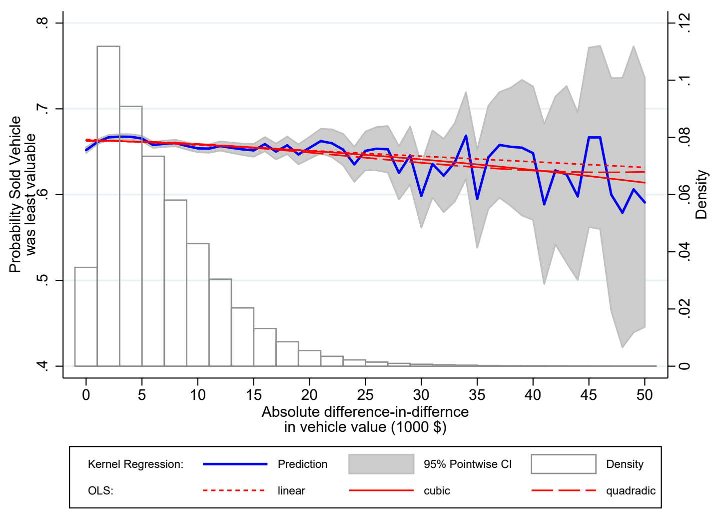
Predicted average change in fuel consumption resulting from an exogenous decrease in kept vehicle GPM of 10% (e.g., from 27.5 MPG to 30.6 MPG) for vehicle purchases. Change in fuel economy expressed as percentage of annual gasoline consumption of the kept vehicle.

Figure 1: Instrumental Variables Reduced Form Relationships

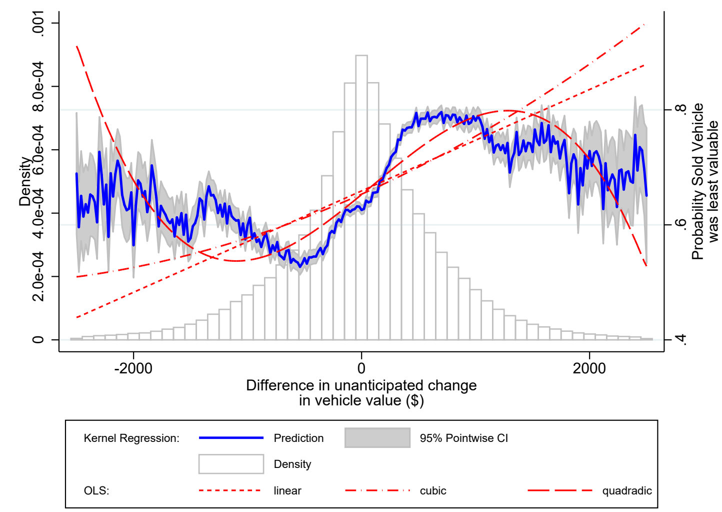
(a) Price Difference IV



(b) Price DiD IV



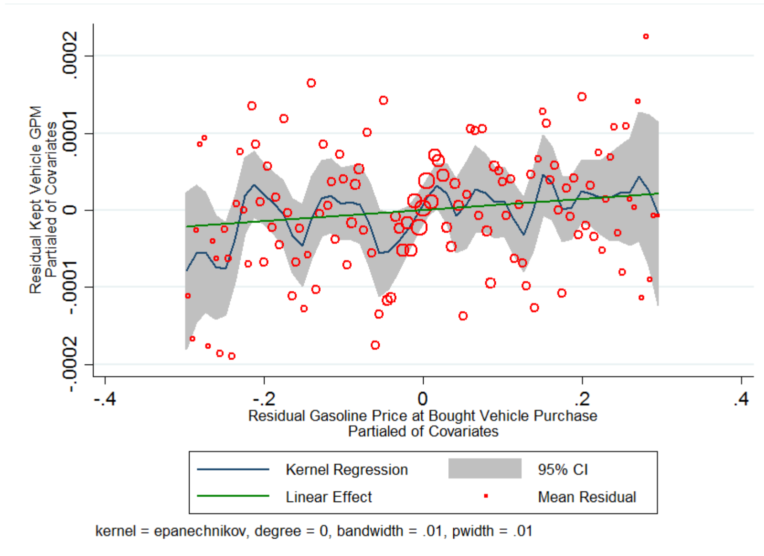
(c) Price Deviation from Trend DiD IV



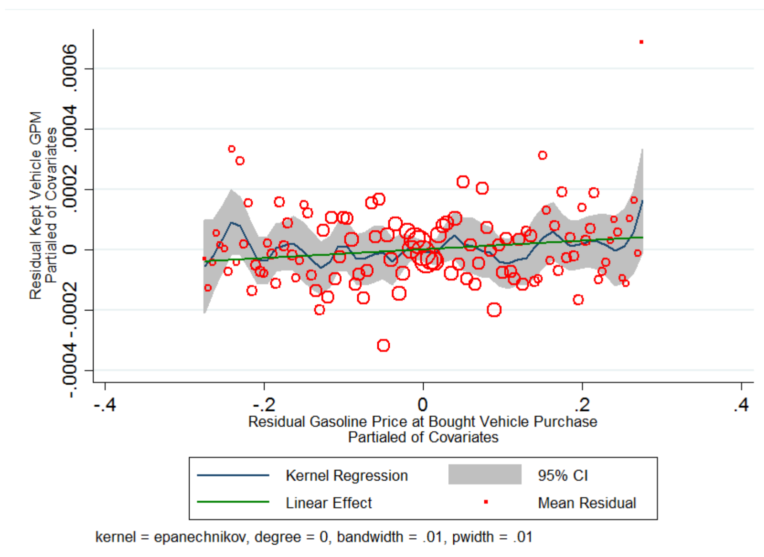
All 2x2 households. Probabilities conditional on a vehicle purchase (new or used) estimated within $1,000 bins. Binomial 95% confidence intervals shown in dashed lines. Values of the instruments in the Price Difference IV and Price DiD IV less than or greater than zero perfectly predict the least valuable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



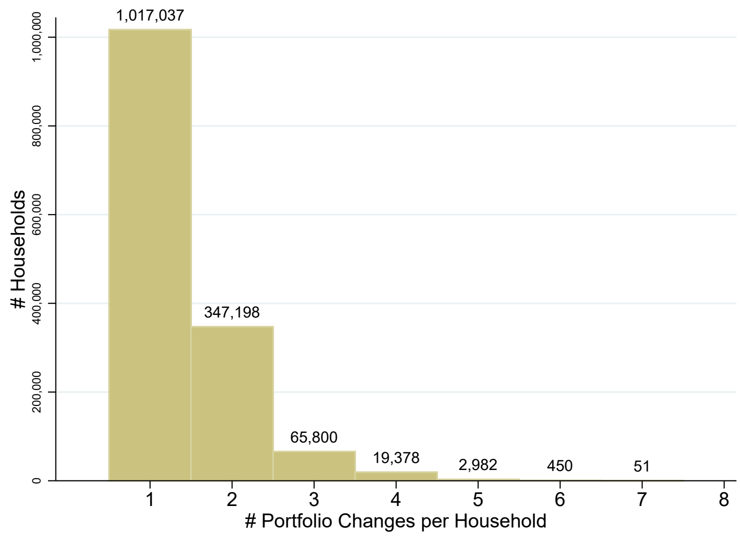
(b) Used Vehicle Purchases



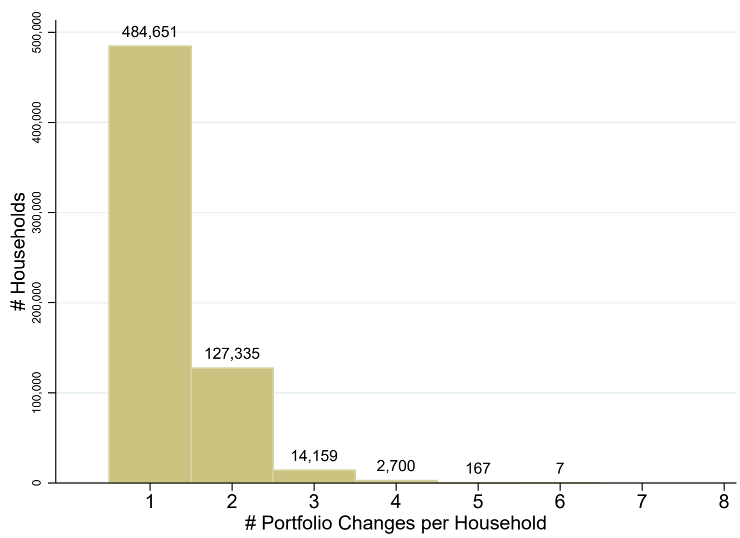
Plot of the reduced-form relationship between gasoline price at the time of kept vehicle purchase and the fuel economy (in GPM) of the purchased vehicle. Both variables are partialed of all other regression covariates. Graphs are limited to the 1st through 99th percentiles of residual kept vehicle gasoline 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

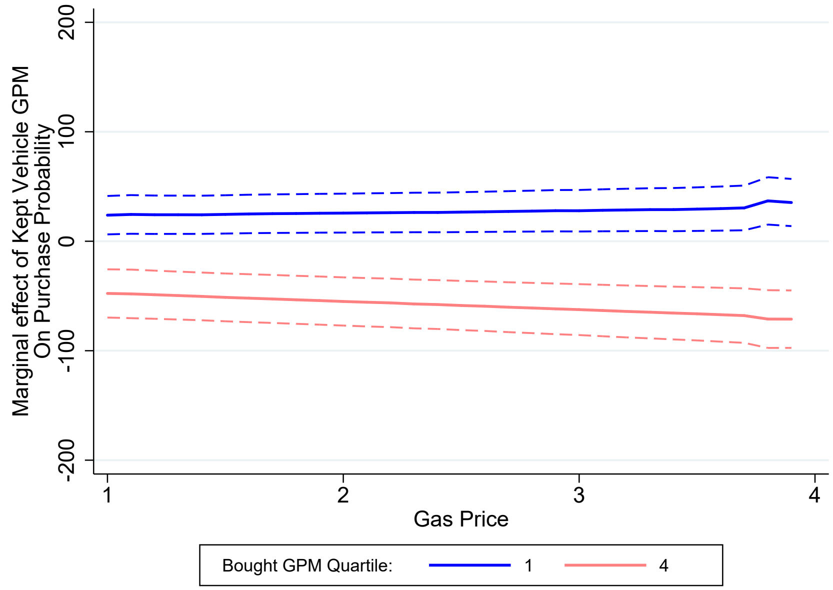


(b) IV Sample



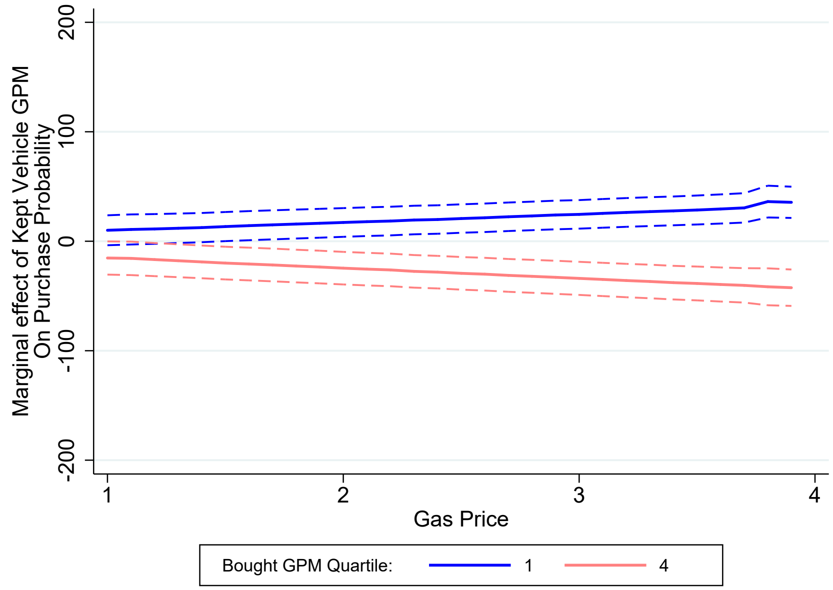
Distribution of the total number of observed vehicle transactions for each household from 2001 to 2007 for (a) the full sample of 2x2 replacement households and (b) households for which the data support deploying our IVs. In specifications including household fixed effects the fixed effect perfectly 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 - New Vehicle Purchases



Population average marginal of the kept vehicle GPM on the probability a household purchases a vehicle in the 1st (blue) or 4th (red) quartile of the GPM distribution for new 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.

Figure 5: Marginal Effect of Kept Vehicle GPM on Bought Vehicle GPM - 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.