FUEL CONSUMPTION AND GASOLINE PRICES:
THE ROLE OF ASSORTATIVE MATCHING BETWEEN HOUSEHOLDS AND AUTOMOBILES

H. Spencer Banzhaf
Taha Kasim

Fuel Consumption and Gasoline Prices: The Role of Assortative Matching between Households and Automobiles
H. Spencer Banzhaf and Taha Kasim
NBER Working Paper No. 22983
December 2016
JEL No. Q4,Q5

ABSTRACT

Analyses of polities to reduce gasoline consumption have focused on two effects, a compositional effect on the fuel economy of the automotive fleet and a utilization effect on how much people drive. However, the literature has missed a third effect: a matching effect, in which high-utilization households are matched to fuel-efficient vehicles in equilibrium. We show that higher gas prices should lead to stronger assortative matching. Empirical estimates using US micro-level data are consistent with this hypothesis. We find the effect of a gas tax through the matching effect is larger than the compositional effect, with a $1 tax saving 1.7% of gas consumption.

H. Spencer Banzhaf
Department of Economics
Andrew Young School of Policy Studies
Georgia State University
P.O. Box 3992
Atlanta, GA 30302
and NBER
hsbanzhaf@gsu.edu

Taha Kasim
Dept. of Economics, Furman Univ.
3300 Poinsett Hwy.
Greenville, SC 29613
mohammed.taha@furman.edu

Fuel Consumption and Gasoline Prices:
The Role of Assortative Matching
between Households and Automobiles[*]

H. Spencer Banzhaf[†]
Taha Kasim[‡]

December 2016

Abstract. Analyses of polities to reduce gasoline consumption have focused on two effects, a *compositional effect* on the fuel economy of the automotive fleet and a *utilization effect* on how much people drive. However, the literature has missed a third effect: a *matching effect*, in which high-utilization households are matched to fuel-efficient vehicles in equilibrium. We show that higher gas prices should lead to stronger assortative matching. Empirical estimates using US micro-level data are consistent with this hypothesis. We find the effect of a gas tax through the matching effect is larger than the compositional effect, with a $1 tax saving 1.7% of gas consumption.

I. Introduction

Both to reduce dependency on petroleum imports and to reduce air pollution, many countries seek to reduce gasoline consumption from passenger vehicles. In the United States, passenger vehicles account for 16% of all greenhouse gas emissions (US EPA 2015 p. 2-28). Additionally, they are a major source of conventional pollutants such as carbon monoxide, ozone, and fine particulates. These pollutants can cause such health effects as headache, respiratory disease, heart disease and premature mortality. They may be particularly potent in such vulnerable populations as the elderly and young children. Indeed, studies have shown that *in-utero* and childhood exposure to these gasses have the potential to affect various health outcomes in the short- and the long-run (Currie et al. 2014).

[†] Professor, Dept. of Economics, Georgia State Univ., hsbanzhaf@gsu.edu.

[‡] Visiting Assistant Professor, Dept. of Economics, Furman Univ., mohammed.taha@furman.edu.

Policies to reduce gasoline consumption generally fall into two main categories: price-based policies and performance-based policies. Price-based policies such as a gasoline tax work by making driving more expensive. Among OECD countries, the US has the lowest gasoline taxes save Mexico, at $0.49 per gallon (average state plus federal tax) as of 2011, compared to $4.10 in the UK and $3.95 in Germany (US DOE 2011). Performance-based policies mandate minimum average fuel economy standards for the fleet. Examples include the US Corporate Average Fuel Economy (CAFE) standards and similar policies in Europe, Japan, and elsewhere.

Traditionally, the literature assessing the effectiveness of such policies has decomposed their effects into two channels: a *compositional effect*, relating to the average fuel efficiency of the fleet, and a *utilization effect*, relating to how much people drive (e.g. Bento et al. 2009; Goldberg 1998; Li, Timmins, and von Haefen 2009; Puller and Greening 1999). By increasing the cost of driving, gas taxes create an incentive to drive less as well as to obtain a more fuel efficient car. Thus, they operate through both channels. Performance standards like CAFE mandate more fuel efficient cars (on average), but do not create incentives to drive less—indeed, by putting people in more fuel efficient cars they reduce the cost of driving and increase utilization (the so-called "rebound effect"). Bento et al. (2009) and Gillingham (2012) measure both the compositional and the utilization effects using structural models in which household simultaneously choose an automobile and their level of driving in a consistent framework. They find that the utilization effect is bigger than the compositional effect.

In this paper, we show that this large literature has overlooked a third effect, a sorting or *matching effect*. In theory, even if the compositional and utilization effects are held to zero— that is, even if the fleet remains exactly the same and each individual drives exactly the same number of miles annually—an increase in gasoline prices could still reduce gasoline consumption if it leads to stronger assortative matching between high-utilization households and the fuel efficiency of the cars they drive. Intuitively, if a household does little driving, the fuel efficiency of its car has little impact on overall consumption (at the extreme, a car parked in the garage burns no gas regardless of its fuel efficiency), whereas if a household drives great distances the fuel efficiency of its car matters a good deal. Conversely, if a car is very fuel efficient, its utilization has little impact on overall consumption (at the extreme, the utilization of a bicycle would not matter), whereas if a car is fuel inefficient its utilization matters a good deal.

To our knowledge, this matching effect has been completely neglected in the literature. Yet it is not trivial. To gauge the potential importance of the matching effect, if household demand for VMT and vehicle fuel efficiency were perfectly matched in 2009, gasoline consumption would have been 15% lower, even with no change in the composition of the fleet and no change in individuals' utilization.

In addition to pointing out the importance of this effect, in this paper we show that, *a priori*, there are good reasons to believe that higher gasoline prices increase the degree of matching. The intuition is as follows. Increases in gasoline prices increase demand for fuel-efficient cars relative to inefficient ones, because the monetary savings of owning an efficient vehicle increase. But there is exogenous heterogeneity among households in their utilization. When gasoline prices rise, the savings of owning an efficient vehicle increase more for higher-utilization households than for low-utilization households. Consequently, high-utilization households are now comparatively more likely to outbid low-utilization households for fuel-efficient cars (ceteris paribus). Thus, when gas prices increase, high-utilization households tend to be allocated more fuel-efficient cars than before, while low-utilization households are allocated less efficient cars. Accordingly, in equilibrium, after an increase in the gasoline prices, there should be a stronger assortative matching from households, based on the amount of driving they are likely to do, to the fuel economy of the cars they own.

Finally, we test this hypothesis empirically. Using data from the US National Highway Transportation Survey, we adopt a difference-in-differences framework, comparing the way (exogenously) high-utilization households change the fuel economy of their car when gas prices increase relative to the way low-utilization households do. We find a statistically significant short-run matching effect from gas prices that is robust to a variety of identification strategies, including ordinary least squares (OLS) and different ways of instrumenting for utilization. Specifically, we find that when gasoline prices increase by $1, a household responds on average by driving a vehicle that gives 0.01 to 0.03 more miles per gallon for every 1,000 miles it drives in the short run. Using variation only in gasoline taxes, we find an effect of 0.03 to 0.06 miles per gallon for every 1,000 miles. This matching affect results in an additional savings of about 1.7% of fuel consumption, or 2.1 billion gallons and 92 million tonnes of carbon dioxide (CO_2) annually. These effects are likely to be bigger in the long run.

In so far as it focuses on households' choices for fuel efficiency at the extensive margin and how such choices are affected by gasoline prices, our paper relates to the large literature on automobile demand and the compositional effect, including Allcot and Wozny (2014); Austin and Dinan (2005); Bento et al. (2009); Busse, Knittel, and Zettelmeyer (2013); Feng, Fullerton, and Gan (2013); Gillingham (2012); Goldberg (1998); Jacobsen (2013); Klier and Linn (2010); Li, Timmins, and von Haefen (2009); Li, Linn, and Muehlegger 2014); Sallee, West, and Fan (2016); and Puller and Greening (1999. Overall, this literature has found that an increase in gasoline prices increases the demand for more fuel efficient cars, which in equilibrium modestly nudges the composition of the automobile fleet toward fuel efficiency. In particular, new cars entering the fleet become more fuel efficient while older inefficient cars are scrapped more rapidly. Interestingly, Allcot and Wozny (2014); Busse, Knittel, and Zettelmeyer (2013); and Sallee, West, and Fan (2016) find that, in between, the shift in demand manifests in price rather than quantity effects, with the prices of fuel-efficient used cars rising relative to inefficient cars when gas prices rise. Our work can be viewed as exploring the heterogeneity in these demand effects at the extensive margin, showing that they are larger for (exogenously) higher-utilization households than lower-utilization households. In equilibrium, this gives rise to more assortative matching, which in turn becomes a third channel through which gas prices affect gasoline consumption.

The rest of the paper is structured as follows: Section II explains the matching effect using a theoretical model; Section III introduces the basic empirical approach and describes the data; Section IV presents the results; Section V briefly discusses the implications; and Section VI concludes.

II. Theoretical Framework

A. Conceptual Basics

Consider an economy with a set $\mathcal{C}$ of cars indexed by c and a set $\mathcal{I}$ of individuals indexed by i. These sets have the same measure, which we normalize to one, and in equilibrium there is a bijective mapping of cars to individuals. Individual i's fuel use is equal to its utilization, measured in vehicle miles travelled (VMT), times the inverse of the fuel efficiency of its car, or gallons per mile (GPM):

$$FUEL_i = VMT_i \times GPM_{c(i)}, \tag{1}$$

where $c(i)$ represents the car assigned to individual i.

From this simple tautology, it would seem that changes in fuel use can be decomposed into two effects, a change in VMT (utilization) and a change in fuel economy (GPM). While this is true at the individual level, to extend it to the over-all economy is to commit the fallacy of composition. Overall fuel usage is the sum over all individuals of these products:

$$TOTAL\ FUEL = \int_{\mathcal{I}} VMT_i \times GPM_{c(i)} f(i) di, \tag{2}$$

where $f(\ )$ is the density. This sum of products can be rewritten as

$$TOTAL\ FUEL = \overline{VMT} \times \overline{GPM} + \text{COV}(VMT, GPM). \tag{3}$$

Thus, changes in total fuel consumption can be decomposed into *three* effects: the utilization effect measured in changes in average VMT, the compositional effect measured in changes in the average GPM of the fleet, and the matching effect which represents changes in the covariance between a households VMT and the GPM of the car assigned to it in equilibrium. Though seemingly obvious when written this way, to our knowledge the existence of this third effect has been overlooked in the literature.

Our hypothesis is that when gasoline prices increase, high-GPM cars (i.e. fuel-inefficient ones) will tend more strongly to be assigned to low-VMT households. Thus, the covariance between VMT and GPM will decrease and total fuel usage will fall, even holding constant average VMT and average GPM (indeed, even holding constant every individual's VMT and every vehicle in the fleet).

The argument for this hypothesis comprises two basic steps. First, when gasoline prices increase, the demand for high-GPM cars falls relative to the demand for low-GPM cars, and— more to our point—this shift is greater for high-VMT households. Second, in equilibrium, this shift in demands leads to a more systematic assignment of high-GPM cars to low-VMT households. Because it is a simpler point to make, we begin with the second point first.

B. The Matching Effect

A partial equilibrium model is sufficient to explain how this shift in relative demands leads to the matching effect. For simplicity, let there be two types of households that differ only in terms of what Gillingham (2012) calls their "utilization type," or their demand for VMT for reasons exogenous to the choice of car: thus, there are high-utilization households (H) and low-utilization households (L). Let there be two car types: fuel-inefficient (I) and fuel-efficient (E). In Figure 1 Panel A shows the demand for high-GPM cars relative to low-GPM cars. Heterogeneity in the relative demand for such cars is reflected by D^H and D^L, which indicate the relative marginal willingness to pay (WTP) of each respective household type. This WTP could be positive or negative, depending on which car type is preferred. Additionally, we take no stand on whether the high-VMT types or the low-VMT types have a higher relative WTP for such cars: high-VMT types may want them more than low-VMT types for their comfort when driving, or want them less than low-VMT types because of their operating cost. As an illustration, Figure 1 shows an example of a baseline scenario where $D_0^H = D_0^L$. The market demand is D_0 and the relative price is P_0. The supply elasticity plays no role in the argument; in this example, the market supply curve is perfectly inelastic, as might be consistent with the market for used cars too young to be scrapped.

Panel A depicts a situation where gasoline prices increase, so the relative WTP for high-GPM vehicles falls. The new market demand is D_1 and the price falls to P_1 in the vehicle market. This is consistent with the findings of Allcott and Wozny (2014); Busse, Knittel, and Zettelmeyer; and Sallee, West, and Fan (2016). Our insights hinge on the fact that the relative WTP for a high-GPM automobile will particularly decrease for the high-utilization types, as higher gasoline prices translate into greater costs for them. This is shown in this figure where D_1^H is now lower than D_1^L. As a consequence, in equilibrium, the low-VMT types now have a greater share of the high-GPM vehicles than before ($\frac{A}{A+B}$ versus $\frac{1}{2}$). By the same logic, as shown in Panel B, the high-VMT types now have a greater proportion of low-GPM vehicles relative to the baseline case.

Consequently, the correlation between VMT and fuel inefficiency is now lower, so that the covariance in Equation (3) is now lower with higher gas prices. (Equivalent, the correlation between VMT and fuel economy is higher.) In the empirical portion of this paper, we test this

hypothesis directly by looking at the matching of car types to utilization types under variation in gasoline prices.

Before getting to that empirical work, however, we must complete the argument by showing that the demand for fuel inefficient cars does shift back more among high utilization types than low utilization types when gas prices rise.

C. Relative Shifts in Demand

We begin by generalizing our two-car, two-household model to continuums of types. In particular, let $\mathcal{C}$ be a simply ordered set of cars such that for any two elements $c, c' \in \mathcal{C}$, $c' > c \leftrightarrow GPM_{c'} > GPM_c$. In differentiating cars only by their GPM, we do not imply that mileage is the only relevant factor. To the contrary, we allow for the possibility that as GPM increases, the car becomes more desirable in other ways such as increased horsepower and/or comfort.

Additionally, we assume that individuals can be ranked according to their utilization type. In particular, let $\pi(i, c, VMT)$ be individual i's inverse demand curve, or marginal WTP, for VMT *conditional* on being in car c. We make the following assumption about the ordering of $\mathcal{I}$:

Assumption A1. For any two elements $i, i' \in \mathcal{I}$, if $i' > i$ then $\pi(i', c, VMT) > \pi(i, c, VMT)$.

That is, as we go through the index i, the demand curves for VMT, conditional on any car c, shift up. Note Assumption A1 does *not* necessarily imply that $VMT_{i'} > VMT_i$. For one thing, i and i' may face different gas prices. Moreover, they may choose different cars, which also affects their VMT choice. Assumption A1 only implies that, if they were in the same arbitrary car and faced the same prices, i' would choose higher VMT than i.

With these preliminaries, consider now the WTP of some household i for an arbitrary car c, relative to having no car at all, which is a function of gasoline prices p^G:

$$
\begin{aligned}
WTP(c, i, p^G) = f(c, i) + \int_0^{VMT^*(i,c,p^G)} \pi(i, c, VMT) dVMT \\
- VMT^*(i, c, p^G) GPM(c) p^G.
\end{aligned}
\tag{4}
$$

The individual's WTP for car c comprises three terms. The first term $f(\ )$ represents a value unrelated to actually driving the car. It represents an option value of having it in the garage, a collec-

tor's value, or some similar utility value. The remaining two terms represent the consumer surplus from driving the car. Integrating under $\pi(\ )$ up to the optimal level of driving given gasoline prices, $VMT^*(i, c, p^G)$, gives the total (gross) value of driving car c. From this, we subtract the cost of driving car c when gasoline prices are p^G, which is $VMT^*(i, c, p^G)GPM(c)p^G$.   In equilibrium, an individual buys the car with the highest consumer surplus, after subtracting out the price of purchasing the car.

Our model depends on the triple difference among car types, utilization types, and gasoline prices—that is, on how WTP differs by car type, how that difference differs by utilization type, and how that difference-in-difference differs with gasoline prices. For simplicity, assume that all the functions in Equation (4) are twice differentiable in all their arguments.[1] In Appendix A, we derive the following relationship on the third partial cross derivative of WTP:

$$\frac{\partial^3 WTP(i, c, p^G)}{\partial i \partial c \partial p^G} = \begin{aligned} & -\frac{\partial VMT^*(i, c, p^G)}{\partial i}\frac{dGPM(c)}{dc} \\ & +\frac{\partial^2 VMT^*(i, c, p^G)}{\partial c \partial p^G}\frac{\partial \pi(i, c, VMT^*)}{\partial i} \\ & +\frac{\partial VMT^*(i, c, p^G)}{\partial p^G}\frac{\partial^2 \pi(i, c, VMT^*)}{\partial i \partial c} - \frac{\partial^2 VMT^*(i, c, p^G)}{\partial i \partial c}GPM \\ & +\frac{\partial^2 VMT^*(i, c, p^G)}{\partial i \partial p^G}\left[\frac{\partial \pi(i, c, VMT^*)}{\partial c} - \frac{dGPM(c)}{dc}p^G\right]. \end{aligned} \quad (5)$$

Note that here we are taking the car type c as exogenous, so there are no effects of p^G or utilization type on the optimal choice of c; instead, in taking the partial derivative with respect to c, we consider an exogenous change of car type.

Conditional on this car, the individual does optimally choose *VMT*. At equilibrium, the individual sets marginal WTP for an additional mile equal to the cost, so $\pi(i, c, VMT^*) = GPM(c)p^G$. As we take the derivatives of the two sides of this expression with respect to c, we get two effects. The marginal WTP for an addition mile increases, $\partial \pi / \partial c > 0$, through a utility or *comfort effect*, as driving becomes more comfortable and/or more fun. But at the same time, the actual price paid for an additional mile increases through the *GPM effect*, that is $GPM'(c)p^G$

---

[1] This assumption is not necessary but simplifies the presentation.

> 0. Thus, what we call the overall *car effect on VMT*, i.e. $\partial VMT^*/\partial c$, cannot be signed *a priori*. However, it is the opposite sign as the "rebound effect" in the literature, which is generally estimated to be positive (Gillingham et al. 2013, Linn 2016, Small and van Dender 2007, West et al. 2013). A crucial question when evaluating Equation (5) turns out to be the difference in the car effect on VMT between different values of i. Consider the following Assumption:

Assumption A2. If $i' > i$ and $c' > c$, then $[\pi(i', c', p^G) - \pi(i', c, p^G)] - [\pi(i, c', p^G) - \pi(i, c, p^G)] \geq 0$.

Or, if π is twice differentiable, then $\partial^2 \pi / \partial i \partial c \geq 0$. Because the effect on operating costs is the same at a given car for all household types, A2 implies that the car effect on VMT is non-decreasing in i, $\partial^2 VMT / \partial i \partial c \geq 0$. Thus, as i exogenously moves from c to c' it might decrease its VMT, whereas a higher utilization type i' will decrease VMT less or perhaps increase it. Assumption A2 is stronger than A1 but is consistent with the idea that demand for driving is increasing in i. For example, Assumption A2 is consistent with inverse demand curves for VMT that are proportional in c and i.

Assumption A2 is sufficient for the following proposition:

Proposition: $\frac{\partial^3 WTP(i,c,p^G)}{\partial i \partial c \partial p^G} < 0$ so long as either (i) the car effect on VMT is sufficiently small or (ii) it is the same sign as $\partial^2 VMT^*/\partial i \partial p^G$.

A proof of the proposition can proceed by inspecting each line on the right-hand side of Equation (5). The first line is negative, since by definition of our ordered sets, $\frac{\partial VMT^*}{\partial i} > 0$ and $\frac{dGPM}{dc} > 0$. This line, the lowest-order term in Equation (5), represents the basic intuition for why the triple difference is negative. As gas prices increase, operating high-GPM cars becomes more expensive, but this is especially true for high-utilization types.

Completing the proposition requires considering the higher-order terms in Equation (5). The second line also is negative. $\partial \pi / \partial i$ is positive by Assumption A1. $\partial^2 VMT / \partial c \partial p^G$ is negative because, by the law of demand, VMT is decreasing in the effective price of driving, which is $GPM(c)p^G$, and the effect of a given change in gasoline prices on the effective price of driving is greater for higher-GPM vehicles.

The third line is non-positive by Assumption A2. In the first term, Assumption A2 states

that $\partial^2\pi/\partial i\partial c$ is non-negative, and it is multiplied by $\partial VMT^*/\partial p^G$ which is negative by the law of demand. In the second term, $\partial^2 VMT^*/\partial i\partial c$ also is non-negative by Assumption A2 and of course GPM is positive. This third line is the only place we use Assumption A2. Notice that it is actually a sufficient—not necessary—condition: It would suffice for $\partial^2 VMT^*/\partial c\partial i$ to be "not too negative" rather than non-negative.

This then leaves the last line of Equation (5) as the only term that cannot be signed definitively *a priori*. The Proposition requires either that this line is small or negative. The first term in the product, $\partial^2 VMT^*/\partial i\partial p^G$, cannot be signed definitively. $\partial VMT^*/\partial i$ is positive by Assumption A1, but whether VMT becomes more or less responsive to changes in gasoline prices with utilization type is another question. However, if VMT becomes less price elastic with increasing utilization types, as seems plausible, then this term is positive. The term in square brackets is the same sign as the car effect on VMT, which is the negative of the rebound effect. $\partial\pi/\partial c$ represents the comfort effect while $dGPM(c)/dC$ represents the GPM effect. The literature on the rebound effect suggests this entire term is probably small and certainly negative (Gillingham et al. 2013, Linn 2016, Small and van Dender 2007, West et al. 2013). This logic suggests the terms in the last line are unlikely to reverse the other effects present in Equation (5). For it to do so, it would have to be the case that as people are exogenously placed in larger cars, they actually want to increase their VMT because of the comfort effect, but only high-utilization types are willing to continue purchasing that VMT as gasoline prices increase, making them outbid the low types for the car. This seems unlikely to be the case and even more unlikely to overcome the lower order effects.

Under Assumption A2 (and the caveats in the previous paragraph), the Proposition states that as gasoline prices increase, the demand for high-GPM cars falls relative to the demand for low-GPM cars more so for high-utilization types than for low-utilization types. This is the key result of this sub-section. Given the equilibrium results in the previous sub-section, the implication is that as gasoline prices increase, the correlation between utilization and GPM decreases (or increases between utilization and MPG).

III. Empirical Strategy and Data

A. *Empirical Strategy*

Our basic strategy for testing our hypothesis is to run the following difference-in-differences regression. The model compares the fuel economy of cars matched to high-utilization households to those matched to low-utilization households among period-places with comparatively high and low gasoline prices. Our preferred specification is:

$$MPG_{icmt} = \alpha + \beta Price_{mt}^{gas} + \gamma VMT_{icmt} + \theta \left[Price_{ct}^{gas} \times VMT_{icmt} \right]$$
$$+ \mathbf{x}_{imt}{}'\boldsymbol{\delta} + \boldsymbol{\eta_m} + \lambda_t + \varepsilon_{icmt}, \tag{6}$$

where $MPG_{icmt} = 1/GPM_{icmt}$ is miles per gallon of vehicle c owned by household i in metro area m in year t, $Price_{mt}^{gas}$ is the per gallon price of gasoline in metro m in year t, VMT_{icmt} is VMT of vehicle c owned by household i, $\mathbf{x}_{imt}$ are demographic variables of household i, $\boldsymbol{\eta_m}$ are MSA dummies, λ_t is a year dummy and ε_{icmt} is an idiosyncratic error term. The matching effect of higher gasoline prices is represented by θ and our hypothesis is that $\theta > 0$. In Equation (6), λ_t controls for time-varying unobservables and η_m controls for time-invariant city-specific unobservables.

Four points about our empirical strategy deserve further comment. First, because we include year and city dummies, and because (as discussed below) we use contemporaneous gas prices, we view our estimates as short run effects. Given the transaction costs involved in changing cars, we would expect the effects to be higher in the long run.

Second, we assume gasoline prices are exogenous. In doing so, it is important to emphasize that we are using micro-level data with city and year dummies. Thus, conditional on city and year, our assumption is that gas prices are unaffected by individuals' choices of fuel economy. Thus, a year's deviation in tastes for MPG from the city's long-run average (insofar as they differ from national annual deviations) are uncorrelated with city-year gas prices. In an alternative specification, we use city-month dummies, reducing the error in Equation (6) to the individual-level only.

Third, in our basic approach using OLS, we also assume that VMT is exogenous to the

fuel efficiency of the car. However, as discussed above, VMT and MPG likely are chosen simultaneously. If the MPG effect outweighs the comfort effect, households with more fuel efficient vehicles drive more, which is consistent with the rebound effect. With the exception of Linn (2016), recent evidence has suggested that this effect is fairly small (Gillingham et al. (2013), Small and van Dender (2007), West et al. 2013). If so, this endogeneity problem is probably small enough to ignore and OLS will be unbiased. Accordingly, we first use OLS to estimate Equation (6). However, in the following subsection we also consider a variety of instrumental variables (IV) strategies, including interactions of household characteristics with gasoline prices and nonlinearities inherent in the structural model of Bento et al. (2013). We generally find our results are robust to these approaches.

Fourth, in the data (described below) VMT is observed on the *car*, not the individual. Thus, VMT_{icmt} is the VMT of the car, not an individual or household. In households with only one car, these would be observationally equivalent. In households with multiple cars, this raises the question of how cars are assigned to individuals. In our base model, we effectively treat the data as if multi-car households not only have an exogenous total VMT, but an exogenous split of VMT into baskets to be assigned to cars. For example, if a household has two cars with 10,000 and 5,000 VMT respectively, we treat the household as first exogenously splitting its total VMT into one basket of 10,000 miles and one of 5,000. However, we treat the assignment of cars to these VMT baskets as an endogenous choice. Thus, whether the 10,000 mile basket goes to one car and the 5,000 to the other or vice versa is treated in our approach as a choice which can change under differing economic conditions. This approach allows for some within-household matching insofar as a high-VMT individual in a household can be more likely to be assigned the high-MPG car in the household as gasoline prices increase. However, it is conservative in that it does not allow VMT to otherwise be shifted to the high-MPG car, as it forces the split to stay constant (10,000 and 5,000 in this example). As an alternative, we also consider each household, rather than each car, as a unit, using only one car per household or alternatively per-car VMT and average fuel economy of all cars in the household. This approach is even more conservative as it removes all within-household matching.

B. Data

Our data can be divided into three categories: (i) a confidential version of the 2001 and the 2009

National Household Travel Survey (NHTS), which surveys a random sample of US households about the cars they own and their VMT for each car; (ii) quarterly gasoline prices from 2001-2 and 2008-9 at the city level, purchased from the American Chamber of Commerce's ACCRA Cost of Living Index (COLI) database; and (iii) data on the fuel economy of automobiles by make and model, from the U.S. Department of Energy. These data were merged to produce a distinct data-set which provides information on driving habits, household characteristics, average gasoline prices paid and the fuel economy of vehicles owned by these households.

The 2001 and 2009 NHTS samples consists of 69,815 households and 150,147 households respectively, from both urban and rural areas in the US. The 2001 NHTS was conducted from March 2001 through June 2002 and the 2009 NHTS was conducted between March 2008 and May 2009. Both data sets contain information on household driving patterns, including VMT and the type of car(s) owned by the household (make, model, vintage).[2] Furthermore, the survey includes demographic data about the household (income, number of household members, age distribution, race, work status). The confidential NHTS data to which we had access provides more specific geographic data on each household than the public version, down to its zip code. After cleaning the data, we are left with 23,739 households in the 2001 sample and 137,610 households in the 2009 sample.[3]

To each household, we match quarterly gasoline prices in the household's metropolitan statistical area (MSA) at the time of the NHTS survey. Using contemporary gasoline prices, ra-

---

[2] NHTS collected VMT data through (i) direct questioning; (ii) estimations based on odometer readings (only for 2001 NHTS); and (iii) estimations based on travel diaries. NHTS data reports two different measures of VMT, "reported VMT" and "bestmiles." Reported VMT is based only on the direct survey questions. Bestmiles is based on regression techniques combining information from all three reporting techniques. Since, bestmiles requires more information, this variables had more missing values compared to VMT. Following Bento et al. (2009), we used reported VMT in our central estimates, using bestmiles wherever reported VMT is missing. As sensitivity analyses, we also considered using only reported VMT or only bestmiles. Our results are qualitatively similar when using those data instead.

[3] Both reported VMT and besmiles ranged from 0 to 200,000 miles. However, following Bento et al. (2009), mileage below 100 and above 60,000 indicate possible coding errors and so were dropped. This resulted in 7,139 (15%) and 18,922 (7%) observations dropped from the 2001 and 2009 NHTS data, respectively. Additionally, vehicles that were more than 24 years old were not identified in the 2009 NHTS sample and so were dropped from the 2001 sample for consistency. Finally, diesel and electric vehicles were dropped. The results with these assumptions relaxed are similar to the main results and are available upon request.

ther than longer-term averages, is appropriate if households assume gasoline prices take a random walk, so that current prices represent the best long-term forecast. In previous work, Bento et al. (2009), Berry et al. (1995), Gillingham (2012), and Goldberg (1995) all assume that prices take a random walk and that only current prices matter. David and Hamilton (2004) and Geman (2007) have shown that gasoline prices do actually follow a random walk. Li, Timmins, and von Haefen (2009) also find the lagged gasoline prices have little impact on purchasing decisions. Finally, Anderson et al. (2011) and Anderson, Kellogg, and Sallee (2013) provide empirical evidence suggesting that on average expected real gasoline prices as reported in surveys are equal to current prices. Accordingly, we match gasoline prices at the time of the survey. Spatially, ACCRA provides gasoline prices for more than 300 MSAs. Using the geographical indicators of the NHTS data, we matched these prices to each household in the NHTS data-sets, using the US Department of Housing and Urban Development (HUD) USPS ZIP Code Crosswalk Files.[4] Gasoline prices are converted to 2009 dollars using the Consumer Price Index (CPI).

Finally, the fuel economy data was obtained from the Department of Energy's website www.fueleconomy.gov. This website provides the fuel economy data by vehicle make, model and year and these variables were used to match the vehicle fuel economy with the NHTS data.[5]

Table 1 provides summary statistics of household responses in the NHTS. The first column reports the descriptive statistics for the entire sample. This is broken down by NHTS survey years in columns 2 and 3, respectively. Column 4 compares columns 2 and 3 and reports the differences and standard errors in parentheses. Table 1 shows that the average real gasoline price increased from $1.62 in 2001 to $3.14 in 2009 ($1.35 to $3.17 using nominal dollars). Moreover, the heterogeneity in gasoline prices increased, with the standard deviation increasing from 0.22 to 0.65. One reason for this is major fluctuations in gasoline prices during the period 2008-

---

[4] For households in the NHTS outside these 300 metropolitan areas, we used NHTS-reported gasoline prices. The NHTS prices are at a more aggregate geographic level, namely Petroleum Administration for Defense Districts (PADDs). These districts are similar to Census regions in that each state is assigned to a given region, but the regions are drawn up in such a way as to maximize gasoline distribution. ACCRA and NHTS gasoline prices are highly correlated. We regressed ACCRA gasoline prices on NHTS gasoline prices and used predicted gasoline prices for those NHTS MSAs that could not be matched with ACCRA data.

[5] We use a combined MPG which is a weighted average based on the standard assumption (used by the US EPA and others) that the automobile was operated 55% under city driving conditions and 45% under highway conditions.

2009. Gasoline prices continued to increase during the first two quarters of 2008, reaching a peak of $4.60 and then started to decline before escalating again in the beginning of 2009. Panel A in Figure 2 shows the kernel densities of (real) gasoline prices for the two survey years. It shows very little overlap in the distributions, with the 2009 distribution being both higher and more diffuse. Panel B of Figure 2 plots average MSA gasoline prices in 2009 against average MSA gasoline prices in 2001. The figure shows that there is a good deal of cross-sectional variation in changes in gasoline prices, as well as both cross-sectional and temporal variation in levels of gasoline prices.

Along with this increase in gasoline prices, Table 1 shows that average VMT declined about 12% and average MPG improved by about 1%. Figure 3 plots the distribution of annual VMT and MPG for both surveys. The figure illustrates the slight leftward shift in the distribution of VMT over time and little detectable change in that for MPG.

IV. Results

A. *Patterns in the Raw Data*

We begin our analysis of the matching effect by looking at patterns in the raw data. In Figure 4 Panel A, the expected value of miles per gallon (MPG) conditional on VMT is plotted for the years 2001 and 2009 using linear regression. Overlayed on each plot is a similar counter-factual plot of what the relationship would look like, given the actual distribution of VMT and the actual composition of the fleet in each year, if there were perfectly assortative matching, so that the lowest-VMT individual had the lowest MPG (highest GPM) car, etc. The figures show that the observed assignment is far from perfectly assortative. A shift from the empirical relationship to perfect matching is one way to gauge the potential for the matching effect. If household demand for VMT and vehicle fuel efficiency were perfectly matched in 2009, gasoline consumption would have been 15% lower, even with no change in the composition of the fleet and no change in individuals' utilization.

Our hypothesis does not center on how strong the matching effect should be in any particular cross section. Rather, our hypothesis is that when gas prices are higher, the matching effect should be stronger. Since gas prices increased from an average of $1.62 per gallon in 2001 to $3.14 in 2009 (in real terms), we would expect more assortative matching over time. The raw data reveal this pattern. Figure 4 Panel A shows that, between 2001 and 2009, as gasoline prices

increased, the relationship between these two variables went from slightly decreasing to slightly increasing. Panel B displays the CDF of car fuel economy among individuals in the highest quartile of VMT and lowest quartile of VMT, for 2001 and 2009 respectively. It shows that in 2001, the distribution of MPG among the lowest VMT quartile actually first-order stochastically dominated the distribution among the highest quartile. That is, the low utilization types had cars with better gas mileage than the high utilization types. This is consistent with the downward slope in Panel A for 2001. In 2009, when gas prices were higher, the situation reverses, with the distribution of MPG among the highest VMT quartile first-order stochastically dominating the distribution among the lowest quartile. This is consistent with the upward slope in Panel A for 2009. The switch from 2001 to 2009 in these patterns is consistent with our hypothesis, with high utilization types being more likely to have cars with better gas mileage when gas prices are high.

B. OLS

As discussed in the previous section, our base model uses OLS to estimate the difference-in-differences model. Table 2 shows the results of this model under different variants of Equation (6). The first column of the table only includes the main effects of VMT (in thousands of miles) and gasoline prices as well as their interaction. The interaction between VMT and gasoline prices represents the matching effect, that is, the change in correlation between MPG and VMT when gasoline prices increase. Standard errors clustered by state are reported in parentheses. The second column adds MSA dummies to control for time-invariant unobservables between the survey years, such as the availability of mass transit, urban sprawl, and differences in average demands across cities. The third column adds year dummies to control for national time-varying effects. These could include the change in the fleet composition between years, other technological changes, changes in lifestyle and preferences, changes in commuting patterns, and so forth. Columns 4-6 parallel the first three specifications, but with the addition of household-level controls. These controls include number of household members, number of adults in the households, number of workers in the household, the fraction of children (age <18) and women in the household, and household income quartiles. Column 6 represents Equation (6) as it is written and is our preferred specification. The seventh column replaces the city and year dummies with city-month dummies. Finally, the eighth column includes interactions between the demographic variables and gasoline prices, capturing additional heterogeneity in the effect of

gasoline prices on VMT, beyond that already captured in the interaction with VMT. For all these regressions, the matching effect is positive and statistically significant and quite robust across specifications.

The results imply that as gasoline prices increase, households that drive more are driving a more fuel efficient vehicle relative to households that drive less. Specifically, the results of Table 2 indicate that for a $1 increase in gasoline prices, households now choose cars with about 0.02 greater MPG per 1000 miles they drive. Holding constant the mean fleet MPG at 2001 levels, when gasoline prices increase from $1.62 to $3.14 the model predicts an individual driving 4,000 miles (at the 25th percentile of VMT) switches on average from a car getting 25.2 MPG to one getting 25.0, while an individual driving 14,000 miles (at about the 75th percentile) on average switches from a car getting 25.2 MPG to one getting 25.4. In other words, when gas prices increase, fuel-efficient cars are more likely to be allocated to high-VMT types, while inefficient cars are more likely to be allocated to low-VMT types, consistent with the matching effect.

As noted previously, one potential concern with this model is the endogeneity of gasoline prices. Our preferred specification in column 6 includes year and MSA dummies. Thus, any endogeneity of gasoline prices would have to come from city-year shocks in the average demand for (or supply of) vehicle types that drive city-year variation in gasoline prices, after controlling for time-invariant city effects and city-invariant year effects. This seems unlikely. However, in column 7 we include city-month dummies. This removes all variation in gasoline prices, so the mean residual in Equation (6) only relates to the individual's departure from the city-month average residual. Any city-time variation in average MPG that could affect gasoline prices is absorbed in the fixed effect. The results remain robust to this specification, indicating little concern about endogeneity of gasoline prices. Finally, we also consider instrumenting for gasoline prices in additional results presented below.

C. Instrumenting for VMT

A potentially more serious issue with the OLS model is that VMT may be endogenous to MPG. VMT should affect the MPG of the chosen car, but by the same token the MPG of the chosen car also affects VMT, either decreasing it through the comfort effect or increasing it through the GPM effect.

To overcome this simultaneity problem, we use a variety of instrumental variables strategies. Our first strategy is to use the *interactions* of demographics with gasoline prices as instruments for VMT. In particular, consider the following model:

$$MPG_{icmt} = \alpha + \beta Price_{mt}^{gas} + \gamma VMT_{icmt} + \theta\left[Price_{mt}^{gas} \times VMT_{icmt}\right]$$
$$+\mathbf{x}_{imt}{}'\boldsymbol{\delta} + \eta_{m0} + \lambda_{t0} + \varepsilon_{icmt}, \tag{7}$$

$$VMT_{icmt} = a_1 + c_1 Price_{mt}^{gas} + \left[Price_{ct}^{gas} \times \mathbf{x}_{imt}\right]'\boldsymbol{d_1} + \mathbf{x}_{imt}{}'\boldsymbol{b_1} + \eta_{m1} + \lambda_{t1} + \mu_{icmt1} \tag{8}$$

$$Price_{mt}^{gas} \times VMT_{icmt} = \begin{aligned}&a_2 + c_2 Price_{mt}^{gas} + \left[Price_{mt}^{gas} \times \mathbf{x}_{imt}\right]'\boldsymbol{d_2}\\&+\mathbf{x}_{imt}{}'\boldsymbol{b_2} + \eta_{m2} + \lambda_{t2} + \mu_{icmt2}\end{aligned} \tag{9}$$

where, $\mathbf{x}_{imt}$ is a vector of household characteristics. The interactions of these characteristics with gasoline prices serve as instruments. These characteristics are the same as those used as controls in the OLS model.[6]

The identifying assumptions of this strategy deserve some discussion. Note the level effects of $\mathbf{x}_{imt}$ are *not* excluded from the outcome equation, Equation (7). Thus, demographics are allowed to affect the choice of car. The identifying assumption is only that gasoline prices do not affect car choice differently for different demographic types *except* through VMT. Demographics do drive heterogeneity in VMT demand (at a given price), which in turn affects heterogeneous sensitivity of MPG to gasoline prices. All this is consistent with our model of WTP expressed in Equation (4), where individual demographics can affect the taste for car types at the extensive margin *f*( ) as well as the WTP for travel conditional on car type $\pi(i, c, VMT)$, but the first order effect of gasoline prices on car demand only operate through VMT (as seen in the last term of the expression). Our first IV strategy effectively ignores the second order effects related to how heterogeneity in $\partial VMT^*/\partial Price^{Gas}$ (after conditioning on a *level* of VMT^*) feeds back on the optimal choice of MPG. This restriction is embedded in a number of other models of car demand as well, including Feng, Fullerton, and Gan (2013), Gillingham (2012), and Goldberg

---

[6] Thus, the interaction of household income quartile, number of household members, number of adults, number of workers, and fraction of children and women in the household with gasoline prices serve as instruments for VMT and [VMT × Pricegas]. These variables were selected based on the household determinants of VMT established by the previous literature (Bento et al. 2009, Small and Van Dender 2007, Li et al. 2014).

(1998).

Table 3 shows the results for Equation (7) of the IV model. Besides the endogenous VMT terms, the only controls in the first column of Table 3 are gasoline prices, predicted VMT and demographics. MSA and year dummies are added successively to columns 2 and 3. Finally, MSA-month dummies are included in column 4, which eliminates variation in gasoline prices in levels. (The specifications follow the pattern in Table 2 but always include demographics.) The matching effect is positive and statistically significant in all but the first column (with no controls). The effect in Table 3 ranges from 0.013 to 0.022. Overall, the estimated matching effect is similar to the OLS estimates, especially when minimal controls are included, indicating little concern about the potential endogeneity of VMT and MPG after conditioning on these effects.

As noted previously, the above exclusion restrictions are motivated by our theoretical model, in which the heterogeneous effects of gasoline prices on MPG work only through heterogeneity in utilization. Nevertheless, it might be desirable to include interactions of demographics with gasoline prices in the outcome equation in other ways, beyond the utilization channel, for additional flexibility. For example, perhaps different demographic groups form different expectations about future gasoline prices or have differences in how salient gasoline prices are to them. To accommodate this possibility, we consider two very different alternative strategies. One is to add additional economic structure, with nonlinearities in the model that we can leverage for identification. The other strategy is simply to loosen the causal interpretation of VMT in the model and look at heterogeneous effects of prices as they are correlated with VMT.

Consider first the strategy to impose additional economic structure, and to leverage non-linearities to identify heterogeneous effects of gasoline prices on MPG choice. If VMT is a non-linear function of household characteristics, we can include $Price_{mt}^{gas} \times \mathbf{x_{ict}}$ in Equation (7). We use a functional form consistent with the structural model of Bento et al. (2009), in which VMT and MPG are simultaneously determined and in which such nonlinearities arise through the economics of the model. Bento et al. (2009) use the following indirect utility function:

$$V_{ic}' = -\frac{1}{\lambda_i} \exp\left(-\lambda_i \left(\frac{Y_i/T_i - r_{ic}}{p_{imt}^x}\right)\right) - \frac{1}{\beta_{ic}} \exp\left(\alpha_{ic} + \beta_{ic} \frac{p_{ic}^M}{p_{imt}^x}\right) + \tau_{ic}, \tag{10}$$

where $\alpha_{ic} = \widetilde{\boldsymbol{\alpha}}_i' \mathbf{z}_{ic}^{\alpha}$; $\beta_{ic} = -\exp\left(\widetilde{\boldsymbol{\beta}}_i' \mathbf{z}_{ic}^{\beta}\right)$; $\lambda_i = \exp(\widetilde{\boldsymbol{\lambda}}_i' \mathbf{z}_{ic}^{\lambda})$; $\tau_{ic} = \widetilde{\boldsymbol{\tau}}_i' \mathbf{z}_{ic}^{\tau}$; r_{ic} is vehicle rental price; p_{ic}^M is vehicle utilization price; Y_i is the household income; T_i is a fixed number of choice occasions set equal to the number of adults in the household plus one; and p_{imt}^x is the Hicksian composite commodity price.[7] $\mathbf{z}_{ic}^{\alpha}$, $\mathbf{z}_{ic}^{\beta}$, $\mathbf{z}_{ic}^{\tau}$ are vectors of automobile characteristics (age, make, class dummies) interacted with household demographics and $\mathbf{z}_i^{\lambda}$ are household characteristics.[8] Using Roy's identity, VMT demand for household i conditional on car c is:

$$VMT_{ic} = \exp\left(\alpha_{ic} + \beta_{ic} \frac{p_{ic}^M}{p_{imt}^x} + \lambda_i \left(\frac{Y_i/T_i - r_{ic}}{p_{imt}^x} \right) \right) \tag{11}$$

One can predict $\widehat{VMT}_{ic(i)}$ by using the above model, but $\widehat{VMT}_{ic(i)}$ still would be endogenous to MPG through the vehicle characteristics $\mathbf{z}$, which are simultaneously chosen. However, we instead can predict VMT of household i at a fixed set of characteristics. In particular, for each household i, we predict VMT for each car type (i.e. make-model-year) c. We then take a weighted average of those VMTs, using the proportion of type c in the sample as weights. Using this $\widehat{VMT}_{i\bar{c}mt}$ as an instrument, we then estimate the following model:

$$MPG_{icmt} = \alpha + \beta Price_{mt}^{gas} + \mathbf{x}_{imt}\boldsymbol{\delta} + \left[Price_{mt}^{gas} \times \mathbf{x}_{imt} \right]\boldsymbol{\rho} + \pi \widehat{VMT}_{icmt} \tag{12}$$
$$+ \theta \left[Price_{mt}^{gas} \times \widehat{VMT}_{icmt} \right] + \eta_m + \lambda_t + \varepsilon_{icmt}$$

$$VMT_{icmt} = a + b\widehat{VMT}_{i\bar{c}mt} + cPrice_{mt}^{gas} + \left[Price_{mt}^{gas} \times \widehat{VMT}_{i\bar{c}mt} \right]d + \eta_m + \lambda_t$$
$$+ \mu_{icmt} \tag{13}$$

In Equation (13), VMT is a function of gasoline prices and $\widehat{VMT}_{i\bar{c}mt}$, the latter being computed

---

[7] If the number of adults is less than the number of household vehicles, T_i equals the number of household vehicles. p_{imt}^x is computed using the ACCRA cost-of-living index for each year, with adjustments over time using the CPI.

[8] Bento et al. (2009) categorized households into twelve different strata based on employment status, average household age and the number and age distribution of children in the household. Vehicles are categorized by vehicle class. For more information, see Bento et al. (2009).

from the model estimated by Bento et al. (2009) averaged over the automobile fleet. This varia-ble is a highly non-linear function of demographics, including household members, household income quartile, number of children and women in the household, household age distribution, number of adults in the households, and vehicle class, vehicle age and driving cost per mile of the average vehicle. In Equation (12), we then estimate the main equation of interest using this constructed variable as an instrument for VMT. The matching effect is again given by θ: the co-efficient associated with the interaction term between gasoline prices and $\widehat{VMT}_{i\bar{c}mt}$. The model given by Equations (12) and (13) overcomes the issues associated with the previous two models. An advantage of this particular model over the model given by Equations (7)-(9) is that now we can include $Price_{mt}^{gas} \times \mathbf{x}_{imt}$ in the second-stage. These will capture heterogeneous effects of a change in gasoline prices on MPG.

Table 4 shows the results of the second stage of the IV model given by Equation (12)-(13). In the first column, only demographic controls are included. In the second and third col-umns, city and time dummies are successively added. In the fourth column, city-month dummies replace the city and year dummies. As with the previous model using linear instruments, the matching effect is positive and significant in all but the first model (with no controls). The point estimates are slightly bigger than the previous models, but of the same magnitude. Overall then, these results are consistent with our previous estimates. In the last column, the interactions of demographic controls with gasoline prices, one of the chief motivations for using these instru-ments, are added. Unfortunately, the standard errors are much higher using this approach. Alt-hough the estimated point effect is much larger than in the previous models, with the large stand-ard errors for this model we cannot reject the earlier point estimates. It may be too great a challenge to separately identify heterogeneity across utilization types in the effect of gasoline prices on MPG from other such heterogeneity.

As an alternative to this additional structure, we also consider a simple reduced form ap-proach that measures the correlation between heterogeneity in the effect of gasoline prices on MPG and heterogeneity in VMT. Consider the following two equations:

$$MPG_{icmt} = \alpha + \mathbf{x}_{imt}{}'\boldsymbol{\beta} + \delta Price_{mt}^{gas} + \left[Price_{mt}^{gas} \times \mathbf{x}_{imt}\right]'\boldsymbol{\theta} + \eta_{m1} + \lambda_{t1} + \varepsilon_{icmt} \qquad (14)$$

$$VMT_{icmt} = a + \mathbf{x}_{imt}{}'\mathbf{b} + cPrice_{mt}^{gas} + \left[Price_{mt}^{gas} \times \mathbf{x}_{imt}\right]'\mathbf{d} + \eta_{m2} + \lambda_{t2} + \mu_{icmt}. \tag{15}$$

Because the effects of VMT and **x** cannot be identified separately, Equations (14) and (15) represent reduced form regressions in which $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ (and $\boldsymbol{b}$ and $\boldsymbol{d}$) capture direct effects of demographics and effects through VMT (which is a linear function of demographics). MPG and VMT are simultaneously determined, so $E[\varepsilon\mu] \neq 0$, but because VMT does not appear in the reduced form equation for MPG, the equations can be estimated with OLS. Then, the estimated heterogeneous effects of price on MPG can be correlated with VMT, to see whether it is the high- or low-VMT types that are more responsive to gasoline prices. In other words, we give up on identifying the heterogeneous effects of gasoline prices on MPG across utilization types per se, but still identify the heterogeneity across demographic groups, and we can see how those demographic groups are correlated with utilization.

To accomplish this one could take the following steps:

1. Estimate Equation (14) by OLS and let $\hat{T} = \partial\widehat{MPG}/\partial Price^{gas} = \hat{\delta} + \mathbf{x}'\hat{\boldsymbol{\theta}}$ be the estimated treatment effect of a change in gasoline prices on MPG for a household with characteristics **x**.

2. Estimate Equation (15) by OLS and let $\widehat{VMT} = \hat{a} + \mathbf{x}'\widehat{\boldsymbol{b}} + \hat{c}\overline{Price^{gas}} + [\overline{Price^{gas}} \times \mathbf{x}]'\widehat{\boldsymbol{d}}$, where $\overline{Price^{gas}}$ is the average gasoline price in the data. These are predicted household VMT at the average gasoline price. Average gasoline prices are used to isolate shifts in the demand curves of different utilization types, evaluated at the same price.

3. To estimate the correlation between $\partial\widehat{MPG}/\partial Price^{gas}$ and $\widehat{VMT}$, run the following regression:

$$\hat{T} = \pi_0 + \pi_1\widehat{VMT} + \sigma, \tag{16}$$

where the coefficient of interest is π_1.

As we show in Appendix B, this whole process is equivalent to estimating the following model:

$$MPG_{icmt} = \alpha + \beta Price_{mt}^{gas} + \mathbf{x}_{imt}'\boldsymbol{\gamma} + \theta\left[Price_{mt}^{gas} \times \widehat{VMT}_{icmt}\right] + \eta_m + \lambda_t + \varepsilon_{icmt}, \qquad (17)$$

where $\widehat{VMT}$ is the predicted VMT from Equation (15), but at average gasoline prices. The matching effect in Equation (17) is given by θ. However, this strategy merely provides a correlation between the predicted VMT and the effect of $Price^{gas}$.

The results for Equation (17) with corrected standard errors are shown in Table 5. The first column omits any controls except demographics. The second and third columns add MSA dummies and year dummies respectively. The fourth column includes MSA-month dummies. The estimated effects continue to be similar to the OLS and IV models, with a statistically significant matching effect (again in all but the first model) of roughly 0.015 to 0.02. Thus, even if we cannot say that the causal effect is through utilization *per se*, we can say that the types of people who do drive more are allocated more fuel-efficient cars when gasoline prices are higher, relative to those who drive less. This reduced form correlation is enough to trigger the matching effect and to provide additional savings in fuel consumption when gasoline prices rise.

Figure 5 summarizes the results of these four models, using our preferred specification with year and city dummies and demographic controls. It shows the changing in the matching relationship predicted from various changes in gasoline prices ($0.25, $0.55, $1, and $3.50).

D. Robustness Checks

The above models all use miles per gallon as the dependent variable. Table 6 replicates the main results with gallons per mile as the dependent variable instead (in thousands). Thus, the coefficient estimates indicate the change in gasoline consumption (in thousands of gallons) per mile when there is a small change in the explanatory variables. The matching effect is negative for all the models, implying that when gasoline prices increase, a household that drives more drives a vehicle that consumes less gasoline, relative to a household than drives less. The bottom row of the table converts these estimates to an equivalent effect on MPG at the mean of the data. The results are generally similar to the corresponding estimates from the main tables, indicating little sensitivity to this change of specification. The matching effect corresponds to about 0.01 to 0.05 MPG per dollar gasoline price per 1000 miles VMT.

As noted previously, we considered a variety of alternative models as well. In particular, we considered alternative measures of VMT as well as alternative ways of trimming the variable

(see notes 2 and 3 above). We also considered nominal gasoline prices instead of inflation-adjusted prices. For our IV models, we considered GMM estimation as well as two-stage least squares. In all models, we computed standard White heteroskedastic-robust standard errors as well as clustered standard errors. Our basic findings are unchanged with all these models, which are available upon request.

E. Within- and Between-Household Matching

Thus far, we have estimated the matching effect for all vehicles in the household. As noted previously, households with N vehicles are treated as if they have N exogenous bundles of trips which they then allocate to vehicles. Continuing with this assumption, we can focus just on within-household matching effects, in which we look at how such bundles are allocated to vehicles within the household under different gasoline prices. Figure 6 shows, on the vertical axis, the difference in MPG of the car allocated to the highest VMT bundle in the household relative to the lowest, plotted against the respective differences in VMT. Panel A does this for two-car households, Panel B for three-car households. (Notice some differences in VMT are quite large, but the left-hand side of the figure holds most of the density.) The figure shows that, in 2008-9, when gas prices averaged $3.14, the difference is positive, meaning that the higher-MPG car was allocated to more VMT. Moreover, it is increasing in VMT especially on the left side which holds most of the data, indicating that as the split in miles between the two bundles of trips increases, the difference in MPG grows. In 2001-2, when gas prices averaged $1.62, the difference was negative for two-car families and decreasing. For three-car households in 2001-2, the difference is close to zero with little trend. In either case, the difference is higher in 2008-9, when gas prices average $3.14, than 2001-2. This suggests households match their high-MPG car to more VMT when gas prices are higher.

To further explore these within-household effects, we repeat the OLS regression in Table 2, column 2 on multi-car households, but also include household fixed effects (which supersede MSA dummies and household demographics). We find a matching effect of 0.0350, statistically significant at the 1% level. This effect is somewhat bigger than the effect using all the data, which is not surprising given that it would be easier to re-match vehicles to trips within a household than between households. We view these results as suggestive, but caution that they

depend on our simplifying assumption that we can take baskets of trips as exogenously determined prior to being matched to vehicles.

We similarly can focus on between-household effects by reducing the data set to one observation per household. This between-household analysis is interesting in itself, because it focuses on the transactions that take place as households sell old cars and buy new ones and the reallocation that emerges. It has the further virtue of eliminating the necessity of our assumption about the exogenous division of trips within the household. We conduct such a between-household analysis three ways. Our preferred approach is to take the average MPG of all cars in the household, weighted by each car's VMT, to represent a household-level average MPG. Likewise, we take the total VMT of the household divided by the number of adult, to obtain the household's per-capita VMT. This approach removes within-household variation while giving some weight to all its decisions and makes no assumptions about how total VMT is allocated to cars. The results are displayed in Panel A of Table 7 (which for brevity shows only the coefficients from the matching effect). The columns correspond to our preferred specification (with year and MSA dummies and demographic characteristics) for the OLS, linear IV, IV with the non-linear instrument, and non-causal models respectively. The results remain similar to the previous findings, but slightly smaller, which is consistent with the idea that the transaction costs of re-matching to automobiles is more difficult between households than within them.

As an alternative to taking the household-level average, in Panel B we take only the car with the highest VMT in the household ("top car"). The results are similar to Panel A. This second approach has the disadvantage of taking the VMT assigned by multi-car households to the top car as exogenous. Finally, in Panel C, we take only one-car households.[9] While this approach has the advantage of sidestepping the question of how VMT is allocated across cars, it has the disadvantage of created a more selected and less representative sample (as well as dropping over half the data). For example, the average one-car household has 1.8 people and drives 8,964 miles per year, whereas the average household in the data has 2.4 people and drives its top car 13,027 miles per year. Nevertheless, the estimates using this approach are similar to the

---

[9] We also include households in which the second car is effectively unused, with less than one thousand miles.

other two approaches, except for the non-linear instruments, which were ineffective in this sub-sample (with the instruments failing to pass most tests of relevance).

F. Gasoline Taxes

Our analysis to this point has used variation in the overall price of gasoline. However, energy policies typically influence gasoline prices through gasoline taxes. Moreover, households might respond differently to gasoline taxes than other components of the price. For example, they might view them as more permanent than variation in oil prices or refinery costs or more salient. Speaking to this point, Li, Linn, and Muehlegger (2014) recently found that gasoline price elasticities are larger for changes in the gasoline tax than from other sources of variation in prices. They also found that the MPG of the average car responded more to taxes. This raises the possibility that the matching effect might also be stronger from changes in gasoline taxes than other price changes.

To follow up on this issue, we gathered data on gasoline prices by state from the American Petroleum Institute's (API's) *State Motor Fuels Tax Reports*. These data are at the monthly level for 2008-9 (from which we constructed quarterly averages to match our quarterly gasoline price data) and the annual level for 2001-2. Weighting by our data, the average gasoline tax increased 6 cents between survey waves. The variability ranged from increases of more than 10 cents in Texas, Oregon, and Maine to small decreases in Virginia and California.

Following Li, Linn, and Muehlegger (2014), we then reran our OLS models using both variation in gasoline taxes and net-of-tax price variation. This approach allows us to separately identify the two effects. As an alternative, we continue to employ our basic model with gasoline prices in the equation of interest, but instrument for gasoline prices with gasoline taxes. This approach obviously limits the source of variation in the data to the taxes.[10]

Panel A of Table 8 shows the results of the first approach, in which we separately estimate the effects of gasoline taxes and net-of-tax gasoline prices, showing only the respective matching effects. The estimates suggest the effect of the gasoline tax is three times the effect of

---

[10] We also considered instrumenting for VMT using the above strategies, however, the presence of multiple nonlinear endogenous variables (for the first strategy, VMT interacted with taxes and VMT interacted with net-of-tax prices; for the second, VMT interacted with gasoline prices with both treated as endogenous) made it impossible to identify all the structural effects without additional instruments.

prices generally, with a statistically significant matching effect of about 0.06 (versus about 0.02 in our previous models). Net-of-tax prices have a slightly smaller effect than the over-all price effect, of about 0.017. (Note because they are a much larger share of overall prices they need not move as much.) Panel B shows the results in which we instrument for gasoline prices using gasoline taxes. The point estimates using this approach range from 0.03 to 0.04, still higher than our earlier estimates of about 0.02 using raw gasoline prices. These findings are consistent with the notion that gas taxes are more salient than other shocks and with previous results from Li, Linn, and Muehlegger (2014). They also suggest that gasoline taxes can have a more potent effect than our earlier estimates using only gas prices would suggest.

V. Implications

In the introduction, we noted that an increase in gasoline prices should have three effects: a utilization effect on VMT, a compositional effect on the average fuel economy of the fleet, and a matching effect. Our model does not estimate the utilization effect, but does capture the combined composition and matching effects. It also provides a way to decompose the two effects. Using our model, the change in household fuel economy due to a change in gasoline prices is given by $\partial MPG_j / \partial P^{gas} = \hat{\beta} + \hat{\theta} VMT_j$, where $\hat{\beta}$ and $\hat{\theta}$ are the estimated parameters on Price and Price*VMT respectively. The proportion of the change in gasoline consumption through the matching channel can be measured by Equation (18):

$$
\begin{array}{c}
\text{Proportion of predicted total change} \\
\text{from matching effect}
\end{array}
= 1 - \left(\frac{\sum_j \left(\frac{VMT_j}{\overline{MPG_j} + \hat{\beta} + \hat{\theta} \cdot \overline{VMT}} - \frac{VMT_j}{\overline{MPG_j}} \right)}{\sum_j \left(\frac{VMT_j}{\overline{MPG_j} + \hat{\beta} + \hat{\theta} \cdot VMT_j} - \frac{VMT_j}{\overline{MPG_j}} \right)} \right), \quad (18)
$$

where $\frac{VMT_j}{MPG_j}$ measures the gasoline consumption of vehicle j and $\overline{VMT}$ is the average VMT.

Using our preferred specification with MSA and year dummies and demographic controls, we predict that a $1 increase in gasoline taxes would result in a savings of 3.5%, or 4.4 billion gallons a year (column 4, Table 8, Panel A). This represents a short-run elasticity of 0.03 through the composition and matching effects combined. This elasticity is somewhat smaller than the short-run estimate of 0.06 in Li, Linn, and Muehlegger (2014) and the medium-run (two year) estimate of 0.09 in Gillingham (2012). Using the above decomposition, we estimate that

roughly half (49%) of this affect is through the matching channel. Thus, we predict that a $1 increase in gasoline taxes would result in a savings of 2.7% of US gasoline consumption in the short run from the matching affect alone, or 2.1 billion gallons annually.[11] These gasoline savings correspond to 92 million metric tonnes of CO_2 annually. Using the US social cost of carbon of $36/tonne and Parry and Small's (2005) estimate of $0.16/gallon of damages from conventional pollution, these gasoline savings are equivalent to $3.7 billion in benefits from avoided pollution per year, just through the matching channel—i.e., with no change in the automotive fleet and no change in total miles driven. Moreover, these short-run effects are likely to be lower bounds on the long-run effects. Using the ratio of Gillingham's (2012) elasticity of 0.09 to our estimate of 0.03 as a rough gauge, and assuming the proportion of the matching-plus-composition effects due just to matching stays constant at about one-half, these savings would be scaled up by a factor of 3x over time.

VI. Conclusion

Combustion of gasoline has detrimental environmental and health impacts. To reduce these effects, one policy tool is to increase gasoline prices. Two effects of gasoline prices have been identified in the literature, the utilization effect and the compositional effect. The former implies that an increase in gasoline prices results in a decrease in the demand for VMT. The latter suggests that it causes an evolution of fleet towards fuel efficiency. As we have shown in this paper, there is a potential third effect of gasoline prices that has been overlooked in the previous studies. Due to heterogeneity in demand for VMT, increase in gasoline prices should particularly affect vehicle choices of households that drive more. In the presence of higher gasoline prices, the costs borne by high-VMT households is greater and hence they become more to outbid low-VMT households for a fuel efficient vehicle. In equilibrium, there will be matching of fuel efficiency among high-VMT households and fuel inefficiency among low-VMT households. This is the first study that analyzes such assortative behavior.

To understand the potential importance of the matching effect, if fuel economy was assigned perfectly to households based on their VMT demand, US gasoline consumption would be

---

[11] As the models are linear in prices, the effects of smaller changes in taxes can be computed as any proportion of $1. Similarly, if one prefers to use the estimates from prices in Tables 2-4 rather than taxes, one could proportionately adjust these effects by about one-fourth to one-half.

15% lower, without changing the fleet or VMT. Our study exploits variation in gasoline prices in the last decade to estimate how gasoline prices affect vehicle choice. Using our model based on taxes and our preferred specification, if gasoline taxes were $1 higher, gasoline consumption would have been about 2.7% lower due to the matching effect alone. These additional benefits of higher prices further favor price-based policies.

REFERENCES

Allcott, Hunt, and Nathan Wozny. 2016. "Gasoline Prices, Fuel Economy, and the Energy Paradox." *Review of Economics and Statistics* 96(5):779-95.

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: 383-403.

Anderson, Soren T., Ryan Kellogg, James M. Sallee, and Richard T. Curtin. 2011. "Forecasting Gasoline Prices Using Consumer Surveys." *American Economic Review* 101(3):110–14.

Austin, David, and Terry Dinan. 2005. "Clearing the air: The Costs and Consequences of Higher CAFE Standards and Increased Gasoline Taxes." *Journal of Environmental Economics and management* 50(3): 562-82.

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): 667-99.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63(4): 841-890.

Brown, Jennifer, Justine Hastings, Erin T. Mansur, and Sofia B. Villas-Boas. 2008. "Reformulating Competition? Gasoline Content Regulation and Wholesale Gasoline Prices." *Journal of Environmental Economics and Management* 55(1): 1-19.

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): 220-56.

Currie, Janet, Joshua G. Zivin, Jaime Mullins and Matthew Neidell. 2014. "What Do We Know About Short and Long Term Effects of Early Life Exposure to Pollution?" *Annual Review of Resource Economics* 6: 217-247.

Davis, Michael C., and James D. Hamilton. 2004. "Why Are Prices Sticky? The Dynamics of Wholesale Gasoline Prices." *Journal of Money, Credit, and Banking* 36: 17-37.

Davis, Lucas, and Lutz Kilian. 2011. "Estimating the Effect of a Gasoline Tax on Carbon Emissions." *Journal of Applied Econometrics* 26(7): 1187-1214.

Feng, Ye, Don Fullerton, and Li Gan. 2013. "Vehicle Choices, Miles Driven, and Pollution Policies." *Journal of Regulatory Economics* 44(1): 4-29.

Geman, Hélyette. 2007. "Mean Reversion Versus Random Walk in Oil and Natural Gas Prices." In M. Fu, R. Jarrow, J. Yen, and R. Elliott, eds. *Advances in Mathematical Finance*. Birkhauser: Boston, 219-228.

Gillingham, Kenneth. 2012. "Selection on Anticipated Driving and the Consumer Response to

Changing Gasoline Prices." Mimeo.

Gillingham, Kenneth, Matthew J. Kotchen, David S. Rapson and Gernot Wagner. 2013. "Energy Policy: The Rebound Effect is Overplayed." *Nature* 493: 475-476.

Goldberg, Pinelopi Koujianou. 1995. "Product Differentiation and Oligopoly in International Markets: The Case of the US Automobile Industry." *Econometrica* 63(4): 891-951.

Goldberg, Pinelopi Koujianou. 1998. "The Effect of Corporate Average Fuel Efficiency Standards in the US." *Journal of Industrial Economics* 46(1): 1-33.

Graham, Daniel J., and Stephen Glaister. 2002. "The Demand for Automobile Fuel: A Survey of Elasticities." *Journal of Transport Economics and Policy* 36(1): 1-26.

Hughes, J.E., Christopher R. Knittel, and Daniel Sperling. 2008. "Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand." *The Energy Journal* 29(1): 113-134.

Jacobsen, Mark R. 2013. "Evaluating US fuel economy standards in a model with producer and household heterogeneity." *American Economic Journal: Economic Policy* 5(23): 148-87.

Klier, Thomas, and Joshua Linn. 2010. "The Price of Gasoline and the Demand for Fuel Efficiency: Evidence from Monthly New Vehicle Sales Data." *American Economic Journal: Economic Policy* 2(3): 134–53.

Klier, Thomas, and Joshua Linn. 2011. "Corporate Average Fuel Economy Standards and the Market for New Vehicles." *Annual Review of Resource Economics* 3: 445-62.

Li, Shanjun, Christopher Timmins, and Roger H. von Haefen. 2009. "How Do Gasoline Prices Affect Fleet Fuel Economy?" *American Economic Journal: Economic Policy* 1(2): 113-37.

Li, Shanjun, Joshua Linn, and Erich Muehlegger. 2014. "Gasoline Taxes and Consumer Behavior." *American Economic Journal: Economic Policy* 6(4): 302-42.

Linn, Joshua. 2016. "The Rebound Effect for Passenger Vehicles." *Energy Journal* 37(2): 257-88.

Parry, Ian W.H., and Kenneth A. Small. 2005. "Does Britain or the United States Have the Right Gasoline Tax?" *American Economic Review* 95(4): 1276-89.

Puller, Steven L., and Lorna A. Greening. 1999. "Household Adjustment to Gasoline Price Change: An Analysis Using 9 years of US Survey Data." *Energy Economics* 21: 37-52.

Sallee, James M., Sarah E. West, and Wei Fan. 2016. "Do Consumers Recognize the Value of Fuel Economy? Evidence from Used Car Prices and Gasoline Price Fluctuations." *Journal of Public Economics*, forthcoming.

Small, Kenneth A., and Kurt Van Dender. 2007. "Fuel Efficiency and Motor Vehicle Travel:

The Declining Rebound Effect." *Energy Journal* 28(1): 25-52.

United States Department of Energy (US DOE). 2011. Fuel Taxes by Country. http://www.afdc.energy.gov/data/10327. Last accessed 11 July, 2016.

United States Environmental Protection Agency (US EPA). 2015. *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2013*. EPA 430-R-15-004.

West, Jeremy, Mark Hoekstra, Jonathan Meer, and Steven L. Puller. 2013. "The Impact of Increased Fuel Efficiency on Driving: Evidence from a Regression Discontinuity Design." Mimeo.

Figure 1: The Matching Effect
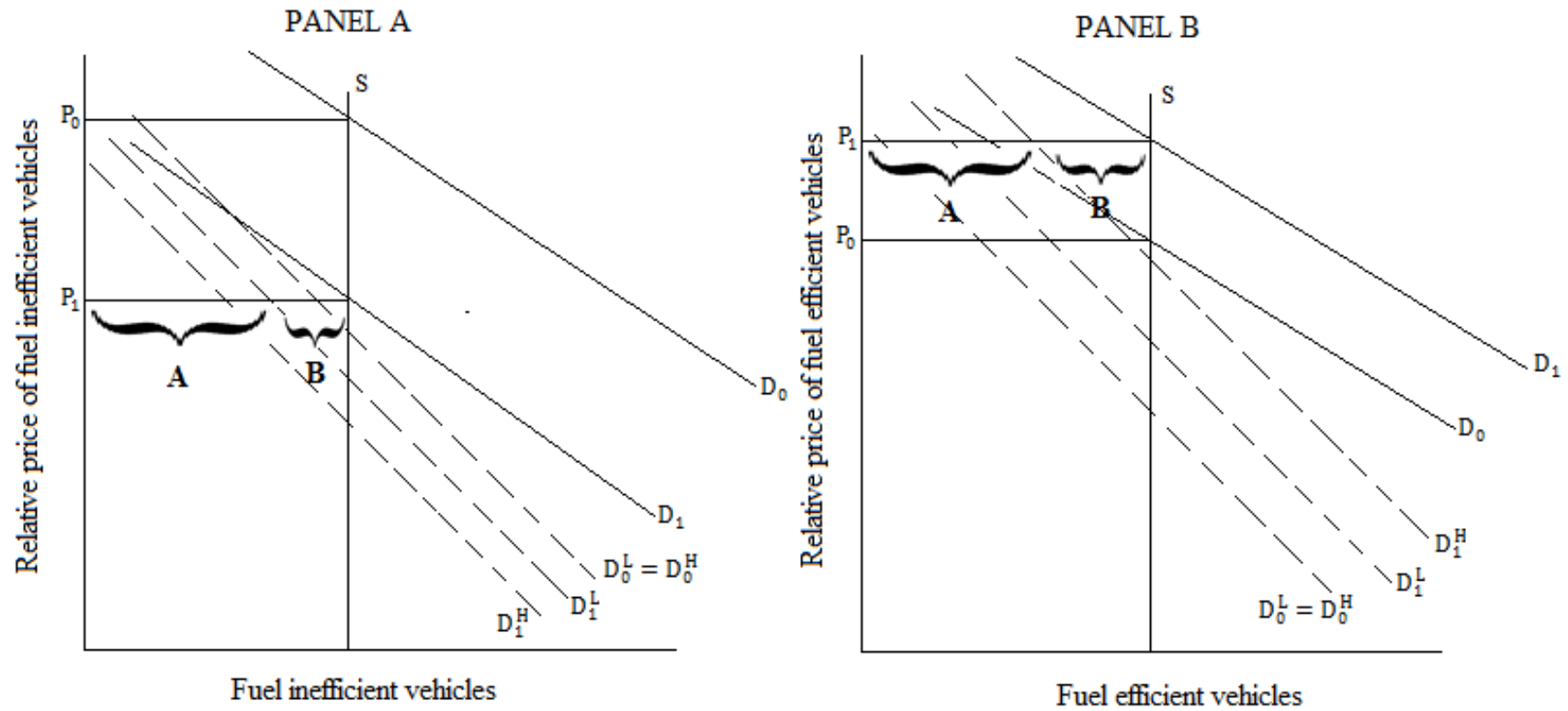
Notes: In these diagrams partial-equilibrium models illustrate the matching effect. There are two types of households: high-VMT and low-VMT designated by H and L. There are two types of vehicles: high-GPM and low-GPM. The demand curves D^H and D^L show the relative willingness to pay (WTP) of the two household types for different fuel economies. In both the panels above, the baseline case is given by a scenario where $D_0^H = D_0^L$ and the baseline market demand curve is given by D_0 for a relative price P_0. After an increase in gasoline prices, the relative WTP for inefficient vehicles decreases. The new market demand curve is shown by D_1 for a relative price P_1. However, due to the heterogeneity in demand among agents for VMT, the decrease (increase) in the relative WTP for the high-VMT types for inefficient (efficient) vehicles is greater than for the low-VMT types.

Figure 2: MSA 2009 Gasoline Prices v. 2001 Gasoline Prices (in 2009 Dollars)

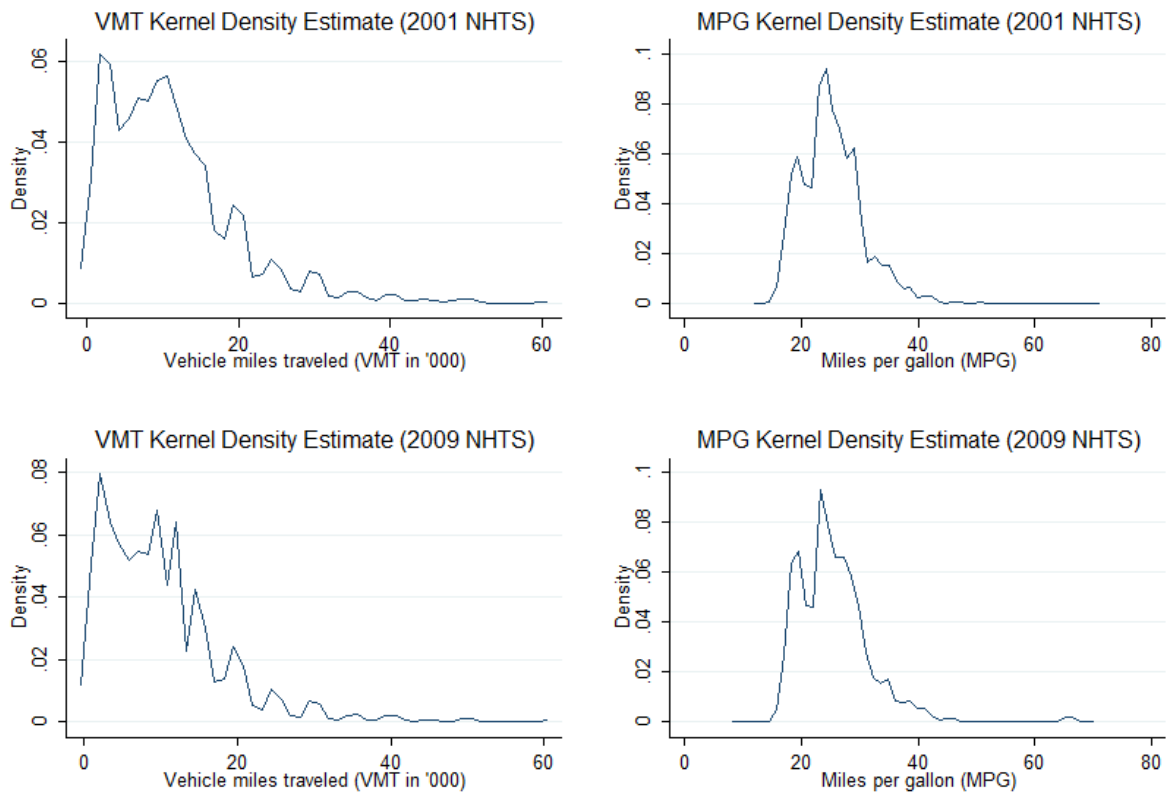**Panel A: Gasoline Prices Kernel Densities**
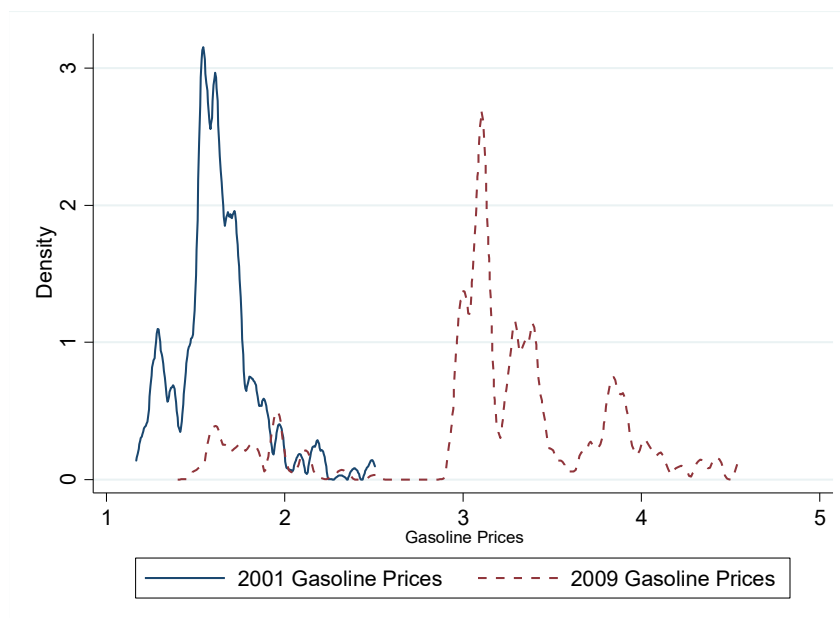
Panel B: 2009 Prices against 2001 Prices
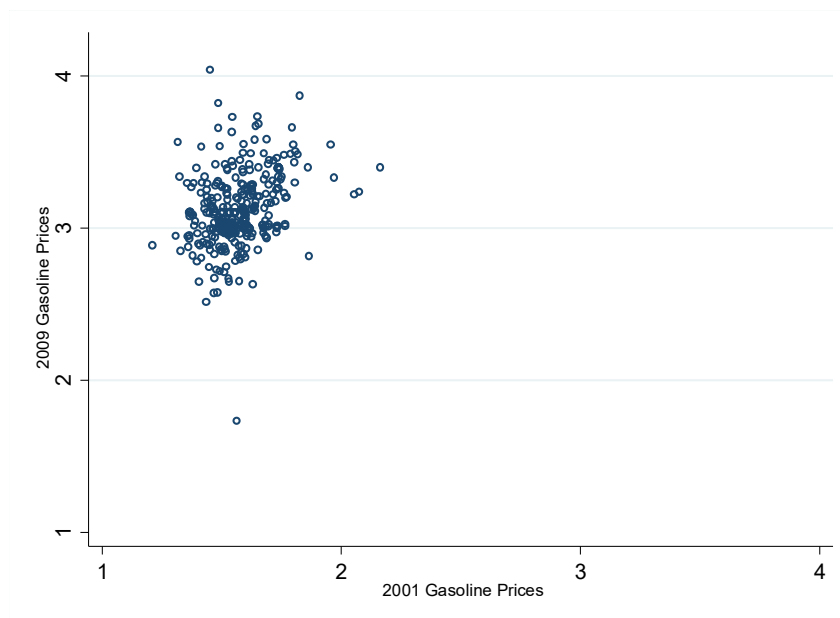
Notes: The graph in Panel A shows the kernel densities of gasoline prices for the 2001 (solid line) and the 2009 (dashed line) ACCRA gasoline prices. Panel B plots 2009 average MSA gasoline prices against 2001 average MSA gasoline prices.

**Figure 3: Kernel densities of VMT and MPG**

Source: 2001 & 2009 NHTS and Department of Energy

Notes: These graphs show the kernel densities of reported vehicle miles traveled (VMT) in thousands and miles per gallon (MPG) per vehicle for the 2001 and the 2009 *National Household Travel Survey* (NHTS).

Figure 4: Is there a matching effect?

Panel A. Matching in Raw Data vs. Perfect Matching, 2001 and 2009
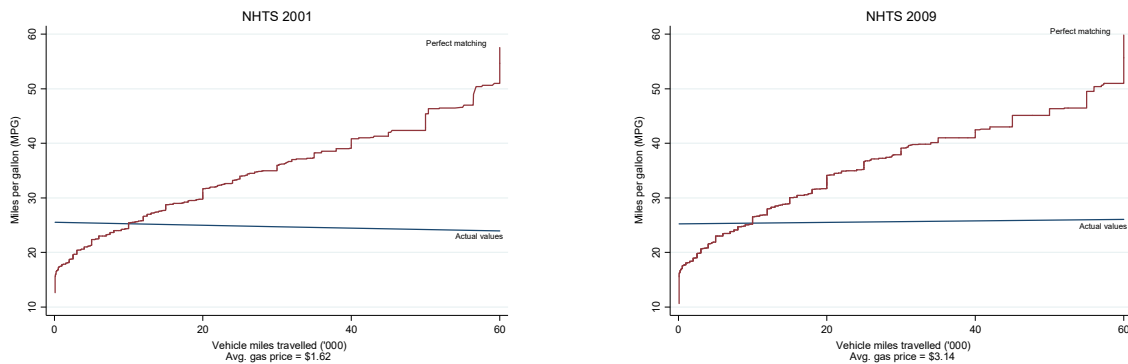
Panel B. CDF of MPG, by Top and Bottom VMT Quartile, 2001 and 2009
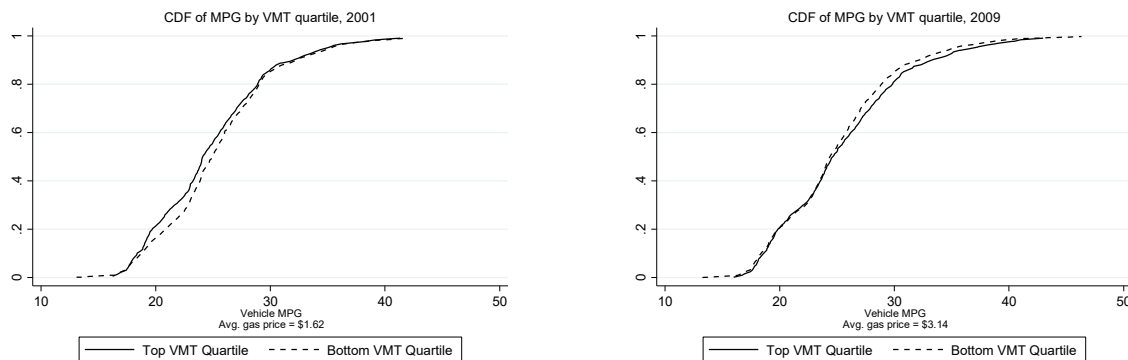
Notes: In Panel A, actual values in both graphs represent a linear regression of fuel economy on vehicle miles traveled (VMT). For perfect matching, vehicles were re-assigned to households based on their respective VMT, so that their ranks are the same. Panel B shows the distribution of fuel economy among those in the highest and lowest quartile of VMT. In 2001 low-VMT households have more fuel-efficient cars than high-VMT households (in the sense of first-order stochastic dominance of the distribution); in 2009, the reverse is true.

Figure 5. MPG-VMT relationship for different tax regimes (Year = 2009)

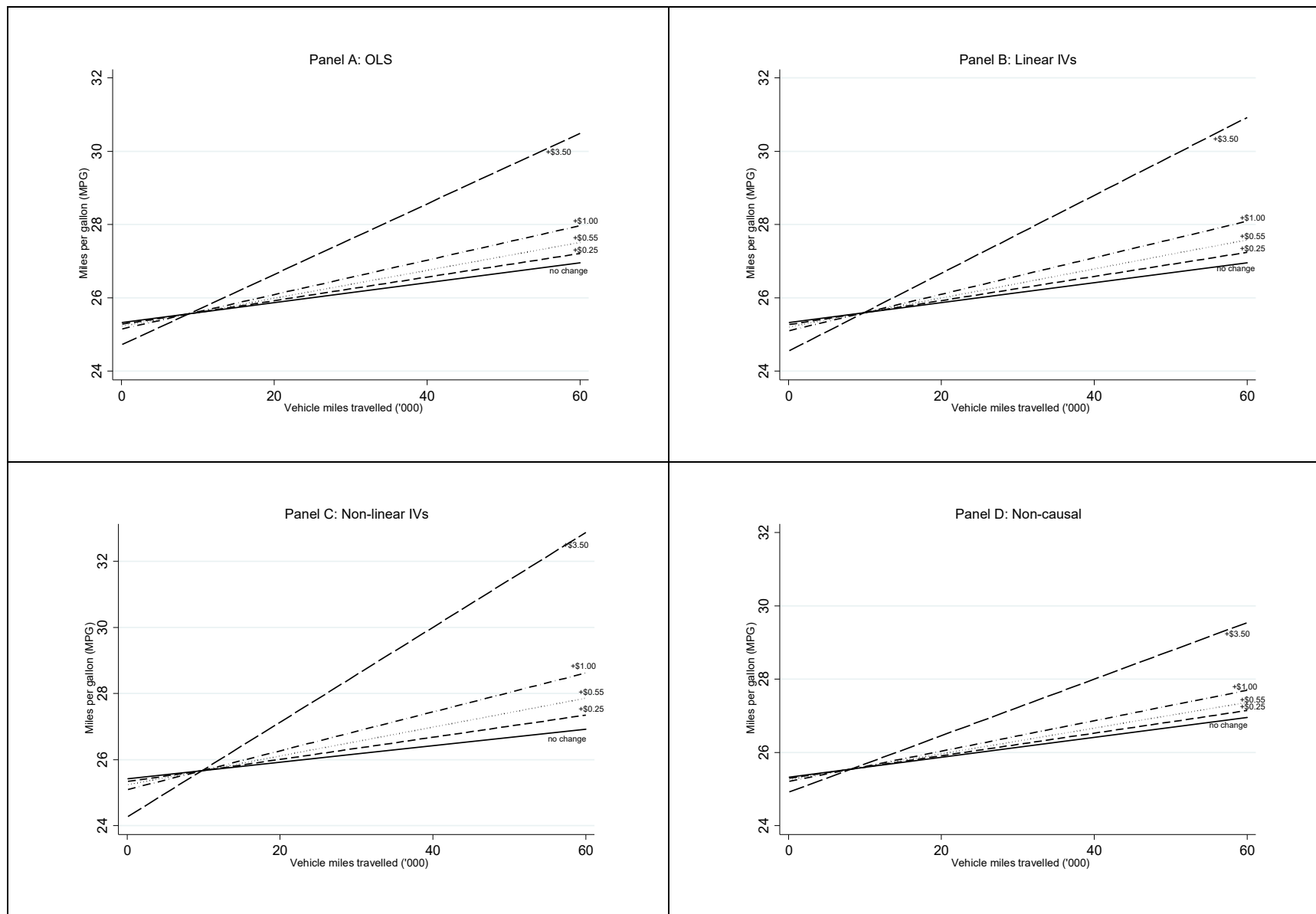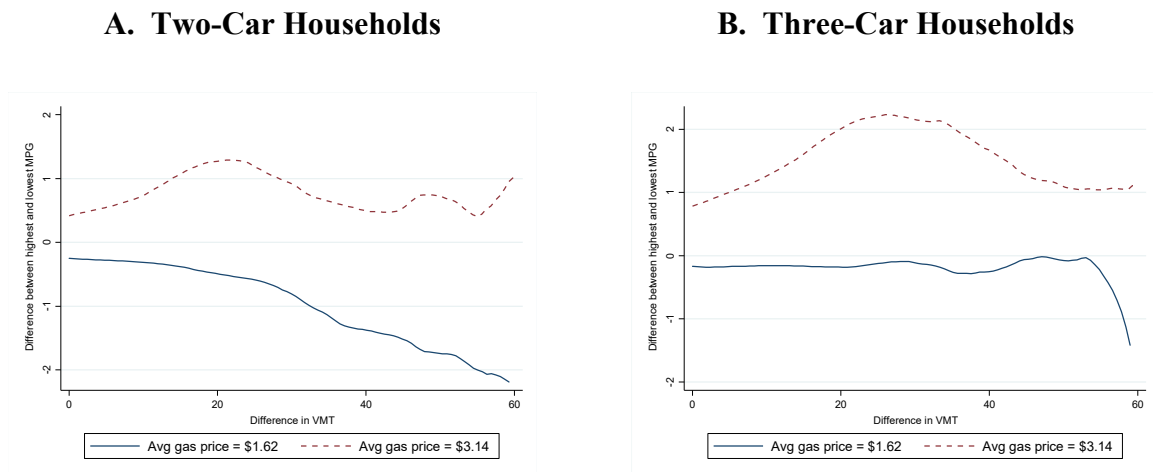**Figure 6: Within-Household Matching**

A. Two-Car Households   B. Three-Car Households

Note: The figures show the difference in MPG of the car allocated to the highest VMT bundle in the household relative to the lowest, plotted against the respective differences in VMT.

Table 1: Summary Statistics: 2001 & 2009 National Household Travel Survey and Matched Gasoline Prices

Variable	Full sample	2001 NHTS	2009 NHTS	Difference
Gasoline price (2009 dollars)	2.902	1.618	3.137	1.520***
	(0.906)	(0.219)	(0.651)	(0.003)
Vehicle miles traveled (VMT) in '000	10.160	11.244	9.974	-1.271***
	(8.259)	(8.881)	(8.132)	(0.045)
Combined unadjusted fuel economy (MPG)	25.545	25.224	25.600	0.377***
	(6.193)	(5.296)	(6.333)	(0.028)
Household size	2.648	2.847	2.614	-0.233***
	(1.280)	(1.370)	(1.260)	(0.007)
Number of workers in household	1.218	1.579	1.156	-0.424***
	(0.970)	(1.011)	(0.949)	(0.005)
Number of adults	2.090	2.112	2.087	-0.026***
	(0.713)	(0.743)	(0.708)	(0.004)
Fraction of children	0.133	0.175	0.126	-0.049***
	(0.212)	(0.231)	(0.208)	(0.001)
Fraction of women	0.520	0.510	0.522	0.013***
	(0.229)	(0.233)	(0.228)	(0.001)
Household income <$25,000	0.136	0.162	0.132	-0.030***
	(0.343)	(0.368)	(0.339)	(0.002)
Household income <$50,000 and ≥ $25,000	0.262	0.323	0.251	-0.072***
	(0.440)	(0.468)	(0.434)	(0.002)
Household income <$75,000 and ≥ $50,000	0.200	0.217	0.197	-0.020***
	(0.400)	(0.412)	(0.397)	(0.002)
Household income ≥ $75,000	0.402	0.298	0.420	0.122***
	(0.490)	(0.458)	(0.494)	(0.002)
White, only	0.879	0.849	0.885	0.035***
	(0.326)	(0.358)	(0.320)	(0.002)
African-American, only	0.049	0.050	0.049	-0.001
	(0.216)	(0.218)	(0.216)	(0.001)
Asian, only	0.020	0.019	0.020	0.002**
	(0.141)	(0.135)	(0.142)	(0.001)
Number of households	161,349	23,739	137,610	

Notes: The main entries in columns 1, 2 and 3 report the mean level of the variables with standard deviations in parentheses. Column 4 reports the differences and standard errors.
***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.

Table 2. Difference-in-Differences Ordinary Least Squares Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Price^{gas}	0.04101 (0.0648)	-0.10211*** (0.0242)	-0.16228*** (0.0231)	0.04038 (0.0622)	-0.10129*** (0.0220)	-0.17382*** (0.0239)	N/A	-0.19544** (0.0955)
VMT	-0.03866*** (0.0074)	-0.03780*** (0.0069)	-0.03548*** (0.0074)	-0.03173*** (0.0074)	-0.02843*** (0.0073)	-0.02564*** (0.0078)	-0.02551*** (0.0090)	-0.02609*** (0.0075)
Matching effect (VMT x Price^{gas})	**0.01954*** (0.0032)**	**0.01972*** (0.0027)**	**0.01907*** (0.0028)**	**0.01991*** (0.0032)**	**0.02053*** (0.0027)**	**0.01972*** (0.0028)**	**0.01998*** (0.0029)**	**0.01991*** (0.0029)**
MSA dummies		X	X		X	X		X
Year dummies			X			X		X
MSA-Month dummies							X	
Demographic controls				X	X	X	X	X
Price^{gas} × Demographic controls								X
N =	307,977	307,977	307,977	287,463	287,463	287,463	287,463	287,463
R^2	0.002	0.024	0.024	0.014	0.038	0.038	0.062	0.038

Notes: The estimates are associated with the OLS model represented by Equation (6). The first column shows the results of the basic model with no controls. MSA dummies are added in column 2 and year dummies are added in column 3. Columns 4-6 repeat the first three columns, but add demographic controls. Column 6 represents the model in Equation (6) and is our preferred specification. Column 7 replaces the MSA and year dummies with MSA-month dummies, which also removes variation in gas prices. Column 8 includes interactions between demographic controls and gasoline prices. Demographic controls are number of household members, number of adults in the households, number of workers in the household, the fraction of children (age <18) and women in the household and household income quartile. Standard errors clustered at the state level are in parentheses.
***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.

Table 3. Difference-in-Differences with Linear Instruments — 2nd Stage Results

	(1)	(2)	(3)	(4)
Price^{gas}	0.06378 (0.1066)	-0.14567* (0.0786)	-0.21433** (0.0908)	N/A
VMT	-0.21694 (0.1580)	-0.27532* (0.1615)	-0.43277*** (0.1345)	-0.42052*** (0.1359)
Matching effect (VMT x Price^{gas})	**0.01257 (0.0090)**	**0.02030*** (0.0074)**	**0.02182*** (0.0082)**	**0.01625** (0.0076)**
MSA dummies		X	X	
Year dummies			X	
MSA-Month dummies				X
Demographic controls	X	X	X	X
Price^{gas} × Demographic controls	Instr.	Instr.	Instr.	Instr.
N =	287,463	287,463	287,463	286,872
F-statistics of excluded instruments				
VMT	11.9	14.1	6.5	9.4
VMT × Price^{gas}	427.0	252.0	225.7	237.2

Notes: The estimates are associated with the coefficients estimates from the IV model represented by Equations (7) – (9), using demographics interacted with gas prices as excluded instruments. Column 1 omits MSA and year dummies, which are successively added in columns 2 and 3. Column 4 replaces the MSA and year dummies with MSA-month dummies, which also removes variation in gas prices. Demographic controls are number of household members, number of adults in the households, number of workers in the household, the fraction of children (age <18) and women in the household and household income quartile. Standard errors clustered at the state level are reported in parentheses.
***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.

Table 4. Difference-in-Differences with Non-Linear Instruments — 2nd Stage Results

	(1)	(2)	(3)	(4)	(5)
Price^{gas}	0.41909** (0.1845)	-0.20999 (0.1352)	-0.32972** (0.1451)	N/A	-0.71365 (0.5910)
$\widehat{VMT}_{i\bar{c}mt}$	0.30908*** (0.1171)	0.03266 (0.0816)	0.00542 (0.0717)	-0.03144 (0.0730)	-0.11501 (0.1905)
Matching effect (VMT x Price^{gas})	**-0.01013 (0.0138)**	**0.03070** (0.0130)**	**0.03383** (0.0135)**	**0.03835** (0.0152)**	**0.07954 (0.0881)**
MSA dummies		X	X		X
Year dummies			X		X
MSA-Month dummies				X	
Demographic controls	X	X	X	X	X
Price^{gas} × Demographic controls					X
N =	251,960	251,960	251,960	251,960	251,960
F-statistics of excluded instruments					
$\widehat{VMT}_{i\bar{c}mt}$	17.1	16.3	18.7	15.2	18.1
$\widehat{VMT}_{i\bar{c}mt}$ × Price^{gas}	109.8	90.2	83.5	101.3	17.7

Notes: The estimates reported are coefficients estimates for the second stage of the model represented by Equation (13). $\widehat{VMT}_{i\bar{c}mt}$ as predicted from the structural model of Bento et al. (2009) serves as the instrument for VMT. Column 1 omits MSA and year dummies, which are successively added in columns 2 and 3. Column 4 replaces the MSA and year dummies with MSA-month dummies, which also removes variation in gas prices. Column 5 includes demographics interacted with price in the structural equation. Demographic controls are number of household members, number of adults in the households, number of workers in the household, the fraction of children (age <18) and women in the household and household income quartile. Standard errors clustered at the state level are reported in parentheses.
***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.

Table 5. Non-causal estimation of the matching effect

	(1)	(2)	(3)	(4)
Price^{gas}	0.12163 (0.1344)	-0.06784 (0.0730)	-0.11667 (0.0732)	N/A
VMT		*Not identified*		
Matching effect ($\widehat{\text{VMT}}$ x Price^{gas})	**0.01172 (0.0095)**	**0.01705** (0.0069)**	**0.01398* (0.0073)**	**0.01517** (0.0075)**
MSA FE		X	X	
Year FE			X	
MSA-Month FE				X
Demographic controls	X	X	X	X
N =	287,463	287,463	287,463	287,463
R^2	0.013	0.035	0.036	0.060

Notes: The estimates are associated with coefficients estimates of the model represented by Equation (17). Predicted VMT is a function of average gasoline prices, demographic characteristics, interactions between these characteristics and average gasoline prices, year dummies and MSA dummies. Corrected standard errors are reported in parentheses.
***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.

Table 6. Difference-in-Differences using Gallons-per-Mile

	(1)	(2)	(3)	(4)
	OLS	IV-lin	IV-nonlin	Non-causal
Price^{gas}	0.18681***	0.16861	0.45591*	0.04648
	(0.0368)	(0.1561)	(0.2464)	(0.1220)
VMT	0.03618***	0.60231***	-0.01816	
	(0.0104)	(0.1927)	(0.1037)	
Matching effect	**-0.02445***	**-0.01995**	**-0.04937**	**-0.01051**
(VMT x Price^{gas})	**(0.0040)**	**(0.0140)**	**(0.0228)**	**(0.0118)**
MSA FE	X	X	X	X
Year FE	X	X	X	X
Demographic Controls	X	X	X	X
N =	287,463	287,463	251,960	287,463
Instruments:	--	Price^{gas} × Demographic controls	$\widehat{VMT}_{i\bar{j}}$; $\widehat{VMT}_{i\bar{j}}$ × Price^{gas}	--
Compare to:	Table 2 Col. 6	Table 3 Col. 3	Table 4 Col. 3	Table 5 Col. 3
Equivalent value of matching effect for MPG at mean	0.0144***	0.0118	0.0292**	0.0062

Notes: The dependent variable in these regressions is 1000 × GPM. The columns correspond to selected models from Tables 2, 3, 4, and 5 respectively. In Column 3, $\widehat{VMT}_{i\bar{c}mt}$ is estimated using Bento et al. (2009) structural model. Standard errors clustered at the state level are reported in parentheses. The unit of observation is the vehicle.
***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.

Table 7. Between-Household Analysis

	(1)	(2)	(3)	(4)
	OLS	IV-lin	IV-nonlin	Non-causal
A. Avg. MPG				
Matching effect	0.00887***	0.01583***	0.02722***	0.01840***
(N = 149,376)	(0.0022)	(0.0054)	(0.0082)	(0.0047)
B. Top Car				
Matching effect	0.01254***	0.02304***	0.03318**	0.01960***
(N = 149,376)	(0.0025)	(0.0047)	(0.0147)	(0.0050)
C. One-Car Households				
Matching effect	0.01141***	0.01453**	0.00133	0.01583**
(N = 70,197)	(0.0032)	(0.0065)	(0.0197)	(0.0062)
MSA FE	X	X	X	X
Year FE	X	X	X	X
Demographic Controls	X	X	X	X

Notes: Each household is one observation. Panel A uses the households average MPG. Panel B uses the MPG of the highest-VMT car. Panel C uses only one-car households. All specifications control for year dummies and MSA dummies and household characteristics.
***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.

Table 8. Matching Effect of Gasoline Taxes

	(1)	(2)	(3)	(4)	(5)
Panel A. OLS.					
Matching effect (VMT x Tax)	0.06760*** (0.0118)	0.06808*** (0.0121)	0.06686* (0.0367)	0.06664* (0.0368)	0.06696* (0.0373)
Matching effect (VMT x Net-of-Tax P)	0.01700*** (0.0018)	0.01740*** (0.0018)	0.01753*** (0.0043)	0.01667*** (0.0044)	0.01688*** (0.0046)
Panel B. IV using gas tax as instrument					
Matching effect (VMT x Price^{gas})	0.03037** (0.0145)	0.02986** (0.0141)	0.03684*** (0.0123)	0.03643*** (0.0124)	0.03723*** (0.0131)
MSA dummies			X	X	
Year dummies				X	
MSA-Month dummies					X
Demographic controls		X	X	X	X
N =	304,496	284,159	284,159	284,159	284,159

This table shows the matching effect from gasoline taxes. Panel A separately estimates the matching effect from taxes from the net-of-tax portion of gas prices. Panel B estimates the effect of gasoline prices when gasoline taxes are used as an instrument.

Appendix A. Derivation of Equation (5).

We begin by taking the derivative of the expression for WTP in Equation (4) with respect to the car type c. Making use of Leibniz's rule when differentiating VMT^* in the limits of integration, we obtain:

$$\frac{\partial WTP(c,i,p^G)}{\partial c} = \begin{aligned} & \frac{\partial f(c,i)}{\partial c} + \int_0^{VMT^*(c,i,p^G)} \frac{\partial \pi(c,i,VMT)}{\partial c} dVMT \\ & + \frac{\partial VMT^*(c,i,p^G)}{\partial c}[\pi(c,i,VMT^*) - GPM(c)p^G] \\ & - VMT^*(c,i,p^G)\frac{\partial GPM}{\partial c}p^G. \end{aligned} \tag{A1}$$

This expression shows how the WTP for a car changes with the car type, as a function of the individual type and gasoline prices. Next, we take the cross partial derivative with respect to the individuals' utilization type.

$$\frac{\partial^2 WTP(c,i,p^G)}{\partial c \partial i} = \begin{aligned} & \frac{\partial^2 f(c,i)}{\partial c \partial i} + \int_0^{VMT^*(c,i,p^G)} \frac{\partial^2 \pi(c,i,VMT^*)}{\partial c \partial i} dVMT \\ & + \frac{\partial VMT^*(c,i,p^G)}{\partial i}\left[\frac{\partial \pi(c,i,VMT^*)}{\partial c} - \frac{dGPM(c)}{dc}p^G\right] \\ & + \frac{\partial VMT^*(c,i,p^G)}{\partial c}\frac{\partial \pi(c,i,VMT^*)}{\partial i} \\ & + \frac{\partial^2 VMT^*(c,i,p^G)}{\partial c \partial i}[\pi(c,i,VMT^*) - GPM(c)p^G]. \end{aligned} \tag{A2}$$

Now taking the third derivative with respect to p^G:

$$\frac{\partial^3 WTP(c,i,p^G)}{\partial c \partial i \partial p^G} = \begin{aligned} & \frac{\partial VMT^*(c,i,p^G)}{\partial p^G}\frac{\partial^2 \pi(c,i,VMT^*)}{\partial c \partial i} - \frac{\partial VMT^*(c,i,p^G)}{\partial i}\frac{dGPM(c)}{dc} \\ & + \frac{\partial^2 VMT^*(c,i,p^G)}{\partial i \partial p^G}\left[\frac{\partial \pi(c,i,VMT^*)}{\partial c} - \frac{dGPM(c)}{dc}p^G\right] \\ & + \frac{\partial^2 VMT^*(c,i,p^G)}{\partial c \partial p^G}\frac{\partial \pi(c,i,VMT^*)}{\partial i} - \frac{\partial^2 VMT^*(c,i,p^G)}{\partial c \partial i}GPM \\ & + \frac{\partial^3 VMT^*(c,i,p^G)}{\partial c \partial i \partial p^G}[\pi(c,i,VMT^*) - GPM(c)p^G]. \end{aligned} \tag{A3}$$

But by the first-order condition for optimal VMT conditional on c, $\pi(c,i,VMT^*) = GPM(c)p^G$; that is, the marginal WTP for VMT equals the marginal cost. Thus, the last term in (A3) drops out and we are left, after re-arranging some terms, with:

$$\frac{\partial^3 WTP(c,i,p^G)}{\partial c \partial i \partial p^G} = \begin{aligned} &-\frac{\partial VMT^*(c,i,p^G)}{\partial i}\frac{dGPM(c)}{dc} \\ &+\frac{\partial^2 VMT^*(c,i,p^G)}{\partial c \partial p^G}\frac{\partial \pi(c,i,VMT^*)}{\partial i} \\ &\frac{\partial VMT^*(c,i,p^G)}{\partial p^G}\frac{\partial^2 \pi(c,i,VMT^*)}{\partial c \partial i} - \frac{\partial^2 VMT^*(c,i,p^G)}{\partial c \partial i}GPM \\ &+\frac{\partial^2 VMT^*(c,i,p^G)}{\partial i \partial p^G}\left[\frac{\partial \pi(c,i,VMT^*)}{\partial c} - \frac{dGPM(c)}{dc}p^G\right]. \end{aligned} \qquad (A4)$$

Appendix B. Proof of the non-causal model

Consider the following two simultaneous equations:

$$y_1 = \alpha + \beta x_1 + \delta x_2 + \theta[x_1 \cdot x_2] + \varepsilon \tag{A.1}$$

$$y_2 = a + b x_1 + c x_2 + d[x_1 \cdot x_2] + \mu \tag{A.2}$$

where $E(x_1\varepsilon) = E(x_2\varepsilon) = E(x_1\mu) = E(x_2\mu) = 0$. Let $\mathbf{X_1} = [1 \quad x_1]$, $\mathbf{X_2} = [x_2 \quad x_1 \cdot x_2]$, $\mathbf{X} = [\mathbf{X_1} \quad \mathbf{X_2}]$, $\mathbf{A_1} = [\alpha \quad \beta]'$ and $\mathbf{A_2} = [\delta \quad \theta]'$. Given the orthogonality conditions, (A.1) and (A.2) can be estimated by OLS. By Frisch-Waugh-Lovell (FWL) theorem:

$$\widehat{\mathbf{A_2}} = (\mathbf{X_2'M_1X_2})^{-1}\mathbf{X_2'M_1}y_1 \tag{A.3}$$

where $\mathbf{M_i} = \mathbf{I} - \mathbf{P_{X_i}} = \mathbf{I} - \mathbf{X_i}(\mathbf{X_i'X_i})^{-1}\mathbf{X_i'}$. $\mathbf{P_{X_i}}$ is the projection matrix onto the space spanned by the columns of $\mathbf{X_i}$ and $\mathbf{X_i'M_i} = \mathbf{0}$. $\frac{\partial E(y_1|X)}{\partial x_2} = \hat{\delta} + \hat{\theta}x_1 = \mathbf{X_1}\widehat{\mathbf{A_2}}$. Call this $\hat{T}$; this is the estimated treatment effect of x_2. To estimate how ∂x_2 is correlated with $\widehat{y_2}$, estimate the following regression:

$$\hat{T} = \pi_0 + \pi_1\widehat{y_2} + \omega \tag{A.4}$$

where $\widehat{y_2} = \hat{a} + \hat{b}x_1 + \hat{c}\overline{x_2} + \hat{d}[x_1 \cdot \overline{x_2}]$, the structural parameters are estimated by OLS and $\overline{x_2} = \sum_{i=1}^N \frac{x_{2i}}{N}$. Let $\mathbf{X_3} = [1 \quad \widehat{y_2}]$ and $\mathbf{\Pi} = [\pi_0 \quad \pi_1]'$.

$$\widehat{\mathbf{\Pi}} = (\mathbf{X_3'X_3})^{-1}\mathbf{X_3'}\hat{T}$$

$$\widehat{\mathbf{\Pi}} = (\mathbf{X_3'X_3})^{-1}\mathbf{X_3'X_1}\widehat{\mathbf{A_2}}$$

$$\widehat{\mathbf{\Pi}} = (\mathbf{X_3'X_3})^{-1}\mathbf{X_3'X_1}(\mathbf{X_2'M_1X_2})^{-1}\mathbf{X_2'M_1}y_1 \tag{A.5}$$

Now consider the following model:

$$y_1 = \alpha + \beta x_1 + \pi_0 x_2 + \pi_1[\widehat{y_2} \cdot x_2] + \xi \tag{A.6}$$

where $\mathbf{X_4} = [x_2 \quad \widehat{y_2} \cdot x_2]$. By Frisch-Waugh-Lovell (FWL) theorem:

$$\hat{B} = (X_4' M_1 X_4)^{-1} X_4' M_1 y_1 \qquad (A.7)$$

Note that the projection of y_1 onto the space defined by X_1 will be the same for (A.1) and (A.6). This implies that $M_2 = M_4$: the partial effects of X_1 in (A.1) can be obtained when the residuals from a regression of y_1 on X_2 are regressed on the set of residuals obtained when each column of X_1 is regressed on X_2. The partial effects of X_1 in (A.6) can be obtained when the residuals from a regression y_1 on X_4 are regressed on the set of residuals obtained when each column of X_1 is regressed on X_4. Hence, in both cases the residual-maker is the same and $M_2 = M_4$. Given this results, the fact that $X_i' M_i = 0$ and using (A.5) and (A.7) we have,

$$(X_3' X_3)^{-1} X_3' X_1 (X_2' M_1 X_2)^{-1} X_2' M_2 M_1 y_1 = (X_4' M_1 X_4)^{-1} X_4' M_4 M_1 y_1 \qquad (A.8)$$

$$(X_3' X_3)^{-1} X_3' X_1 (X_2' M_1 X_2)^{-1} X_2' M_2 M_1 y_1 = (X_4' M_1 X_4)^{-1} X_4' M_2 M_1 y_1$$

$$(X_3' X_3)^{-1} X_3' X_1 (X_2' M_1 X_2)^{-1} X_2' M_1 y_1 = (X_4' M_1 X_4)^{-1} X_4' M_1 y_1$$

$$\hat{\Pi} = \hat{B} \quad \blacksquare$$