Environ Resource Econ (2014) 59:633–657
DOI 10.1007/s10640-013-9753-9

The Effect of Gasoline Taxes and Public Transit Investments on Driving Patterns

Elisheba Spiller · Heather Stephens · Christopher Timmins · Allison Smith

Accepted: 23 December 2013 / Published online: 10 January 2014

Abstract This paper analyzes how driving patterns are affected by gasoline taxes and the availability of a substitute for driving—public transportation. We develop a measure of transportation substitutability based on the difference between individuals' predicted commute times by private and public transit, conditional upon their demographic characteristics and geographic location. Improved substitutability decreases annual vehicle miles traveled (VMT) by inducing modal shifts to public transit, though gasoline taxes are found to have a much larger impact on VMT. Our results imply that a policy that raises gasoline taxes and recycles the revenues into public transit improvements can have even larger impacts on driving patterns than either policy alone.

Keywords Driving patterns · Elasticity of demand for driving · Gasoline prices · Public transportation · Sorting

1 Introduction

Though gasoline consumption has recently decreased due to economic slowdown, high pump prices, and increased fuel efficiency (EIA 2012), light duty vehicles still account for 17 %

E. Spiller (✉)
Environmental Defense Fund, 257 Park Avenue, South New York, NY 10010, USA
e-mail: espiller@edf.org

H. Stephens
California State University, Long Beach, CA, USA
e-mail: Heather.Stephens@csulb.edu

C. Timmins
Duke University, Durham, NC, USA
e-mail: timmins@econ.duke.edu

A. Smith
University of Virginia School of Law, Charlottesville, VA, USA
e-mail: absmith0919@gmail.com

of U.S. greenhouse gas emissions. Given the threat of climate change, policy makers have attempted to reduce these emissions with a variety of policy tools, including increased fuel efficiency standards, gasoline taxes, and public transportation investments. However, these alternative policy approaches are not equal in many respects. For example, increasing gasoline taxes can both reduce vehicle miles traveled (VMT) and shift consumer behavior towards more fuel efficient vehicles without imposing costs on automobile producers. It can also induce individuals to substitute away from private modes of transportation, increasing public transit usage, reducing accidents, and curbing congestion.

At the same time, public transportation investments may help to make public transit more attractive to drivers, inducing modal shifts. In this paper, we analyze the potential for gasoline taxes and public transportation investments to induce households to shift from private transportation to public alternatives, and we recover their impacts on overall driving and gasoline consumption. By so doing, we seek to determine whether increasing the accessibility of public transit systems along with increasing gasoline taxes can be an effective combined strategy to reduce gasoline consumption and VMT.

Understanding how individuals respond to gasoline prices is crucial in this endeavor. While the price elasticity of demand for gasoline has received a great deal of attention in the literature, to the best of our knowledge, measures of public transit accessibility have not been included in the estimation of gasoline demand elasticity. This is problematic, because not taking into account the ability of individuals to shift away from private driving to public transportation modes, especially in urban areas with highly developed transit systems, ignores a potentially important pathway to reducing VMT and emissions. To address this issue, we allow both the availability of public transportation and gasoline prices to affect modal choice, and we show that the switch to public transportation has an important impact on reducing VMT.

In addition, existing studies have focused on data measured at the state or national level, obscuring what could be important local differences due to varying urban structures and public transportation systems. We extend the literature by merging data from the National Household Travel Survey with public-use microdata from the U.S. Census, allowing us to incorporate a measure of the accessibility of public transit in a metropolitan area into a model of demand for VMT. Accessibility is a crucial component of the effective price of using public transit; using variation in this measure therefore allows us to determine the extent to which public transit can substitute for private transportation. In particular, we use the difference between predicted private vehicle and public transit commute times faced by a particular individual. We hypothesize that those facing time tradeoffs that are more favorable for public commuting will be more willing to switch modes and will therefore drive less. Moreover, with this variable we can test whether gasoline prices affect VMT both directly and indirectly (i.e., through mode choice decisions).

We find that, as predicted, those for whom public transportation is a good substitute for private transit (given similar transit times) drive less and are more likely to choose public transit than those who do not have access to similarly efficient public transportation options. However, improvements in relative public transit commute times are found to have as lower impact on modal shifting than increases in gasoline prices. We demonstrate that a policy to improve public transit times relative to private transit times can have an economically important impact on overall driving only when implemented jointly with a policy to increase gasoline taxes.

The paper is organized into the following sections. Section 2 presents a review of the existing literature. Section 3 discusses the choice of commuting and how to control for public transit availability. This is followed by a discussion of the data in Sect. 4. Section 5 presents the empirical methodology used in the analysis and Sect. 6 discusses the regression results. Section 7 addresses unobserved household heterogeneity. Lastly, Sect. 8 concludes.

2 Existing Literature

The literature on price elasticity of demand for driving or gasoline is extensive, both for data within the US as well as in Europe. However, none (to the best of our knowledge) have included measures of public transit access or incorporated the modal choice between private and public transit into the elasticity estimates. Instead, information about the household and the vehicle stock has been used to estimate changes in behavior given gasoline price changes. Multiple surveys of the literature demonstrate a wide range of elasticity estimates, depending on many different factors including econometric tool, data, country, etc. For example, the meta-analyses of the elasticity literature describe estimates found with cross-sectional data as long-run, given that variation in gasoline prices arises from comparisons between equilibrium outcomes across different cities (rather than changes in gasoline prices within a city over time). Hence, cross-sectional estimates tend to be much higher than those based on panel data even within the same methodology (see Baltagi and Griffin 1983; Dahl and Sterner 1991; Johansson and Schipper 1997; Graham and Glaister 2002 for a discussion of cross-sectional data impacts on elasticity estimates).

Sterner et al. (1992) conduct a survey of gasoline price demand elasticity estimates taken from countries worldwide, finding an average of −0.24 in the short run to −0.79 in the long run. Long-run estimates allow the vehicle stock to adjust to gasoline price, providing the vehicle owner with another avenue through which to respond to high gasoline prices by purchasing a more fuel efficient vehicle, thus increasing the elasticity of demand for gasoline.[1] Generally, elasticities tend to be higher in the OECD than in the US: Graham and Glaister's (2002) extensive survey of the literature finds that the range of long-run gasoline price elasticities in European countries can be as high as −1.35 compared to −.23 in the US.[2] This is unsurprising as gasoline taxes have been consistently much higher in Europe, leading to greater adoption of fuel efficient vehicles. While household adoption of efficient vehicles given gasoline price increases tends to result in higher long run estimates of elasticities,[3] Blum et al. (1988) estimate large short-run elasticities in Germany and Austria, ranging from −0.25 to −0.83. These generally larger elasticity estimates, especially in the long run, have important policy implications, as gasoline taxes can have measurable impacts on gasoline consumption if individuals are in fact, responsive to prices. Sterner (2007) demonstrates that fuel consumption would have been 35 % less if all of the OECD had priced gasoline as high as in the UK, and 30 % higher if the prices were as low as in the US.

While it is clear that the elasticity of demand for gasoline and driving will crucially affect the ability to combat climate change given the tax policy chosen by the country, the impact of public transit investments in these outcomes has not been fully studied. In fact, one possibility consistent with our results is that higher quality public transit in European

[1] For example, Fullerton and Gan (2005) find that a 1 % increase in the price of an SUV causes shifts in purchasing away from bundles of vehicles with SUVs in favor of bundles with only cars.

[2] For other extensive surveys of the literature, see Dahl (1986), Oum (1989), Dahl and Sterner (1991), Goodwin (1992), Espey (1998), Basso and Oum (2007).

[3] This difference in estimated elasticities can be large; for example, Dahl (1978) finds a-long run elasticity for US households of −0.78 versus a short-run elasticity of −0.44. However, both Haughton and Sarkar (1996) and Sipes and Mendelsohn (2001) find smaller discrepancies of 0.1 to 0.2 between long-run and short-run estimates.

countries could facilitate households' abilities to reduce VMT when gas prices rise; failing to account explicitly for the quality of public transit may make it look as though Europeans are simply more sensitive to changes in fuel prices. Our paper extends upon the elasticity literature by focusing on how gasoline and VMT demand are affected by access to public transportation, a clear gap in the current literature.

A related body of literature has researched the impact of the costs of transportation modes on public transit utilization. Su and DeSalvo (2008) examine the effects of transport subsidies on the spatial size of a metropolitan area theorizing that residents will choose the mode, either public or private (assumed to be an automobile), that has a lower cost. They assume that the marginal cost of using public transit is primarily a time cost, as waiting for the transit vehicle to arrive is more costly to travelers than being stuck in traffic in a private vehicle (U.S. Department of Transportation 1986). Increases in the price of automobile travel (such as would be caused by increases in gasoline prices) lead individuals to substitute toward public transit, increasing the area (relative to the city center) over which residents would choose to utilize public transit, thus increasing overall usage and extending urban sprawl. Grazi et al. (2008) find a similar relationship between urban form and transportation choices. Greater urban density decreases the probability of using a car and is related to a shorter commuting distance. In more densely populated areas, workers shift away from automobiles as the time-cost of public transit is "substantially reduced"(p. 109). The authors argue that this is due to both a better public transit network and frequency of service as well as greater traffic congestion associated with high-density areas. They also cite parking costs in high-density areas as a motivation for the modal shift. Our paper incorporates the findings from this literature by controlling for the endogeneity of public transit use and its effect on driving and estimating both the direct impact of gasoline prices on VMT along with the indirect effects of higher prices on mode choice. Furthermore, we seek to understand how the accessibility of public transit affects individuals' sensitivity to gasoline prices, driving patterns and mode choice.

A final strain of literature considers transportation mode choice through the use of discrete choice models. In his pioneering work, McFadden (1974) considered the choice by households about whether to drive or take public transit to work and then used the results to predict future household use of public transit. McFadden's results suggest a tradeoff between public and private transit that depends on differences in the relative costs in both time and money across the modes. Train (1980) examined the joint decision of how many automobiles to purchase and which transportation mode (private vehicle, bus, or train) to use for commuting. His results suggest that increasing the cost of driving leads to a shift in mode choice toward public transit, and a subsequent reduction in automobile ownership. However, he cautioned that his results may not be predictive or applicable to other areas beyond the specific neighborhoods in San Francisco that he examines. Bhat (1997) added the stops made for non-work activities into his analysis of commuting patterns. Using data from the Boston metropolitan statistical area (MSA), he found that demographic characteristics predicted the number of non-work stops and mode choice. In related work, Bhat and Sardesai (2006) examined the commuting patterns of households in Austin, TX, and found that both non-work stops and waiting time (the full time cost of public transit) were important determinants in which transportation mode was used for commuting. In all cases, these papers are limited to one MSA and only look at one commuting trip; they do not examine the full commuting patterns of households. Our paper extends upon this literature by examining the overall commuting patterns for individuals across urban areas in the U.S., rather than focusing on only one randomly chosen trip.

3 Measuring Public Transportation Availability

A main goal of this paper is to create a measure of the relative tradeoffs between public and private transportation. As much of the transportation literature has demonstrated, one of the most important factors in determining mode choice is the time tradeoff faced by the individual. Following upon this fact, we simplify our analysis by comparing relative travel times under different modes. However, two practical issues affect this comparison. First, no dataset exists that contains both private and public travel times at the individual level. This impacts our estimation technique; we describe how we address this issue more in Sect. 5. Second, as it is important to compare similar trips across households, we have chosen to compare travel times for commutes to work.

In addition to their similarity of purpose across households, there are other reasons to consider work commutes. In the short run, these trips cannot be eliminated from a household's daily miles traveled in order to avoid higher gasoline prices. Also, they cannot be easily altered in distance. Driving is an integral part of the modern lifestyle and workweek; according to the 2000 Census, 87.9 % of the United States population commute to work in a car, and, of these, 75.7 % drive alone (Reschovsky 2004). The means by which workers can most quickly, easily, or cheaply get to work could have a significant effect on the demand for gasoline. Similarly, the availability of alternative means of commuting, such as public transit, should affect the demand for gasoline. Thus, a metropolitan area (or sub-area within a metropolitan area) with superior public transportation options should exhibit lower average VMT than an otherwise comparable city. When public transportation options are good, utility maximizing individuals may be more able and willing to switch to alternative modes rather than absorb the increased costs from higher gasoline prices or continuing to travel solely by private vehicle. Furthermore, commuting trips may be easier to switch to public transit than other vital household trips, such as shopping for groceries and driving children to activities, as commuting is done at regular times and public transit is likely to be more available at peak commuting times.[4] This, in addition to the regularity of commuting, motivates its use as the type of trip utilized in the creation of this comparison measure.

Our analysis is based on a theoretical framework in which households are assumed to choose their VMT and mode of transit so as to maximize their utility. Parry and Small (2009) contend that this utility is affected by the total cost of travel, which, depending on the mode, includes components such as service frequency, speed, wait time at transit stops, the externalities of pollution and accidents inflicted by other drivers, and direct monetary costs. In most situations, travel by public transit should be expected to take longer because of the wait time for the transit vehicle to arrive, slower speed of travel of the vehicle, and the number of stops made en route. Individuals therefore not only compare the difference in monetary costs, such as the fare price versus gasoline and parking costs, but also consider the time and hassle associated with the trip. The possibility of switching modes without significant sacrifices of time or convenience is likely the strongest incentive to public transportation utilization. Thus, we simplify the analysis and utilize the time trade-off between public and private modes as our measure of public transit accessibility.

While we utilize the difference between commuting times by different travel modes as a proxy for overall transit accessibility, our analysis of overall driving includes non-commuting driving. Our analysis of VMT is based on reported yearly VMT at the individual level,

[4] Subway and buses run more frequently at peak times in most cities within the US, which may not be true in other parts of the world.

allowing us to look at how changes in gasoline prices and accessibility to public transit affects overall driving patterns, as opposed to just commuting trips.

4 Data

The main data used are from the 2001 National Household Travel Survey (NHTS). This survey is conducted by the U.S. Department of Transportation's Federal Highway Administration and the Bureau of Transportation Statistics. It quantifies the travel behavior of the American public by gathering data on long-distance and local travel. In addition to the trip-related data, information on demographic, geographic, and economic characteristics is gathered. The survey is presented in five data panels, representing (i) the household, (ii) each person within the household, (iii) each household vehicle, (iv) characteristics of each day trip, and (v) characteristics of each long trip made by a person within the travel period.[5]

Our analysis is performed with the individual as the unit of analysis, where the dependent variable is the annual VMT for each individual. All other variables are also measured at the individual level; the only exception to this is the measure of average fuel efficiency of vehicles in each household. For multi-vehicle households, these total miles could occur in different vehicles. While the data match vehicles to individuals, the match is not exact, and many individuals are assigned as the main driver of more than one vehicle. To address this problem of vehicle selection the individual is ascribed the average fuel efficiency measure of all the vehicles in the individual's household.

The full 2001 NHTS contains a total of 160,758 observations, although we limit our analysis to those in key metropolitan areas as described below. Some alterations to the data were made in order to facilitate compatibility with data from the 2000 Household Census, which were also used. Household family incomes were re-coded using the mid-point of reported ranges, but the highest range was top-coded at $100,000. A similar re-coding was performed for education level, collapsing the detailed categories used by the NHTS into four codes representing less than high school, high school graduate, some college (including technical school and associate degrees), and college degree (or higher). Similarly, the detailed options for the method of travel to work were condensed into categories for private vehicle (car, SUV, van, pickup truck, other truck, RV, and motorcycle) and public transit (public bus, commuter train, subway/elevated rail, and street car).[6]

We use the confidential NHTS data[7] that include an individual's geographic location in a census tract within an MSA. As defined by the U.S. Office of Management and Budget, an MSA consists of a core area containing a certain population (50,000 for the 2000 Census definitions used in the 2001 NHTS) and surrounding communities that are highly related to the core through strong economic and social integration. A census tract is a smaller subdivision that usually has between 2,500 and 8,000 individuals.

[5] For a detailed description and summary of the dataset, see Hu and Reuscher (2004).

[6] Hotel and airport shuttles, limousines, taxis, private boats, and private airplanes were not included as fitting either of these two definitions; however, this excluded a negligible portion of the survey responses. Out of the 91,742 individuals who reported a method of commute, all of the above represented a combined 118 observations. Walking and biking, which constituted a total of 1.4 % of those surveyed, were also excluded. Likewise, no distinction was made for those who carpool, representing 3.2 % of those who commuted by private vehicle.

[7] While much of the National Household Travel Survey data is publicly available, geographic locations of individuals are only available via a confidential agreement with the U.S. Department of Transportation.

While presumably differing in urban structure from those cities offering only bus routes, some metropolitan areas in our analysis also feature subway/elevated rail and commuter rail. The ultimate purpose of this paper is to discern if a difference exists among commuters in their choice of transportation mode depending on the ease of use and relative cost of public transit as compared to driving a private vehicle. For many commuters in large cities that contain public transit beyond bus routes, these alternative forms may be the most efficient and preferable option. Therefore, their exclusion could obscure an important difference across metropolitan areas. Since the predicted commute times are created for each individual, the inter-city differences will not affect how the predictions are formed. See Table 1 for a list of the included metropolitan areas and the presence of subway or rail in any of the cities within the MSA, indicated by an “X” in the second column.

Unfortunately, the NHTS does not include sufficient information on comparable public transit access for the drivers in the survey given low numbers of public transit users in this dataset. Thus, Public Use Microdata Samples (PUMS) from the 2000 Census are also utilized. The Census data provide many of the same demographic and geographic variables of interest included in the NHTS surveys. These data are used in the analysis to predict commute times for each individual in the NHTS based on these common variables (see Sect. 5). While the NHTS was conducted one year after the Census, this one-year difference should not significantly alter the values of the variables enough to invalidate the compatibility of these datasets.

The Census data provide the first constraint on which geographical areas are included in the analysis. When predicting transit times (see Sect. 5), we map both NHTS and Census data to its Public Use Microdata Area (PUMA). PUMAs are the smallest level at which the Census PUMS data are available; they contain about 100,000 residents. In order to predict public transit times for individuals in a certain PUMA, it is necessary that there are enough observations per PUMA to successfully estimate a regression equation. We include PUMAs from the Census that have at least 20 observations of public transit utilization in order to satisfy the ordinary least squares (OLS) order condition. This reduces the available areas in the final sample, as some PUMAs do not have sufficient public transit observations to allow for predictions.

The other important data needed for this analysis are gasoline prices. The ACCRA Cost-of-Living Index provides the local price of unleaded, self-service (where available) gasoline on a quarterly basis for about 300 cities. Gasoline prices were collected for 2001, and the prices for the core city in the MSA were used and averaged across the year. These prices are also presented in Table 1. Our use of yearly average MSA-level gasoline prices follows the strategy employed in much of the elasticity literature (including Wheaton 1982; Goldberg 1998; West 2004; Bento et al. 2005; Li et al. 2009). As we do not have access to a larger sample of more detailed, time-varying gasoline price data, we are not able to include in the regressions MSA-level dummy variables, which would account for any location-specific unobservable determinants of driving behavior. Instead, we include census division dummies in order to account for regional variations in driving patterns and access to public transit. Furthermore, our public transit access measure is created at the PUMA level, which increases the amount of variation within the MSA.[8]

[8] We are unable to use MSA dummies in the regression, as they would be correlated with the MSA price variable. We tested the implications of this by using free monthly MSA gasoline price data available from the DOE for a select sample of six MSAs and including MSA dummies in the regression. The results with month-MSA level prices are statistically insignificant, and they are also not statistically different from the results on the subsample of the same MSAs with yearly level MSA prices making us more confident in using

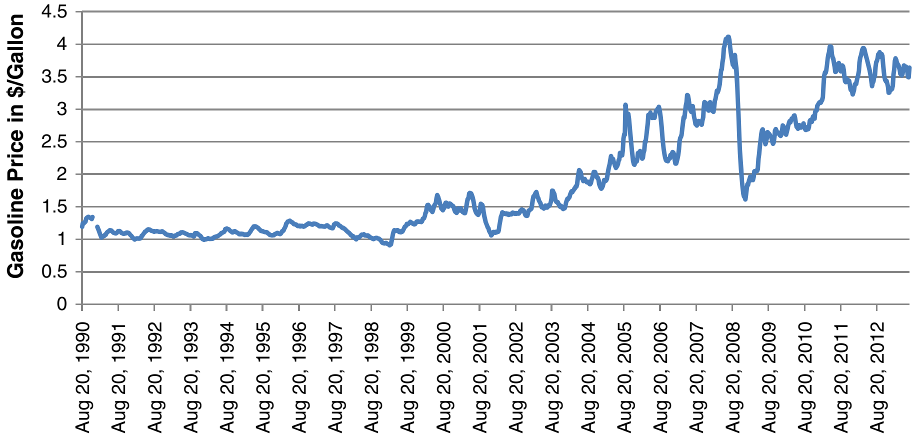

Fig. 1 Weekly U.S. regular all formulations retail gasoline prices ($/Gallon) *Notes* Gasoline price data show the weekly price averaged over the entire US; series taken from Energy Information Administration webpage "Weekly U.S. Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)": http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?f=W\&n=PET\&s=EMM_EPMR_PTE_NUS_DPG

Given the small amount of variation in gasoline prices over time during this period and the years prior, it is likely individuals are making decisions on driving given a longer time period of gasoline trends rather than month-by-month variation in prices. Figure 1 demonstrates that national average gasoline prices were relatively stable until after 2003, further reinforcing the argument that the use of average yearly gasoline prices may reflect the true price factor that affected households' decisions during this period.[9] Figure 2 shows weekly prices during the sample period for our analysis (March 2001–May 2002); average national prices are generally stable and hover around $1.40.

Because the proxy for public transit quality is created from predicted commute times, it is only valid for those who are in stable, full-time jobs with a regular commute. To make the NHTS comparable to the PUMS data, individuals in the NHTS data are therefore only included if they are employed, self-reported as a driver, and over the age of 25. The reported mode of commuting is for the week prior to the survey, which we assume is the individual's regular choice for mode of transit. The individual also provides information on use of transit (not exclusively for commuting), ranging from never having used public transit in the previous month to more than twice per week utilization. We aggregate these responses to look at mode choice; those who ever used public transit in the last month are coded as "public transit users", which is our discrete variable used for mode choice. Non-response and missing data for certain key variables require the exclusion of numerous observations from the analysis.

In order to account for the fact that household location can result in different driving patterns for each individual, we measure the distance that each individual would have to drive to reach the nearest urban center. Specifically, we calculate distances to the population

Footnote 8 continued
our annual MSA gasoline price data. Unfortunately, time-varying data are expensive and we do not have funding to purchase the time-varying data for the full sample of MSAs.

[9] As our data spans through the terrorist attacks in September 2001, factors such as transit disruptions in New York and Washington, DC, higher gasoline prices, etc., may have affected the results. Thus, we test our results excluding the observations after 9/11 and our results were not statistically significantly different from not excluding them; suggesting the 9/11 disruptions do not drive our results.

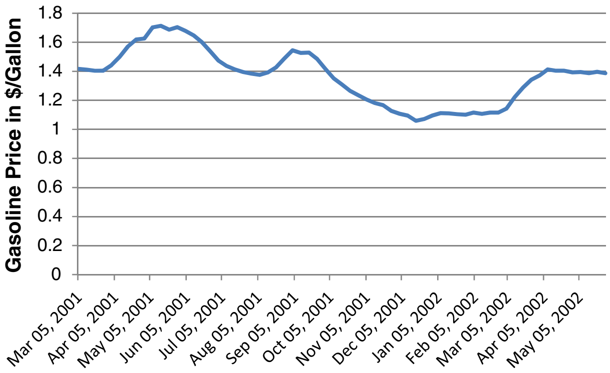

Fig. 2 2001–2002 Weekly U.S. regular all formulations retail gasoline prices ($/Gallon) *Notes* Gasoline price data show the weekly price averaged over the entire US during the NHTS sample time period; series taken from EIA webpage "Weekly U.S. Regular All Formulations Retail Gasoline Prices (Dollars per Gallon)": http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?f=W\&n=PET\&s=EMM_EPMR_PTE_NUS_DPG

weighted centroid of each MSA from the population weighted centroid of the census tract where each NHTS individual is located using ArcGIS and the 2000 Census boundary data. This allows us to control for differences in relative remoteness of census tracts.

The subset of the data used for the analysis contains 9,677 usable observations, representing 45 MSAs. Summary statistics by metropolitan area are shown in Table 1.[10] Table 2 compares the final NHTS sample with three groups: (i) an urban subset of the NHTS (which includes individuals in an MSA with a population greater than one million but does not account for the other cleaning steps), (ii) the entire NHTS survey, and (iii) the 2000 Census data used in the analysis.

As Table 2 indicates, there are some differences between the full NHTS dataset and the final sample. Average time to work in the final sample is greater than in the full NHTS but is more similar to what is observed for the same region in the 2000 Census data. Distance to work and number of workers in the family is similarly greater in the final sample, while the number of household vehicles is similar to that in the full NHTS and smaller than but closer to that in the Census. The final sample is somewhat wealthier and drives more than the urban subset but has VMT similar to that of the full NHTS. Thus, individuals in the excluded metropolitan areas have lower incomes and shorter commute times and distances than those in the included MSAs. We therefore proceed with the caveat that the potential for sample selection bias may exist, but the data do not allow us to detect substantial observable demographic differences.[11] Nevertheless, our results should still provide some insight into how public transit accessibility and gasoline prices affect households in our study area.

5 Empirical Methodology

As no dataset has both public and private commute times for each individual, our goal is to predict public commute times for individuals who currently commute by private means

[10] Table 1 includes an X for each city that has a subway system. However, we do not control explicitly for the existence of a subway system in our estimation. See Sect. 5 (Footnote 17) for a discussion of why we do not need this control and of how our model does control for spatial variation in the quality of public transit.

[11] At a minimum our results should be robust for the population and areas included in our data.

Table 1 Summary statistics by metropolitan area (mean attributes)—final sample

Metro area	Subway/ rail	2001 gas price	Distance to work	Time to work	VMT	HH size	HH income	NHTS obs.
Atlanta, GA	X	1.28	10.13	23.30	12, 496.39	2.63	62,587.72	57
Austin–San Marcos, TX		1.31	7.54	17.55	10, 597.32	2.35	59,891.30	23
Boston–Worcester–Lawrence, MA–NH–ME–CT	X	1.5	13.85	27.10	11, 640.31	2.97	72,243.35	263
Buffalo–Niagara Falls, NY	X	1.48	7.65	17.29	8, 007.12	2.62	55,476.19	42
Charlotte–Gastonia–Rock Hill, NC–SC	X	1.32	8.29	19.88	10, 382.26	1.38	54,062.50	8
Chicago–Gary–Kenosha, IL–IN–WI	X	1.54	15.00	31.03	10, 011.72	3.00	70,935.11	393
Cincinnati–Hamilton, OH–KY–IN		1.36	10.47	19.31	8, 724.91	2.61	61,352.46	61
Cleveland–Akron, OH	X	1.39	11.45	21.30	11, 074.79	2.82	60,117.92	106
Columbus, OH		1.4	6.59	15.41	8, 910.27	2.48	63,812.50	40
Dallas–Fort Worth, TX	X	1.37	12.98	25.21	11, 399.75	2.47	73,458.90	73
Denver–Boulder–Greeley, CO	X	1.51	11.67	24.34	11, 239.55	2.63	67,823.13	147
Detroit–Ann Arbor–Flint, MI	X	1.39	9.37	19.45	8, 430.94	3.71	46,041.67	24
Grand Rapids–Muskegon–Holland, MI		1.42	10.08	20.69	13, 160.88	3.46	70,000.00	13
Greensboro–Winston-Salem–High Point, NC		1.31	16.65	34.60	12, 847.35	2.00	37,500.00	5
Hartford, CT		1.57	9.88	20.63	7, 373.00	2.88	76,562.50	8
Houston–Galveston–Brazoria, TX	X	1.36	13.38	26.72	11, 735.50	2.79	66,603.05	131
Indianapolis, IN		1.33	3.88	11.43	7, 512.35	2.00	45,000.00	9
Jacksonville, FL	X	1.4	20.00	30.00	27, 386.13	2.00	52,500.00	2
Kansas City, MO–KS		1.29	7.37	15.63	9, 413.80	3.00	60,555.56	9
Los Angeles–Riverside–Orange County, CA	X	1.64	13.41	26.89	10, 921.49	2.84	69,673.25	329
Louisville, KY–IN		1.36	7.06	13.11	9, 686.37	3.04	74,565.22	23
Memphis, TN–AR–MS		1.31	8.31	19.15	11, 092.98	2.12	58,088.24	17
Miami–Fort Lauderdale, FL	X	1.42	12.78	27.68	10, 559.34	3.38	51,388.89	63
Milwaukee–Racine, WI		1.41	8.47	18.55	9, 014.12	2.90	64,656.86	51
Minneapolis–St. Paul, MN–WI	X	1.45	12.79	23.13	11, 583.09	2.82	70,126.15	218

Table 1 continued

Metro area	Subway/rail	2001 gas price	Distance to work	Time to work	VMT	HH size	HH income	NHTS obs.
Nashville, TN		1.36	4.53	12.75	5,732.66	2.50	59,375.00	4
New Orleans, LA	X	1.49	7.14	18.93	9,216.15	2.56	53,281.25	16
New York–Northern New Jersey–Long Island, NY–NJ–CT–PA	X	1.57	17.44	33.22	9,773.12	3.04	76,746.03	693
Norfolk–Virginia Beach–Newport News, VA–NC	X	1.37	10.42	18.88	11,093.16	2.84	59,300.00	25
Orlando, FL		1.33	11.25	22.37	10,011.51	2.62	61,071.43	21
Philadelphia–Wilmington–Atlantic City, PA–NJ–DE–MD	X	1.56	22.42	24.41	13,436.91	2.64	64,659.09	22
Phoenix–Mesa, AZ	X	1.5	13.20	24.05	10,474.68	2.65	53,669.35	62
Pittsburgh, PA	X	1.36	10.66	23.39	9,456.10	2.73	63,181.82	88
Portland–Salem, OR–WA	X	1.55	16.90	22.51	8,799.30	2.69	65,396.04	101
Raleigh–Durham–Chapel Hill, NC		1.39	11.16	19.83	14,980.81	3.03	74,318.18	33
Sacramento–Yolo, CA	X	1.69	11.95	22.00	12,023.19	2.92	73,103.45	87
St. Louis, MO–IL	X	1.45	14.43	25.39	12,240.14	2.68	66,089.74	78
Salt Lake City–Ogden, UT	X	1.46	17.29	25.16	10,162.58	3.26	61,814.52	62
San Antonio, TX		1.24	9.45	18.21	9,459.15	2.89	55,462.96	27
San Diego, CA	X	1.79	13.42	23.80	11,327.42	3.03	70,767.86	140
San Francisco–Oakland–San Jose, CA	X	1.92	12.04	25.42	8,745.99	2.84	78,882.35	170
Seattle–Tacoma–Bremerton, WA	X	1.54	13.03	28.44	11,534.52	2.81	68,903.23	155
Tampa–St. Petersburg–Clearwater, FL		1.36	11.44	20.45	11,649.36	2.29	61,785.71	21
Washington–Baltimore, DC–MD–VA–WV	X	1.56	16.45	32.71	12,007.25	2.82	77,300.72	276
West Palm Beach–Boca Raton, FL		1.49	7.72	19.17	15,123.85	3.00	70,312.50	8
National		1.45	11.59	22.45	10,943.32	2.73	63,565.42	4,204

Each number in this table is the average value for all NHTS households living in the particular MSA in question. *Distance to Work* is reported in miles, *Time to Work* is reported in minutes, *VMT* is the household level vehicle miles travelled, and *HH Income* is the aggregated family income in dollars

Table 2 Comparison of summary statistics by sample (mean attributes)

	2001 NHTS (final sample)	2001 NHTS (urban)	2001 NHTS (full sample)	2000 census
Observations	9,677	63,132	160,758	14,081,466
Time to work	25.78	26.56	21.65	24.60
Distance to work	14.09	13.47	12.17	
# HH vehicles	2.37	2.06	2.22	2.54
Income	$66,737	$58,419	$53,082	$61,816
HH size	2.99	3.23	3.18	3.19
Number of adults	2.13	2.15	2.11	
Number of workers	1.96	1.60	1.59	
Education level	Some college	Some college	Some college	HS graduate
Population density	4,919.3	6,732.84	5,016.68	
Annual VMT	15,624.03	12,730.64	16,413.74	

The first column (final sample) refers to the average values in the NHTS dataset used for our estimation; the urban sample is the average of all households in the NHTS dataset living in an urban region as defined in the NHTS

(and vice versa). To do so, it is imperative to have a large enough sample of public transit commuters. While the NHTS does survey some individuals who commute by public transit, the percentage (1.8 %) and overall quantity (2,256) of public transit commuters is quite low—an insufficient number to inform the above prediction. However, the Census PUMS data have a larger percentage (5 %) and overall quantity (141,597) of public transit commuters. This motivates a combination of the two datasets: we predict public transit commute times for individuals in the NHTS based on the commute time of those in the Census with similar demographic characteristics who live in the same part of the city (i.e., within the same PUMA) and commute by public transit. Since the actual commute time by public transit cannot be observed for those taking private transit, conditioning upon the PUMA and observable individual attributes is meant to provide a better estimate of the public commute time than simply using city averages or another homogenous measure.

In order to compare public and private transit, we also predict commute times by private transit for each individual in the NHTS using Census data from the same PUMA.[12] In estimating both private and public commute times, we include a full set of demographic characteristics such as the individual's vehicles, home and work location, financial and time constraints, and family characteristics to control for factors that likely affect transit times. These predicted times are then used to create a measure of public transit substitutability, defined as the difference between commute times by private and public modes of commuting for each individual.

We use the difference in public and private transit times as a proxy for transit accessibility given that opinion surveys have found the most onerous parts of travel by public transit to be the time spent waiting and the slower speed of travel (Su and DeSalvo 2008). These concerns are more salient than the monetary cost arising from fares. When the travel times of public

[12] Certain transportation models calculate driving times by public and private times, though there are several drawbacks to these models. For example, a Stata program that uses Google maps does not calculate optimal public transit times (which may involve travel by combinations of different transit modes). Other existing models only work for one MSA and thus would not allow for a calculation of elasticities given our data constraint on gasoline prices.

and private transportation are more similar, we hypothesize that households will be more likely to shift toward public transit for commuting.[13]

Equation (1) allows us to estimate private and public transit commute times for each individual in the 2001 NHTS final dataset using the 2000 Census data.[14] These predictions are obtained by using observable characteristics that have been shown to correlate with public transit use and work distance, based on choice of neighborhood. In a first stage, our regression is at the PUMA level (the smallest geographic area for which the 2000 Census microdata are available). It is assumed that individuals with similar observable characteristics will sort and locate to parts of the PUMA where access to public transit is similar. This is a far less restrictive assumption than requiring the same be true over the entire MSA; thus our approach controls for within-PUMA differences in access to public transportation. The theory behind sorting models predicts that similar individuals commute over a comparable section of the PUMA and therefore face similar public transit (or private transit) commute times.[15] Thus, we predict travel times for both types of transit modes for each individual in the 2001 NHTS final dataset based on their PUMA of residence. The use of predicted values for public and private commuting measures the expected trade-off based on the experiences of a large sample of representatives who actually utilize one of the forms of transit in each individual's location.

In the first stage, we use the following regression model with data from the Census to generate our predictions, first for private transit and then for public transit in each PUMA j:

$$\begin{aligned} TRANTIME_{i,j} = {} & \beta_{0,j} + \beta_{1,j} HHINCOME_{i,j} + \beta_{2,j} HHVEHCNT_{i,j} + \beta_{3,j} AGE_{i,j} \\ & + \beta_{4,j} MALE_{i,j} + \beta_{5,j} WHITE_{i,j} + \beta_{6,j} EDUC_{i,j} + \beta_{7,j} HHSIZE_{i,j} \\ & + \beta_{8,j} L\_W\_PUMA_{i,j} + \beta_{9,j} HHSIZE_{i,j} * HHINCOME_{i,j} \\ & + \beta_{10,j} HHSIZE_{i,j} * HHVEHCNT_{i,j} \\ & + \beta_{11,j} HHVEHCNT_{i,j} * HHINCOME_{i,j} \\ & + \beta_{12,j} AGE_{i,j} * HHINCOME_{i,j} + \beta_{13,j} AGE_{i,j} * AGE_{i,j} + \varepsilon_{i,j} \end{aligned} \tag{1}$$

where $TRANTIME_{i,j}$ is the time it takes for individual i in PUMA j to commute to work by either private or public transit; $HHINCOME_{i,j}$ is the total income of the household; $HHVEHCNT_{i,j}$ is the number of vehicles owned by the household; $AGE_{i,j}$ is the individual's age; $MALE_{i,j}$ is a dummy variable equal to one if the individual is male; $WHITE_{i,j}$ is a dummy variable equal to one if the respondent is white; $EDUC_{i,j}$ is a dummy for having completed college; $HHSIZE_{i,j}$ is the total number of people in the individual's household; $L\_W\_PUMA_{i,j}$ is a dummy variable that equals one if an individual lives and works in the same PUMA; and ε_{ij} is the error term. We include interaction terms in order to allow the functional form of the prediction model to be more flexible. The model is estimated separately for each PUMA, so all parameters are indexed by j. Once these parameters have been estimated, we project the PUMA-level coefficients on the NHTS data and create our predicted travel times.

Figure 3 displays the distributions of these commute times. The mean of the observed private commute time is similar to the mean of the predicted private transit time—24.5

[13] Although other factors (such as cleanliness, privacy, and security) may likely factor into the decision on whether to take public transit, it is difficult to quantify these aspects with data. Therefore, our analysis ignores other reasons why individuals may choose not to use public transit.

[14] We also conduct matching techniques instead of predictions in the first stage. These alternate techniques did not provide us with meaningful estimates; results are not shown here.

[15] For more about this, see Sect. 7.

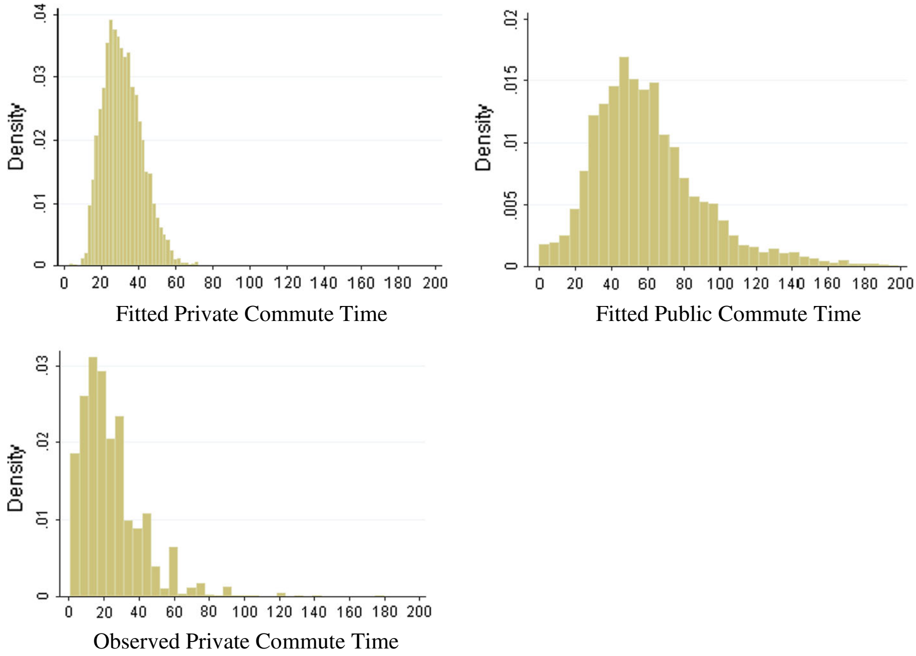

Fig. 3 Distributions of commute times. *Notes* The fitted private and public commute time graphs show the distribution of predicted commute times for NHTS individuals based on Eq. (1). The observed private commute time shows the distribution of observed commute times for NHTS individuals who commute to work by private modes

and 31.6 min, respectively.[16,17] We then create our measure of public transit accessibility by subtracting the predicted public transit time from the predicted private transit time, and include it in the main regression. This measure is defined as *Time Difference*, where an increase in this measure implies an increase in substitutability between the two modes: private transit times increase relative to public transit times. We hypothesize that an increase in public transit accessibility (i.e., *Time Difference*) would lead to a modal shift towards public transit, thereby decreasing VMT.[18]

Equation (1), the estimated commute time equation, uses observable characteristics to model mode-specific commute times for each individual in the NHTS, and thus the predictions may not pick up individual idiosyncrasies. However, as noted before, the predicted measures will control for the expected trade-off between the two forms of transit in a particular PUMA.

16 A *t*-test reveals these two means are not statistically different from each other.

17 Our fitted results include no observations below zero even though we utilize OLS, as none of the observed private or public times are negative or zero.

18 Utilizing the difference in predicted public and private transit times allows us to capture many factors including speed, frequency and reliability that would affect the decision to take public transit. Existence of a subway system also affects public transit times; however, we do not include this as a regressor, as the first stage prediction regressions are conducted at the PUMA level, where it is a reasonable approximation to consider access to the subway as uniform (including distances to subways would provide within-PUMA variation, however we do not have access to this sort of data). Therefore, our estimates do control implicitly not only for the existence of a subway or rail system in cities that have them, but for some degree of within-MSA heterogeneity in access to those systems as well.

Given that we are predicting public transit times for individuals who do not utilize public transit (and vice versa), we may be facing a selection issue that could potentially cause a bias in our predicted travel times. Essentially, if those who take public transit do so for reasons that we are unable to observe (such as living directly on a bus line), then we may underestimate the time it would take for a regular car driver to commute by public transit. This would cause an upward bias on the difference between predicted public and private times. However, if the overestimation of public transit times is systematic, and everyone's public transit times are underestimated, then our final specification would not be biased. Alternatively, if we are not necessarily underestimating public transit times, but instead are overestimating for some and underestimating for others, then this would be a potential source of bias. We attempted to control for this potential bias by conducting two selection correction techniques described in Heckman (1979) (utilizing exclusion restrictions such as number of vehicles and education) and in Lee (1978). However, since there are only a handful of individuals taking public transit in most MSAs, the initial probit model was not well specified in many of the PUMAs; hence, we were unable to use those techniques. Regardless, for those PUMAs with a large enough number of individuals taking public transit, the coefficient on the selection correction was insignificant, demonstrating that selection may not be such an important issue in our application. We control for unobserved heterogeneity by conducting our estimation at the PUMA level. For more discussion of this, see Sect. 7.

Once these predicted commute times have been constructed, it is possible to estimate the VMT equation. Annual VMT is a function of gasoline prices, the choice of transit mode, and demographic characteristics of the household which affect their driving. We cannot, however, simply estimate this equation using ordinary least squares (OLS). Since the choice of transit mode is endogenous in the vehicle miles traveled equation; changes in access to public transit or improvements in public transit commute times would affect both the choice to use public transit and overall VMT. This endogeneity can bias our results if not dealt with explicitly.

To address this, we utilize our public transit accessibility measure (*Time Difference*) as an instrument for mode choice. The assumption here is that public transit accessibility only affects VMT through the individual's modal choice. Since our choice of mode includes taking public transportation for any reason (not just for commuting) this assumption is reasonable: an individual would only choose to drive less due to an increase in public transit availability because that individual would choose to take public transit.[19] As we consider the individual's total VMT in Eq. (2), we also include the observed commute distance for the individual as a separate regressor to account for idiosyncratic differences from average commutes in the PUMA.

The model therefore includes two different choices: VMT and mode. The first choice is reflected in Eq. (2):

$$\begin{aligned}\ln(VMT_i) = {} & \alpha_0 + \alpha_1 PRICE_i + \alpha_2 TRANSITUSER_i \\ & + \alpha_3 MPG_i + \alpha_4 DISTMSA_i + \alpha_5 COMMUTEDIST_i \\ & + \alpha_6 X_{1,i} \ldots + \alpha_{11} X_{5,i} + \mu_i \end{aligned} \tag{2}$$

where the independent variable, $\ln(VMT_i)$, is the natural log of the annual VMT by individual i; $TRANSITUSER_i$ is a dummy for whether the individual is a public transit user (as defined in the NHTS by having taken public transit more than once in the last month), and is endogenous;

[19] If public transit accessibility affected VMT directly (not just through mode choice) it would mean that individuals with better access drive less without shifting modes, which does not seem likely. While it is possible that there could be unobserved factors that lead people to avoid taking public transit while driving more, we assume that once we have controlled for other factors affecting VMT that this is not the case.

$PRICE_i$ is the local retail price of gasoline in individual i's MSA; MPG_i is the average miles per gallon of all vehicles in an individual's household; $DISTMSA_i$ is the distance to MSA from the center of the individual's census tract; $COMMUTEDIST_i$ is the individual's reported commute distance to work; X_1 through X_5 represent five demographic and household descriptive variables (age, race, size of household, total household income, and the number of vehicles in the household); and μ_i is an individual specific idiosyncratic error term. *VMT* measures all travel made by private vehicle (including, but not limited to, commute length), so a switch to public transit from private vehicle travel will not be obscured by using this dependent variable.

Unfortunately, given the unavailability of the primary vehicle, we cannot calculate an elasticity of demand for gasoline directly and have to proxy for that with an elasticity of demand for VMT. In particular, we use the NHTS Person dataset, which indicates which vehicle in the household is primarily used by the individual. However, many individuals in our sample are matched to more than one vehicle, which can occur if certain households share vehicles or have individuals who own more than one vehicle. Thus, it is not possible for us to match the actual vehicle used by the individual without throwing out a substantial portion of the dataset. We believe that, even if we could do so, it would ignore the individual's ability to substitute between the vehicles in his or her garage when relative operating costs change. Other studies have found evidence of within-bundle re-optimization under increased gasoline prices. For example, Spiller and Stephens (2012) measure the elasticity of demand for gasoline given household heterogeneity while allowing for optimization within the bundle of vehicles the household already owns. They find that households engage in within-bundle substitution and utilize their more fuel-efficient vehicles when faced with high gasoline prices. Furthermore, their results show that households with more than one car are substantially more elastic than those with only one vehicle.[20] While we cannot explicitly measure the individual's trade-off between his or her vehicles, our specification allows the individual to optimize over all the vehicles that he or she owns. Therefore, we include as a regressor the number of vehicles owned by the household along with the average MPG of those vehicles. This provides the best estimate available without a more complicated model of this joint decision. Despite using VMT instead of gallons of gasoline, the basic negative relationship between gasoline consumption and VMT holds even if the assumption of consistent choice of vehicle is violated for a portion of an individuals' yearly travel. This implies that a decrease in VMT corresponds to a decrease in gasoline consumption.

Similarly, we do not interact MPG_i with the price variables to get a measure of dollars per mile due to the nature of the data (i.e., the VMT data is at the individual level, *not* at the vehicle level, while MPG is at the vehicle level). This implies that we would only be able to interact the average MPG of all the vehicles in an individual's garage with gasoline price, which is therefore less precise. While it would be beneficial to regress dollars per mile instead of just gasoline price alone, the data do not allow us to do so.

If Eq. (2) were to be estimated directly, the endogeneity of mode choice would cause bias in the parameters. We expect the sign of the bias to reduce the gasoline price coefficient and increase the mode choice coefficient when not controlling for endogeneity: changes in VMT would be attributed more to gasoline prices and less to mode choice, especially if gasoline prices affect mode choice directly. This implies that ignoring the endogeneity issue may cause the researcher to misestimate the elasticity of demand for gasoline.

[20] Feng et al. (2005) also find that households with more than one vehicle are more elastic than households with one.

In order to account for the endogeneity of mode choice,[21] we model the choice to use public transit in Eq. (3):

$$\begin{aligned} TRANSITUSER_i = {} & \beta_0 + \beta_1 PRICE_i + \beta_2 TIMEDIFF_i \\ & + \beta_3 MPG_i + \beta_4 DISTMSA_i + \beta_5 COMMUTEDIST_i \\ & + \beta_6 X_{1i} + \cdots + \beta_{11} X_{6i} + \beta_{12} CDDUM_{1i} \\ & + \cdots + \beta_{19} CDDUM_{8i} \\ & + \beta_{20} WINTER_i + \beta_{21} SPRING_i + \beta_{22} SUMMER_i + \nu_i \end{aligned} \tag{3}$$

where *TIMEDIFF* is our constructed measure of public transit accessibility (*Time Difference*); $CDDUM_{1,i}$ to $CDDUM_{8,i}$ represent the nine Census Divisions in the sample (with one excluded[22]); $WINTER_i$, $SPRING_i$, and $SUMMER_i$[23] are indicator variables for when the individual was interviewed by NHTS (this captures differences in seasons that may affect willingness to wait for the bus, for example); and ν_i is an idiosyncratic error term. Thus, our exclusion restrictions are *Time Difference*, the census division dummies (which pick up regional differences in public transit utilization), and the seasonal dummies.[24]

As public transit accessibility improves, the travel times for public and private transit become more similar, implying that buses, subways, and other forms of public transit are better substitutes for private transit for that individual, providing the individual with more options for travel. We hypothesize that a relative improvement in public transit times will increase the likelihood of shifting modes away from private driving, and thus will decrease vehicle miles traveled. Furthermore, we expect that gasoline prices affect modal shifting. The policy implication here is that an increase in gasoline price (such as through a tax) occurring simultaneously with an increase in access to public transit, or having the tax revenues recycled into these improvements, would further reduce total vehicle miles traveled relative to a gasoline tax or public transit improvement alone.

We estimate these models jointly through one-stage instrumental variables (or generalized instrumental variables). Our predicted "second stage" parameters (A) are estimated in the following way:

$$\hat{A}_{IV} = \left[X'Z\left(Z'Z\right)^{-1}Z'X\right]^{-1} X'Z\left(Z'Z\right)^{-1}Z'Y \tag{4}$$

where A is the vector of predicted parameters from Eq. (2), the X's are the variables in Eq. (2), the Z's are the full set of instruments in Eq. (3), and Y is logged VMT. We calculate standard errors through 1,000 bootstrap repetitions to account for the constructed *Time Difference* being used as an instrumental variable in Eq. (4). This corrects the standard errors from the two-stage process[25] without having to specify asymptotic standard errors.

[21] The Durbin-Wu-Hausman test for endogeneity produces a p-value of 0.000, significantly proving the endogeneity of mode choice and invalidating OLS.

[22] Census Divisions include Pacific, Mountain, West North Central, East North Central, West South Central, East South Central, East North Central, Middle Atlantic, South Atlantic, and New England; we exclude the South Atlantic division.

[23] Fall is the excluded dummy variable. The seasons are generated in the following way—winter: December to February; spring: March to May; summer: June to August.

[24] We do not include census division dummies in the VMT equation, as this masks much of the gasoline price variation in our data.

[25] See Freedman and Peters (1984) and Efron and Tibshirani (1993) for a discussion on the validity of bootstrapping standard errors in multiple stage estimation techniques.

Table 3 Regression results

Dependent variable: Log(VMT)	OLS	OLS	IV
Transit user (mode choice)	–	−.228***	−.468***
		(.029)	(.237)
Price	−.607***	−.498***	−.334**
	(.119)	(.081)	(.156)
MPG	−.002	−.003	.0003
	(.008)	(.005)	(.008)
Age	−.009***	−.009***	−.011***
	(.001)	(.001)	(.001)
Number of vehicles	.091***	.079***	.049**
	(.024)	(.013)	(.026)
White	.187***	.189***	.197***
	(.036)	(.032)	(.039)
Household size	−.040***	−.044***	−.049***
	(.012)	(.010)	(.015)
Distance to nearest MSA	3.92e−06***	3.39e−06***	2.77e−06**
	(7.73e−07)	(6.88e−07)	(1.08e−06)
Distance to work	.009***	.009***	.009**
	(.001)	(.003)	(.003)
Household income	3.90e−06***	4.37e−06***	4.85e−06***
	(5.75e−07)	(4.69e−07)	(7.25e−07)
Constant	9.93***	9.87***	9.71***
	(.267)	(.175)	(.273)
Observations	4,230	4,230	4,230
R-squared	0.078	0.088	0.077
Price elasticity of demand for VMT (mean)	−.929	−0.762	−0.511

The dependent variable is logged individual VMT, using NHTS data. The first column corresponds to the naïve estimation without a dummy variable for transit user; the elasticity estimate reflects the missing variable. The second column includes the dummy variable for transit user, while the third column instruments for the mode choice of commuting. Statistical significance: *** 99 %, ** 95 %, * 90 %

6 Results

We present three different sets of results with bootstrapped standard errors in Table 3: a naïve OLS specification without controlling for mode choice, the OLS regression of Eq. (2), and the IV regression of Eq. (4).

Across the three columns, many of the coefficients are intuitive: higher household income, more household vehicles, longer commute distance, and longer distance from the center of the nearest MSA are all associated with an increase in VMT. MPG is not statistically significant, most likely due to the aggregation of vehicles within the household.[26]

Turning to the effect of mode choice on VMT, we first look at how the inclusion of mode choice affects the price coefficient. If we run the VMT regression without controlling for mode choice, the price coefficient is larger than when we include mode choice in the

[26] The MPG coefficient remained statistically insignificant even in alternative specifications of the model.

Table 4 Endogenous variable probit regression

Probit estimation dependent variable: transit user	
Price	2.523*** (0.264)
Time difference	.0035*** (.001)
MPG	−.0094 (.0127)
Age	−.0072*** (.0020)
Number of vehicles	−.1293*** (.0383)
White	.0393 (.0553)
Household size	−.0689*** (.0187)
Distance to nearest MSA	−9.7e−6*** (1.3e−6)
Distance to work	.0035*** (.0011)
Household income	5.7e−6*** (8.8e−7)
Constant	−3.126*** (.4864)
Census dummies	Included
Seasonal dummies	Included
Observations	4,230
F-statistic to test for weak instruments[a]	20.27

This model estimates the probability for each individual in the NHTS of commuting to work by public transit. Statistical significance: *** 99 %, ** 95 %, * 90 %

[a] The F-statistic is taken from an OLS regression, rather than the probit regression, to show evidence of a strong set of instruments

OLS specification, but the difference is not statistically significant (p-value of difference in means is 0.416). There is, however, a statistically significant difference in the price coefficient estimates between the IV and OLS models. This demonstrates how the (direct) impact of price on VMT can be overestimated when not accounting for the endogeneity of mode choice.

Furthermore, relative to IV estimates, OLS significantly underestimates the effect of mode choice on VMT (p-val: 0.000). This implies that ignoring the endogeneity of mode choice leads to an underestimation of the direct impact of public transit utilization on VMT and understates the impact of public transit accessibility on driving patterns.

As hypothesized, the parameter on the endogenous variable (TRANSITUSER) is negative and significant, demonstrating that public transit use has a large and negative impact on VMT. However, the results from Table 3 do not reveal how mode choice itself is directly affected by gasoline prices or public transit accessibility. Thus, Table 4 presents the results from a "first stage" regression, where we consider the probability of choosing public transit as a function of public transit accessibility, gasoline prices, and other explanatory variables included in the set of instruments from Eq. (4).

This probit regression demonstrates that both improvements in public transit and increased gasoline prices induce modal shifts, though the impact of gasoline prices on mode choice is much larger than the impact of public transit accessibility on mode choice; the percent change in mode choice probabilities given a 1 % improvement in *Time Difference* is .3 % compared to 1.2 % for price. These results imply that improving the public transit system by making it faster and more convenient will have a much smaller impact on mode choice (and thus, driving) than a gasoline tax.[27]

[27] These estimates demonstrate the impact of marginal changes in both gasoline price and public transit accessibility; large changes in public transit (such as from the implementation of a new subway or light rail line) may have large external impacts from new sorting behavior. Our model does not attempt to evaluate large changes in public transit infrastructure; instead we analyze the impact of a marginal change, such as an increase in the number of buses on a currently existing route. We do not expect these marginal changes to have large cascading impacts on sorting behavior. Furthermore, though our model implicitly finds a short-run estimate (in that we hold fixed vehicle stock and housing location), we expect that, in the long-run, major adjustments

Given that price affects VMT both directly and indirectly through mode choice, it is necessary to calculate the full elasticity of demand for VMT in the following way:[28]

$$\begin{aligned}\varepsilon &= \frac{\partial VMT}{\partial price}\frac{price}{VMT}\\ &= \left(\beta_{price} + \beta_{TU}\frac{\partial TransitUser}{\partial price}\right) price \end{aligned} \tag{5}$$

where β_{price} is the coefficient on price, β_{TU} is the coefficient on TRANSITUSER. We calculate the marginal effect of price on mode choice by using the results from the probit estimation:

$$\frac{\partial TransitUser}{\partial price} = \beta_{price}\phi\,(XB) \tag{6}$$

where $\phi\,(\cdot)$ is the normal probability density function (pdf), and the parameters B and variables vector X correspond to those estimated and utilized in the probit model. We find the average marginal effect of mode choice with respect to price to be .749, leading to elasticities of demand for VMT being significantly underestimated when not addressing the endogeneity issue. In particular, the full elasticities with respect to gasoline price corresponding to columns 2 and 3 in Table 3 are −1.03 and −1.06, respectively.

As the marginal effect of gasoline price on mode choice (demonstrated in Eq. (6)) depends on the normal pdf (which is itself a function of all explanatory variables), increased gasoline prices lead to an increase in the marginal effect of price on mode choice. For example, individuals facing high gasoline prices (with gasoline prices one standard deviation above the mean) have 40 % larger marginal effects than those with gasoline prices one standard deviation below the mean (0.88 vs. 0.63). The marginal effect of gasoline price on mode choice is also higher for those with good public transit access, demonstrating that improvements in public transit access have an impact on mode choice through increasing gasoline price sensitivity. Similarly, the marginal effect of public transit availability on mode choice is also 50 % higher for those with high gasoline prices than for those with low gasoline prices (0.0012 vs. 0.0008). These results imply that areas with higher gasoline prices are going to see more of a shift towards public transit utilization than areas with low gasoline prices, and this mechanism works through both public transit accessibility and gasoline prices. The policy result is that a gasoline tax enacted jointly with an improvement in public transit accessibility will have an even larger negative impact on driving patterns in these urban areas than either policy alone.[29]

Our elasticity estimate is on the higher end of the short-run elasticity estimates in the literature, especially compared to those found using individual level data. However, there are several reasons why this would be the case.

Footnote 27 continued
would be small under marginal changes. Though minor adjustments will likely occur with marginal changes, estimating these sorting adjustments would require much more extensive data which we do not have access to, and hence, this estimation is outside the scope of this project.

[28] Given that the dependent variable in Eq. (2) is logged VMT, when calculating the elasticity of VMT with respect to price, VMT falls out of the equation.

[29] These results may imply that public transit improvements today would be more effective than in 2001, given the large increase in average prices over the past decade. This is supported by analytical evidence: Klier and Linn (2010) demonstrate that the demand response is greater with higher gasoline prices. Analyzing the impact of public transit accessibility using more current price data to resolve this issue would be a worthy goal for future work; however, it is outside the scope of this paper. Therefore, our use of earlier data likely underestimates the impact of public transit on driving in later years.

First, while the household data are at the individual level, the gasoline price data are cross-sectional. As described in the literature review section, cross-sectional gasoline data causes our elasticity result to be interpreted more as a long-run estimate, which would be higher than a short-run estimate. However, this study holds fixed households' choice of vehicle bundles in response to a change in commute times or gas prices, as we do not have a model of vehicle choice. If we allowed households to reoptimize their choice of vehicles when faced with higher gasoline prices, on average, individuals would likely purchase more fuel efficient vehicles or ride public transportation more frequently, thus further reducing overall gasoline demand and increasing the elasticity estimate.[30] Given that we focus on marginal changes, however, ignoring this complication should not have an important impact on our results.

Second, we are looking at a distinct group of individuals—those with access to public transit—and, therefore, it may not be representative of the entire country. To test this given our selected urban sample, we run the naïve OLS model (as in column 1) with a larger, non-urban group of individuals (61,825 observations) from the NHTS.[31] We find that the resulting elasticity estimate is still larger than the average in the literature (−.744), though quite a bit smaller than that found in our sample.

Finally, our estimates (even those including the non-urban individuals) may be higher than in the literature is due to the fact that we are not identifying the elasticity by assuming only one vehicle per individual. This elasticity of driving instead reflects the individual's optimization over all vehicles in her garage. This will cause the individual to appear more elastic, since she is able to substitute between the vehicles in her garage as relative operating costs change.[32] Though we do not model this trade-off between the vehicles explicitly, by using overall yearly driving at the individual level, we are in essence identifying the ability of individuals to shift to more efficient vehicles when gasoline prices increase. Therefore, it is natural that our elasticity estimate be larger than the average found in the literature.[33]

7 Unobserved Heterogeneity

A key component of our approach is that we attempt to control for the fact that within a certain geographical location, there may be substantial unobserved heterogeneity with regard to public transit times. Moreover, this heterogeneity may increase with the size of the location. For example, assume we were to aggregate and conduct this analysis at the MSA level. Within the New York–Northern New Jersey–Long Island, NY–NJ–CT–PA MSA, an individual living in Manhattan would presumably face a much different public transit time than a similar individual living on Long Island. If we had aggregated our data across these two individuals and simply predicted their public transit times based on demographic characteristics, we would likely misestimate the public transit time for either or both of them.

30 Even if individuals choose to drive slightly more with a more fuel efficient vehicle (i.e., the rebound effect), this effect will mitigate the decline in consumption but will not completely eliminate it. Gillingham et al. (2013) argue that while the rebound effect exists, it is too small (generally less than 10 %) to overturn energy savings entirely. An older study, Small and Van Dender (2005), also find small rebound effects, both in the long and short-run.

31 We are not able to use the full sample of the NHTS for this estimation given crucial missing data on gasoline prices and MPG.

32 See Spiller (2012) for a discussion of how the explicit modeling of vehicle substitution affects the elasticity estimate.

33 In fact, Lin and Zeng (2012) find that the elasticity of VMT is three times higher than the elasticity of gasoline price.

This sort of geographic heterogeneity introduces a form of measurement bias that we deal with by performing our predictions at a fine level of geography (i.e., the PUMA). Additionally, our approach helps to alleviate the potential for selection bias, where those who take public transit have unobserved lower transit times than those who take private transit.

By obtaining access to the confidential NHTS data, which identifies individuals at the census tract level, and using a PUMA-census tract crosswalk, we were able to match individuals from the NHTS (who commute to work in a private vehicle) to observably similar individuals from the Census data who live in the same part of the same metropolitan area. The hope is that the latter group will provide a valid counterfactual for the commute times for an individual in the NHTS. Put differently, within a PUMA, we assume that access to public transit is sufficiently homogenous that predicted commute times for those taking public or private transit would be a good predictor for demographically similar individuals. Two comparable individuals living in the same area should not expect vastly different public transit times. The similarity of their observed attributes, as sorting models demonstrate, helps predict work location and thus commute times. Therefore, we minimize selection bias by decreasing the size of the geographical area of analysis and matching along observable characteristics.

If we failed to address selection or measurement bias, we would likely miscalculate the estimated coefficients on price and mode choice. In particular, the impact of these sources of bias on the elasticity will depend on the correlation between the error term and the mode choice variable. We demonstrate its effect by widening our geographical area of analysis and estimating the public transit times at the MSA and national levels (instead of at the PUMA level). In predicting whether or not individuals take public transit, we find that the coefficient of mode choice becomes larger, while the price coefficient becomes significantly smaller at both the MSA and national levels.[34] Therefore, not accounting for regional heterogeneity underestimates the direct impact of gasoline prices on VMT due to the correlation with unobservables.

This demonstrates how important it is to accurately identify the level of transit substitutability by minimizing unobserved heterogeneity that is entered into the model. While it is arguable that this could be further controlled for by matching at the census tract level, the PUMS Census data do not permit us to do so, and it is unclear whether confidential Census data would even have enough observations at that level to allow for this sort of estimation. Thus, we attempt to control for this by estimating using observable characteristics. Nevertheless, we recognize that there still could be some selection bias for which we are unable to control.

8 Conclusion

The literature describing the sensitivity of driving behavior to gasoline prices does not generally consider the role that access to other transportation options may play in consumers' driving decisions. This study estimates the price elasticity of driving for urban households when controlling for the substitutability of public and private transit. This substitutability is estimated as the difference between the commute times an individual is predicted to face when traveling by private versus public modes. These predicted commute times incorporate observable attributes from detailed Census information to develop more accurate predictions and control for sorting.

[34] Though we do not present the results here, they are available from the authors upon request.

The analysis supports the theory that individuals facing a better public transit system will substitute away from private transit, decreasing both their yearly miles traveled in a private vehicle, and consequently their gasoline consumption. We find that public transportation infrastructure can have an important effect on individuals' choice of yearly miles traveled beyond the change induced by the gasoline price alone.

Though public transit access has a negative impact on overall driving, the effect is not large. We find that a gasoline tax would have greater efficacy in inducing mode choice shifts, reducing miles driven and gasoline consumed. However, recycling the revenues from a gasoline tax into public transportation improvements would have an extra impact on decreasing VMT overall. Furthermore, not accounting for the endogeneity of mode choice in urban areas causes the model to be misspecified, and makes individuals appear more directly price sensitive than they truly are. In reality, much of that price sensitivity really reflects modal switching and depends upon the quality of the public transportation network.

Our findings indicate that government stimulus money, such as that included in the 2009 American Recovery and Reinvestment Act, used for public transit investments may not be the most effective way of achieving environmental improvements (although we make no comment on its effectiveness in stimulating the economy). Instead, allocating money for public transit projects from gasoline tax revenues could help reduce gasoline consumption and driving without imposing large costs on the government and deficit. High gasoline taxes can have multiple benefits, both in terms of decreasing gasoline consumption and shifting modes to public transit. In this sense, the United States could take a lesson from other parts of the world, such as Europe, where high gasoline taxes go along with extensive public transit networks. Our analysis indicates that the quality of public transit plays a role in driving decisions, and future work on this issue could provide important insights into the value of constructing convenient public transit systems.

Acknowledgments We thank seminar participants at Resources for the Future and Arizona State University; and conference participants from the AERE session at the 2012 Allied Social Science Association Meeting, the 13th UC Santa Barbara Occasional Workshop, the 2012 AERE Summer Meetings, and the 2012 Urban Economics Association Conference.

References

Baltagi BH, Griffin JM (1983) Gasoline Demand in the OECD: an Application of Pooling and Testing Procedures. Eur Econ Rev 22(2):117–137

Basso L, Oum TH (2007) Automobile fuel demand: a critical assessment of empirical methodologies. Transp Rev 27(4):449–484

Bento A, Goulder LH, Henry E, Jacobsen M, von Haefen RH (2005) Distributional and efficiency impacts of gasoline taxes: an econometrically based multi-market study. AEA Papers Proc 95(2):282–287

Bhat C (1997) Work travel mode choice and number of non-work commute stops. Transp Res Part B Methodol 31(1):41–54

Bhat CR, Sardesai R (2006) The impact of stop-making and travel time reliability on commute mode choice. Transp Res Part B 40(9):709–730

Blum UCH, Foos G, Gaudry MJI (1988) Aggregate time series gasoline demand models: review of the literature and new evidence for west Germany. Transp Res A 22A(2):75–88

Dahl CA (1978) Consumer adjustment to a gasoline tax. Rev Econ Stat 61(3):427–432

Dahl C (1986) Gasoline demand survey. Energy J 7(1):67–82

Dahl C, Sterner T (1991) Analysing gasoline demand elasticities: a survey. Energy Econ 13:203–210

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman and Hall, New York

Espey M (1998) Gasoline demand revisited: an international meta-analysis of elasticities. Energy Econ 20:273–295

US Energy Information Administration (2012) Estimated US gasoline consumption low compared to five-year average. http://www.eia.gov/todayinenergy/detail.cfm?id=7510
Feng Y, Fullerton D, Gan L (2005) Vehicle choices, miles driven, and pollution policies. Working Paper 11553. NBER Working Paper Series, Cambridge, MA
Freedman DA, Peters SC (1984) Bootstrapping an econometric model: some empirical results. J Bus Econ Stat 2(2):150–158
Fullerton D, Gan L (2005) Cost-effective policies to reduce vehicle emissions. AEA Papers Proc 95(2):300–304
Gillingham K, Kotchen M, Rapson DS, Wagner G (2013) Energy policy: the rebound effect is overplayed. Nature 493:475–476
Goldberg PK (1998) The effects of the corporate average fuel efficiency standards in the US. J Ind Econ 46(1):1–33
Goodwin PB (1992) A review of new demand elasticities with special reference to short- and long-run effects of price changes. J Transp Econ Policy 26:155–164
Graham D, Glaister S (2002) The demand for automobile fuel: a survey of elasticities. J Transp Econ Policy 36:1–26
Grazi F, van den Bergh CJM, van Ommeren J (2008) An empirical analysis of Urban form, transport, and global warming. Energy J 29(4):97–122
Haughton J, Sarkar S (1996) Gasoline tax as a corrective tax: estimates for the United States, 1970–1991. Energy J 17(2):103–126
Heckman J (1979) Sample selection bias as a specification error. Econometrica 47:153–161
Hu PS, Reuscher TR (2004) Summary of travel trends: 2001 National Household Travel Survey. U.S, Department of Transportation, Federal Highway Administration, Washington, DC
Johansson O, Schipper L (1997) Measuring the long-run fuel demand for cars: separate estimations of vehicle stock and mean annual driving distance. J Transp Econ Policy 31:278–292
Klier T, Linn J (2010) The price of gasoline and new vehicle fuel economy: evidence from monthly sales data. Am Econ J Econ Policy 2(3):134–153
Lee LF (1978) Unionism and wage rates: a simultaneous equations model with qualitative and limited dependent variables. Int Econ Rev 19(2):415–433
Li S, Timmins C, von Haefen RH (2009) How do gasoline prices affect fleet fuel economy? Am Econ J Econ Policy 1(2):113–137
Lin C and Zeng J (2012) The elasticity of demand for gasoline and the optimal gas tax for China. Working Paper, under review at Energy Policy, accessed 6/7/12
McFadden DM (1974) The measurement of urban travel demand. J Public Econ 3:303–328
Oum TH (1989) Alternative demand models and their elasticity etimates. J Transp Econ Policy 23:163–187
Parry I, Small K (2009) Should urban transit subsidies be reduced? Am Econ Rev 99(3):700–724
Reschovsky C (2004) Journey to Work: 2000, Census 2000 Brief. U.S, Department of Commerce, Census Bureau, Washington, D.C.
Sipes K, Mendelsohn R (2001) The effectiveness of gasoline taxation to manage air pollution. Ecol Econ 36:299–309
Small K and Van Dender K (2005) The effect of improved fuel economy on vehicle miles traveled: estimating the rebound effect using U.S. state data, 1966–2001. Policy and Economics, University of California Energy Institute, UC Berkeley
Spiller E (2012) Household vehicle bundle choice and gasoline demand: a discrete-continuous approach, Working Paper. Available at SSRN: http://ssrn.com/abstract=2129396
Spiller E, Stephens HM (2012) The distributional concerns of gasoline taxes: why where we live matters. Resources for the Future Discussion Paper No. 12–30
Sterner T (2007) Fuel taxes: an important instrument for climate policy. Energy Policy 35(6):3194–3202
Sterner T, Dahl C, Franzen M (1992) Gasoline tax policy: carbon emissions and the global environment. J Transp Econ Policy 26:109–119
Su Q, DeSalvo JS (2008) The effect of transportation subsidies on urban sprawl. J Reg Sci 48(3):567–594
Train K (1980) A structured logit model of auto ownership and mode choice. Rev Econ Stud 47(2):357–370
West Sarah E (2004) Distributional effects of alternative vehicle pollution control policies. J Public Econ 88:735–757
Wheaton W (1982) The long-run structure of transportation and gasoline demand. Bell J Econ 13:439–454

Data Sources

U.S. Department of Commerce, Census Bureau. 2000 Census data, 5% sample. Selected variables extracted through University of Minnesota, Minnesota Population Center, Integrated Public Use Microdata Series (IPUMS). Available from http://ipums.org/

U.S. Department of Transportation, Federal Highway Administration. 2001 National Household Travel Survey. Available from http://nhts.ornl.gov

ACCRA Cost of Living Index data for gasoline prices in the chosen metropolitan areas. Available from http://www.coli.org/