Leveraging Inspection Records for Vehicle Miles Traveled Estimates and Analysis of Mileage-Based User Fees in Pennsylvania

Chenyu Yuan
Department of Civil and Environmental Engineering,
Carnegie Mellon University, Pittsburgh, PA
ORCiD: 0000-0002-3821-3314

Zhufeng Fan
Department of Civil and Environmental Engineering,
Carnegie Mellon University, Pittsburgh, PA
ORCiD: 0000-0001-9040-9114

Lin Lyu
Department of Civil and Environmental Engineering,
Carnegie Mellon University, Pittsburgh, PA
ORCiD: 0000-0002-6571-3883

Prithvi S. Acharya
Department of Engineering and Public Policy,
Carnegie Mellon University, Pittsburgh, PA
ORCiD: 0000-0002-5557-7523

H. Scott Matthews[1]
Department of Civil Engineering,
Carnegie Mellon University, Pittsburgh, PA
hsm@cmu.edu
ORCiD: 0000-0002-4958-5981

Word Count: 5,683 + (4 tables x 250) = 6,683
Submitted July 31, 2020

[1] Corresponding Author: 5000 Forbes Ave, BPH123A, Pittsburgh, PA 15214

ABSTRACT

An increasing number of jurisdictions are considering mileage-based user fees (MBUFs) to replace fuel taxes, to fund transportation infrastructure. To support the design and evaluation of MBUF programs, and compare them to the existing fuel tax, we leverage over 119 million records across a fifteen-year period, from annual vehicle inspections in Pennsylvania, to develop high-resolution estimates of the annual cost to vehicle owners of fuel taxes, and of MBUF's at various rates. Applying numerous data cleaning and analytical methods, we use odometer readings from subsequent vehicle inspection records to assess annual vehicle miles travelled (VMT) per vehicle aggregated at the state, county, and ZIP code level. Web-scraping was used to assess the fuel economy of each vehicle in the records and develop estimates for fleetwide fuel economy in each area. Based on these estimates, we find that fees for passenger vehicles would vary by county and ZIP code between ¢2.4 and ¢3.2 per mile, to cost vehicle-owners the same as the existing fuel tax. We also find that vehicles registered in urban areas travel 10-30% fewer miles per year and tend to consume about 10% less fuel per year than average. Our results show that a shift to MBUF's will in general lead to drivers in urban areas, and drivers of hybrid electric vehicles, paying a higher amount than they currently do, while drivers in suburban and rural counties will spend less each year.

1. INTRODUCTION

Since being introduced in 1932, gasoline taxes have been the primary means of funding the U.S. road network. The federal fuel tax has been stagnant for nearly three decades ($0.185/gal since 1993) and has not been adjusted for increases in construction and maintenance costs despite a 240% increase in the construction cost index in that time (1). Several concerns have been raised around the lower funding levels, most notably that it may lead to deferred or delayed maintenance, which would in turn lead to reduced safety, and lower resilience of infrastructure over time. The significant decline in the purchasing power of fuel taxes has been compounded by fuel-efficient (and electric) vehicles, which effectively pay lower (or no) taxes per mile driven (2).

Various states have been considering alternatives to motor fuel taxes by directly charging owners of vehicles based on driving activity like miles driven instead of indirectly through fuel consumption. Vehicle miles traveled (VMT) fees, road user charges (RUCs), or mileage-based user fees (MBUFs) have been proposed as a successor to fuel taxes. Despite some differences in implementation, in the rest of this paper we refer to these interchangeably as MBUFs.

While several pilots (and programs) have been deployed in the U.S., MBUFs are yet to be deployed at scale. Many of these deployments have been a result of collaboration between multiple state agencies—most notably, the RUC West consortium and the Eastern Transport Coalition's pilot programs across many states (3). Over 10 state and local transportation departments are also members in the Mileage-Based User Fee Alliance, a "a national non-profit organization that brings together government, business, academic, and transportation policy leaders to conduct education and outreach on the potential for" MBUF's (4).

States and other jurisdictions use a variety of methods to estimate VMT for planning and analysis purposes. Top-down methods like randomized surveys, such as the National Household Transportation Survey (NHTS) have been conducted for about 50 years (5). However, these methods rely on relatively small sample sizes (in some cases as few as 200 for a state) and may result in estimates with high standard deviations. Travel diaries have also been used to build bottom-up estimates of VMT and contain additional information such as trip types. Some studies have also evaluated registration data, but these studies can suffer from time lags.

Other emerging methods seek to scavenge data from existing datasets such as inspection records (6). In this paper, we leverage odometer readings from consecutive periodic safety and emissions inspections for passenger vehicles to estimate miles driven during that period. Given millions of inspections per year from across the state, it is possible to have large sample sizes that allow for improved and higher resolution VMT estimates across a jurisdiction.

We apply these high-resolution estimates of VMT to produce vehicle-level comparisons of what drivers are currently paying in fuel taxes versus what they may pay in hypothetical MBUFs. Specifically, we first aggregate inspection records for each individual vehicle (after conducting data cleaning and filtering as described below), to estimate vehicle-level VMT. Next, we estimate VMT distribution at the state, county, and ZIP level. Based on this, we calculate the annual fuel tax that each vehicle is likely currently paying, based on these VMT estimates and fuel economy data scraped from numerous websites as described below. Finally, we provide balance points between fuel taxes and MBUFs for each county, and each ZIP code area. To show these results at a vehicle-level, we also provide eight example cases and discuss their circumstances under a transition from fuel tax to different MBUF rate settings.

2. REVIEW OF LITERATURE

Researchers have discussed and modeled the influence of a transition from fuel tax to MBUF in different ways. Sorenson et al. (7) demonstrate how MBUFs would not only provide stable revenues and equitable distribution of tax burden, but also, transitioning from fuel taxes to MBUF can reduce traffic, emissions, and road wear. While positing these benefits, they also acknowledge the lack of public uptake for the idea of MBUF, and argue that jurisdictions must ensure clear messaging on these fees, especially in the context of the urban-rural divide and that MBUF programs must consider factors like household income and variations in fleet fuel economy.

Given that any inequitable distribution of costs and benefits between urban and rural areas are an important factor for policymakers, Fitzroy et al. (8) estimated revenue-neutral MBUF for urban and rural households. They use U.S. Department of Agriculture Rural-Urban Commuting data to classify eight western states into ten different commuting areas, and estimate VMT in each area based on demographics (household size and income, etc.) as developed by the National Household Travel Survey (NHTS) (5,9). Their analyses also decoded VINs through the NHTSA web Application Programming Interface (API), acquired fuel economy information, estimated the effects of mileage-based fees, and compared their percent change with current fuel tax in each state.

Instead of focusing on factors such as commuting time and household demographics, Matteson et al. (10) develop another fuel consumption and mileage allocation comparison model to support revenue-neutral MBUF in the state of Washington. They also use a VMT database and fuel consumption records with VIN configuration data provided by state authorities. Applying a statistical matching process to leverage EPA fuel economy data, they adjusted the VMT distributions and calculated a fuel tax and MBUF at the county level. They use several comparison bases such as geography, vehicle type, vehicle age range and fuel economy to discuss the impact of transitioning from fuel taxes to RUC in each county.

It is clear from the literature that the design and establishment of MBUF programs must consider social, geographical, and economic factors to ensure that programs are equitable and sustainable. It is particularly important to consider rural-urban differences and sweeping changes in fleet fuel economy across time and space. Building on the literature, we develop a data-driven, granular assessment of the impact of a potential transition in Pennsylvania, conducting individual vehicle-level analyses, and aggregating these at the state, county, and ZIP code level.

3. DATA SOURCES

The data used in this analysis are vehicle safety and emissions inspection records from Pennsylvania, provided by the state Department of Transportation (PennDOT) and by a private safety inspection software company, CompuSpections Inc. Table 1 shows a summary of all raw records by year and source. Pennsylvania has a decentralized inspection program that requires annual safety inspections for all vehicles in all counties, in addition to annual emissions inspections in a subset of counties with air quality non-attainment issues, mostly near urban areas (11). PennDOT's data (2000-2016) consist of three types of records: Emissions (for all vehicles in the state that underwent an emissions test), Exempt (for all vehicles that were exempted from an emissions test due to age or lack of mileage but some information is still collected), and Safety. While emissions inspection and exemption records are assumed to be exhaustive, safety inspection data from PennDOT was only available for vehicles receiving a safety inspection at a station that voluntarily paid a fee to report the inspection result to the state. As shown in Table 1, very few safety inspections are voluntarily reported to the State. CompuSpections does not serve all safety inspection stations in the state, so data is a sample of vehicle inspection records from 1999 to 2018. In recent years, a significant share of the safety inspection records for the state's 6 million vehicles are recorded. While all datasets contain many variables for our analysis, we require only the Vehicle Identification Number (VIN), inspection date, and the vehicle's odometer reading on that date. We augmented station addresses not provided in the data with supplemental information from PennDOT.

Table 1: Raw Inspection Data Record Count by Data Source and Inspection Year.

Year \ Origin	**CompuSpections**	**Exempt**	**Safety**	**Emissions**	**Overall**
1999	2	/	/	/	2
2000	1	780,800	/	3,000,804	3,781,605
2001	1	728,400	/	3,057,150	3,785,551
2002	7	658,378	/	3,103,306	3,761,691
2003	5	662,381	/	3,151,591	3,813,977
2004	15	984,715	/	5,562,887	6,547,617
2005	17,324	1,172,051	/	5,611,680	6,801,055
2006	59,622	1,243,627	/	5,494,224	6,797,473
2007	143,859	1,078,123	28,336	5,450,212	6,700,530
2008	193,770	1,382,762	215,787	5,511,450	7,303,769
2009	308,012	1,465,220	288,678	5,544,118	7,606,028
2010	655,482	1,687,027	315,102	5,599,702	8,257,313
2011	857,507	1,705,361	353,423	5,507,609	8,423,900
2012	1,040,980	1,736,269	372,393	5,479,813	8,629,455
2013	1,183,380	1,781,218	570,497	5,558,013	9,093,108
2014	1,320,397	1,831,794	602,998	5,578,552	9,333,741
2015	1,548,213	/	625,877	5,250,120	7,424,210
2016	1,680,438	/	637,209	6,477,578	8,795,225
2017	1,639,841	/	/	/	1,639,841
2018	660,365	/	/	/	660,365
Invalid Date	286,278	0	0	0	286,278
Overall	**11,595,499**	**18,898,126**	**4,010,300**	**84,938,809**	**119,442,734**

For the MBUF analysis, we also require information on fuel economy for each vehicle. Since inspection records do not include this information, four website sources were scraped to 'decode' VINs to acquire details (such as fuel economy) for specific make/model/year/trim levels of vehicles in the inspection records—decodethis.com (12), vinquery.com (13), fueleconomy.gov (14), and the National Highway Traffic Safety Administration (NHTSA) API (15).

The decodethis.com and vinquery.com sites decode VINs with diverse vehicle characteristics, including vehicle make, model year, transmission type, number of cylinders and fuel economy (i.e., separate city and highway fuel economy estimates). However, these two websites contain data on a limited model year range and require full 17-digit VIN codes to decode. We also decoded VINs using a combination of information from the EPA fueleconomy.gov and NHTSA website APIs. Instead of using 17-character unique VINs, NHTSA API requires 'squish VIN' to get detailed vehicle information. The definition of squish VIN is the first 11 digits of unique VIN, except the 9th digit which is a check digit. By using squish VIN, we can aggregate multiple unique VIN's to avoid a large volume of repetitive retrieval (e.g., identical model year and trim level Honda Accords that differ only by their serial numbers). EPA's fueleconomy.gov provides fuel economy data about different vehicle type (e.g., a 2012 4-cylinder Honda Accord). As needed, we use key characteristics acquired from the NHTSA API to match with fueleconomy.gov.

4. DATA CLEANING AND MANIPULATION

In the datasets described above, inspection records for each vehicle are distributed irregularly over time. While inspections are required to be annual, it is exceedingly rare for consequent records to be exactly one year apart. To estimate VMT for each vehicle, we use Python's dictionary data structure (key-value pair) to store each vehicle's VIN as a key and store this key's odometer readings and inspection dates as value pairs in a list. In other words, we transform four datasets into a single VIN-based database, so each vehicle's odometer readings and inspection dates can be retrieved by its unique VIN. The raw datasets contain records with missing, manually entered, or corrupted data, and records with typographical errors (for VINs and inspection dates). Therefore, prior to the development of a vehicle-level key-value database, it was necessary to filter out a small percentage of records based on the criteria listed below:

1. VINs not 17 digits (no-entry; invalid VIN's; old vehicles manufactured before 1981 when the 17-digit standard began),
2. invalid odometer readings (no entry, non-numeric entry), and,
3. invalid date (no entry; invalid date format).

Table 2 shows a summary of the remaining inspection records and unique VINs in the datasets through every filtering step. After filtering, there were 118,107,460 (96%) records and 22,041,151 (98%) unique VINs preserved (an average of about 5 inspection records per VIN).

Table 2: Summary of Records During Data Cleaning Process

Data Source	Cleaning Step	Data Records	Percent (%)	Unique VIN's	Percent (%)
All	Before Cleaning	119,442,734	100	22,386,103	100
Emissions	Before Cleaning	84,938,809	100	18,861,333	100
	VIN Length = 17	84,695,893	99.71	18,694,080	99.11
	Odometer Reading is Numeric	84,695,892	99.71	18,694,080	99.11
Exempt	Before Cleaning	18,898,126	100	9,301,598	100
	VIN Length = 17	18,539,714	98.1	9,206,091	98.97
	Odometer Reading is Numeric	18,352,239	98.06	9,140,539	98.27
Safety	Before Cleaning	4,010,300	100	2,086,635	100
	VIN Length = 17	3,950,055	98.5	2,043,710	97.94
	Odometer Reading is Numeric	3,885,014	96.88	2,011,443	96.4
CompuSpections	Before Cleaning	11,595,499	100	4,779,092	100
	VIN Length = 17	11,244,110	96.97	4,766,240	99.73
	Odometer Reading is Numeric	11,174,328	96.37	4,751,444	99.42
	Valid Date Format	11,174,315	96.37	4,751,443	99.42
All	After Cleaning	118,107,460	95.93	22,041,151	98.46

The cleaned dataset contains over 22 million inspection records over time. Using this database, we estimate a distribution for vehicle-level VMT for the state of Pennsylvania. Each vehicle's annual VMT was calculated through normalized daily VMT using equation 1, which requires a unique VIN to have (at least) two (odometer reading; inspection date) value pairs. However, we consider vehicles that have more than 3 pairs, as two pairs can generate only one VMT estimate. There were 14 million vehicles with ≥3 pairs (64% of all in dataset), 12 million with ≥ 4 (54%), 10 million with ≥5 (45%), and 8.5 million with ≥6 (40%). To ensure every vehicle has sufficient inspection data, the subsequent analysis will use a threshold of ≥3 pairs, meaning every vehicle has at least three pairs, to generate two annual VMT estimates using equation 1.

$$annual\ VMT = 365\ days \times \frac{Odometer\ Reading_{latter} - Odometer\ Reading_{former}}{Inspection\ Date_{latter} - Inspection\ Date_{former}} \quad (1)$$

We also consider the intervals between inspection dates for vehicles. While subsequent inspections are expected about one year apart, this was not always the case. We filtered data for the interval between subsequent inspections to be less than a threshold, to ensure that VMT for each pair of records were assigned to an appropriate year. We estimated the distribution of day intervals between all inspection dates in the database. The 5^{th} percentile of day difference is 146 days, and the 25^{th}-50^{th}-75^{th} to be 347, 369, and 395, respectively (the average is 391 days). We find inspection date gaps are inconsistent, and large date intervals account for a small amount of records but may be far away from the median value, so we use the interquartile range (IQR) measure of statistical dispersion, which can be calculated by the difference between the third and first quartiles (Q3 – Q1). We treat large day differences as outliers if a value is larger than Q3 + 1.5 IQR. Consequently, a date difference value of 467 (about 90th percentile) is picked as a threshold for large inspection dates (outliers). In other words, we do not generate a VMT estimate for a vehicle when the inspections differ by more than 467 days.

Before applying this threshold, we also consider the relationship between odometer readings and inspection dates for a vehicle. For some inspection records, the odometer reading on the latter date is recorded as being lower than the former, indicating a likely error either in the odometer reading or date on record (e.g., from manual data entry). These records were exempted from our analyses even if only one of the pairs is invalid. Besides, there might exist some vehicles that have different odometer readings on the same date which come from different data sources. We also regard these vehicles as invalid. Overall, 921,438 vehicles were removed and 13,088,212 reserved in this check process, as indicated in Table 3.

For date series whose days interval are acceptable, i.e., less than 467 days, there were three typical cases:

Case 1. Purely Consecutive Years, i.e., there are no gap years in the series, or there are multiple inspections are within one year, e.g., ['10/27/2009', '9/21/2010', '11/2/2011', '11/28/2012', '11/6/2013']

Case 2. Consecutive Years, i.e., there are no gap years in the year cohort, but multiple inspections are in one year. Case (1) is included in this case, e.g., ['8/30/2011', '10/3/2011', '1/11/2012', '4/10/2012', '6/1/2013']

Case 3. Acceptable Gap Years, i.e., there may be a year gap between neighboring dates, but this gap is within the 467 days threshold, e.g., ['10/27/2009', '10/9/2010', '1/9/2012', '3/1/2013']

To assess which of these cases to use when developing VMT estimates, we considered three options, and totaled the number of records which would fit the requirements of each. Table 3 shows a summary of vehicles remained in database when applying for valid check and different restrictions conditions. Option 3 has only 1% more data than Option 2, so, rather than considering these two options separately, we chose to use Option 2, which is less restrictive than Option 1 and is still able to use half of the 13 million records, after filtering out vehicles that have invalid (odometer reading; inspection date) pairs.

Table 3: Vehicle count by inspection information conditions.

Option	Description	Count	Percent (%)
	Dataset from previous step	14,009,650	
	Valid VINs and odometer reading–inspection date pairs	13,088,212	100
1	Purely Consecutive Years AND Max Days Difference < 467	4,984,900	38.1
2	Consecutive Years AND Max Days Difference < 467	6,861,493	52.4
3	(Consecutive Years OR Acceptable Gap Years) AND Max Days Difference < 467	7,009,612	53.6

Although we have defined a method to calculate daily VMT, each vehicle's inspection records generally straddle calendar years, so differences in odometer readings provide estimates of driving across two years, which need to be allocated to derive estimates per calendar year. For instance, given an odometer reading R1 at year Y1, and an odometer reading R2 at year Y2 (Y1 < Y2), we can calculate daily VMT using equation 1, but does this VMT belong to Y1 or Y2? We set four allocation options here and discuss the difference among them.

- Assumption 1 – Calculate annual VMT, then allocate it to Y1 (former year).
- Assumption 2 – Calculate annual VMT, then allocate it to Y2 (latter year).
- Assumption 3 – Allocate VMT on both years. First calculate daily VMT, then multiply this daily VMT by number of days in each year.
- Assumption 4 – Similar to Assumption 3 but add a supplementary assumption: if an inspection date is less than 30 days from the beginning or end of a year, we do not include these days' miles within that year.

Using these assumptions, we generate four versions of annual VMT estimates, for several years. Figure 1 show histograms after filtering out outliers larger than 3IQR for the four allocation methods, using year 2005 and 2015 as examples. The density plot shows that the distribution of VMT under each of the four allocations are remarkably like each other. However, the plots with Assumption 3 and Assumption 4 are smoother and always in between the lines of Assumption 1 and 2. Adding a 30-day threshold in Assumption 4 seems not to bring any difference. Consequently, we conclude that the allocating method used in Assumption 3 can better reflect how VMT estimates distribute from 2000 to 2018 and use this assumption for the remaining work.

Given millions of VINs, finding additional vehicle information via web scraping would be time consuming and complex. After the filtering step, the next data preprocessing and cleaning step was the use of 'squish VINs' to aid in web scraping. As described in an earlier section, a 'squish VIN' is a compressed version of unique VIN code. The 14 million vehicles mentioned in Table 3 were mapped to only 88,000 unique squish VINs. After mapping all VINs into their corresponding squish VIN group, the following methods were used to exclude squish VIN that violate naming regulation: (1) illegal alphanumeric characters, (2) other vehicle type (trucks, motorbikes, and buses). For the first case, illegal characters in squish VIN include any lower-case letters, capital letters 'I', 'Q', 'O', 'U', 'Z' and number '0', which are not allowed in a VIN. Only 181 squish VIN with illegal alphanumeric characters were found in the dataset. 3,513 squish VINs were excluded when other vehicle types were filtered out.

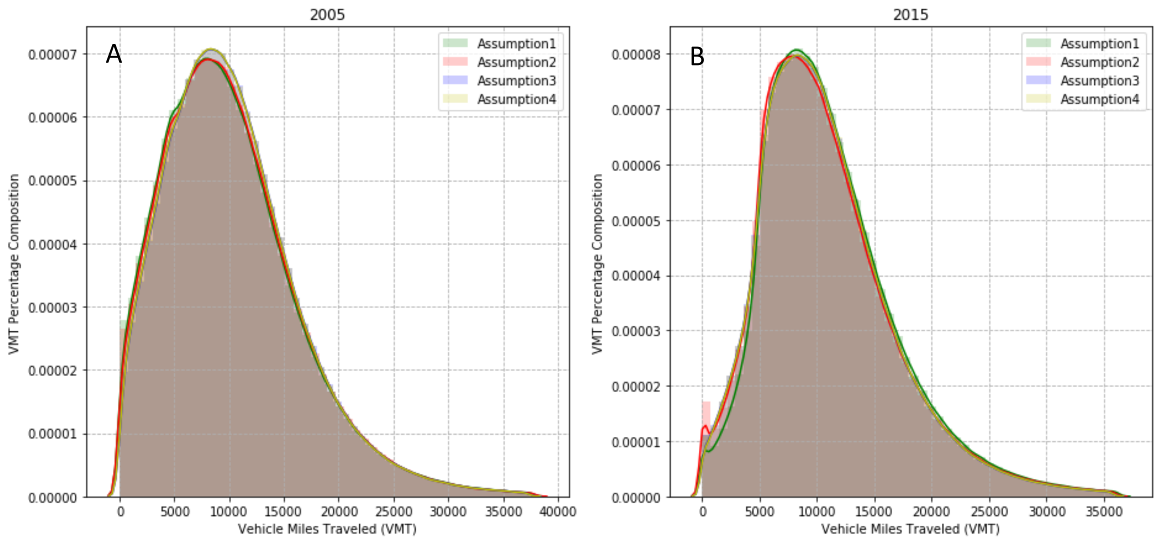

Figure 1: VMT distribution for vehicles in Pennsylvania, (A) 2005, (B) 2015.

After data cleaning and pre-processing, two web scraping methods were designed to collect fuel economy data. The first one used decodethis.com and vinquery.com website to decode each unique VIN; approximately 10% of our records could be decoded from these sources. The second method was a two-step method using the NHTSA API and the EPA Fuel Economy website. Data on model year, make, model name, transmission type, number of cylinders and engine displacement for each squish VIN were first retrieved from the NHTSA API, which does not provide information on fuel economy. We thus supplemented with information from the EPA Fuel Economy website. By querying that website with key characteristics (model year, model, make transmission type and so on) provided by NHTSA API, we collected city and highway fuel economy, in miles per gallon (MPG) for each squish VIN.

5. VMT ESTIMATION

Before developing higher geographic resolution estimates, we assessed temporal changes to estimated annual VMT for the state of Pennsylvania from year to year. Since the same vehicle may appear in multiple records, the number of records is higher than the number of registered vehicles. Figure 2 shows density plots of VMT and number of vehicles used for calculation from 2000 to 2016 (excluding outliers larger than 3IQR), giving an explicit interpretation of distribution for all years analyzed. Year 2017 had only 227 vehicles and year 2018 has only 3 in our dataset, so these two years are not plotted. The VMT appears to have an average of around 10,000 miles (per vehicle, per year), with a left-skewed distribution. We also find that that of more recent years, 2014 has more records than later years.

To provide higher resolution fuel consumption estimates, we calculated annual VMT at the county and ZIP code level. For this purpose, we assumed each vehicle to be based in the same county or ZIP code as the station where it was inspected, since no additional geographic information was available in the inspections database. If available, registration data could have improved this location assignment. The PennDOT Safety and emissions datasets, as well as the CompuSpections data, have ZIP codes for the inspection sites, whereas the Exempt dataset did not. PennDOT separately provided location information for all inspection stations, including all stations' unique identifier (OIS), county and ZIP code. Nevertheless, not every station appearing in the inspection records could be successfully matched, so a small number of records still lack location information. We recognize that vehicles may move (be inspected in several locations over time). Of the vehicles in our database, 71% were inspected in the same county over time across all their inspection records, and 10% inspected in the same ZIP code.

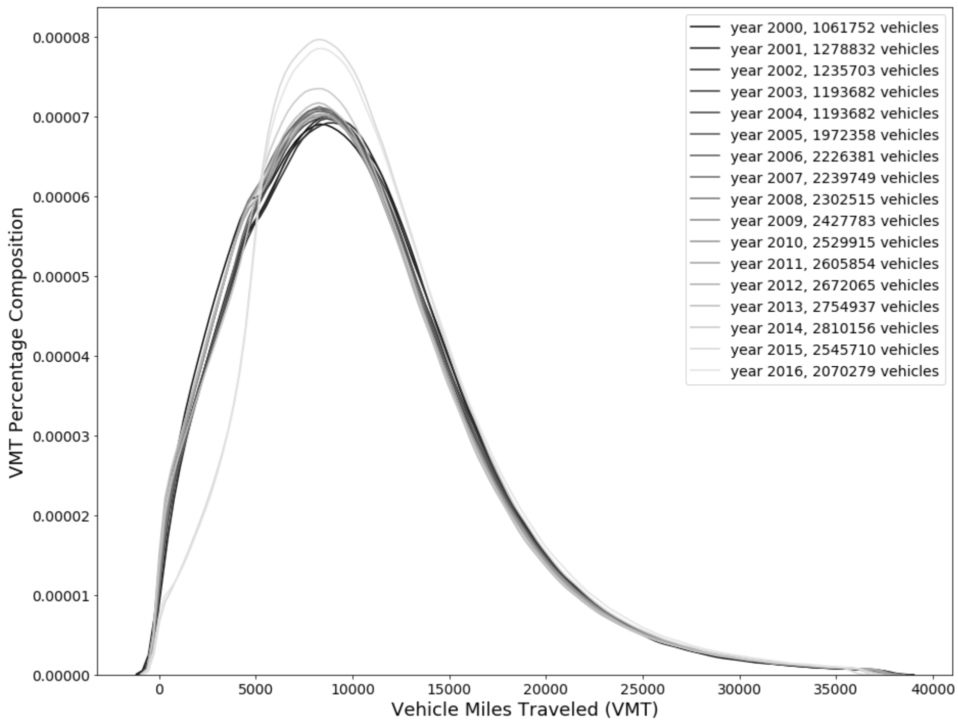

Figure 2: Annual statewide VMT Distribution from 2000 to 2016.

To assess how urbanity affected VMT and fleet fuel economy, we pick five counties: Allegheny, Dauphin, Erie, Lehigh, and Montgomery, which have a mix of urban, suburban, and rural ZIP code areas as case studies. Figure 3 shows the average VMT of these five counties separated into urban, rural, and suburban regions in each county. The x-axis ticks show the years and the corresponding number of unique vehicles included in the estimation for that year. We find that annual VMT increases slightly over time in nearly all regions, with vehicles in rural areas typically travelling more miles per year than those in urban areas. Some fluctuations in the trendlines (e.g., Erie rural 2003, Lehigh urban 2003, Dauphin rural 2017) are explained by there being a miniscule number of records for these regions, leading to unreliable estimations of average VMT. Moreover, given that year 2014 has the most records of the most recent five years, we choose to use year 2014 as representative to estimate and compare fuel taxes and MBUFs in the subsequent analysis.

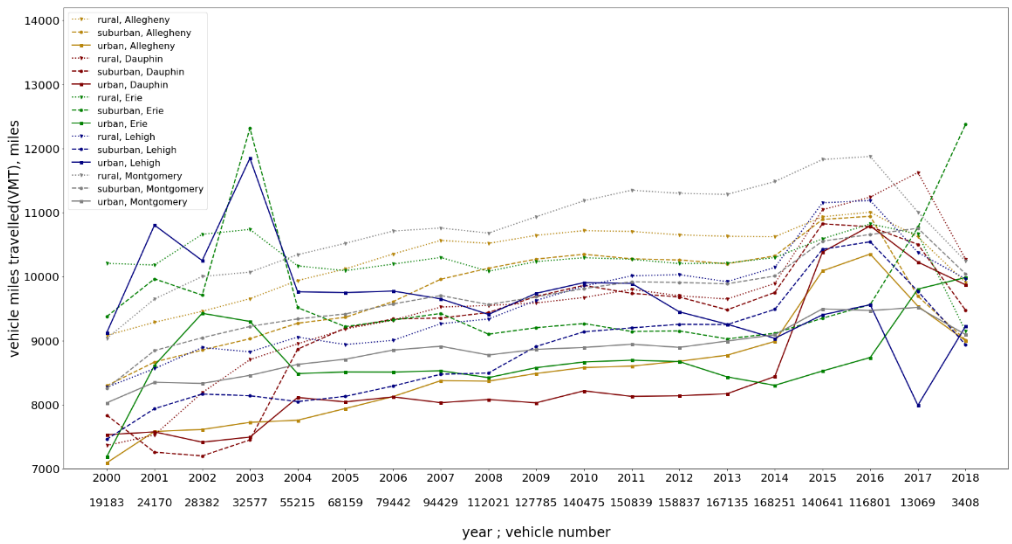

Figure 3: Average VMT by counties and urbanity in 2000-2018.

6. COMPARING THE USER COSTS OF FUEL TAXES AND MBUF

To better understand the differences between fuel taxes and MBUFs, and their impact on vehicle owners, we compared the annual consumer cost of the two. As discussed in Section 5, we chose to conduct our analyses on data from the year 2014, but updated estimates can easily be derived by applying these methods by re-running code to new data as it becomes available. In contrast to a flat fuel tax per gallon, the equity of MBUF programs are an ongoing concern, especially when considering their disparate effects on rural or urban drivers, and the potential penalization of lower income households. We show this effect by calculating "balance points"—i.e., the MBUF rate at which 50% of drivers in a region would pay less under an MBUF system, while the other half would pay more, than they currently spend in fuel taxes. Additionally, we compared county and ZIP code level estimates of VMT and fuel consumption with the statewide average estimates. County-level estimates were not calculated in counties with sparse (fewer than 20 total inspection records) data, which not coincidentally are also low population counties and not subject to emissions inspections. We also show examples of eight sample vehicles, and how policy transitions would affect them. To estimate fuel taxes, we used the current tax rate in Pennsylvania, \$0.576 per gallon, as standard. This rate includes the federal and state fuel tax (but not sales tax). Despite the VMT data being for 2014, we believe this to be an appropriate comparison given the small changes in VMT from year to year as shown in Figure 2. To calculate fuel taxes per year, we calculated the fuel economy and VMT for each vehicle. As described in Sections 3 and 4 we used web scraping to collect city and highway fuel economy for each VIN. We calculated a 'combined fuel economy' by applying the EPA's formula of a geometric mean, which weighs the city and highway fuel economies (in MPG) at 0.55 and 0.45 respectively (16).

After estimating the fuel taxes (by using VMT and combined fuel economy estimates to calculate fuel consumption, and hence fuel taxes), we calculated MBUF at the vehicle level by using the same VMT estimates. The annual MBUF rates considered in these analyses were between ¢1 and ¢3 per mile. For the rough state average annual VMT of 10,000 miles, a vehicle owner would pay \$100 a year at an MBUF rate of ¢1 per mile, and \$300 for ¢3 per mile. For vehicles driven near the upper end of our distribution

(35,000 miles annually) the cost would be $350 a year at the lower rate, and more than $1,000 at the higher rate. *Figure 4* illustrates the effect of changing the MBUF rate on the number of passenger vehicles in Pennsylvania which would spend more or less on taxes per year in Pennsylvania, and shows how the "balance point" of around ¢2.7 per mile, was estimated for the statewide fleet. Next, we applied a similar calculation to study the "balance point" for each county.

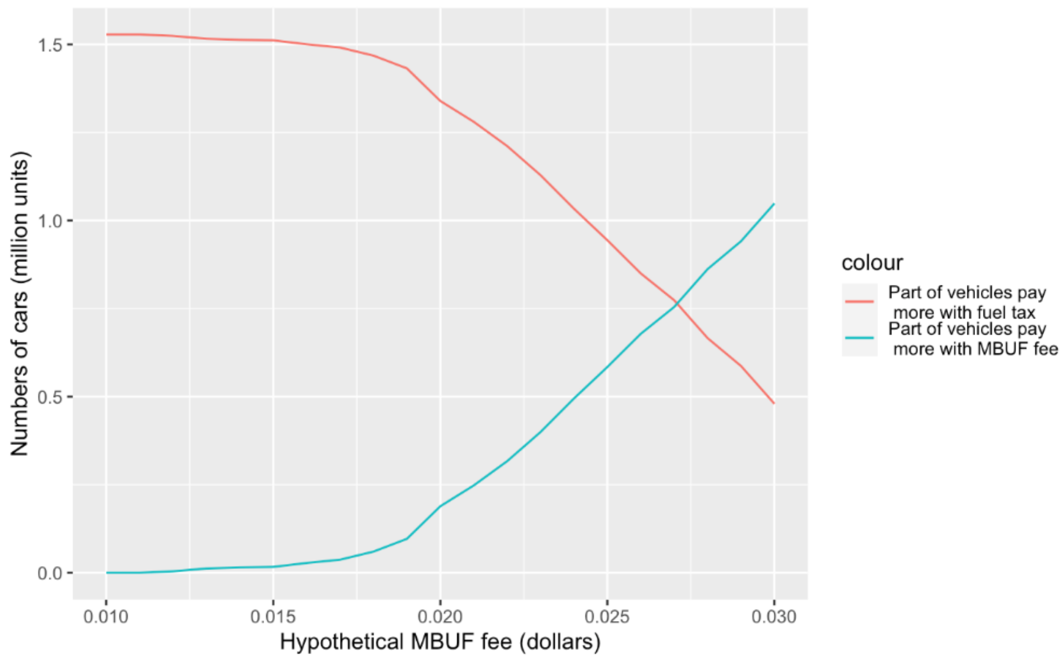

Figure 4: Comparison of fuel tax and hypothetical MBUF fee with combined fuel tax in state level.

Before calculating the county "balance points", we assessed how VMT and fuel economy vary between counties. *Figure 5* shows how these estimates vary between each county and the statewide mean. In urban and suburban counties (denoted by grey borders), VMT and fuel consumption are at, or slightly less than, average. However, in rural areas, the situation differs by county. Rural counties like Schuylkill, Venango, Greene and Northumberland county have an average VMT and fuel consumed significantly below average. However, other rural counties remain at average, or slightly higher. *Figure 6* shows the "balance points" for each county. In most counties, the "balance points" are around ¢2.7 per mile.

As a proof-of-concept for even more granular estimates, we also estimated ZIP code level VMT, fuel economy, and MBUF balance points, for Allegheny County (which includes the city of Pittsburgh in the center, and several suburban and rural areas). In 2014, the county had a mean VMT of 10,100 miles, and estimated fuel consumption of 492 gallons per vehicle. Figure 7 shows that, most urban ZIP code areas in the county (surrounded by bold grey lines), have VMT and fuel consumption below the county average, whereas in suburban and rural areas in the county, VMT and fuel consumption are higher than average. In Figure 8, we see that in some urban areas such as downtown and the east suburbs, the balance points of fuel tax and MBUF are lower than ¢2.75 per mile, whereas in the relatively more rural northeastern part of the county, the balance points are as high as ¢3.25 per mile.

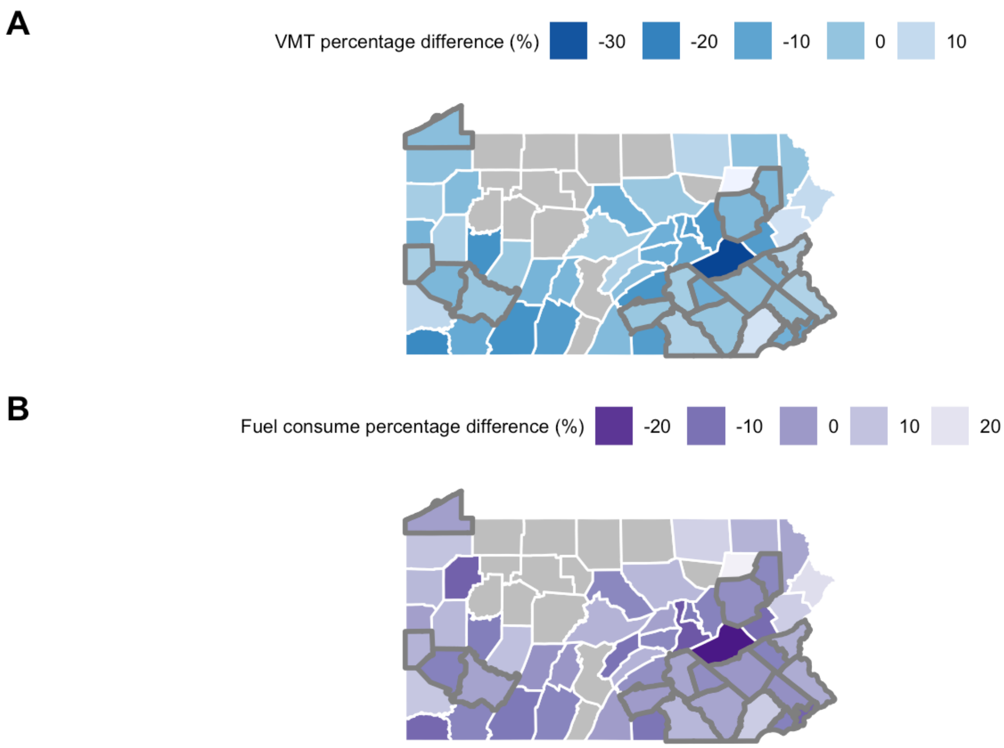

Figure 5: CMT and fuel tax percentage differences in each county in Pennsylvania in year 2014, A: VMT percentage difference (%), B: Fuel consumed percentage difference (%)

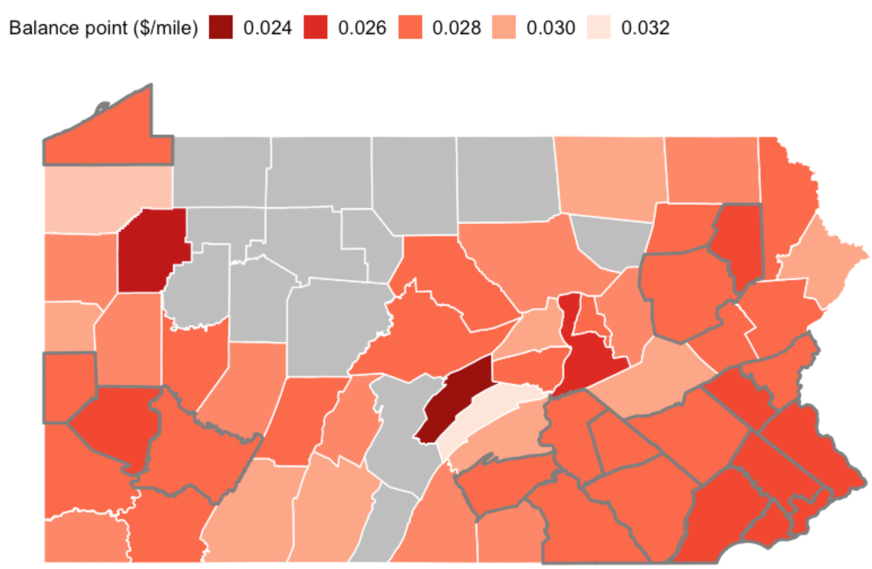

Figure 6: Balance point between MBUF and fuel tax in each county in Pennsylvania.

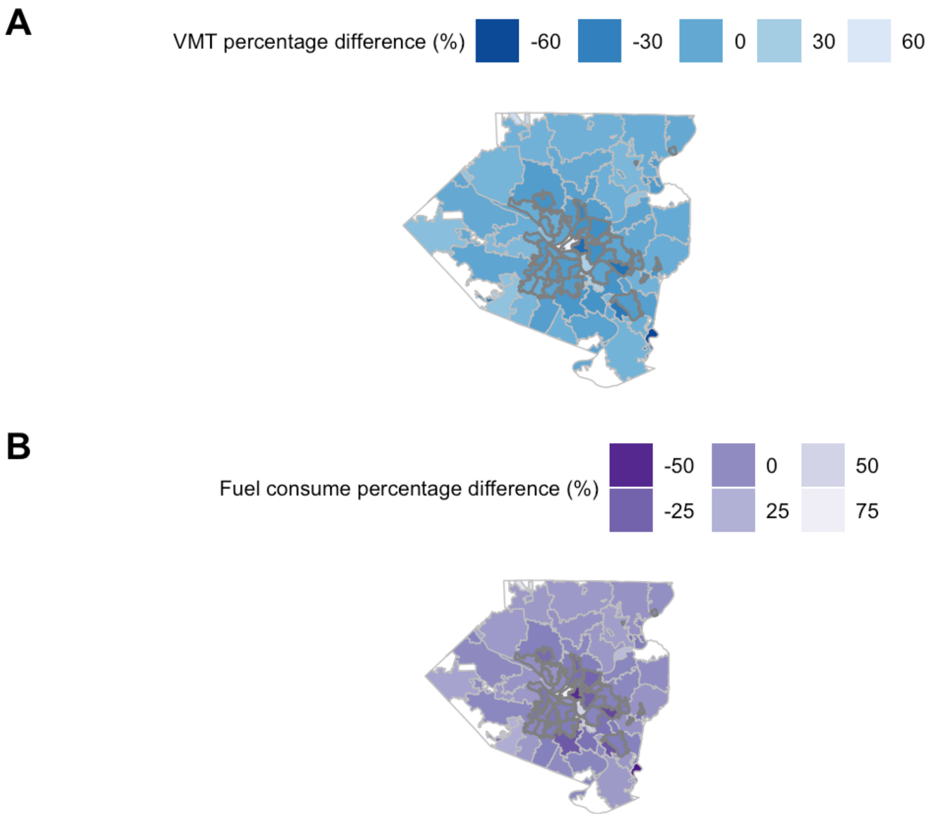

Figure 7: VMT and fuel tax percentage differences in each ZIP code area in Allegheny County in year 2014, A VMT percentage difference (%), B Fuel consumed percentage difference (%).

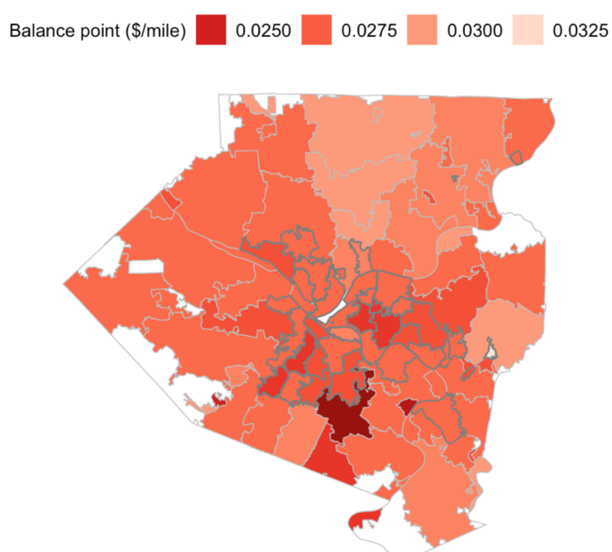

Figure 8: Balance point between MBUF and fuel tax in each ZIP code area in Allegheny County (urban area border in County drawn in bold grey line).

The state, county, and ZIP code level analysis shows the high geographical variation between counties, in estimated annual VMT, fuel economy, and MBUF "balance point". This can also be demonstrated by example at the individual vehicle level. Eight vehicles of four types with high and low VMT have been selected as examples; Table 4 shows how much more or less their owners would spend if transitioning from the current fuel tax to an MBUF at various rates. High fuel economy vehicles with high VMT will pay more with MBUF than a fuel tax. Hybrid vehicles (such as the Toyota Prius) may cost over 150% more with a ¢3 per mile MBUF than a fuel tax. Low fuel economy vehicles (such as a Honda Pilot, or most pickup trucks) will likely have a lower overall annual cost with MBUF than with the existing per-gallon tax. Overall, this emphasizes the point that fuel economy and VMT are both very important in determining whether a vehicle owner would be better or worse off with a MBUF system.

Table 4 Vehicle level analyses show that hybrid electric vehicles may be penalized by MBUFs.

Make/Model	Vehicle Type	VMT (miles/year)	Fuel economy (MPG)	Annual Fuel Tax ($)	Change in cost between fuel tax and MBUF of:		
					¢1/mi	¢2/mi	¢3/mi
2013 Chevrolet Malibu	Sedan	13,277	26	292	-55%	-9%	36%
2013 Ford Fusion	Sedan	33,399	36	541	-38%	23%	85%
2013 Honda Insight	Hybrid Sedan	16,323	42	222	-27%	47%	120%
2010 Toyota Prius	Hybrid Sedan	6,364	50	74	-14%	72%	158%
2008 Honda Pilot	SUV	35,348	18	1159	-70%	-39%	-8%
2011 Toyota RAV4	SUV	8,121	24	196	-59%	-17%	24%
2012 Ford F-150	Pickup	27,162	16	985	-72%	-45%	-17%
2010 Dodge Ram 1500	Pickup	9,266	15	359	-74%	-48%	-23%

7. DISCUSSION AND CONCLUSIONS

Our analyses estimate the "balance point" MBUF rate at the county and ZIP code level by leveraging 100+ million inspection records to assess variations in annual VMT and fleet fuel economy. As a result, we found the "balance point" MBUF to be about ¢2.75 per mile for the state of Pennsylvania, but to vary substantially (from ¢2.4 to ¢3.2 per mile) between counties. From these results, it also appears that drivers in rural counties will generally pay a higher per-mile fee for this balance point to be achieved. This was true in both the county and ZIP code level results. We also show that these analyses can be conducted for individual vehicles, and quantify how hybrid vehicles, and those with higher fuel economy will likely pay more in taxes annually, with a flat MBUF than with per-gallon fuel taxes. Such results would be useful to present to the public in informational materials about such a transition.

Policymakers must consider the implications of potentially penalizing smaller, and higher fuel efficiency vehicles in developing MBUF rate structures, as energy and environmental concerns have led agencies to promote and subsidize them. It would be easy to tailor slightly lower (but non-zero) MBUF fees for such vehicles without compromising the need to gain incremental transportation revenue. There are also large possible equity implications associated with the resulting changes in fees paid per vehicle given urban and rural, vehicle age, ability to live near work, and other differences that will need to be studied more before pursuing such efforts.

What is also evident from these results is that jurisdictions will require to carefully tailor marketing around an MBUF program, to ensure that vehicle owners' perceptions are not swayed by generalized statements about specific types of counties or ZIP codes paying a higher cost. Our results show that the difference between fuel tax and MBUF costs vary significantly based on the individual driving characteristics of each vehicle, and that these broad generalizations are likely to be misleading.

ACKNOWLEDGEMENTS

The authors thank Bernie Elder of CompuSpections, Michael Smith of PennDOT, and Mark Colangelo of Parsons for providing the raw data and approving its use for this project. We thank MBUFA for connections and networking. We also thank the US DOT Mobility21 Center, the US Department of Energy Vehicle Technologies Office (Award DE-EE0008466), and the Department of Civil and Environmental Engineering at Carnegie Mellon University for funding. The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

AUTHOR CONTRIBUTIONS

CY and LL developed algorithms and wrote Python code to estimate vehicle-level fuel economy. ZF developed algorithms and wrote Python code to estimate annual VMT per vehicle and by region. CY and ZF developed a method to combine these estimates assess the impact of MBUFs and estimate the balance point. PSA provided feedback on research methods and conclusions. HSM devised the project, acquired data, and provided critical feedback on methods and analyses. All authors wrote and edited the manuscript.

REFERENCES

1. RSMeans. Historical Cost Indexes. 2020.
2. Dumortier J, Zhang F, Marron J. State and federal fuel taxes: The road ahead for U.S. infrastructure funding. Transp Policy. 2017 Jan 1;53:39–49.
3. The Eastern Transportation Coalition. Findings-MBUF Pilot [Internet]. 2020 [cited 2020 Jul 31]. Available from: https://tetcoalitionmbuf.org/findings/
4. Mileage-Based User Fee Alliance [Internet]. 2020 [cited 2020 Jul 31]. Available from: http://www.mbufa.org/alliance.html
5. Federal Highway Administration. National Household Travel Survey [Internet]. 2020 [cited 2020 Jul 31]. Available from: https://nhts.ornl.gov/
6. Krishnappa V, Matthews HS, Liu Y. Data-Driven Analysis to Support Revised Tire Tread Inspection Standards. Transp Res Rec J Transp Res Board. 2019 Nov 16;2673(11):517–28.

7. Sorensen P, Ecola L, Wachs M. Emerging Strategies in Mileage-Based User Fees. Transp Res Rec J Transp Res Board. 2013 Jan 12;2345(1):31–8.
8. Fitzroy SS, Schroeckenthaler K. Estimating Revenue-Neutral Mileage-Based Fees for Urban and Rural Households in Eight Western States. Transp Res Rec J Transp Res Board. 2018 Dec 9;2672(4):23–31.
9. Methodology - Local Area Transportation Characteristics for Households | Bureau of Transportation Statistics [Internet]. Bureau of Transportation Statistics. 2019 [cited 2020 Jul 31]. Available from: https://www.bts.gov/archive/subject_areas/national_household_travel_survey/methodology
10. Matteson M, Crocker N, Martin-Mahar L, See C. Development of a Fuel Consumption and Mileage Allocation Model to Support Mileage-Based User Fee Research in Washington State. Transp Res Rec J Transp Res Board. 2016 Jan 1;2597(1):11–9.
11. PA DEP I/M [Internet]. Pennsylvania's Vehicle Emission Testing Program. [cited 2020 Jul 31]. Available from: http://www.drivecleanpa.state.pa.us/
12. decodeTHIS: The Universal VIN Decoder [Internet]. [cited 2020 Jul 31]. Available from: https://www.decodethis.com/
13. Home - VINquery.com [Internet]. [cited 2020 Jul 31]. Available from: https://www.vinquery.com/
14. Fuel Economy [Internet]. U.S. Department of Energy. [cited 2020 Jul 31]. Available from: https://fueleconomy.gov/
15. Vehicle API [Internet]. National Highway Traffic Safety Administration. [cited 2020 Jul 31]. Available from: https://vpic.nhtsa.dot.gov/api/
16. U.S. Environmental Protection Agency. Questions and Answers Fuel Economy Testing and Labeling. 2014.