Driver Behavior and Fuel Efficiency*

Karl W. Dunkle Werner†

University of Michigan

Undergraduate Honors Thesis

April 5, 2013

Abstract

Gasoline consumption is an important policy issue, with major impacts on pollution, climate change and global trade. Previous research has focused on vehicle choice and distance traveled as the determinants of total gasoline consumption, treating the fuel economy of the vehicle as a fixed parameter. However, even for the same vehicle, driver behaviors cause large differences in efficiency. I use a rich dataset of naturalistic driving to analyze second-by-second fuel consumption and find that total gasoline use would fall by 17–26% if all drivers behaved like the most efficient individuals.

*Many thanks to the wonderful people who have helped me on this project. First and foremost, my advisor Shaun McRae has been invaluable throughout the process. He provided insightful guidance, a realistic perspective and incredibly patient assistance. Jeffrey Smith, Jeremy Fox, Earl Werner, Ruth Dunkle and Dave Childers have been amazing resources for all of my statistics and economics questions. Any errors, econometric or otherwise, are entirely my own. Stata `.do` files and the LaTeX source files are available upon request.

†University of Michigan, Karldw@umich.edu

Contents

> The car as we know it is on the way out. To a large extent, I deplore its passing, for as a basically old-fashioned machine, it enshrines a basically old-fashioned idea: freedom. In terms of pollution, noise and human life, the price of that freedom may be high, but perhaps the car, by the very muddle and confusion it causes, may be holding back the remorseless spread of the regimented, electronic society.
>
> — J. G. Ballard

1 Introduction

Driving is a core part of the American ethos, with many costs and inefficiencies. According to the US Environmental Protection Agency, the transportation sector accounts for approximately 31% of national greenhouse gas emissions.[1] A quarter of world oil production, approximately 22 million barrels per day, goes to make consumer gasoline.[2]

There are many reasons society and individuals might choose to reduce gasoline consumption, from pollution to financial savings. Burning gasoline releases carbon dioxide (CO_2), which contributes to global climate change. Estimates of the short-term marginal damage of CO_2 are usually \$5–25 per ton.[3,4] Those marginal damage estimates are consistent with a least-cost approach to stabilize CO_2 at atmospheric concentrations of 550 ppm and realize an average warming of 2.9 °C.[5] If the governments of the world decide to set tighter limits, e.g. 450 ppm, that decision implies a higher estimate of damages and engenders higher costs of abatement. For reference, current CO_2 concentrations are 396.8 ppm.[6] Damages of \$5–25/ton of CO_2 translate to 4.8–24¢ per gallon of gasoline.[7]

Beyond global warming, burning fuel has a number of undesirable effects, both environmental and economic. Automobiles burning gasoline and diesel are

---

[1] US EPA, OAR, Climate Change Devision (2013). Inventory of US Greenhouse Gas Emissions and Sinks: 1990–2011.
http://www.epa.gov/climatechange/ghgemissions/usinventoryreport.html

[2] US Energy Information Administration (2011, 2012).
http://www.eia.gov/tools/faqs/faq.cfm?id=307&t=11
http://www.eia.gov/cfapps/ipdbproject/IEDIndex3.cfm?tid=5&pid=5&aid=2

[3] In this case, "short term" means within the next 100 years, a brief period for the global climate but a very long one for economists.

[4] Estimates of future damages are discounted to present values. Indeed, most of the variance in estimated damages is from the assumed discount factor of future generations' utility rather than disagreement on the costs of damages (Metcalf, 2009).

[5] 550 parts per million by volume of CO_2 equivalent: mainly carbon dioxide but also including methane, nitrous oxide, HFC-23, HFC-134a and sulfur hexafluoride.

[6] Tans, P. (2013, March). Trends in Atmospheric Carbon Dioxide.
http://www.esrl.noaa.gov/gmd/ccgg/trends/

[7] US Energy Information Administration (2012).
http://www.eia.gov/tools/faqs/faq.cfm?id=307&t=11

responsible for problems of local air pollution, smog and particulates, estimated to cause $71–277 billion of damage in the US each year (Muller and Mendelsohn, 2007). Oil and gasoline consumption have a real, negative impact on the US and other countries. It is worth examining ways to reduce the amount of gasoline consumed while retaining the convenience offered by automobiles. There are three approaches to reduce gasoline consumption: change vehicles, drive fewer miles or modify driving behavior. I will examine the variation in driving behavior and find enormous heterogeneity between drivers.

Driver behaviors impact gasoline consumption substantially through their behaviors and driving decisions. Previous research has focused on two major areas: buying durable goods and driver response to gasoline prices. Both vehicle purchase and the gasoline price elasticity are certainly important considerations. I believe that my approach, focusing on the role of driver behavior, has been under-explored in the existing literature.

A very detailed dataset, with six weeks of observations from 108 drivers, allows me to examine driver behavior and fuel consumption in a novel way. Using second-by-second recordings, I calculate the effects drivers have on their vehicles' fuel consumption. I estimate that if every driver behaved like the most efficient driver, the fuel savings would be 17% for highway driving and 26% for city driving. These savings are as large as removing a fifth of all vehicles from the road.[8] Changes of that magnitude would be important on the microeconomic level, reducing household gasoline expenditure, and at the macroeconomic level, reducing US oil consumption and imports.

In Section 2 I discuss the data collection, cleaning and calculation procedures. Section 3 provides a theory of gasoline consumption and details a model of driver heterogeneity. Section 4 discusses results and Section 5 provides a conclusion.

2 Methods

2.1 Data Collection

The University of Michigan Transportation Research Institute (UMTRI) used records from the Michigan Secretary of State to select a random sample of drivers in southeast Michigan (LeBlanc et al., 2010). The UMTRI researchers provided a vehicle to each of the 117 study participants. The vehicles were almost identical versions of the Honda Accord SE from the 2006 or 2007 model year. (The only differences between the Accord 2006 and 2007 model years were minor cosmetic changes.) The US Environmental Protection Agency's fuel

[8]In fact, removing a fifth of all vehicles would result in a smaller decrease in fuel use in the general equilibrium because drivers would use the other 80% of cars more.

ratings are the same for the two model years, at 13.1 liters per 100 kilometers (L/100km) or 18 miles per gallon (mpg) for city driving and 9.05 L/100km (26 mpg) for highway driving.[9] The vast majority, more than 90%, of driving occurred within southeast Michigan, a relatively flat region that contains urban and rural areas.

UMTRI conducted the study with the primary purpose of testing a crash warning system. The system was installed in every car and would beep when it detected a dangerous situation. The crash monitoring system provided infrequent haptic and audio warnings to drivers, but did not control acceleration, braking or steering. I will assume that the crash warning system did not significantly impact the distribution of drivers' fuel use behavior. Drivers may have behaved differently in the UMTRI vehicles than they would in their own. The change in behavior decreases the external validity of the following analysis to the extent that the behavioral changes systematically impact fuel consumption.

The cars were extensively instrumented, with global positioning system (GPS) receivers, a compass, an external thermometer, internal and external cameras, a crash warning system, a radar system and systems to record the output of the vehicle computer. The vehicle itself reported transmission gear, engine speed, air conditioner use, windshield wiper use and, most importantly, fuel. The data from the onboard computer were recorded ten times per second, and one observation per second was extracted for analysis. Fuel consumption was measured at a resolution of 0.2 milliliters (mL) or 0.00676 fluid ounces.

Of the 117 drivers, 9 did not follow the experimental protocols. I have excluded them from further analysis, using only the 23,177,377 observations from the 108 compliant drivers. All estimates have been appropriately weighted to account for the amount of time each driver spent driving. In the regressions that address highway and non-highway driving separately, I have weighted by the inverse of the total time each driver spent on that road type.

Possible sources of vehicle heterogeneity are differences in maintenance history, tire pressure or manufacture. Though it is possible to include dummy variables for each vehicle in the model, doing so is problematic because most drivers used only one vehicle and the estimated vehicle coefficients would unduly influence the estimated driver effects. Recall that the EPA fuel efficiency ratings for these vehicles are identical and the vehicles were well maintained by UMTRI. Henceforth, I neglect differences between vehicles, making it possible to directly examine the influences of each driver's choices. As a check for robustness, I tested regressions with errors clustered by driver and by vehicle and found the results very similar. Other factors, including cargo weight, should not be thought of as vehicle heterogeneity, but are instead a component of the driver effect.

[9] EPA (2013). Compare Side-by-Side Fuel Economy.
http://www.fueleconomy.gov/feg/Find.do?action=sbs&id=21962
http://www.fueleconomy.gov/feg/Find.do?action=sbs&id=23569

The next data consideration is the fuel measurement system. The fuel gauge reported cumulative fuel consumption of each trip with a precision of 0.2 mL. The measurements are therefore multiples of the form $x = 0.2n$ mL; $n \in \mathbb{Z}$. The true fuel use corresponding to a measurement of x could be anywhere in the range $[x - 0.1,\ x + 0.1]$, which is somewhat coarse for the second-by-second calculations detailed below. Most of the fuel use is on the order of 0–1 mL/s.[10] I further assume that for any measured x, real fuel use is approximately uniformly distributed in the range $[x - 0.1,\ x + 0.1]$. Though this assumption obviously does not hold at $x = 0$, it seems reasonable over the remaining range of fuel use. While the car is running, fuel use is non-zero. However, for 1,763,246 periods, representing 9.86% of the remaining data, the difference in fuel use between two seconds is observed to be 0.0 mL. If possible, I would like to avoid losing these data, one of the factors that informs the choice of units in Section 2.4.

2.2 Data Cleaning

Originally, there were 23,177,377 one-second observations in the study, excluding the non-compliant participants.[11] Mechanistically, the first and last one-second observations of each trip are lost to calculate the differences between adjacent observations ($n = 49{,}550$). Next, I drop observations for which the vehicle was moving slower than 5 kilometers per hour (kph) or 3.11 miles per hour (mph) or speed data were missing, 4,590,558 seconds and 348 seconds, respectively. Dropping speed as in this way *is* fundamentally non-random, and the low speed behavior of a vehicle can be important.[12] However, most speed and acceleration decisions occur when driving at some positive speed. Data about a non-moving vehicle are problematic for models that examine the relationship between speed, acceleration and fuel use.

Anomalously high values of fuel consumption, measured in liters per 100 kilometers (L/100km), can result from using very large amounts of gasoline or traveling very small distances. All of the analysis that follows depends on a valid measure of fuel consumption. For very small differences in distance, the calculated fuel consumption becomes unreliable (approaching $+\infty$). Despite these concerns, L/100km is still a more appropriate unit than miles per gallon (mpg), as detailed in Section 2.4. The graph in Figure 1 shows the observations with a very small differences in distance, after dropping observations with reported speed below 5 kph. The vertical line indicates represents the difference in distance (over one second) that corresponds to a speed of 5 kph. To ensure the validity of the fuel consumption calculation, I am dropping observations with a

[10] The possible use for a measured value of $x = 0$ is the range $(0,\ 0.1]$.

[11] Before cleaning there are 6438 driver–hours of data.

[12] For example, hybrid vehicles are able to substantially reduce their fuel use by switching to an electric motor while moving at low speed.

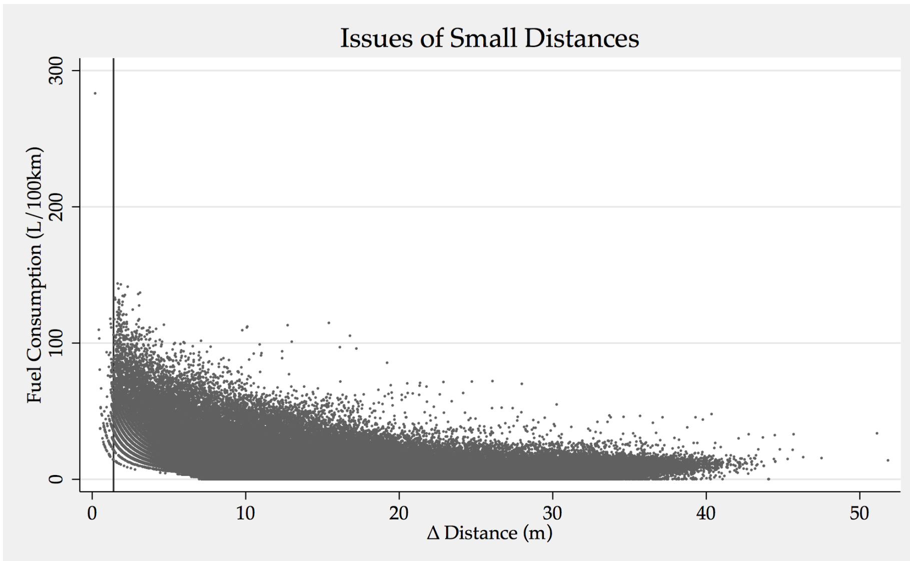

Figure 1: The graph above represents the points with *reported* speed above 5 kph. To avoid the issues of invalid fuel consumption, I will drop all points to the left of the vertical line, which represents a calculated speed of 5 kph.

difference in distance smaller than 1.389 meters (4.557 ft) per second. The observations plotted in that graph are a random sample of 1% of the data, though other random draws yield similar results.

The UMTRI dataset includes a speed variable, measured by the cars' speedometers. Speed is the derivative of distance traveled with respect to time, expressed as $v = \frac{dx}{dt}$ in standard notation.[13] Therefore, I can calculate speed from distance traveled. For the vast majority of observations, the two measures agree well, but there is a spurious trend, shown in Figure 2. That figure graphs the difference between speed, as provided by UMTRI, and the derivative of position. Looking at the difference, some noise is expected, but the positive linear trend is problematic. Therefore, I am dropping all cases where calculated speed and provided speed differ by more than 5 kph (3.11 mph). The diagonal line of Figure 2 has a slope of 0.50. Figure 2 is a random subset of 0.1% of the dataset, though the results are robust to different random draws.

Next, there is cause to worry about anomalous position data: latitude, longitude, altitude and change in altitude. A few observations for latitude and longitude are obviously wrong. In lieu of a formal analysis based on global information system (GIS) calculations, I drop all data points that are outside

[13] More precisely, velocity is derivative of position with respect to time and speed is the absolute value of velocity. I follow the physics convention of using v as the symbol for both.

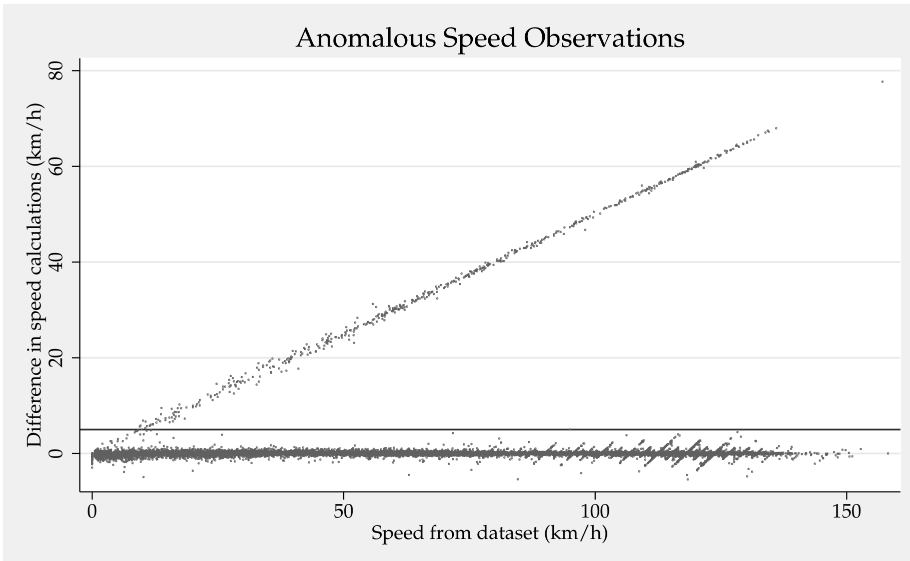

Figure 2: Some values of speed provided in the dataset do not agree with the speed calculated as the derivative of position. Values that differed by more than 5 kph (horizontal black line) were dropped.

the contiguous 48 states.[14] For the purposes of this study, data outside of the range [-125, -70] °E and [24, 49] °N are dropped.[15,16]

Though it is possible for a GPS receiver to calculate altitude directly, the UMTRI researchers gathered elevation data by matching the recorded two-dimensional GPS data with GIS maps. For the most part, these matched data agree with observations on Google Earth for the same latitude and longitude, but there are a few anomalies. The minimum altitude observation of the original data is miles below sea level and the maximum is taller than Mt. Everest. Because change in altitude *is* a component of the model, I must drop erroneous values of altitude and Δaltitude. To apply some logical bounds, the highest road in the US has an elevation of 4,345 m (14,255 ft)[17] and the lowest -86 m (-282.2 ft), so I drop any observations outside of the range [-86, 4345].[18] Fi-

[14] While I do not include latitude or longitude directly in any of my models, impossible observations are always a cause for concern. A more sophisticated approach would use GIS software to eliminate any impossible data points: those in the middle of the ocean or too far from preceding observations.

[15] Southeast Michigan is within driving range of Ontario, Canada. Any trips to southern Canada are retained, since most destinations in eastern Canada are south of the 49$^{\text{th}}$ parallel.

[16] Coordinates obtained from Google Maps (2013). https://maps.google.com/

[17] Dailey, D. (2007). Discover America's Highest Road. http://suite101.com/article/discover-americas-highest-road-a22242

[18] As with the latitude and longitude data, a far more thorough way to clean the altitude data would involve GIS.

Table 1: Excluded data

Variable	Criteria to drop	No. dropped[a]
Differencing[b]	First and Last	49,550
Speed[c]	< 5 kph or missing	4,590,906
Distance difference[d]	< 1.38889 m	24,616
Speed agreement	$\lvert \frac{dx}{dt} -$ speed$\rvert \geq 5$ kph	460,445
Latitude	< 24 or > 49 °N	209,956
Longitude	< -125 or > -70 °E	76
Altitude	< -86 or > 4345 m	533,794
Δ Altitude	$\lvert \frac{dh}{dt} \rvert > 8$ m/s	674
Sum		5,870,017 (25.3%)

[a]The number dropped for each criterion, after the previous exclusion criteria.
[b]Because of differencing, the first and last observations of each trip are lost.
[c]Only 348 observations are missing speed data.
[d]A distance of 1.389 m in one second is equivalent to traveling at 5 kph.

nally, I check that the data for change in altitude are reasonable. A road grade of 30% is quite steep, representing a rise of 30 m for every 100 m traveled. Slopes this severe are rare, and almost never found on highways. Assuming a maximum non-highway speed of 26.8 m/s (60 mph), the maximum elevation change should be no more than 7.7 m/s. I exclude absolute values of Δaltitude greater than 8 m/s (26.3 ft/s). In total, problems with reported position force me to drop 744,400 observations. The final cleaned dataset has 17,264,635 observations from 108 drivers, approximately 70.2% of the original data collected.

2.3 Descriptive Statistics

Descriptive statistics for the cleaned data are listed in Table 2. Notable in the table are the large standard deviations in many of the variables. Speed, acceleration and fuel consumption all have large standard deviations, illustrating the wide variety of drivers, behaviors and driving circumstances in the study. After dropping the anomalous values noted in Section 2.2, the speed reported in the dataset and speed calculated as a derivative of distance agree very well.

Driver characteristics are included in the dataset and this summary table, though the nature of the analysis described in Section 3.3 precludes me from including the sex, age or income variables in my regressions. These variables would be perfectly collinear with the individual driver coefficients I include in the models.

Extrema of latitude, longitude and altitude are excluded from the summary table for privacy reasons. These data could theoretically identify a point in space, which would compromise the privacy of the study participants.

Table 2: Summary statistics, weighted by total driving time

Variable	Mean	S. D.	Min.	Max.
Latitude (°N)	42.217	0.966	†	†
Longitude (°E)	−83.675	1.115	†	†
Altitude (m)	205.23	54.23	†	†
Δ Altitude (m/s)	−0.001	0.265	−7.984	7.946
Odometer (km)	34,809	15,825	9,545	84,246
Engine speed (RPM)	1752	539	0	6915
Outside temperature (°C)	14.38	10.84	−17.22	46.11
AC use (0/1)	0.215	0.411	0	1
Trip duration (min)	15.32	18.20	0	260.75
Trip distance (km)	14.09	25.45	0	495.22
Speed[a] (kph)	69.98	34.86	5	214.22
Distance rate[b] (kph)	69.97	34.85	5	214.24
Acceleration (kph/s)	−0.016	2.281	−25.01	16.91
Fuel rate (mL/s)	1.778	1.316	0	26.9
Fuel consumption (L/100km)	11.775	12.468	0	541.397
Sex (M=0, F=1)[c]	0.405	0.491	0	1
Household median income[c, d] ($)	64,354	26,925	19,141	155,282
Household mean income[c, d] ($)	77,815	32,518	24,188	189,285
Total time driven[c] (hr)	52.33	19.55	15.00	97.66
Highway time[c] (hr)	12.90	8.237	0.2892	36.17
Non-highway time[c] (hr)	39.43	17.73	9.558	84.94

[a]Speed, provided in the dataset.
[b]Speed, calculated as the derivative of distance.
[c]Driver characteristics not weighted by driving time.
[d]Based on participants' ZIP codes, not survey data.

2.4 Choosing Units for Fuel Use

There are two widely used standards for measuring the fuel use of a vehicle, miles per gallon (mpg) and liters per 100 kilometers (L/100km), respectively called "fuel economy" and "fuel consumption". Unlike most conversions between metric and imperial units, mpg and L/100km are not linear transformations of one another. Miles per gallon is a measure of the distance that can be traveled for a given quantity of fuel, while liters per 100 kilometers is a measure of the fuel required to travel a certain distance. If the quantity of fuel used is small, the calculation of mpg will be very sensitive to errors in fuel measurement. Similarly, if the distance traveled is small, the calculation of L/100km will be sensitive to small errors in distance measurement.

Fortunately, distance traveled can be measured *very* accurately, so the error in distance is small. However, as mentioned before, the fuel gauge measurement precision was only 0.2 mL, implying a root mean squared error of $(10\sqrt{3})^{-1} \approx 0.0577$ mL.[19]

Both fuel use and distance traveled are recorded cumulatively over the entire trip. Fuel consumption is the difference in fuel use divided by the difference in distance traveled:

$$FuelCons_{dts} \equiv \frac{fuelused_{dts+1} - fuelused_{dts-1}}{10\,(dist_{dts+1} - dist_{dts-1})} \tag{1}$$

Where $FuelCons$ is fuel consumption, measured in L/100km. The subscripts d, t and s are a hierarchical ordering of driver, trip and second, respectively. Because $fuelused$ is measured in mL and $dist$ in kilometers, a factor of 10 is necessary in the denominator to convert $FuelCons$ to L/100km.

$$\begin{aligned} FuelCons &\approx \frac{d\,fuel}{d\,dist} = \left(\frac{d\,fuel}{d\,time}\right) \cdot \left(\frac{d\,dist}{d\,time}\right)^{-1} \\ &= \left(\frac{fuel_{s+1} - fuel_{s-1}}{2\Delta\,time}\right) \cdot \left(\frac{2\Delta\,time}{dist_{s+1} - dist_{s-1}}\right) + O\left((\Delta\,time)^2\right) \end{aligned}$$

Equation 1 is an approximation for the derivative of fuel use with respect to dist traveled, which may be expressed: The center difference formula of Equation 1 is more accurate than the forward difference:

$$\left(\frac{fuelused_{s+1} - fuelused_s}{10\,(dist_{s+1} - dist_s)}\right)$$

Or the backward difference:

$$\left(\frac{fuelused_s - fuelused_{s-1}}{10\,(dist_s - dist_{s-1})}\right)$$

Because both the forward and backward difference calculations have errors on the order of $(\Delta\,time)$ rather than $(\Delta\,time)^2$.

One way to decrease the sensitivity of fuel consumption to small errors would be to average over longer time periods. In a longer time, the car would travel further and consume more fuel. Therefore, the errors in measurement would have smaller effects.[20] The downside of averaging over a longer time period is the loss of granularity. Because drivers make acceleration and speed decisions on a second-by-second basis, I would like to preserve as tight a focus as possible. I have tested different averaging periods and found the estimated results to decrease slightly in magnitude with longer averaging periods.

Figure 3 and Figure 4 represent the distributions of fuel consumption for highway and non-highway driving, respectively. Most of the observations in

[19] Assuming a uniform distribution within each 0.2 mL interval.
[20] Assuming that the errors in measurement of fuel use and distance are homoskedastic, a reasonable guess for GPS systems and fuel gauges.

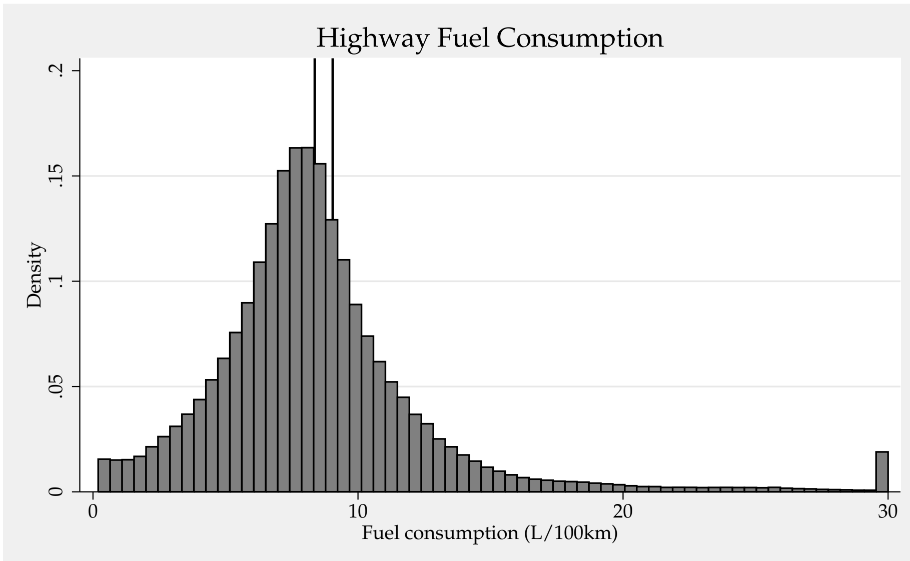

Figure 3: Fuel consumption, calculated on a second-by-second basis, for highway driving. The EPA estimate of 9.05 L/100km is marked by the right vertical line. Average performance, indicated by the line on the left, is only slightly more efficient than EPA estimates.

both graphs occur *below* their respective EPA estimates for highway and city driving, indicating that for most seconds the car is operating more efficiently than the EPA estimate. In Figure 3 the EPA estimate is the line on the right and average consumption is line on the left. In Figure 4 the EPA estimate and average fuel consumption lines overlap. The high efficiency calculated here is an artifact of using speeds above 5 kph, which avoids fuel-intensive idling.

3 A Theory of Fuel Economy Differences

3.1 The Economics of Driving

Drivers do not consume gasoline. At least, they have no interest in consuming gasoline directly. Drivers really consume *transportation*, the ability to get where they want to go. In turn, the demand for transportation translates to a demand for driver–kilometers (also called vehicle–kilometers). Drivers only consume gasoline to move their vehicle where they want to travel.

Of course, driving a car comes at some cost. Each kilometers driven requires both gasoline and the driver's time, as well as wear and tear on the

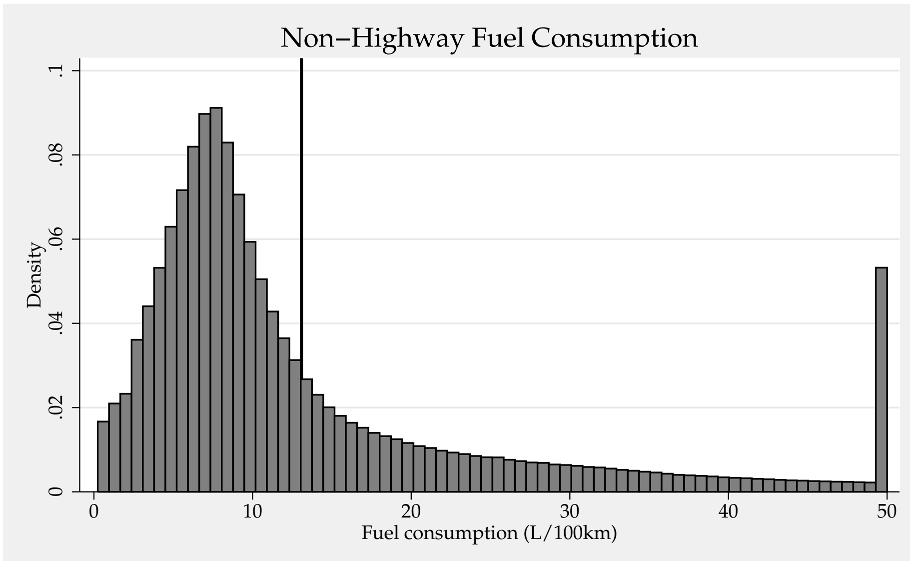

Figure 4: Fuel consumption, calculated on a second-by-second basis, for non-highway driving. Notably, the bulk of the graph occurs at consumption levels *below* the EPA estimate, marked by a vertical line at 13.1 L/100km, indicating that for most seconds the car is operating more efficiently than the EPA estimate. Average performance is indicated by a vertical bar that overlaps the EPA estimate.

vehicle.[21] The conventional view on driving is that a driver has some fixed cost per mile, and then decides how many miles to drive (how much transportation to consume). In fact, the fuel costs depend on the driver's behavior, both in route planning as well as speed and acceleration choices. Fuel consumption is a variable the drivers choose implicitly through their behavior. The price drivers pay is a function of both time and fuel consumption, and these are often trade-offs of one another. Drivers who value their time highly may drive faster and accelerate more aggressively.[22] Indeed, Wolff (2012) attempts to calculate the value drivers place on their time by measuring highway speeds. People may also derive utility from driving fast or accelerating rapidly, an ancillary benefit to the transportation they consume.

Safety is another concern in drivers' speed choices. Consumers of transportation may choose to drive slower if they believe that doing so is less risky. Preferences for safety may also affect the drivers' vehicle purchase decisions,

[21] The vehicle has a substantial upfront capital cost, and the lifespan of the car is largely determined by use (kilometers driven) rather than the passage of time.

[22] The most efficient speed for these Honda Accords is 98 kph (61 mph), faster than city driving but slower than most highway driving (LeBlanc et al., 2010). Accelerating heavily to reach the most efficient speed is still a fuel-intensive process.

possibly pushing safety-conscious drivers toward heavier vehicles that are both sturdier and less efficient. Jacobsen (2012) found that each one-mpg increase in the corporate average fuel economy (CAFE) standard increases expected US fatalities by 149 per year. Put simply, a more efficient CAFE standard pushes manufacturers toward lighter vehicles, which often provide less protection in crashes.

There is a behavioral economics perspective as well; drivers may plan to drive efficiently, but in the moment may drive faster or slower than they planned. Consumers may also undervalue their fuel savings (overvalue their time) at the moment they make their decisions. In fact, many drivers have a very poor idea how their behaviors impact their fuel consumption.

The decisions drivers make can be broken into three useful categories, divisions that mirror the time scales over which the decisions occur: vehicle demand, demand for vehicle–miles traveled and demand for vehicle performance. The first two of these have been widely studied in the economics literature. Vehicle purchase decisions were not part of this study, therefore participants did not make any strategic decisions. Demand for vehicle–miles traveled also incorporates other characteristics of the route, including hills, traffic and road type. Second-by-second driving decisions that affect vehicle performance include speed, acceleration and air conditioning use. Drivers have continuous control over these three variables and on a second-by-second basis how they want their vehicles to perform.

The longest time scale of decision making is the choice of vehicle, the discrete choice for durable goods. *A priori*, I believe that driver choices in one time scale are correlated with choices in other time scales. The driver who chooses to buy a sports car may tend to drive aggressively, even in a different vehicle. The UMTRI researchers avoided this confounding problem by assigning the same type of vehicle to every driver.

Vehicle maintenance decisions are made on the same time scale as purchase decisions, and for simplicity I group both together. This study is particularly useful in that each driver was given a well-maintained, identical vehicle and allowed to drive naturally for a substantial period of time, avoiding the issue of vehicle purchase. (See section 2 for more information about the vehicles used in the study.)

Much of the economics literature has focused on the discrete choice demand for durable goods as the major determinant of fuel consumption (and fuel efficiency). When shopping for a car, drivers decide what make and model to buy, weighing a number of vehicle parameters. From the inconsequential, like paint color, to more relevant questions of carrying capacity, engine power and fuel economy, drivers choose the characteristics that matter to them. Assuming a rational actor model, drivers also include fuel efficiency and projected fuel costs in their evaluation of each vehicle. Higher fuel economy, more efficient vehicles often require more advanced technologies and are therefore more expensive, all else equal. Drivers weigh the current cost of capital (an efficient car) against expectations of future gasoline costs. Sivak and Schoettle (2012) discuss the importance of vehicle choice, noting that the least efficient car of model year 2011 gets 21 L/100km (11 mpg). While the most efficient model achieves 6.5 L/100km (36 mpg).[23]

The net present cost of a vehicle depends on the initial cost of the car, the expected gasoline costs, the depreciation rate of the car and the discount rate of future costs and savings. Several studies have investigated the implicit discount rates inferred from consumers' automotive purchases. Very high discount rates would indicate little regard for the future, pushing the consumer to buy a cheap, inefficient car now and pay higher gas costs later. In fact, consumers appear to have foresight in their vehicle purchases. Espey and Nair (2005) found *very* low implied discount rates, approximately 4 percent. Their analysis used contemporaneous gasoline prices of \$1.50–2.00. Current gasoline prices of approximately \$3.50–4.00 may influence consumers to discount differently. More recently, Busse et al. (2013) examined the prices paid for new and used cars of different efficiencies and found that consumers discount at rates similar to auto loan interest rates. Discounting at the rate of borrowing is the rational behavior, indicating that consumers are not myopic.[24]

In another recent study, Allcott and Wozny (2012) approach the question from a different angle, setting the discount rate according to consumers' borrowing costs (or savings returns) and calculating how buyers behavior compares to the theoretical interest rate. Allcott and Wozny find that consumers value future fuel savings 26% less than they should, given their discount rates.

Of course efficiency and cost minimization are not the only, or even the most important factors in an automotive purchase. For many drivers, the size, shape, carrying capacity, seating capacity, safety, luxury or performance of a car will be more important than the fuel efficiency, costs of gasoline or upfront vehicle cost.

Vehicle maintenance also occurs on long time frames, months to years. According to the Environmental Protection Agency, performing maintenance on a severely out-of-tune car will improve fuel economy by approximately 4%.[25,26] The vehicles used in this study were well maintained and thoroughly checked between drivers (LeBlanc et al., 2010).

[23] Including only traditional gasoline internal combustion engines, not hybrid, electric or alternative fuel vehicles.

[24] If consumers pay for a car from savings, the appropriate discount rate is the interest they would earn on those savings had they forgone their vehicle purchase.

[25] In some rare cases vehicle maintenance can improve fuel economy by 40% or more, particularly repairing a broken oxygen sensor.

[26] EPA (2011). Gas mileage tips – driving more efficiently.
http://www.fueleconomy.gov/feg/drivehabits.shtml

The next time step after vehicle purchase is route choice, demand for vehicle miles traveled, that occurs on timescales of minutes to days. Drivers decide where they want to go, then pick routes to maximize their utility.[27] The benefit a route offers is a function several related factors, including time of day, traffic, scenic views, duration of the trip, fuel spent and others.

Drivers may employ "trip chaining" as a route planning technique to reduce total travel time and fuel consumption. Under a trip chaining scheme, drivers will travel to several destinations in a row rather than returning to their home between each stop. This behavior often decreases the total distance traveled, and therefore the gasoline used. Fuel usage depends strongly on route choice. Perhaps the most obvious factor is the distance traveled, but there are other important aspects. In Section 3.2 I discuss how climbing and descending hills increases fuel consumption. Similarly, a route with many stop lights or stop signs will increase fuel consumption, all else equal, because each time the car has to accelerate, the engine consumes a substantial amount of fuel.

The final and finest time step involves driving choices, demand for vehicle performance, including how fast to drive and how hard to accelerate. These choices are made on time scales ranging from a fraction of a second to several minutes. Obviously, the choices of automobile and route affect operational decision making; a van cannot accelerate like a sports car and no one drives 30 kph (19 mph) on the expressway.[28] By using a naturalistic design, the study was able to control vehicle choice and collect very rich data on demand for vehicle–miles and vehicle performance.[29] Study participants were given a car and allowed to drive as they preferred for six weeks.

Researchers have conducted other naturalistic driving studies, but these are limited to small numbers of drivers traveling along fixed routes. Evans (1979) and Lenner (1995) are examples of studies that allowed drivers to use instrumented vehicles, but only along controlled, predetermined routes. Ishiguro (1997) conducted a similar study in which drivers drove heavy vehicles. The present study conducted by UMTRI is a useful opportunity because of the large number of drivers ($n = 108$), the freedom the drivers were given and the level of detail in the data.

Evans (1979) allowed drivers to make speed and acceleration decisions normally, but used a small set of routes and a small sample of drivers. Collected

[27] In reality, drivers may choose their destination based on their route, for example the driver who takes a scenic drive and decides to stop for lunch. In any case, drivers pick a route, then make vehicle performance decisions along that route.

[28] Except in congested traffic.

[29] The term "naturalistic driving" refers to a study where drivers are allowed to drive and make decisions as they wish, while researchers maintain some control over drivers' actions. "Naturalistic" is distinct from natural driving, where data is recorded about drivers as they behave normally. Studies that examine patterns of natural driving are limited. in that they do not follow individual drivers, and cannot control for vehicle choice.

three decades earlier than the UMTRI study, Evans' data were not nearly as precise or fine-grained. Because I have access to drivers' second-by-second decisions, it is possible to draw statistically powerful inferences from detailed models.

Evans et al. investigated the tradeoff between fuel consumption and trip time and found that a 1% increase in trip time (and therefore a 1% reduction in average speed) caused a 1.1% increase in fuel consumption. However, he noted that highway driving occurs above the most efficient speed, while most city driving occurs below the most efficient speed.

3.2 The Physics of Driving

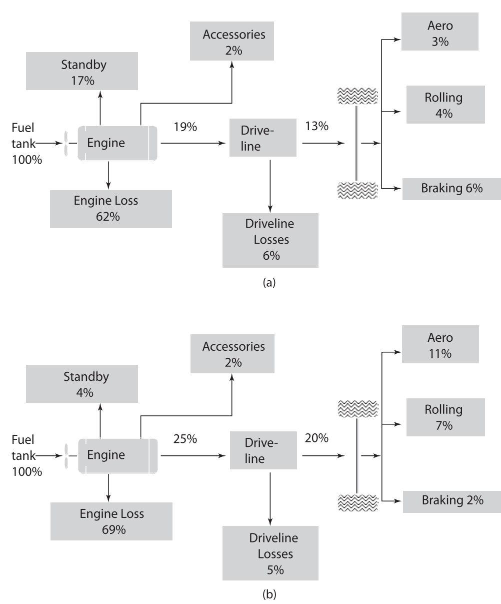

Figure 5: Power flows for (a) city and (b) highway driving. Note the large standby losses for city driving and the large aerodynamic losses on the highway. I excluded the standby losses from my analysis by dropping observations with a speed below 5 kph.

Source: Transportation Research Board (2006). Tires and Passenger Vehicle Fuel Economy.

To build a model of the physics of driving, one first needs to understand the power flows in an automobile. I define P_{load} to be total vehicle load, neglecting minor effects like wind and road curvature (Ross, 1997). The car provides a power P_{load} to the wheels:

$$P_{load} = P_{tires} + P_{air} + P_{inertia} + P_{accessory} + P_{hill} \tag{2}$$

Where P_{tires} is the power used to overcome rolling resistance, P_{air} is the power used to overcome air resistance and $P_{inertia}$ is the power used to accelerate the car. $P_{accessory}$ is the power consumed by accessories, notably air conditioning. P_{hill} is the power to move the automobile up a slope, or the power reclaimed as the vehicle moves down.

P_{tires}, P_{air} and $P_{accessory}$ are always non-negative, as rolling resistance, air resistance and accessories can only cost energy.[30] $P_{inertia}$ and P_{hill} can be negative if the car is slowing or descending a hill, respectively.

However, there are additional inefficiencies in the engine and drive train, as illustrated in Figure 5. The energy contained in the fuel burned is much less than the energy delivered to the wheels ($P_{fuel} > P_{load}$). Instead, the engine and drivetrain create thermodynamic and mechanical inefficiencies, some of which are inescapable features of heat-based engines. Ross (1997) provides a more involved discussion of heat engines, pressure–volume charts and thermodynamic work than would be appropriate here.

It is worth keeping in mind that power is simply the time-derivative of energy; $\frac{dE}{dt} = P$. A fuel tank is full of energy stored as gasoline, and one liter of gasoline contains approximately 8.787 kilowatt-hours (kWh) of energy.[31] A gasoline fuel rate of 1 mL/s represents approximately 31.6 kW of chemical power.[32] Power is measured in units of energy per time while fuel consumption is measured in energy per distance. Therefore, the factor relating power and fuel consumption has units of time per distance, or the reciprocal of speed.

$$Fuel\ consumption = \frac{energy}{distance} = \frac{energy}{time} \cdot \frac{time}{distance} = \frac{power}{speed} \tag{3}$$

Therefore fuel consumption is a function of power divided by speed.

3.2.1 Air Resistance

P_{air} is a function of the size (A) and shape (C_D) of the automobile, the density of air (ρ) and the cube of speed (v^3) (Ross, 1997).[33]

[30] While it is technically possible that wind pushes a car forward, the typical speeds of wind are small relative to the typical speeds of automobiles. I will neglect this technicality.

[31] Pure gasoline contains about 9.7 kWh/L, but the addition of various additives lowers the volumetric energy content. The total energy also fluctuates seasonally with different fuel blends.
US EPA: Office of Air and Radiation (1995). Fuel Economy Impact Analysis of RFG.
http://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100B3FL.txt

[32] 1 mL/s of gasoline also represents 42.4 horsepower. It turns out that the output of one horse over a sustained period is approximately 1 horsepower (Stephenson and Wassersug, 1993).

[33] Other equations are appropriate for low speeds, where air flow is said to be "laminar" rather than "turbulent". Because I include velocity flexibly in section 3.3, the specifics of the air drag function are not particularly important as long as the function is well approximated by a Taylor series (Weisstein, E. W. (2013). Taylor Series.)
http://mathworld.wolfram.com/TaylorSeries.html

$$P_{air} = \frac{1}{2000}\,\rho \cdot C_D \cdot A \cdot v^3$$
$$\frac{P_{air}}{v} = \frac{1}{2000}\,\rho \cdot C_D \cdot A \cdot v^2 \tag{4}$$

To the extent that air density fluctuates with temperature, air resistance is also a factor of temperature. Air density at 30 °C is 1.164 kg/m^3, while at -20 °C it is 1.395 kg/m^3, 20% denser. [34] Speed exerts a much stronger influence than temperature, though for the sake of completeness I include inverse temperature in the model.

3.2.2 Rolling Resistance

As the vehicle rolls forward, the tires deform slightly. The bottom of a tire flattens out, then springs back as the wheel rotates further. Each time the wall of the tire deforms and returns, it costs a small amount of energy. The number of cycles of deformation is a linear function of the distance traveled, and therefore the energy expended to overcome rolling resistance is approximately proportional to the distance traveled (Transportation Research Board, 2006). The resistance also depends on the design of the tire, ambient temperature, tire temperature and tire pressure.[35]

The National Research Council of the National Academies investigated the potential efficiency gains from changes in rolling resistance and tire choice. Their policy suggestion in 2006 was a 10% reduction in rolling resistance over the following decade (Transportation Research Board, 2006). The report projected a 1–2% increase in fuel economy (miles per gallon) from such a change. (An increase of 1–2% in fuel economy represents a 1–2% decrease in fuel consumption.)

$$P_{tires} = C_R \cdot M \cdot g \cdot v$$

Where C_R is a coefficient of rolling resistance specific to the tire, M is the mass of the car, $g = 9.8\,m/s^2$ is the acceleration due to gravity and v is the speed of the vehicle Ross (1997).

Using Equation (5), the contribution of rolling resistance should be incorporated in the constant in the fuel consumption model, as C_R, M and g are (almost) constants.[36]

$$\frac{P_{tires}}{v} = C_R \cdot M \cdot g \tag{5}$$

[34] WolframAlpha (2013).
https://www.wolframalpha.com/input/?i=air+density+at+-20C
https://www.wolframalpha.com/input/?i=air+density+at+30C

[35] There is also a some energy expended to compress the road surface. On paved roads, the deflection is very small, and this is not a major energy expenditure.

[36] C_R will vary slightly with temperature and M is impacted by vehicle load. The gravitational parameter g is almost constant over the surface of the earth.

3.2.3 Inertia, Speed and Acceleration

One of the largest power uses, particularly in city driving, is the change in vehicle inertia. Ross (1997) notes that:

$$P_{inertia} = \frac{1}{2}\,\delta \cdot M \cdot \frac{d}{dt}(v^2) = \delta \cdot M \cdot v \cdot \frac{dv}{dt}$$

Which can be simplified to:

$$\frac{P_{inertia}}{v} = \delta \cdot M \cdot \frac{dv}{dt} = \delta \cdot M \cdot a \tag{6}$$

In the above equation, M is the mass of the vehicle, δ is a factor to correct for the rotational and linear inertia of the car and a is acceleration.[37] Therefore, I calculate acceleration and include it in the model specification.

3.2.4 Hills

It takes a strong push to move an automobile up a hill, so it is important to include a parameter for elevation changes.

$$P_{hill} = M \cdot g \cdot v \cdot \frac{d}{dt}(h)$$

And therefore,

$$\frac{P_{hill}}{v} = M \cdot g \cdot \frac{dh}{dt} \tag{7}$$

Where M is the vehicle mass, g is the acceleration of gravity and $\frac{dh}{dt}$ is the change in elevation per second. Since the change in elevation may be positive or negative, hills may increase or decrease fuel consumption in any given second.

Drivers brake to maintain a safe speed while descending a hill, reducing the amount of energy that can be reclaimed from a decrease in elevation. Boriboonsomsin and Barth (2009) found that otherwise equivalent hilly and non-hilly routes created a 15–20% difference in fuel economy.

When specifying the model of driver choices, I must consider the extent to which drivers control their change in elevation. Do changes in elevation happen to a driver or does the driver cause these changes? Unlike ambient temperature, elevation changes are within a driver's control, at least to some extent. Second-by-second changes in elevation depend both on route choice (hilly vs. flat) and driving speed (how fast to ascend a hill). In some situations, change in elevation may act as an unsought proxy for speed. Southeast Michigan has few topographical features, so the proxy effect is less worrisome for this study than it would be in other areas of the US.

[37] δ is a unit-less empirical constant, approximately 1.03–1.04 (Berry, 2010; Ross, 1997).

3.2.5 Air Conditioning

The use of air conditioning (AC), the last form of power from Equation 2, depends strongly on ambient temperature. In this study 47.4% of observations above 25 °C used AC, while only 0.169% of observations included AC use at temperatures below 0 °C.[38] Air conditioning is an energy-intensive accessory. Other accessories also use power, but Farrington and Rugh (2000) found that "the air conditioning system is the single largest auxiliary load on a vehicle by nearly an order of magnitude." Furthermore, most other accessories are necessary for safety e.g. headlights, defrosters and windshield wipers. I am interested in the ways drivers vary in their *voluntary* choices, so these other accessories are of less interest. I do not have reason to believe that use of accessories other than air conditioning is a substantial contribution to between-driver variation.

3.2.6 A Brief Introduction to Internal Combustion Engines

To understand how an engine burns fuel and drives a vehicle, it is helpful to discuss briefly the mechanics of an internal combustion engine. Figure 6 is a diagram showing the movement of a piston in a four-stroke engine, the type used in modern automobiles. At the beginning of the cycle, the piston is at the top of its range and begins to move downward (the first stroke) meanwhile, fuel is injected, as shown in B. The piston begins to move upward, compressing and heating the fuel and air mixture in the cylinder (the second stroke), step C. When the piston reaches the top, the spark plug fires, exploding the fuel in step D. The explosion forces the piston downward, providing power to the car in step E. Finally, the piston moves up again and the exhaust gas is driven out of the cylinder.

One might ask whether fuel use is discretized, as gasoline is only injected once in the six step cycle. In some sense, it is, but these cycles occur so rapidly that I can treat fuel use as a continuous measure.[39] Even at the slow engine speed of 500 revolutions per minute, the six-cylinder engine would create 25 explosions per second, much faster than the data are collected. Fuel is injected in minuscule quantities, approximately 0.02 mL per explosion. Therefore I will not worry further about the discrete nature of fuel injection.

3.2.7 Temperature

Temperature has an enormous impact on the performance of automobiles for myriad reasons in addition to its influences on rolling and air resistance. Below I discuss these effects in the broadest terms. Looking again at Figure 6, if the

[38] Drivers may also use AC to control humidity, regardless of ambient temperature. Running the compressor is still energy-intensive, so AC should be included at any temperature.

[39] Except for the issues of discrete measurement discussed in Section 2.2.

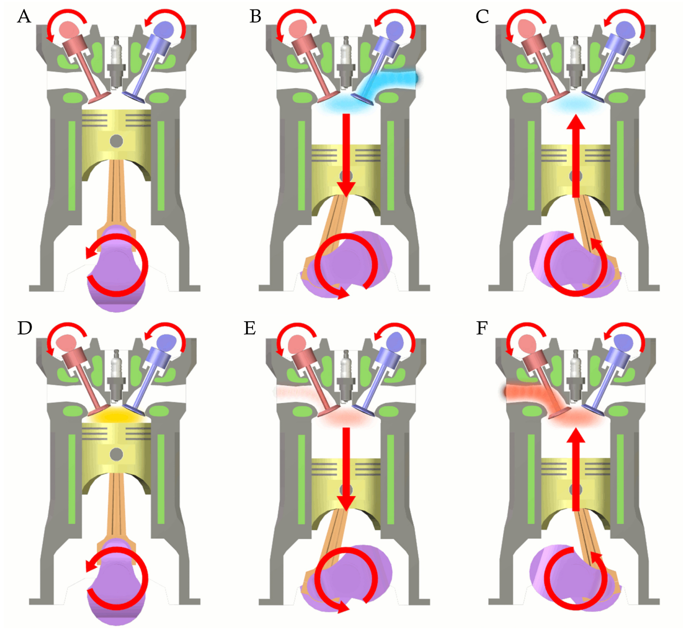

Figure 6: A four-stroke cycle (A) Start of cycle, (B) Fuel intake, (C) Compression, (D) Spark, (E) Power and (F) Exhaust

Source: Wapcaplet (2005, June 9). Four stroke cycle. https://commons.wikimedia.org/wiki/Category:Four-stroke_cycle

walls of the cylinder are cold, much of the energy released in the explosion of step (E) will go to heating the engine rather than driving the piston. The car adjusts by injecting more fuel in each cycle to heat the engine while providing sufficient power. Cold starts, where the engine begins at ambient temperature, are particularly demanding of fuel. To the extent that the heat loss continues to cool the engine below its ideal operating temperature, cold ambient temperatures will require additional fuel throughout the trip.[40]

Additionally friction of all types increases at colder temperatures. The many moving parts in a car each have a slightly harder time moving at low temperatures. A few notable examples are the crankshaft, axles, wheel bearings and tires. Some components will warm up throughout the trip, mitigating the effect of cold ambient temperatures.

Hysteresis is an important factor because an engine that is still warm from the previous trip will behave differently than a cold engine. Unfortunately,

[40] If ambient temperatures are warm, the engine instead has to work to cool off using the car's radiator.

the dataset does not have information on the temperature of the engine block. Instead, I use an interaction of the time between the current and previous trips, the elapsed time for the current trip and the ambient temperature to capture the hysteretic effects.

3.3 Model Specification

3.3.1 Driver-Specific Effects

In the discussion of driver-specific effects it is worth differentiating the degree of control drivers have over each variable in the study, and *when* they have control. Some, like ambient temperatures, are entirely outside a driver's control. Others, like route choice, are within a driver's control at one point, but not at later times (the factors bundled with demand for vehicle–miles). Finally, drivers have second-by-second control of some variables, like speed and acceleration. This analysis focuses on the impact of the demand for vehicle performance, controlling for the decisions made at longer time scales.

In an effort to measure the impacts drivers exert, I specify three models: one with driver effects alone and no controls; one with driver effects and controls for factors outside the drivers' immediate control; and one that controls for the demand for vehicle performance. In each of these models, I will estimate the impact the drivers have on fuel consumption, and note the variation in driver effects decline as decisions of speed and acceleration are included.

Rather than evaluate the many permutations of speed and acceleration to find specific behaviors that impact fuel efficiency, this analysis focuses on the overall impact a driver has on fuel consumption. Broadly speaking, I am interested in how drivers differ in their style and how those styles impact fuel use, without delving into the details of any one style. Calculating how specific vehicles perform under various kinetic schemes is a valuable exercise, one best left to physics and automotive engineering. Though I included a basic discussion of automotive physics in Section 3.2, the economics of driver behavior is the main focus of this thesis.

In principle, it is useful to consider driver specific effects as a form of mixed-effects model. The driver effects and the covariates would fill the roles of random effects and fixed effects, respectively.[41]

$$FuelCons_{di} = \beta X_i + \gamma U_d + \epsilon_{di}$$

Where fuel consumption by driver d for observation i is explained by some vector of fixed effects, X_i, and random effects of each driver U_d. Each observation has

[41] Stata offers the `xtmixed` command, which estimates a mixed-effects model via maximum likelihood estimation. Because the dataset is so large, `xtmixed` and autocorrelated errors require more computational resources than I have, even with use of a powerful workstation.

some error ϵ_{di}. The random effects are specific to each driver, and the drivers in the sample are drawn from a larger population.[42] From a statistical point of view, the drivers should be included as a random affect because they are drawn from the broad pool of all drivers.

Because of the difficulties of implementing a random effects model in such a large data set, I elected to include each driver as a fixed effect instead, with an indicator variable in the model. I interpret the coefficient on a driver's indicator as the "treatment effect" of being that driver.[43] Specifying the fixed-effects model does have advantages over the mixed effects model mentioned: allowing each driver their own coefficient allows for a non-normal distribution of driver effects.[44]

The evaluation follows three major steps. First, I estimate a regression of the fixed effect of each driver with no covariates. That simple model is not terribly informative, but it provides a benchmark for adding the other covariates. Next, I run a regression taking into account variables the driver cannot control: windshield wiper speed, temperature, odometer, altitude, change in altitude, cumulative duration and cumulative distance of the trip. I include these variables in the model very flexibly, including many interactions and quartic terms.[45] Given the many degrees of freedom available, it makes sense to emphasize flexibility over parsimony. Finally, I estimate the driver effects in a model that includes speed and acceleration, variables over which drivers have complete control. I conduct separate analyses for highway and non-highway driving.

$$FuelCons_{dts} = \gamma_d D_d + \epsilon_{dts} \tag{8}$$

$FuelCons$ is fuel consumption, measured in L/100km and indexed hierarchically by driver (d), trip (t) and second (s). D represents an indicator for each driver and γ_d are the associated coefficients. ϵ_{dts} represents the observation specific errors.

Here I present the second equation, with day of week, odometer, altitude, warm start, ambient temperature and windshield wiper variables but without out speed, acceleration or air conditioning:

$$FuelCons_{dts} = \beta_d D_d + \beta_W W_{dts} * warm_{dts} + \beta_A alti_{dts} + \beta_O odom_{dts} + \eta_{dts} \tag{9}$$

$FuelCons$ is fuel consumption, measured in L/100km and indexed by driver (d), trip (t) and second (s). D represents an indicator for each driver and β_d are the associated coefficients. W_{dts} is a vector of weather effects. $warm$ is a

[42] The assumption of random sampling may or may not be valid. The drivers were randomly sampled within southeast Michigan, but there may be response bias.

[43] Wooldridge (2009) notes that the dummy variable regression employed in Equation 9 estimates the same coefficients and standard errors as the fixed effects estimator.

[44] However, I will assume normality in Table 3 to calculate the standard errors.

[45] Though the driver does control the window wipers directly and has instantaneous control, I include windshield wipers as a proxy for other weather variables outside the drivers' control.

function designed to capture engine block warmth (including time since the beginning of the trip, ambient temperature and duration of the previous stop.) The $*$ operator indicates an interaction between every combination of variables in W and $warm$. A function of altitude, including polynomials of level and change per second, is represented by $alti$. Finally, $odom$ is a quartic function of vehicle's odometer reading, included with the goal of capturing any unresolved maintenance issues. The associated vectors of coefficients are β_A and β_O for $alti$ and $odom$, respectively.

The function outlined above contains a good deal more variables than the physical model specified in Section 3.2. I worked to include the data I had in a very flexible way to capture as much of non-driver variation as possible.

In the third model, I compare the remaining inter-driver variation *after* controlling for speed, acceleration and air conditioning use:

$$\begin{aligned} FuelCons_{dts} &= \psi_d D_d + \psi W_{dts} * warm_{dts} * AC_{dts} + \psi_A altitude_{dts} \\ &+ \psi_O odometer_{dts} + \psi_V speed_{dts} + \psi_{acc} acc_{dts} + \nu_{dts} \end{aligned} \tag{10}$$

All of the variables are the same as in Equation 9, with the addition of polynomials in speed and acceleration and the indicator variable for air conditioning. The vectors of coefficients are ψ_V and ψ_{acc}. AC is included in the vector of weather variables.

3.3.2 Highway vs. Non-Highway Driving

Highway driving is a very different activity than driving in a city, requiring a different set of skills and different behaviors. The US Environmental Protection Agency recognizes the difference in fuel consumption patterns between highway and city driving, and the agency publishes two fuel economy rankings for the different road types. As shown in Figure 5, the power flows are substantially different in city driving than highway.

Driving on an expressway offers relatively little opportunity to choose speed or accelerate.[46] Most of the time, drivers pick a speed and stick to it, sometimes using the cars' automated cruise control system.[47] Some drivers choose to exceed the speed limit while others go slower, but in most cases drivers do not stray too far from the posted speeds. In congested traffic, the average speed may be much lower than the posted speed, but the drivers still have very little flexibility in their speed and acceleration decisions.

In contrast, city driving provides much more opportunity drivers to behave differently. Different routes have differing speed limits, which the drivers may

[46] Acceleration on the highway on-ramp is substantial, but acceleration on the highway itself is minimal.

[47] I do not have data on cruise control use.

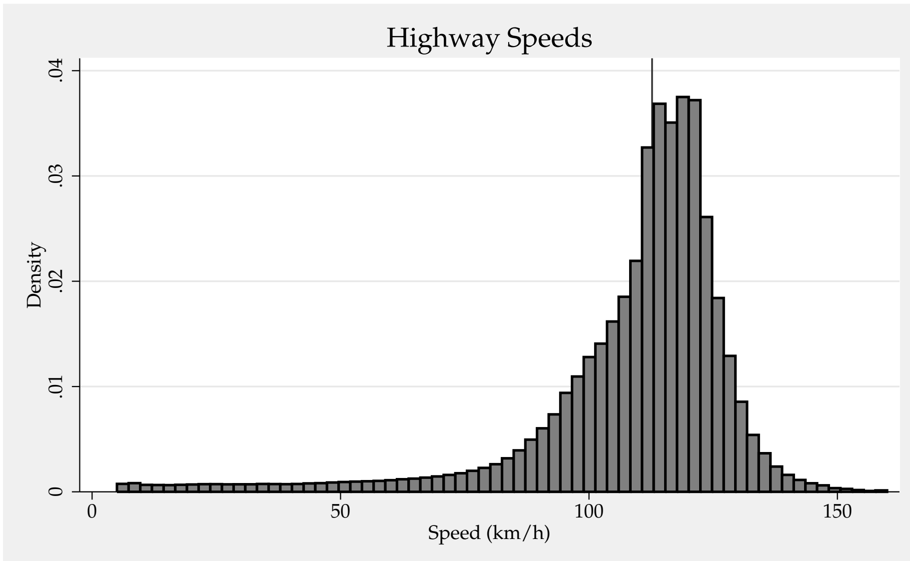

Figure 7: Speeds for highway driving, truncated at 160 kph. Vertical line at 112.7 kph (70 mph), the legal speed limit for many of Michigan's highways.

or may not follow. City driving also requires frequent stops and starts. Each time the car accelerates, the driver is implicitly choosing how much fuel to burn and how quickly to reach the destination. Each time the car stops, it spends some time idling before moving again, one of the substantial fuel costs of city driving.

In light of the substantial differences, I have separated the analysis into highway driving and non-highway driving, using the road classification provided by UMTRI. The study classified road type into six categories: interstate or major highways; major thoroughfares; secondary thoroughfares; minor streets; highway ramps; and out-of-state or unknown. In my analysis I have analyzed the first type as highway driving and all of the others collectively as non-highway driving. The UMTRI researchers used a GIS system to match GPS observations to known roads. Such a matching process requires a degree of "fuzziness" and is prone to some error. I assume that there was no systematic error in road type matching.

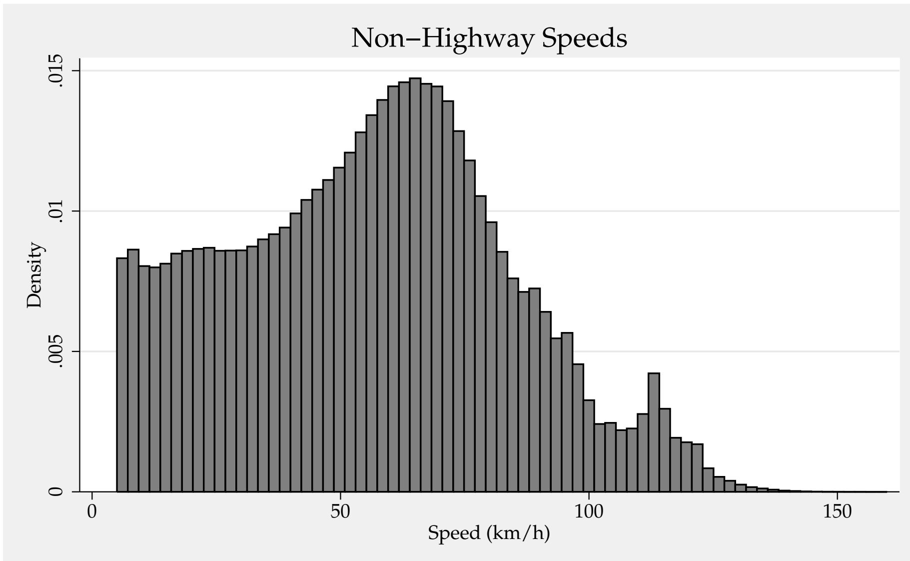

Figure 8: Speeds for non-highway driving, truncated at 160 kph.

4 Results

As mentioned in Section 3.3, I estimated three different models. First, I looked at each driver's estimated effect without any covariates. The distribution of the driver effects are graphed in Figures 9 and 12 for highway and non-highway driving, respectively. In the second model, I examine the driver effects while controlling for factors outside their immediate control. The estimates of these coefficients are represented in Figures 10 and 13 for highway and non-highway driving. Finally, the third model includes speed, acceleration and air conditioning behavior. The driver effects are shown in Figures 11 and 14 for highway and non-highway driving.

Table 3 reports the standard deviation of the observed distribution of driver coefficients. Several features of the table are particularly notable. First, there is greater variation across drivers in non-highway driving. This result makes intuitive sense, as drivers have more 'degrees of freedom' in city driving than they do on the highway. The dispersion in driver coefficients fell, but not to zero. Therefore, there are economically meaningful driver-specific effects on fuel economy even after controlling for speed and acceleration. The histograms in Figures 11 and 14 show the driver coefficients.

Table 3: Standard deviation of estimated driver fixed-effects coefficients

Road type	No covariates (L/100km)	No behavior (L/100km)	With behavior (L/100km)
Highway	0.5523	0.5929	0.3451
Standard errors	(0.00364)	(0.00390)	(0.00227)
Non-highway	1.666	1.794	0.4198
Standard errors	(0.0110)	(0.0118)	(0.00276)

Note:
The standard error of a sample standard deviation is:
$\widehat{\sigma}_s = \sqrt{N - 1 - \frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N}{2}\right)}} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\beta_i - \overline{\beta}\right)^2} \approx 0.006585\, \widehat{\sigma}_\beta$
These errors are underestimates of the true error; a more thorough analysis would take account of the origin of the β_i coefficients, which were estimated by weighted least squares.
Weisstein, E. W. (2013). Standard Deviation Distribution. *MathWorld.*
http://mathworld.wolfram.com/StandardDeviationDistribution.html

The *level* of each driver coefficient does not matter, the variation between them is the result of interest. The estimate for each coefficient is measured relative to Driver 1, and there is no reason to assume that driver is a special benchmark. Therefore, in Figures 9–14 I subtract the median value of the distribution.

4.1 Potential Savings

Suppose all drivers became more efficient, behaving like the drivers with the lowest estimated driver-effect. Table 4 represents the percentage by which fuel use could decrease if all drivers drove like the most efficient, "best", driver or the "good" driver at the 10^{th} percentile of consumption. The savings in the table above simply replace each drivers' estimated fixed effect with the fixed effect of the best (or good) driver. For the purposes of this calculation, I am assuming that *all* of the drivers behave like the best driver or the good driver, even those who were previously more efficient than that individual.

Recall Equation 9:

$$FuelCons_{dts} = \beta_d D_d + \beta_W W_{dts} * warm_{dts} + \beta_A altitude_{dts} + \beta_O odometer_{dts} + \eta_{dts}$$

Then total predicted fuel use is:

$$\widehat{FuelUse} = \sum_{d,t,s} \left(\widehat{FuelCons}_{dts} \cdot \Delta x_{dts} \right)$$

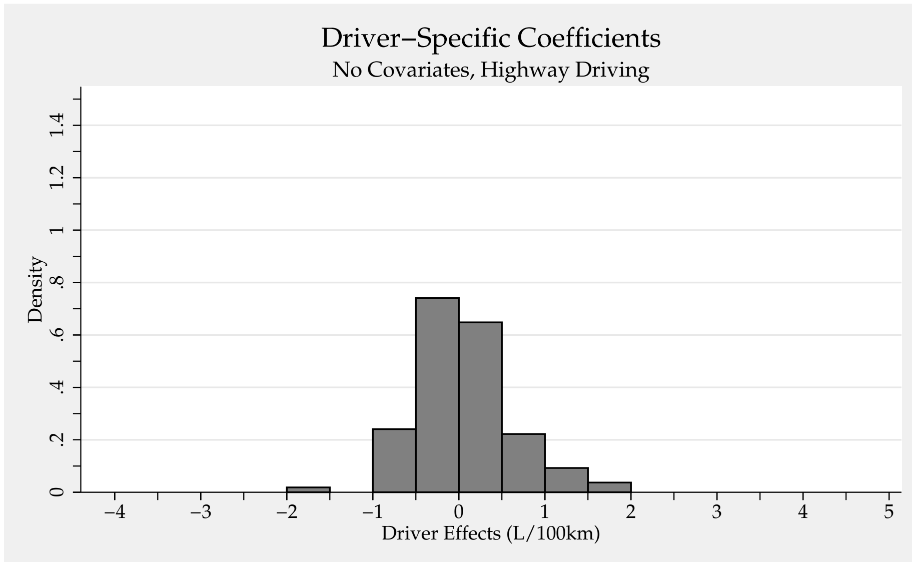

Figure 9: Driver effects for highway driving with no covariates.

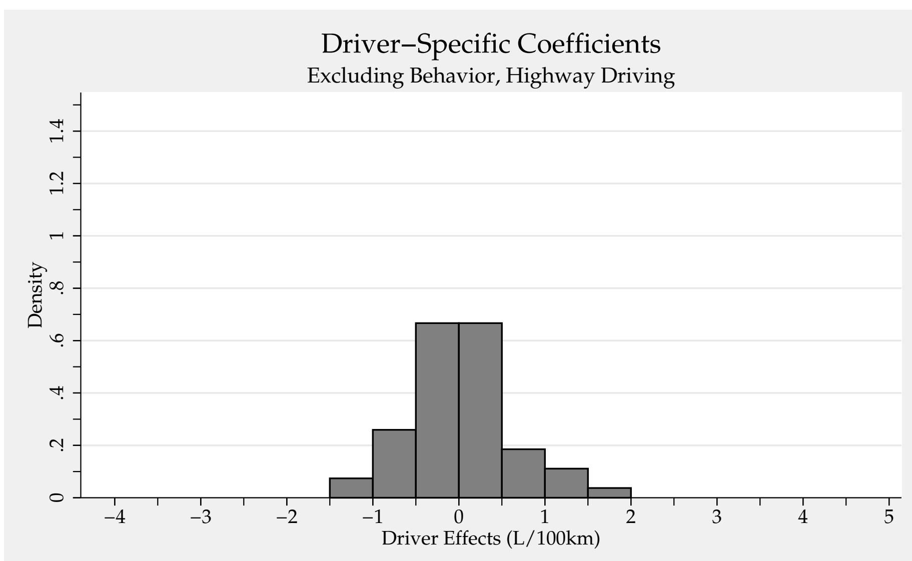

Figure 10: Driver effects for highway driving, including day of week, odometer, altitude, warm start, ambient temperature and rain but not speed, acceleration or AC decisions.

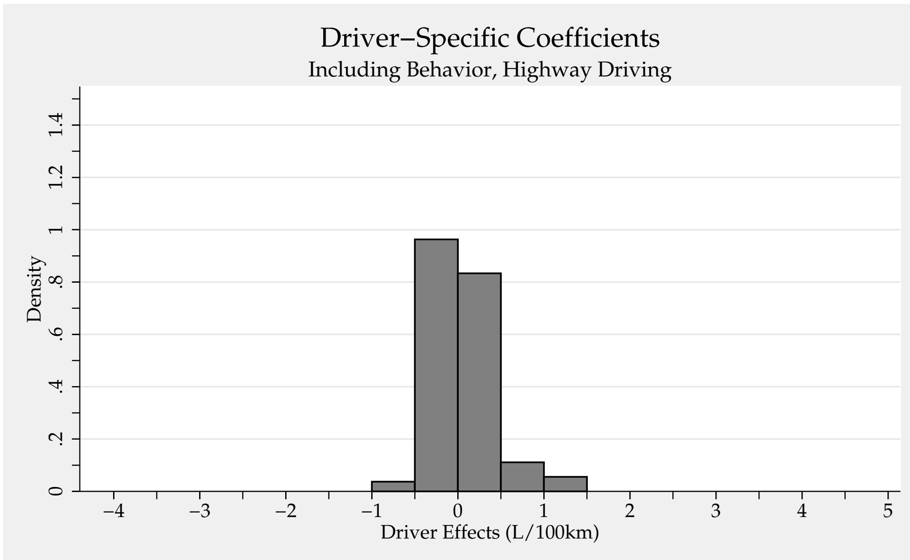

Figure 11: Driver effects for highway driving, including speed or AC decisions, as well as day of week, odometer, altitude, warm start, ambient temperature and rain.

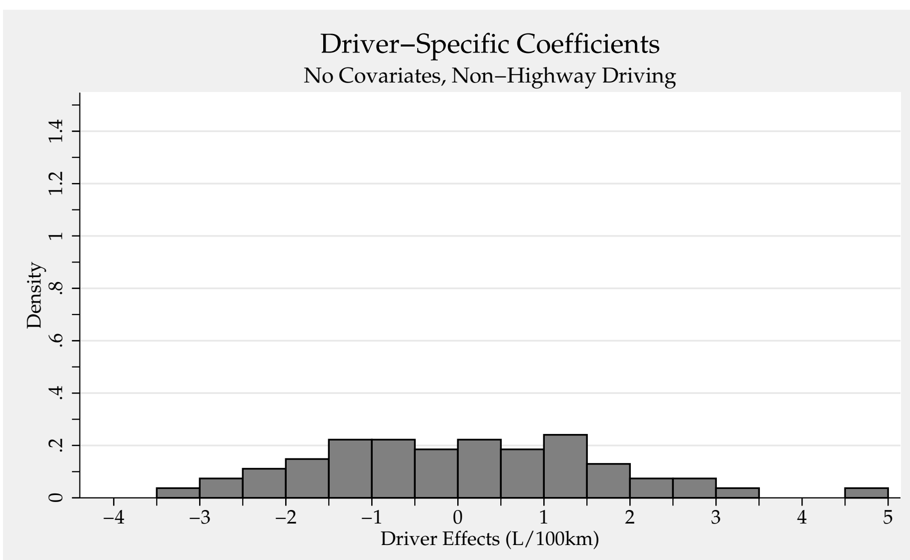

Figure 12: Driver effects for non-highway driving with no covariates.

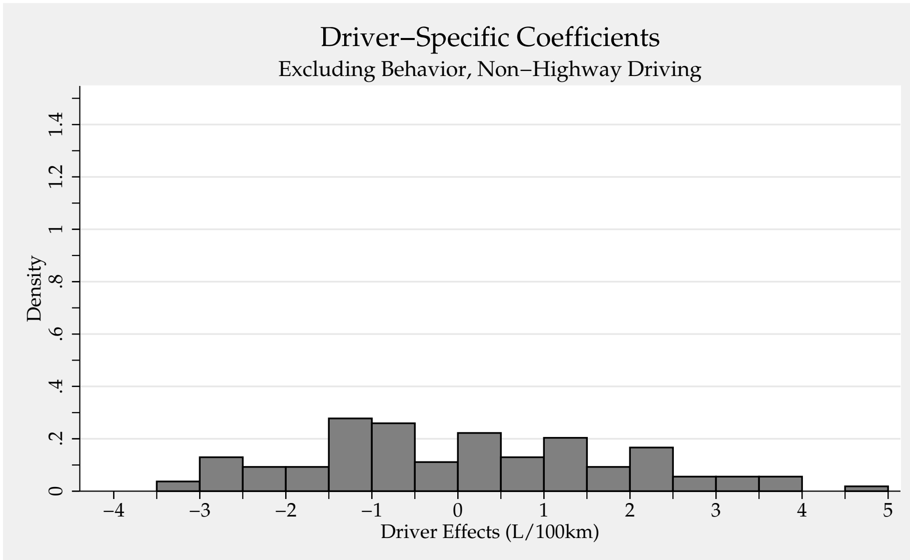

Figure 13: Driver effects for non-highway driving, including day of week, odometer, altitude, warm start, ambient temperature and rain but not speed, acceleration or AC decisions.

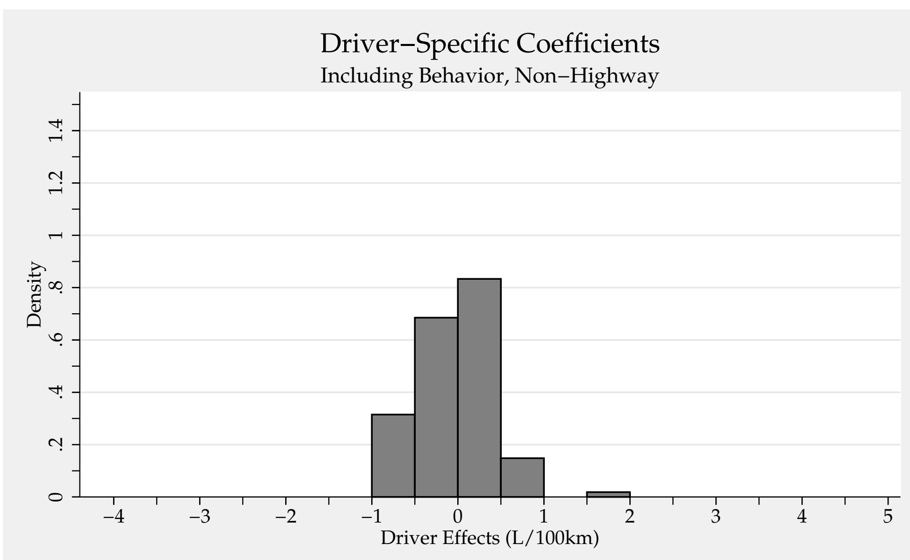

Figure 14: Driver effects for non-highway driving, including speed and AC decisions, as well as day of week, odometer, altitude, warm start, ambient temperature and rain.

Table 4: Estimated fuel savings

Road Type		Preferred (%)	Comparison (%) A	Comparison (%) B
Highway	Best[a]	16.95	19.99	8.27
		(0.115)	(0.116)	(0.0297)
	Good[b]	8.17	8.10	4.67
		(0.124)	(0.128)	(0.0308)
Non-	Best	25.90	22.68	9.52
Highway		(0.345)	(0.334)	(0.0738)
	Good	17.63	14.80	6.11
		(0.354)	(0.344)	(0.0762)

Notes:
The preferred results are from the model of Equation 9, which controls for factors outside a driver's influence.
A corresponds to the model of driver effects with no covariates, Equation 8
B corresponds to the model of driver effects with speed, acceleration and air conditioning, Equation 11
The standard errors in parentheses are calculated by error propagation from heteroskedasticity robust errors clustered by driver.

To calculate hypothetical fuel use, I simply substitute the best (or good) driver coefficient in for the fuel consumption at each second.

$$\widehat{FuelUse}_{best} = \sum_{d,t,s} \left(\left(\widehat{FuelCons}_{dts} - \beta_d D_d + \beta_{best} \right) \cdot \Delta x_{dts} \right)$$

From the values of $\widehat{FuelUse}_{best}$ and $\widehat{FuelUse}_{good}$ I calculate the percentage changes shown in Table 4.

The standard errors are calculated using the propagation of error technique:

$$\sigma^2_{best} = \sum_{d,t,s} \left[(\sigma_{FuelBest} \cdot \Delta x_{dts})^2 + (FuelBest_{dts} \cdot \sigma_{\Delta x})^2 \right]$$

I drop the error in distance measurement, $\sigma_{\Delta x}$, because it is very small relative to $\sigma_{FuelCons}$.

$$\sigma^2_{best} \approx \sum_{d,t,s} \left[(\sigma_{FuelBest} \cdot \Delta x_{dts})^2 \right] = \sum_{d,t,s} \left[\ \left(\sigma^2_{FuelCons} + \sigma^2_{\beta_d} + \sigma^2_{\beta_{best}} \right) \cdot \Delta x^2_{dts} \right]$$

Finally, applying the formula for error in a percentage change:[48]

$$\sigma_{\Delta \% best} = 100 \left| \frac{FuelUse_{best}}{FuelUse} \right| \cdot \sqrt{\frac{\sigma^2_{best}}{FuelUse^2_{best}} + \frac{\sigma^2_{FuelUse}}{FuelUse^2}}$$

[48] US Bureau of the Census. (1999). Percent Changes. http://www.census.gov/acs/www/Downloads/data_documentation/Accuracy/PercChg.pdf

The equations for the fuel savings due to good fuel consumption behavior are very similar to those for best fuel consumption behavior.

The preferred results come from the model of Equation 9, which controls for factors outside the drivers' control, but not for speed, acceleration or AC use. The comparisons A and B are respectively from the model with no covariates (Eqn. 8) and the model with controls for speed, acceleration and AC (Eqn. 11). The comparison models are not useful to calculate potential fuel savings, but they do provide information about the relative importance of speed, acceleration and AC use. Reading from the first row, best highway driving practices would save 16.95% of fuel use, of which we can attribute 8.27% to other driver-specific factors.

Of course, fuel is not the only cost paid by drivers. They also pay in time. If driving more efficiently meant driving slower and trips took proportionately more time, then drivers are paying in time to save on fuel. On the other hand, if drivers could decrease their fuel consumption by driving more rapidly, then they could save both money and time. More efficient driving at higher speeds is a rarity because, while city driving occurs below the most efficient speed, the benefit is outweighed by the cost of accelerating the car after every stop signal.

4.2 Environmental Effects of Fuel Savings

The US Environmental Protection Agency (EPA) estimates that the US emitted 5,732.5 million metric tons (1.264×10^{13} lb) of carbon dioxide equivalent in 2011. Of these emissions, mobile combustion (transportation) accounted for 1,760.5 million metric tons (3.881×10^{12} lb), or 30.7% of the total.[49]

The EPA calculates the combined fuel economy of a vehicle using the formula $MPG_{avg} = 0.55MPG_{city}+0.45MPG_{highway}$.[50] Following their weighting, the best driver condition saves 19.28% of overall fuel and the good driver condition saves 11.61%. In turn, these translate to savings of 339.4 million metric tons (7.483×10^{11} lb) and 204.4 million metric tons (4.506×10^{11} lb) for best and good driving, respectively. Using the estimated marginal damages of $5–25 per ton of CO_2 equivalent, changing to more efficient driving practices would save $1.022–8.845 billion in damages. Damage from other, local forms of pollution would also be reduced 11–20%.

[49] US EPA: Office of Air and Radiation, Climate Change Devision (2013). Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2011
http://www.epa.gov/climatechange/ghgemissions/usinventoryreport.html

[50] US EPA (2013). Gasoline Vehicles: Learn More About the New Label.
http://www.fueleconomy.gov/feg/label/learn-more-gasoline-label.shtml

5 Conclusion

The decreases in fuel consumption calculated in Section 4 are large. Are they estimates beyond reason? It would be extreme to expect every driver in the US to drive efficiently as the best 1%. Nonetheless, the results of this analysis show that there is considerable scope for behavioral changes to reduce drivers' fuel consumption, either through educational campaigns or incentives to drive efficiently. Fuel taxes are one such incentive, albeit a politically unpopular one. If a substantial fraction of the efficiency gains could be realized in the population of American drivers, the fuel savings and externality reductions may be very large. Fuel savings of approximately 12-19% are possible, with the associated decreases in fuel costs and CO_2 emissions.

A decline in fuel use of 12–19% would have general equilibrium effects, which I have not considered. Generally speaking, a decline in fuel demand would lead to a decrease in fuel prices, which would cause some increase in quantity demanded. Incorporating such general equilibrium effects might lead to smaller declines in fuel use than I have predicted.

In the near future, cars might not have drivers making the speed and acceleration decisions. Autonomous vehicles are currently legal in four states of the US: California, Nevada, Florida and Texas.[51] As the routines controlling these cars are written, the programmers and automotive engineers would do well to consider the potential for fuel savings through more efficient driving patterns.

[51] Muller, J. (2012, September 26). With Driverless Cars, Once Again it is California Leading the Way. *Forbes.* http://www.forbes.com/sites/joannmuller/2012/09/26/with-driverless-cars-once-again-it-is-california-leading-the-way/

References

Allcott, H. and N. Wozny (2012). Gasoline prices, fuel economy and the energy paradox. *National Bureau of Economic Research* (Working paper 18583).

Berry, I. M. (2010, February). The effects of driving style and vehicle performance on the real-world fuel consumption of U.S. light-duty vehicles. Master's thesis, Massachusetts Institute of Technology.

Boriboonsomsin, K. and M. Barth (2009). Impacts of road grade on fuel consumption and carbon dioxide emissions evidenced by use of advanced navigation systems. *Transportation Research Record*.

Busse, M. R., C. R. Knittel, and F. Zettelmeyer (2013). Are consumers myopic? evidence from new and used car purchases. *The American Economic Review 103*(1), 220–256.

Espey, M. and S. Nair (2005). Automobile fuel economy: What is it worth? *Contemporary Economic Policy 23*(3), 317–323.

Evans, L. (1979). Driver behavior effects on fuel consumption in urban driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society 21*(4).

Farrington, R. and J. Rugh (2000, September). Impact of vehicle air conditioning on fuel economy, tailpipe emissions, and electric vehicle range. Washington, D.C. Earth Technologies Forum: National Renewable Energy Laboratory.

Ishiguro, S. (1997). Heavy-duty truck fuel economy test in actual road traffic. Technical Report 973183, Society of Automotive Engineers, Warrendale, Pennsylvania.

Jacobsen, M. R. (2012). Fuel economy and safety: The influences of vehicle class and driver behavior. *National Bureau of Economic Research* (Working paper 18012).

LeBlanc, D. J., M. Sivak, and S. Bogard (2010, December). Using naturalistic driving data to assess variations in fuel efficiency among individual drivers. Technical Report UMTRI-2010-34, The University of Michigan Transportation Research Institute, Ann Arbor, Michigan.

Lenner, M. (1995). Measurement by on-board apparatus of passenger cars real-world exhaust emissions and fuel consumption. Technical Report 771A, Swedish National Road and Transport Research Institute, Linkoping, Sweden.

Metcalf, G. E. (2009). Market-based policy options to control U.S. greenhouse gas emissions. *Journal of Economic Perspectives 23*(2).

Muller, N. Z. and R. Mendelsohn (2007). Measuring the damages of air pollution in the United States. *Journal of Environmental Economics and Management 54*(1), 1 – 14.

Ross, M. (1997). Fuel efficiency and the physics of automobiles. *Contemporary Physics 38*(6), 381–394.

Sivak, M. and B. Schoettle (2012). Eco-driving: Strategic, tactical, and operational decisions of the driver that influence vehicle fuel economy. *Transport Policy 22*.

Stephenson, R. D. and R. J. Wassersug (1993, July 15). Horsepower from a horse. *Nature 364*.

Transportation Research Board (2006). Tires and passenger vehicle fuel economy. Technical Report 286, National Research Council of the National Academies, Washington, D.C.

Wolff, H. (2012). Value of time: Speeding behavior and gasoline prices.

Wooldridge, J. M. (2009). *Introductory econometrics: A modern approach* (Fourth ed.). South-Western Pub.