Fuel Price Beliefs and the Consumer Response to Price Fluctuations: The Choice between Gasoline and Diesel Vehicles

Bradley D. Shrago

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics.

Chapel Hill
2017

Approved by:

Brian McManus

Gary Biglaiser

Jane Fruehwirth

Jonathan Williams

Andrew Yates

ABSTRACT

BRADLEY D. SHRAGO: Fuel Price Beliefs and the Consumer Response to Price Fluctuations: The Choice between Gasoline and Diesel Vehicles.
(Under the direction of Brian McManus)

When consumers make fuel choices in vehicle adoption, they must form beliefs regarding each fuel's price over the life-cycle of the vehicle. Whereas prior research has found that current gasoline prices are a reasonable proxy for consumers' gasoline price beliefs, the same may not hold for the price difference between diesel and gasoline, a key determinant of the savings associated with adopting a diesel engine. I consider the market for gasoline and diesel powered pickup trucks in the state of Washington, where the time-series variation in the diesel premium is transient while geographic variation in the premium is relatively persistent. Because time-series variation in the diesel premium exhibits mean-reversion, a forward-looking individual's price beliefs may not respond to such variation. In order to consider the importance of modeling price beliefs in a manner which allows consumers to exhibit different responses to different types of price variation, I develop a two-period model of truck choice and subsequent usage. I employ a rich dataset of vehicle registrations and high-frequency local fuel prices which allows me to separately identify the consumer response to time-series variation in the diesel premium relative to all other sources of fuel price variation. My estimates suggest the response to the former is roughly one-tenth as strong as the latter. Using a variety of counterfactual simulations, I document that modeling price beliefs in a more flexible manner which allows for different responses to different types of fuel price variation increases common measures of the consumer response to price fluctuations by roughly 15-20% in this setting.

Dedicated to the life and memory of Dr. Tiago Miguel Castanheira Correia Costa Pires.

ACKNOWLEDGMENTS

There are several individuals who significantly contributed to my progress as a graduate student, each of whose contributions to my personal and academic life were important in the development of this manuscript. First and foremost, I would like to thank my loving wife, Kayla, for managing the vast majority of our personal lives and efforts towards raising our children during the past six years. I will never forget the sacrifices she made to allow me to continue schooling into my late 20's while working from home with our daughters and (begrudgingly) our dog, Ryder. I thank my two daughters Elodie and Paige, who for five and two years of my graduate studies offered me an outlet to enjoy time away from my work, whether this involved testing the transitivity of their preferences or doing things which children actually enjoy.[1] I thank my parents Andy and Karen for motivating me to find modest enough academic success in high school to allow me to attend a college which allowed me to fulfill my potential, my brother Bryan for providing an academic role model at an impressionable age, and my brother Ben for teaching me that training in mathematics would have a greater impact on my future than training in ice hockey. I am grateful to my grandfather William Shrago and my late grandmother Florette Shrago for their contributions to the education of myself and my father, and thankful for the sacrifices made by my late grandparents Albert Schneider and Marion Schneider Cullen to educate my mother. I thank all of my teachers who showed interest in my academic progress, particularly Mr. Senson and Ms. Pietruszka, the latter of whom unflinchingly believed a third-grader actually meant 'zorilla' and not 'gorilla'. I thank my economics professors Jerry Savitsky, Santhi Hejeebu, and Todd Knoop as well as my mathematics professors Tony deLaubenfels, Steve Bean, and Ann Cannon for giving me exceptional training during my undergraduate years at Cornell College.

Many of the discussions I shared with my friends and colleagues I met in UNC's graduate program were instrumental to my success. I thank all of my peers in my cohort who I worked

[1] $peacock \succ flamingo, flamingo \succ emu \rightarrow peacock \succ emu$

with during our first year who spent evenings doing problem sets in Gardner Hall, and those who contributed to joint projects in my second year. I am indebted to the participants in UNC's Applied Microeconomics seminar series who offered me feedback over the past four years, and am particularly grateful for the experience I gained while presenting my research. To my current coauthors Forrest Spence and Jeff Ackermann, I look forward to continuing to work together in the future. I am also grateful for feedback provided by Eddie Watkins, Quinton White, Julien Isnard, Ray Wang, Chunxiao Li, and Marcela Parada-Contzen during the biweekly UNC Empirical Industrial Organization meeting and the selflessness of Uyen Tran during the 2017 ASSA Annual Meeting in Chicago.

I sincerely thank my advisor, Brian McManus, who undertook the difficult task of developing an aspiring Empirical Industrial Organization economist when he offered to advise me on my dissertation research. He made a continuous effort during my third and fourth years to keep me engaged in the literature, searching for research opportunities, and ultimately helping me to acquire my data and develop my proposal. Oftentimes in these early years, Brian undertook the uncomfortable task of steering me away from research opportunities which would not have been fruitful, and tweaking my suggestions to lead me towards productive research topics. I am grateful for the efforts Brian put into helping me to finalize my job market paper during Summer 2016, and motivating me to continue to improve my paper at times when I was struggling to continue doing so. His intense focus and thoughtful skepticism were necessary to instill within me the skills needed to conduct applied research, and critical to my own focus during my final year at UNC.

My committee members were also highly influential on my development as a researcher, and I am grateful to all of their contributions to my work. I thank Gary Biglaiser for building my understanding of the theory of Industrial Organization and Andy Yates for helping me recognize the value of studying environmental policy through the lens of an Empirical Industrial Organization economist. I thank Jane Freuhwirth for working with me to build and understand sources of identification in empirical models as well as their potential pitfalls, and Jonathan Williams for the efforts he put into helping me improve my coding efficiency and motivating me to be passionate about my research even as I face challenges. In addition to my committee, I would like to thank

Donna Gilleskie and Helen Tauchen for their contributions to all of our graduate students, myself included, and Jeremy Petranka for offering support to my development as an instructor. Finally, I would like to remember the contributions of my former committee member Tiago Pires, who tragically passed away shortly before I defended my proposal. I was blessed with the opportunity to know Tiago during his three years as an Assistant Professor of Economics at UNC, and am indebted to him for the kindness, sincerity, and inquisitiveness with which he approached my work. Tiago was an extraordinary individual and an incredibly bright young economist, and I hope that should I ever stray from the example he set, I can refer back to these acknowledgments and remember the example he provided.

While I was on the job market during early 2017, I spent some time thinking about a list of questions which students had been asked in past interviews, one of which asked me to name three economists who I greatly admire. To an extent, answers to these questions are strategic. Should I simply name three highly respected researchers in my field? Would I be able to provide insight into who I am as a researcher by discussing the work of junior faculty which I have found valuable? While I'd gladly have discussed my admiration for the work of a variety of economists in my field, I know that an honest answer to the question could not be given without mentioning Tiago Pires. He was a bright and promising young economist who undoubtedly would have continued to develop his strong line of empirical research. His passion for academic research shone through in every aspect of our interactions, and his kindness and selflessness was instrumental to the success of myself and many of my fellow graduate students at UNC.

Tiago tirelessly dedicated his time and thoughts to myself and other graduate students in our department as an assistant professor busy with a variety of his own research projects. A rule of thumb held by some of our graduate students was that it was never advisable to wedge a meeting with Tiago in between other obligations. Rarely would our meetings last less than an hour, and in my memory I do not recall a time where Tiago ended a meeting before I had to leave. During my third year (his first in our department), I reached out to Tiago with a few research ideas I'd been working on. This was a difficult time for me, as an aspiring economist working in an area where research ideas are plentiful and data constraints are severe. I did not know him at the time outside of

a few interactions during our seminar series. After asking whether we could meet and sending him a vague research proposal, he responded with a detailed list of comments and offered suggestions for literature which might offer insight into how I could frame the project. Our first meeting lasted somewhere around two hours, and through some bizarre string of events our conversation shifted from a discussion of empirical research on two-sided markets to an hour discussing the difficulties of properly specifying price beliefs in structural models of consumer behavior. While my work on two-sided markets and network effects was left for later, our discussion on price beliefs was instrumental in developing my dissertation.

He was an extraordinary academic with a genuine desire to understand our discipline and contribute to our collective knowledge through his own research as well as the work of coauthors, his fellow faculty, graduate students, and visiting economists who presented their work at UNC. I will always admire the way Tiago would attempt to contribute to projects which were presented at any of our seminar series. Regardless of the stage of a project, he approached seminars with the belief that he could provide feedback which would allow a presenter to move forward and make a valuable contribution to the literature. His critiques were humbly voiced in a respectful manner, often highlighting the opportunity to contribute by solving an empirical challenge, rather than merely noting its existence. I recognize that Tiago was a unique individual, and it will always sadden me to know that future students and faculty at UNC and across our field will not have the opportunity of knowing Tiago. However, I will also never forget that the time we were all able to spend with Tiago unquestionably changed us for the better. For the rest of my life, I will work humbly to exude his kindness, sincerity, and passion for bettering our understanding of the world through the study of economics.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SECTION 1

INTRODUCTION

Understanding consumers' responsiveness to fuel prices and their willingness to substitute into non-gasoline powered engines carries significant importance to both environmental and economic policy. In the United States, the transportation sector accounts for roughly one fourth of greenhouse gas emissions, the majority of which can be attributed to usage of petroleum-based fuels in automobiles and trucks. Overall, energy expenditures have typically represented between 5% and 10% of GDP over the past few decades. In recent years, a variety of policy mechanisms have been proposed and implemented in hopes of reducing environmental damages and reliance on foreign energy sources. Broadly speaking, policy goals tend to be aimed at reducing per-mile fuel consumption from gasoline and diesel-powered engines as well as incentivizing development and adoption of vehicles which operate on renewable fuels.[1] Despite the recent attention to non-gasoline fleet development and recent growth in availability of electric and diesel-powered cars, market shares of non-gasoline powered cars, SUVs, and vans remain low throughout the United States. This poses a challenge for researchers studying fuel choice among U.S. automobile purchasers, as fuel choice has not been a matter of serious consideration for the vast majority of car, SUV, and van purchasers for much of the last century.

In this paper, I consider the market for pickup trucks as a means to study the role of price expectations and heterogeneity in expected use on fuel choice, fuel economy, and subsequent usage decisions. My decision to focus on this market is based on the limited availability and low

---

[1] Examples of such policies include federal and state-level purchase subsidies for electric vehicles, subsidies at the federal and state level granted towards the production of biofuels such as ethanol and biodiesel, and generous treatment of such vehicles under recent Corporate Average Fuel Economy (CAFE) regulations. Diesel-powered cars enjoy a subtler advantage under CAFE regulations, as diesel engines generally enjoy a fuel economy advantage between 25% and 40% over a similarly powered gasoline engine. On a per-gallon basis, diesel requires more petroleum and emits around 15-20% more CO_2 when it is burned, but CAFE regulations currently do not adjust diesel fuel economy for the fuel's higher carbon content.

adoption rates of non-gasoline vehicles in the United States, as the market for pickup trucks is the only segment of the U.S. automobile market where (i) vehicles powered by different fuels have significant market shares and (ii) a large share of vehicles is purchased by individuals rather than fleet managers.[2] I employ a rich dataset of vehicle registrations and high-frequency retailer-level fuel prices in the state of Washington from 2012-2015. My fuel price data, along with the presence of odometer readings attached to a subset of vehicle observations, allows me to account for two econometric challenges facing researchers who study fuel choice: modeling price expectations in a manner consistent with the beliefs of vehicle purchasers, and accounting for the role of selection into fuel type on anticipated usage.

The issue of selection on anticipated usage is well-documented in the literature and summarized by Bento et al. (2011), who show that the energy paradox in fuel economy could be partly explained by unobserved heterogeneity in usage intensity.[3] Selection on anticipated usage is an important consideration for fuel choice in the pickup truck market as well. In this market, the diesel engine option is a costly upgrade (typically several thousand dollars) available on large trucks as a means to reduce operating costs and improve performance.[4] Thus, one might suspect that those individuals with a high level of anticipated usage should have a greater willingness to pay for the engine's attributes, and in turn be more likely to select the upgrade. In order to account for selection on anticipated usage, I use a two-period model of product choice and subsequent usage following Einav et al. (2013) and Gillingham (2013). In my specification, individuals' expected utility from operating the vehicle is scaled by a parameter which represents their usage type at the point of vehicle purchase. For those individuals with a high level of expected usage, it is more

[2]Based on vehicle registration data provided to me by the Washington Department of Licensing (DOL), roughly one fifth of new, personally-registered pickup trucks purchased from 2012-2015 in the state run on diesel, with all but a few dozen of the remaining vehicles operating on gasoline.

[3]The energy paradox refers to the phenomenon wherein individuals seemingly undervalue the present discounted value of energy savings when making purchases.

[4]Until the 2014 model year, the diesel engine had only been available on heavy-duty pickup trucks. Dodge began offering a diesel engine in the 'Full-Size' Ram 1500 in 2014. General Motors began offering a diesel engine in the 'Mid-Size' Colorado (and alternatively badged Canyon) in 2016. Nissan became the first foreign manufacturer to offer a diesel engine in a truck in the United States with their Nissan Titan XD which was released late in 2016.

likely that the utility gains associated with adoption of the more efficient, high-performance diesel engine will outweigh the upgrade cost. By employing odometer readings for a subset of my vehicle registrations, I am able to identify the distribution of usage types and allow adoption decisions to shift based on individuals' anticipated usage levels.

Within any model of vehicle choice the researcher must consider consumers' formation of fuel price beliefs over the life-cycle of the vehicle. Forecasting gasoline and oil prices is a rich literature unto itself, and there is a plethora of studies which attempt to improve upon a no-change forecast, where the expected price of gasoline at any point in the future is equivalent to its current price.[5] Alquist et al. (2011) summarize the predictive power of a variety of methods of forecasting oil prices, and find limited evidence of models which offer a forecast improvement over the no-change forecasting model. This evidence is consistent with consumers' elicited beliefs from surveys as well. Anderson et al. (2013) find that after controlling for inflation expectations, long-term household gasoline price expectations can be reasonably approximated by a no-change forecast. Given this evidence, prior researchers have often specified a no-change forecast for gasoline prices when modeling vehicle adoption, an approach which I also follow for gasoline prices.

The decision of whether to upgrade to a diesel engine requires consumers to forecast operating costs, relative to the engine's gasoline counterpart, over the vehicle's life. While gasoline price beliefs may be reasonably approximated by a no-change forecast, this need not hold for the price differential between the two fuels (referred to as the 'diesel premium' below). In Washington, variation in the diesel premium consists of both geographic price variation which is relatively persistent, as well as time-series variation which exhibits mean-reversion.[6] This suggests that the price beliefs of a forward-looking individual may respond differently to different sources of variation in the diesel premium.[7] As a result, a no-change forecast of the diesel premium over the

[5]I use the term no-change forecast to refer to the family of price forecasts where the forecast price in the future is today's price. As such, a random walk price forecast fits into this family.

[6]In §A.5, I document the mean-reverting nature of the diesel premium.

[7]Li et al. (2014) document that consumers exhibit a stronger response to variation in gasoline taxes than to other sources of variation in gasoline prices, and suggest this may be due to consumers' perceived persistence of tax changes. They also discuss the possibility that this results from greater public awareness of tax changes versus other price

life of a vehicle may be improved upon by forecasting the diesel premium based on the historical average in the individual's locale rather the current diesel premium.

My model of vehicle adoption and subsequent usage offers the opportunity to endow individuals with a variety of price beliefs. Considering the mean-reverting nature of the diesel premium, one approach is to specify expectations based on the average diesel premium in an individual's locale, and assume that individuals' diesel price forecasts fully account for mean-reversion. An alternative approach is to consider the possibility that individuals do not recognize the mean-reverting nature of the diesel premium and assume the same no-change forecast used for gasoline prices is applied to the diesel premium. A third possibility is that neither of these approaches are consistent with individuals' price beliefs, and that individuals' price beliefs do change in response to mean-reverting variation in the diesel premium, but to a lesser extent than the no-change forecast would imply. Rather than making an ex-ante assumption regarding which of the three possibilities represents consumers' price beliefs, I specify price expectations such that a parameter to be identified sheds light on which beliefs are most consistent with consumer behavior.

Identifying the type of price beliefs most consistent with consumer behavior within the confines of my model requires me to separately identify consumers' responsiveness to time-series variation in the diesel premium from other sources of fuel price variation which consumers may perceive as more likely to persist throughout the life of a vehicle. This can be accomplished because of the atypically high level of cross-sectional variation in both gasoline prices and the diesel premium within the state of Washington. Whereas the western portion of the state receives fuel primarily from West Coast fuel transportation networks, the eastern portion of the state receives fuel primarily from the Rocky Mountain region's fuel transportation networks. There is no pipeline which crosses the Cascade mountains, which bisect the state from north to south. Transporting fuel across the state is costly as it must either be done via roadways or marine transport via the Columbia River (EIA, 2015). As a result, gasoline prices tend to be higher and the diesel premium lower to the west of the Cascades relative to prices to the east of the Cascades. Persistently higher levels of

fluctuations.

gas prices and a lower diesel premium result in considerably larger operating cost savings from upgrading to a diesel engine. In terms of the lifetime operating cost differential between a gasoline and diesel powered variant of a particular truck, such cross-sectional variation in fuel prices results in thousands of dollars of variation in savings associated with the diesel option. Given that my data contain upwards of 90,000 new truck purchases across the state, this cross-sectional variation in fuel prices allows me to identify consumers' responsiveness to a relatively persistent source of variation in the diesel premium. Given consumers' responsiveness to this source of price variation, the extent to which consumers respond to time-series variation in the diesel premium allows me to identify a crucial parameter attached to price beliefs.

Estimation results convey the importance of tailoring price expectations to the market of interest when considering the consumer response to fuel price fluctuations when multiple fuels enjoy a significant market share. In the context of fuel choice in pickup truck adoption, this entails allowing consumers to respond differently to time-series variation in diesel prices than cross-sectional variation. In my structural model, results suggest consumers are roughly one-tenth as responsive to time-series variation in the diesel premium as they are to all other sources of fuel price variation. My estimates indicate that failing to decompose price beliefs leads to underestimation of consumers responsiveness to fuel prices by around 15-20% across a variety of different measures. Implicit discount rates, computed based on consumers' implied trade-off between a vehicle's purchase price and its yearly operating cost, are three to four percentage points lower in specifications which decompose price beliefs. Likewise, estimates of the vehicle fleet's elasticity of fuel economy, the elasticity of driving, and the elasticity of fuel consumption are roughly 15-20% higher when price beliefs are decomposed to account for the differential response to time-series variation in the diesel premium. Intuitively, the no-change forecast in this environment treats all sources of price variation in the same manner. However, as estimates suggest, consumers are not particularly responsive to time-series variation in the diesel premium. By including a source of price variation which has little effect on consumers' purchasing decisions in the same manner as sources of price variation which have a more significant impact on consumer behavior, the no-change forecast produces a depressed estimate of the consumer response to price fluctuations.

Although the specific nature of the diesel premium's transitory time-series variation need not apply in other settings, this treatment of price expectations embedded in a structural model of vehicle adoption can be employed in any situation where a no-change forecast may not represent a reasonable proxy for price beliefs, given the proper identifying variation. For example, in the case of electric vehicles, one might surmise based on increasing deployment of charging stations that individuals forecast the operating cost of an electric vehicle to decrease over the vehicle's life-cycle. With adequate variation in both (i) installed charger bases at the time of purchase and (ii) rates of charger deployment across locations, one could estimate the importance of both current charging infrastructure as well as expectations over future infrastructure.

In the next section, I present a brief review of the relevant literature. I proceed by discussing the data used in the project and presenting some descriptive evidence on the importance of decomposing fuel prices and accounting for selection on anticipated usage. I follow my descriptive evidence by presenting and estimating a two-period model of vehicle adoption and subsequent usage. Results and counterfactuals are presented in the penultimate section, after which I conclude with a discussion of caveats and suggestions for future research. For the interested reader, the appendices at the end of the document contain additional information regarding the institutional details of the pickup truck market (A.1), my data cleaning procedure (A.2) and (A.3), fuel prices and their properties (A.4), and a variety of different results produced from alternative model specifications (B, D).

SECTION 2

LITERATURE REVIEW

The automotive industry is one of the largest in the United States and has been studied extensively by economists over the past century. An early literature review on the industry was conducted by Griffin (1928), and it has continued to garner interest from economists over the past century. As such, I limit my review to the areas of literature to which this project most directly contributes. The primary contribution of this project lies at the intersection between two related but largely separate branches of research, as I study the consumer response to fuel economy in an environment where multiple fuels enjoy a significant market share. This paper makes a direct contribution to the literature on the consumer response to fuel price fluctuations as well as the literature on fuel choice. In addition to the primary contributions made across these two branches, the findings in this project also have relevance for the literature on fuel taxation as well as an emerging literature which considers the adoption of larger vehicles, a product class which has garnered limited interest in earlier work on the consumer response to fuel price fluctuations.

2.1 The Consumer Response to Fuel Economy

A wide swath of the literature on automobile demand concerns the role of fuel economy in the automobile industry and the consumer valuation of this product characteristic. This literature is thoroughly reviewed by Helfand and Wolverton (2009), so I focus my discussion on more recent research. One of the primary parameters of interest in this literature is the elasticity of the vehicle fleet's fuel economy, which is the percentage change in average miles-per-gallon of a vehicle fleet corresponding to a percentage change in fuel prices. A variety of researchers have produced an estimate of this figure. Li et al. (2009) estimate the long-run elasticity of fleet fuel economy as roughly 0.20 after the entire vehicle stock is replaced; this estimate includes both the effect of gasoline prices on the new vehicle fleet's fuel economy and the differential effect of fuel prices on scrappage decisions of efficient versus inefficient vehicles. In a paper which is methodologically similar to mine, Gillingham (2013) considers the importance of selection into fuel efficient vehicles on anticipated usage when estimating the elasticity of fleet fuel economy. Using data on odometer readings from California's smog check programs, he is able to model selection on anticipated usage and estimates the elasticity of the new vehicle fleet's fuel economy as 0.09. Klier and Linn (2010) exploit a thirty-year panel of fuel prices and vehicle sales and estimate an elasticity of new vehicle fleet fuel economy of 0.12, whereas Austin and Dinan (2005) produce an estimate of 0.22 in a paper which includes a supply-side model of the automobile industry. Gramlich (2009) likewise incorporates a supply-side model where manufacturers can respond to fuel price fluctuations by selecting the fuel economy of their automobiles, and also produces estimates suggestive of a higher elasticity than those papers which do not include a supply-side model.

A related line of literature considers the energy paradox. In this context, this refers to the possibility that consumers undervalue savings in fuel costs relative to savings on vehicles' purchase prices. Historically, researchers were unable to obtain detailed information on transaction prices, instead relying on MSRP's for new vehicle prices and publicly available data sources such as *Kelley Blue Book* or the *National Automobile Dealers Association*'s used car pricing guides for used vehicle prices (e.g. Kahn, 1986). Because auto dealers can discount inefficient vehicles in response to rising gasoline prices, this results in measurement error which is correlated with fuel prices when MSRP is used as a proxy for transaction prices. As such, studies which lack transaction prices may find evidence that consumers undervalue fuel economy. More recent studies in this area have used transaction-level price data and have found mixed evidence on the energy paradox. Busse, Knittel, and Zettelmeyer (2013) gather a rich database of transaction-level prices for both new and used cars and find little evidence of consumer myopia in either market. Allcott and Wozny (2014) similarly use transaction-level price data and estimate that consumers are indifferent between an increase of $1.00 in discounted future gasoline costs and $0.76 in the vehicle purchase price, results which are indicative of an energy paradox in fuel economy. Most recently, Sallee et al. (2016) use time-series variation in used vehicle prices and estimate that this variation is consistent with full valuation of fuel economy.

2.2 Automobile Fuel Choice

Owing to the dominance of gasoline-powered automobiles in the U.S. and the fact that availability of vehicles with multiple fuel offerings was limited until recent years, the literature on fuel choice in the U.S. automobile market is somewhat limited. Early research studying fuel choice in the U.S. market, such as Brownstone et al. (1996, 2000) and Brownstone and Train (1999), relies in part on stated preference data over hypothetical alternative-fuel vehicles due to the absence of available alternative-fuel vehicles. More recently, researchers have considered the role of network effects and purchase subsidies on the adoption of newer alternative-fuel technologies. Corts (2010) and Shriver (2015) study the impact of flex-fuel vehicle adoption on the subsequent development of E85 retail infrastructure, and estimate that adoption of flex-fuel vehicles increases the number of local fuel retailers offering E85 fuel. However, the market for these vehicles differs from many alternative-fuel adoption choices as the fuel choices made by drivers of flex-fuel vehicles are made at the pump rather than at the point of vehicle purchase. Further, evidence indicates that even among those who do own a flex-fuel vehicle, the fuel is not widely used (Anderson and Sallee, 2011). The role of network effects and purchase subsidies in electric vehicle (EV) adoption is considered by Li et al. (2015), while Holland et al. (2015) document the importance of regional heterogeneity in environmental impacts associated with EV adoption when determining optimal purchase subsidies.

In countries where the penetration of non-gasoline powered personal vehicles is more significant, consumers often do face a fuel choice in automobile adoption. As such, there is a larger literature considering the impact of fuel prices on fuel choice outside of the U.S. market. Yeh (2007) considers various policies which may drive cross-country differentials in natural gas vehicle adoption, estimating that the price of natural gas would need to be roughly 50% cheaper than gasoline or diesel to lead to widespread adoption in the U.S.[1] The choice between gasoline and diesel vehicles has been studied by Verboven (2002) and Linn (2014). Verboven (2002) studies

[1] Historically, natural gas is only about 20% cheaper than gasoline and diesel. Many personal vehicles in the U.S. which run on natural gas are dual fuel vehicles which can also run on gasoline.

quality-based price discrimination in the European automobile market by considering consumers' choices over gasoline and diesel engines. Verboven's work differs from mine for a few reasons. Most importantly, Verboven's work is aimed towards studying price discrimination, where the more efficient diesel engines represent a higher quality variant of the vehicle. While fuel prices enter the model of engine choice, he does not aim to identify the impact of separate sources of price variation or explicitly model price beliefs. Whereas my work relies on both time-series and cross-sectional variation in the incentives to adopt a diesel engine as a means to identify responses to these different types of price variation, Verboven relies on cross-sectional policy differences between different nations to identify consumers' preferences over fuel efficiency. Linn (2014) considers the determinants of cross-country variation in diesel adoption rates, and finds evidence that underlying heterogeneity in consumer preferences over fuel economy explains a significant share of the variation.

2.3 The Elasticity of Demand for Gasoline

There is a rich literature built around estimating the price elasticity of demand for gasoline, owing to the importance of gasoline consumption to climate change, energy security, and its role in the production process of a wide variety of firms. While my work allows me specifically to produce estimates of the elasticity of fuel consumption for pickup truck drivers, it is worthwhile to be aware of broader estimates of this demand elasticity. Espey (1998) conducts a meta-analysis of the elasticity of gasoline demand which acts to summarize studies conducted between 1966 and 1997. She documents a median estimate of -0.23 for the short-run elasticity and a median estimate of -0.43 for the long-run elasticity of demand for gasoline. Recent literature, however, suggests that the demand for gasoline has become more inelastic since the turn of the millennium. Hughes et al. (2008) estimate that whereas the short-run price elasticity of gasoline demand lay between -0.21 and -0.34 between 1975 to 1980, this price elasticity had fallen to an estimated range between -0.03 and -0.08 between 2001 and 2006.

Related to broader work on the elasticity of demand for gasoline, other research considers the elasticity of demand for driving (a key component of the demand for gasoline), and researchers have found this elasticity has decreased as well in recent years. Small and Van Dender (2007) document a declining rebound effect, wherein individuals who purchase more energy-efficient products increase their utilization. Over the course of a thirty-five year panel spanning 1966-2001, they estimate a short-run (long-run) rebound effect of 0.045 (0.222). Estimates from this paper using the higher real income levels and lower real gasoline prices towards the end of the sample yield estimates roughly half the size of their estimates from the full period, again suggesting the consumer response has moderated in recent years. Whereas the rebound effect measures the impact of a decrease in operating costs on driving behavior resulting from higher fuel efficiency, other work considers the impact of decreased operating costs on driving behavior which results from lower gasoline prices. Linn (2013) documents that the magnitude can differ significantly, as he estimates an elasticity of vehicle miles traveled with respect to gasoline prices between -0.09 and -0.20 whereas he estimates a rebound effect between 0.20 and 0.40. In addition to producing an estimate of the elasticity of fleet fuel economy, Gillingham (2013) estimates the elasticity of demand for driving as -0.15 among a large sample of California drivers, whereas Gillingham (2014) produces a demand elasticity of -0.22 while also documenting significant heterogeneity in the elasticity across various demographics.

2.4 Fuel Prices and Behavior of Heavy Vehicle Operators

An additional contribution to the literature is my consideration of vehicles which fall into a class which is not typically studied in the literature on automobile demand. Much of the literature in the U.S. uses fuel economy ratings produced by the EPA, which are unavailable for some pickup trucks, while the literature on trucks is limited to large, commercial vehicles. Although vehicles without a fuel economy rating only represent roughly one quarter of the vehicles in my sample, my use of user-generated fuel economy ratings may be of particular use to those researchers studying behavior of commercial motor vehicle operators. Within this literature, Parry (2008) considers tax policy towards heavy-duty trucks and estimates optimal taxation structures for heavy-duty trucks, while Luechinger and Roth (2016) estimate the effects of a mileage tax implemented in Switzerland on truck traffic. Most closely related to this paper in the literature on truck drivers' behavior is recent work by Leard et al. (2015), who estimate the rebound effect for different types of heavy-duty trucks. Owing to the lack of fuel economy data for their vehicles, they use self-reported measures from the Vehicle Use and Inventory Survey.[2] However, the age of the data, which were last collected in 2002, limits its applicability to future studies in this area. A case-study approach is taken by Klemich et al. (2014) to study the energy paradox among motor carriers. They document a significant use of a wide variety of fuel saving tactics by motor carriers, and offer a discussion of the determinants for whether particular technologies were adopted. One of the most significant challenges in this area is limited *public* availability of data regarding fuel economy and vehicle usage, and while my work does not directly contribute to this line of research on commercial vehicles usage, my use of user-generated data on fuel economy could be of interest to researchers looking to further explore this topic.

[2]Leard et al. do not consider pickup trucks due to the limited sample observed in their micro-data. Therefore, vehicles studied by Leard et al. are almost exclusively large commercial vehicles, whereas I exclude commercial fleets from my sample.

SECTION 3

DATA

Throughout this project, I use two primary datasets which I appended with data from additional sources. The Washington Department of Licensing (DOL) provides a database of all vehicle transactions within the state of Washington from January 1 2012 through December 31 2015. From this database, I extract information on new vehicle purchases, as well as driving intensity from a subset of these purchases. To measure fuel prices, I acquired a database of daily, retailer-level gasoline and diesel prices in the state of Washington from January 1 2012 through December 31 2015 from Oil Price Information Service (OPIS). I appended these two datasets with data from additional sources. I use fuel economy measures provided by two sources, and I present a comparison of the two different measures in §3.3. Local demographics at the zip-code level come from the American Community Survey (ACS) as provided by the American Fact Finder web page.[1] The Western Regional Climate Center provides county-level weather data. For each county in the state of Washington, I extract the average yearly snow and rainfall from a weather station within the county of interest.[2]

3.1 Vehicle Purchases

In the registration database provided to me by the DOL, I observe all vehicle transactions which occur in Washington, hence an observation is generated whenever any of a variety of transactions (new vehicle registration, vehicle transfer, registration renewal, title transfer, etc.) occurs. For each event, I observe the Vehicle Identification Number (VIN), type of owner (e.g. personal, business), zip code and county of registration, and in certain cases an odometer reading at the time of the

[1]Some zip codes are sparsely populated and ACS estimates are unavailable at such a fine level for these zip codes. In these cases, I impute county-level demographics as provided by the ACS.

[2]I use a weather station in the county's largest city if one is provided. If the largest city does not have a weather station which provides data, I use another weather station in the county which is at a similar elevation.

transaction. Due to the nature of identification throughout this manuscript, it is critical that the observed registration date lies close to the date the vehicle was purchased. This is generally the case with new vehicle purchases in Washington, as the dealer is required to file the registration documents with the DOL shortly after the purchase. From this registration database, I attempt to extract all registrations which are associated with a new pickup truck purchase. While the process is thoroughly described in §A.2, I outline the process here, which beings with a database of all VINs which have an associated transaction in the state from 2012 through 2015:

1. I extract all VINs from Model Years 2000-2016 where the DOL has recorded a vehicle model which could be a pickup truck.

2. For each unique VIN stub (the first ten letters of the VIN), I extract one observation and use online VIN decoding software to attach relevant vehicle characteristics to the VIN stub.[3]

3. Vehicle characteristics are merged onto all observations with the same VIN stub.

4. Observations which pertain to a new pickup truck registration are retained.

5. I drop rarely chosen vehicle specifications, vehicles which are registered to locations with insufficient fuel price information, vehicles with an unclear owner type, and vehicles larger than the 'one-ton' pickup class.

I describe this process at length in §A.2. The sample initially contains a total of 128,174 observations, which is reduced to 124,310 observations after dropping 3,864 observations during step (5) of the cleaning process. A wide variety of vehicle characteristics is identifiable through the VIN, and my decoding software populated the vast majority of observations in my sample with a rich set of product characteristics. In the event of decoding errors or missing information, I manually corrected vehicle characteristics based on information provided by a variety of sources

[3]To verify the accuracy of VIN decodes, I manually decoded a subset of observations using the National Insurance Crim Bureau's Passenger Vehicle Identification Manuals.

which cover the automobile industry.[4] A total of 24 different models of pickup trucks were available during this time period, and their market shares for each of the four relevant owner types are listed in Table 3.1. The majority of new pickup trucks are personal registrations (71.3%), with business registrations accounting for the next largest share of purchases (17.9%). Owners with at least five vehicles are allowed to register their vehicles as a fleet, in which case these registrations are recorded as business registrations. Vehicles owned by small businesses which operate fewer than five vehicles are recorded as personal registrations by the DOL. The remaining ten percent of observations are categorized either as leased vehicles or exempt from registration fees, where the latter generally refers to vehicles which are owned by municipal, state, or federal government agencies. The majority of analysis in this project is limited to personal registrations, although descriptive evidence in the next section is presented for the full sample as well. The analysis of business and government registrations is limited due to the fact that many of these vehicles are purchased by fleet managers rather than individuals; analysis of leased vehicle choices is limited because I do not observe a proxy for prices faced by lessees.

Pickup truck models are typically offered in many permutations due to different engine, drivetrain, body style, trim, etc. options. This presents a challenge when it comes to defining choice sets, and in order to facilitate my discrete-choice model in §5.1, I aggregate vehicles up to the level of Model, Series, Generation, Engine, Cab Type, Drivetrain, Wheel Configuration, and Work Truck Trim. Notably, I do not include all trim levels and transmission specifications, as VINs do not always uniquely identify these characteristics. Manufacturer's Suggested Retail Prices (MSRPs) are used as a proxy for transaction prices. I further discuss the implications of using MSRPs rather than transaction prices in §5.2.5.

Tables 3.2 and 3.3 list summary statistics for the estimation sample used throughout §4, presented for both the full sample of registrations as well as the sample of personal registrations. I stratify the sample by broadly-defined truck classes, where the mid-size category includes sport-utility-trucks (SUT) and mid-size trucks, the full-size category includes only half-ton trucks, and

[4]For example, a few vehicles which were decoded were populated with an incorrect engine displacement. The primary sources used to correct this information were autoblog.com and caranddriver.com.

the heavy-duty category includes both three-quarter ton and one-ton trucks. The most popular class of pickup truck is the full-size truck, which includes vehicles such as th Ford F-150 and Ram 1500. Such full-size trucks are usually offered with a range of gasoline-powered engines which allow consumers to weigh the importance of fuel economy against performance and the engine's upgrade cost.[5] These engines are typically more powerful than engines offered in cars and SUVs, as can be noted by the fact that full-size trucks in my full sample produce an average of 355 horsepower and 379 pound-feet of torque. Prior to recent diesel engine offerings in the Ram 1500, Chevrolet Colorado, GMC Canyon, and Nissan Titan, diesel engines had historically only been available as a costly upgrade in heavy-duty trucks.[6] The base engine offering on heavy-duty trucks is typically an eight-cylinder gasoline engine, with a high-performance diesel engine upgrade available at a cost of roughly $6,000 - $9,000. The diesel engines offered in heavy-duty trucks offer a massive performance boost (on average, these engines produce around 800 pound-feet of torque) in addition to improving fuel economy by 15-30% relative to their gasoline counterparts. These engines are popular despite the cost, as roughly three quarters of all heavy-duty trucks purchased in the full sample are equipped with a diesel engine.

Two additional upgrades that are typically selected by consumers are the upgrade to a four-door body and a drivetrain which sends power to all four wheels (4WD). The base body type for most trucks is the 'regular cab', which is a two-door body with either two or three seats in a single row. Roughly 26% of trucks in my full sample are upgraded to the 'extended cab' configuration with a second row of seats, whereas about two thirds are upgraded to the 'crew cab' configuration which offers more legroom in the second row of seats and four full doors.[7] These upgrades carry

[5]For example, the base engine offering on the 2015 Ram 1500 was a V6 which produced 269 pound-feet of torque and yielded 19.0 miles-per-gallon (MPG). For an additional $1,150, consumers could upgrade to a V8 gasoline engine which produces 410 pound-feet of torque but only 15.3 MPG. The Ram 1500 is also the first full-size truck to offer a diesel engine; this upgrade was priced at $4,270 over the base V6, but the diesel V6 produced 420 pound-feet of torque and offered fuel economy of 22.6 MPG.

[6]The Colorado and re-badged Canyon were offered with a diesel engine beginning in late 2015. The most recent generation of the Nissan Titan, which is classified somewhere between the full-size and heavy-duty classes, is also being offered with a diesel engine option.

[7]Many 'extended cab' trucks have front-hinged doors for passengers to access the back seats.

a significant cost; upgrading to an extended cab typically costs around \$2,000 - \$3,000 whereas upgrading to a crew cab typically costs around \$5,000 - \$6,000. The upgrade to a 4WD truck is similarly popular, as roughly 90% of trucks in the full sample are equipped with the upgraded drivetrain despite an upgrade cost which typically ranges between \$3,000 and \$4,000. As shown in Table 3.4, there is significant variation in uptake of the Crew Cab and Four Wheel Drive options across vehicle models.

3.1.1 Odometer Readings

One of the difficulties facing researchers who study fuel choice is the possibility that individuals with a high expected usage level self-select into more fuel efficient engines. Ideally, the researcher would observe high-frequency data on odometer readings, but this data is generally unavailable. In the vehicle registration data provided by the DOL, I observe an odometer reading whenever a transaction other than a registration renewal is made with the DOL. Therefore, for any vehicle which was initially purchased after 01/01/2012 and transferred to a different party by 12/31/2015, I am able to construct a measure of yearly vehicle miles traveled (VMT) while the vehicle was owned by the initial purchaser. I observe a sample of 7,040 odometer readings out of 88,623 personally registered new truck purchases in my sample.

As shown in Table 3.5, the distribution of VMT for trucks varies across truck classes and fuel types (within truck class). For each class-fuel type combination, the distribution of VMT can be reasonably approximated using a lognormal distribution. Figure 3.1 shows the fitted log-normal distribution of VMT for diesel-powered heavy-duty trucks. Consistent with the distribution of VMT observed by Gillingham (2013), there is a high level of variation in usage, with an interquartile range of 8,810 miles to 19,330 miles for drivers of diesel heavy-duty trucks. Kernel density estimates of VMT by fuel type for heavy-duty trucks shown in Figure 3.2 illustrate the role of selection on anticipated usage in fuel choice. On average, heavy-duty trucks powered by diesel engines are driven roughly 25% more miles per year than heavy-duty trucks powered by gasoline engines.

Because of the non-random nature in which odometer readings are observed, one might expect to see differences in product and demographic characteristics between vehicles which carry an odometer reading and those that don't. As shown in Table 3.6, these differences are modest. When the sample is broken into three subsamples based on length of vehicle ownership, vehicle characteristics and demographic characteristics are similar across the three subsamples. It should be noted that heavy-duty trucks, and particularly diesel powered heavy-duty trucks, are somewhat less likely to carry an odometer reading than other trucks. By far the strongest determinant of whether an odometer reading is recorded is the vehicle's initial purchase date, which can be seen in Table 3.7, where I present the results of a logistic regression of an odometer reading dummy variable on a variety of demographic characteristics. This relationship motivates a selection adjustment I use in §5.2.4 to account for non-random entry of odometer readings when estimating my structural model.

3.2 Fuel Prices

Daily, retailer-level fuel prices are provided by OPIS, who use a variety of means to gather data at the retailer-level on fuel prices in the United States, including fleet cards carried by motor carriers, data scraping, and agreements with individual retailers wherein the retailer will update OPIS with their prices on a frequent basis. As a result of their sampling methodology the quantity of retailers which report a price varies across the sample, but reporting rates are reasonably high and improve throughout the course of the sample. I observe an average of 1,968 retailers in a given day reporting a gasoline price. The count of retailers reporting a gasoline price ranges from a minimum of 1,225 to a maximum of 2,184 throughout the sample. Diesel prices are not reported as frequently, but coverage is still sufficient for the purposes of this project. I observe an average of 1,021 retailers reporting a diesel price on a given day, with the observation count ranging from 384 retailers to 1,298 retailers. It should be noted that the proportion of retailers which sell diesel is significantly higher than the ratio of daily gasoline retailers reporting a price to daily diesel retailers reporting a price, and it appears that the vast majority of retailers in the state sell both gasoline and diesel (see §A.3 for more details).[8] The breadth of this data allows me to observe daily fuel prices across the entire state of Washington, which allows me to exploit significant cross-sectional

[8]For this reason, I do not consider the impact of fuel availability on fuel choice throughout the paper.

variation in fuel prices, in addition to time-series variation which is typically seen in a panel of this length. For this project's purposes, I aggregate fuel prices up to the level of daily county averages for both gasoline and diesel. The aggregation process, and my handling of missing data throughout the course of my sample in certain low population counties, is also described further in §A.3 for the interested reader.

Two important phenomena present in the diesel premium are worth discussing as a means to motivate my empirical analysis. First, the unique geography and structure of the refined products pipelines servicing the state of Washington creates significant geographical variation in the diesel premium. Counties to the east of the Cascade Mountains source most of their gasoline and diesel from refineries in the Rocky Mountain region, where gasoline prices are relatively low but the diesel premium is relatively high. Refineries located in Northwest Washington supply most of the gasoline and diesel consumed to the west of the Cascades. In this area of the state, gasoline tends to be relatively expensive while the diesel premium is relatively low. This combination of higher gasoline prices and a lower diesel premium results in significantly larger operating cost reductions associated with upgrading to the diesel engine. For reference, the location of pipelines and refineries in Washington can be seen in Figure 3.3. In Figure 3.4, I present the diesel premium over the course of my sample for two counties on opposite sides of the Cascade Mountains. Whatcom county, located to the west of the Cascade Mountains in the Northwest corner of the state, typically has a significantly lower premium on diesel than Spokane county, which is located to the east of the Cascade Mountains on the Eastern border of the state. Diesel prices are similar in each out these counties, and the difference in the premium is due to lower gasoline prices in Spokane county, as can be seen in Figure 3.5 and 3.6. While these per-gallon differences in the premium between different geographic regions in the state may appear modest, they can result in significant differences in the expected fuel cost savings associated with upgrading to a diesel engine. For example, consider the fuel cost savings over the lifetime of a particular truck which might result from upgrading to a diesel engine. Suppose that an individual faces the choice between a gasoline engine which achieves 12 miles per gallon versus a diesel upgrade which achieves 15 miles per gallon.[9] Figure 3.6 presents the operating cost savings associated with upgrading to the diesel engine and driving 150,000 miles at each county's fuel price over the course of the vehicle's life, were the current prices to persist. Depending on the an individual's location and the date of purchase, operating cost savings associated with adopting the diesel engine range between \$0 and \$8,000. Comparing different regions of the state, an individual living in Whatcom county typically stands to save between \$1,500 and \$2,500 more by upgrading to the diesel engine than an individual in Spokane county.

Second, within each county, the diesel premium (as well as the savings associated with upgrading to the diesel engine) exhibits mean-reversion to the county's average diesel premium. Gasoline and diesel prices are highly correlated over time as the primary determinant of each fuel's price is the price of crude oil. However, the diesel premium fluctuates throughout the year due to factors such as seasonality in diesel and gasoline demand, short-term refinery outages, and a variety of other minor factors. Given the durable nature of automobiles, it stands to reason that a forward-looking individual ought to be relatively irresponsive to time-series variation in the diesel premium. For example, consider a consumer who is deciding between the gasoline and diesel variant of a particular truck in Spokane County during July 2015. At this time, the diesel premium was roughly \$0.00. If this consumer is aware of the history of the diesel premium, her expectation of the diesel premium over the life of the vehicle should not be \$0.00, as it should account for the fact that the average premium in Spokane County has been around \$0.50 per gallon. Simply put, forward-looking individuals ought to respond differently to such fuel price variation than they do to more persistent sources of fuel price variation. In §4 and §5 I illustrate the importance of allowing consumers to exhibit a different response to different sources of price variation. For the interested reader, I document the stationarity of the diesel premium for a significantly longer panel of fuel prices in §A.5.[10]

---

[9]These fuel economy figures are similar to those of the gasoline and diesel variants of the Chevrolet Silverado 2500.

[10]I do not have access to county-level data prior to 01/01/2012, so I use the longer panel which corresponds to

3.3 Comparison of User-Generated Fuel Economy Data to EPA Measures

Researchers studying the automobile market in the United States typical use the Environmental Protection Agency's (EPA) fuel economy ratings as a measure of fuel consumption.[11] For each vehicle measured, the EPA releases a separate estimate for city and highway driving, and reports a combined fuel economy measure which is computed as a weighted harmonic mean using 55% city driving and 45% highway driving. These estimates are meant to inform consumers of the fuel efficiency of vehicles being considered for purchase, and in recent years manufacturers have been required by the EPA to post window stickers containing this information on new vehicles. Over the past few decades, a variety of adjustments have been made in methodology for measuring fuel economy with hopes of providing consumers with estimates that are predictive of the vehicles' fuel consumption. For example, beginning with MY2008, the EPA adjusted their methodology to better account for real-world acceleration, air conditioning usage, and climate conditions. Despite these changes, researchers have documented significant differences between the EPA's fuel economy measures and those recorded by individuals operating the vehicles ('Self-Reported Fuel Economy'). Greene and Lin (2011) consider the relationship between the EPA measures and self-reported measures which were recorded by individuals on the EPA's "My MPG" system. On this platform, users can record their fuel expenditures and compute their fuel economy, along with recording a variety of individual-level characteristics. When comparing self-reported fuel economy to those which use the EPA's 2008 methodology, they note that "...the 2008 estimates underestimate the on-road fuel economy by 14%-16% depending on efficiency level" (Greene and Lin, p. 89).

At the time of this project, the EPA did not produce fuel economy ratings for pickup trucks with a gross vehicle weight rating over 8,500 pounds. As mentioned earlier in this section, roughly one quarter of all new trucks purchased in my sample are heavy-duty trucks which lie above the 8,500 pound threshold and as such, are not rated by the EPA. In order to measure the fuel economy

---

broader geographic areas to illustrate the mean-reverting nature of the diesel premium.

[11] As mentioned in the previous section, Leard et al. (2015) use self-reported fuel economy data when considering the rebound effect for heavy-duty trucks.

of these vehicles, I employ self-reported fuel economy data from fuelly.com, an online platform where individuals can track vehicles' fuel economy. The platform is well populated, and most truck models in my sample have hundreds of vehicles being tracked by consumers. While the EPA's "My MPG" system is still on-line, I chose to use fuelly for two important reasons. First, fuelly has a significantly higher number of pickup trucks being tracked. For example, as of 04/22/2017, a total of 114 MY2016 Ford F-150s equipped with the V8 engine option were being tracked on fuelly, whereas only one such vehicle was being tracked on the "My MPG" platform. Second, the "My EPA" system does not allow drivers to track vehicles which are not tested by the EPA, whereas fuelly allows drivers to track a wide variety of vehicles. This allows me to construct measures of fuel economy for heavy-duty pickup trucks.

For a total of seventy-four unique combinations of make-model-series-generation-engine within my sample from the previous section, I extracted self-reported fuel economy data from fuelly.com during June, 2016. For each make-model-series-model year-engine combination, I recorded the number of vehicles being tracked, the total quantity of miles driven, and the average fuel economy, each of which can be easily scraped from fuelly's interface (for an example, see Figure 3.8).[12] I then aggregated this information up from the level of model year to the level of generation in order to increase the sample size of vehicles being tracked.[13] Due to the lack of large samples of certain trucks equipped with 2WD, I aggregate fuel economy ratings at the level of make-model-series-generation-engine. Rather than disaggregating down the the level of drivetrain to adjust fuel economy for 4WD trucks, I use differences in EPA ratings for 2WD versus 4WD trucks to adjust vehicles' fuel economy accordingly (see §A.3 for details).

Abstracting from differences between 'real-world' driving styles and the EPA's attempts to approximate 'real-world' driving styles in their fuel economy measures, one might imagine two

[12]During MY 2015 and MY 2016, the Ford F-150 was offered with two different 3.5 Liter V6 engines, an entry-level 282 horsepower engine and an upgraded 'EcoBoost' engine which is rated at 365 horsepower. Because fuelly users typically only enter engines at the level of displacement and cylinder configuration, I am unable to separate self-reported fuel economy for these two engines in either the 2015 or 2016 model year. In my next chapter, I use the EPA measurements for these configurations.

[13]Generation is a well-defined concept in the automobile industry. Roughly speaking, each 'generation' corresponds to a significant redesign and adjustment to product characteristics.

primary reasons self-reported fuel economy could differ from estimates produced by the EPA. First, selection into tracking fuel expenditures and fuel economy is presumably not random. For example, those individuals who are more interested in tracking their fuel economy may also be more likely to drive in a manner which increases fuel efficiency.[14] Such a selection mechanism could possibly explain the findings of Greene and Lin (2011), who report that individuals on the "My MPG" platform tend to outperform EPA estimates. Second, the nature of 'real-world' driving conditions can be more challenging to define with pickup trucks than other vehicles. Pickup trucks can be used for a wide variety of pleasure and work-related tasks, and fuel economy can decrease significantly when a truck is used to haul a significant amount of payload or tow a heavy trailer. In particular, we might expect that if the propensity to haul payload or tow is higher for larger pickup trucks than smaller pickup trucks, the self-reported measures might be significantly lower for large pickup trucks than the EPA measures.

A descriptive analysis of my self-reported measures of fuel economy (SRP) suggest that despite the aforementioned concerns, the two different sets of fuel economy measures are strongly correlated. In Figure 3.9, I present a histogram of the differentials between the EPA estimates and the SRP estimates pulled from fuelly. Whether the differentials are computed prior to adjusting for drivetrain differences or after adjusting for differences in drivetrain, the differences are modest. The vast majority of observations lie within ten percent of their corresponding EPA estimates, with no difference more than three miles per gallon. As an additional exercise, I run an OLS regression of the EPA's fuel economy measurements on my SRP measures:

$$SRP_j = \beta_0 + \beta_1 EPA_j + \epsilon_j$$

$$\widetilde{SRP}_j = \beta_0 + \beta_1 EPA_j + \epsilon_j$$

where SRP_j refers to my fuel economy measurement which is not adjusted for different drivetrains

[14]Fuel economy gains from 'hypermiling' techniques such as coasting into turns, slowly accelerating, and driving without air conditioning can range upwards of 25%. In some extreme cases, hypermilers are able to achieve massive efficiency gains. "Lowest fuel consumption driving to 48 contiguous US States (diesel car)." *Guinness World Records*, 07 July 2015, http://www.guinnessworldrecords.com/world-records/lowest-fuel-consumption-48-us-states

and $\widetilde{SRP}_j$ refers to my fuel economy measure which adjusts for different drivetrains. I run each regression for five different sets of observations based on the quantity of vehicles I observed (N_j) and the quantity of miles which were recorded by the vehicles (VMT_j), and present the results in Table 3.8. My measure which adjusts for drivetrain exhibits a mildly higher correlation with the EPA measures, although both measures have a reasonably strong fit, as estimates using SRP_j result in an R^2 between 0.750 and 0.788, and estimates using $\widetilde{SRP}_j$ result in an R^2 between 0.787 and 0.820. Unlike Greene and Lin (2011), my regression coefficients do not show evidence of a systemic difference between the EPA's measures and my SRP estimates. In the five specifications using SRP_j the coefficient on the EPA measure ranges between 0.911 and 0.946 whereas the estimated intercept lies between 0.573 and 1.260. Results are similar when the $\widetilde{SRP}_j$ measure is used, with the coefficient on the EPA measure ranging between 0.951 and 0.975 and the intercept ranging from 0.088 to 0.602. In all ten specifications, ninety-five percent confidence intervals for the coefficients contain both $\beta_0 = 0$ and $\beta_1 = 1$. Thus, the systemic difference between SRP and EPA measures reported in Greene and Lin (2011) does not show up here, possibly allaying concerns about the correlation between drivers' likelihood of tracking their fuel economy and the extent to which they engage in fuel efficient driving practices.

As shown in Figures 3.10 and 3.11, the linear specification in the preceding tables appears to be reasonable, and the variance of the residuals does not vary significantly based on the EPA's fuel economy estimate. Although there is some concern that EPA measures might overestimate the real-world fuel economy of larger trucks if such trucks are more likely to be hauling payload or towing large trailers, such a bias does not manifest itself in my data. While I cannot similarly consider the accuracy of my fuel economy measurements for those trucks which are not tested by the EPA, the evidence presented in this section is suggestive of the possibility that SRP measures could offer a plausible means of capturing fuel economy for such vehicles. For the remainder of this project, I employ the SRP measures which are adjusted for different drivetrains.

Table 3.1: Market Shares (%), by Owner Type

	All	**Personal**	**Business**	**Leased**	**Exempt Owner**
Observations	**124,310**	**88,667**	**22,301**	**9,236**	**4,106**
Avalanche	0.37	0.43	0.28	0.12	0.02
Canyon	0.28	0.29	0.33	0.18	0.00
Colorado	1.57	1.17	1.88	2.65	6.26
Dakota	0.02	0.01	0.06	0.00	0.00
Escalade EXT	0.05	0.05	0.08	0.05	0.00
F-150	19.50	17.64	22.64	26.72	26.35
F-150 SVT	0.64	0.72	0.65	0.12	0.02
F-250	3.87	2.34	6.41	4.60	21.53
F-350	5.09	4.70	6.35	1.78	14.15
Frontier	3.04	2.88	3.52	3.81	1.92
Ram 1500	11.16	11.76	10.21	8.02	10.42
Ram 2500	4.56	5.25	4.07	0.89	0.63
Ram 3500	2.84	3.46	1.99	0.09	0.19
Ranger	0.25	0.24	0.29	0.31	0.05
Ridgeline	0.79	0.98	0.18	0.80	0.00
Sierra 1500	3.18	3.24	3.66	2.79	0.24
Sierra 2500	1.58	1.64	2.13	0.35	0.07
Sierra 3500	0.50	0.56	0.58	0.04	0.00
Silverado 1500	12.64	12.07	15.44	12.57	10.08
Silverado 2500	4.48	3.92	6.93	3.76	4.90
Silverado 3500	2.13	1.94	3.58	0.64	1.88
Tacoma	13.09	15.43	4.88	15.73	1.17
Titan	0.53	0.60	0.52	0.19	0.10
Tundra	7.81	8.68	3.33	13.77	0.00

Table 3.2: Selected Descriptive Statistics: All Registration Types

	All Trucks	**Mid Size**	**Full Size**	**Heavy Duty**
Observations	**124,310**	**24,165**	**68,979**	**31,116**
MSRP	$37,226	$27,079	$36,728	$46,197
Regular Cab	0.07	0.05	0.06	0.09
Extended Cab	0.26	0.31	0.30	0.14
Crew Cab	0.67	0.64	0.64	0.77
Four Wheel Drive	0.90	0.81	0.92	0.93
Dual Rear Wheels	0.01	0.00	0.00	0.06
Work Truck Trim	0.13	0.12	0.13	0.14
Length (Feet)	22.69	20.76	22.83	23.86
Width (Feet)	7.88	7.40	7.97	8.04
Height (Feet)	7.52	6.98	7.57	7.85
Diesel Engine	0.19	0.00	0.02	0.72
Gas: Torque (100/lb-ft)	349.30	250.19	378.79	393.84
Diesel: Torque (100/lb-ft)	783.30	NA	420.00	801.48
Gas: Horsepower/100	326.96	230.71	355.07	374.16
Diesel: Horsepower/100	381.65	NA	240	388.74
Gas: Miles/Gallon	16.20	18.49	15.91	12.11
Diesel: Miles/Gallon	14.55	NA	22.57	14.15

Table 3.3: Selected Descriptive Statistics: Personal Registrations

	All Trucks	**Mid Size**	**Full Size**	**Heavy Duty**
Observations	**88,623**	**19,041**	**48,508**	**21,118**
MSRP	$37,624	$27,388	$37,111	$48,033
Regular Cab	0.03	0.03	0.04	0.01
Extended Cab	0.23	0.30	0.27	0.09
Crew Cab	0.74	0.67	0.69	0.90
Four Wheel Drive	0.94	0.85	0.96	0.99
Dual Rear Wheels	0.01	0.00	0.00	0.05
Work Truck Trim	0.10	0.10	0.11	0.09
Length (Feet)	22.70	20.79	22.89	24.00
Width (Feet)	7.86	7.42	7.97	8.01
Height (Feet)	7.52	6.99	7.57	7.86
Diesel Engine	0.21	0.00	0.02	0.85
Gas: Torque (100/lb-ft)	348.50	253.58	383.55	391.53
Diesel: Torque (100/lb-ft)	782.88	NA	420	801.73
Gas: Horsepower/100	325.18	233.08	358.86	372.01
Diesel: Horsepower/100	379.85	NA	240	387.12
Gas: Miles/Gallon	16.31	18.32	15.78	12.15
Diesel: Miles/Gallon	14.57	NA	22.57	14.16

Table 3.4: Selected Characteristics of Pickup Truck Models

Model	Make	Class	Model Years	$MSRP	Diesel Uptake	Gas MPG	Diesel MPG	Crew Cab	4WD	Footprint (sqft)
Avalanche	Chevrolet	SUT	2011-2013	$48,576		14.31		1.00	1.00	175.05
Canyon	GMC	Midsize	2011-2012, 2014-2016	$33,769		19.39		0.89	1.00	154.24
Colorado	Chevrolet	Midsize	2011-2012, 2014-2016	$27,940		19.87		0.52	0.65	150.84
Dakota	Dodge	Midsize	2011	$27,360		16.18		0.00	1.00	156.66
Escalade EXT	Cadillac	SUT	2011-2013	$69,640		14.20		1.00	1.00	175.60
F-150	Ford	$\frac{1}{2}$-Ton	2011-2016	$37,913		16.28		0.62	0.89	183.05
F-150 SVT	Ford	$\frac{1}{2}$-Ton	2011-2014	$46,161		13.18		0.91	1.00	199.36
F-250	Ford	$\frac{3}{4}$-Ton	2011-2016	$43,080	0.43	12.47	14.50	0.48	0.84	189.27
F-350	Ford	1-Ton	2011-2016	$48,278	0.81	11.32	13.26	0.79	0.92	198.06
Frontier	Nissan	Midsize	2011-2016	$24,752		18.12		0.62	0.78	150.35
Ram 1500	RAM	$\frac{1}{2}$-Ton	2011-2016	$35,923	0.17	15.69	22.57	0.50	0.94	180.98
Ram 2500	RAM	$\frac{3}{4}$-Ton	2011-2016	$46,446	0.82	12.98	15.18	0.97	0.99	188.44
Ram 3500	RAM	1-Ton	2011-2016	$49,034	1.00		13.20	1.00	1.00	191.82
Ranger	Ford	Compact	2011	$22,557		20.22		0.00	0.34	139.45
Ridgeline	Honda	SUT	2011-2014	$35,004		17.37		1.00	1.00	160.97
Sierra 1500	GMC	$\frac{1}{2}$-Ton	2011-2016	$39,972		16.40		0.68	0.95	183.02
Sierra 2500	GMC	$\frac{3}{4}$-Ton	2011-2016	$47,374	0.70	12.13	14.90	0.80	0.99	190.79
Sierra 3500	GMC	1-Ton	2011-2016	$51,618	1.00		14.01	1.00	1.00	192.56
Silverado 1500	Chevrolet	$\frac{1}{2}$-Ton	2011-2016	$36,632		16.40		0.52	0.87	179.68
Silverado 2500	Chevrolet	$\frac{3}{4}$-Ton	2011-2016	$43,390	0.51	12.11	14.90	0.65	0.94	189.29
Silverado 3500	Chevrolet	1-Ton	2011-2016	$46,310	0.79	10.51	13.80	0.72	0.86	194.66
Tacoma	Toyota	Midsize	2011-2016	$26,217		18.56		0.63	0.82	154.01
Titan	Nissan	$\frac{1}{2}$-Ton	2011-2015	$35,554		13.61		0.84	1.00	178.56
Tundra	Toyota	$\frac{1}{2}$-Ton	2011-2016	$33,087		14.69		1.00	0.99	182.83

Table 3.5: Yearly Vehicle Miles Traveled

Class	**Fuel**	**Observations**	μ_{VMT}	σ_{VMT}
Compact	Gasoline	1,465	11,966	6,863
Compact	Diesel	0	NA	NA
Sport Utility Truck	Gasoline	145	12,244	6,053
Sport Utility Truck	Diesel	0	NA	NA
Half Ton	Gasoline	4,087	13,068	7,016
Half Ton	Diesel	5	13,842	11,517
Three-Quarter Ton	Gasoline	214	12,486	7,283
Three-Quarter Ton	Diesel	558	14,946	8,434
One Ton	Gasoline	23	10,828	9,749
One Ton	Diesel	543	14,705	8,506

Table 3.6: Summary Statistics, by Length of Ownership

Variable	μ_A	σ_A	μ_1	σ_1	μ_2	σ_2
Observations	81583	NA	4016	NA	3024	NA
Purchase Date	2.228	1.13	1.281	0.738	0.700	0.482
Elapsed Years	NA	NA	1.472	0.29	2.606	0.45
Odometer Reading Time	NA	NA	2.753	0.728	3.305	0.479
Yearly VMT	NA	NA	13.013	7.297	13.151	7.270
Rain, ft./year	2.735	1.315	2.723	1.376	2.709	1.369
Snow, ft./year	0.766	0.713	0.766	0.696	0.782	0.724
Commute Length, Minutes	2.625	0.632	2.616	0.634	2.632	0.648
Median Household Income	6.339	1.887	6.241	1.829	6.282	1.889
Rural Population Share	0.237	0.3	0.227	0.292	0.228	0.288
Unemployment Rate	9.168	3.21	9.225	3.337	9.214	3.175
Compact	0.2	0.4	0.205	0.404	0.213	0.409
Sport Utility Truck	0.014	0.116	0.019	0.137	0.022	0.148
Half Ton	0.544	0.498	0.585	0.493	0.577	0.494
Three-Quarter Ton	0.134	0.34	0.113	0.316	0.105	0.307
One Ton	0.109	0.312	0.079	0.269	0.083	0.275
Crew Cab	0.74	0.439	0.703	0.457	0.695	0.461
Four Wheel Drive	0.941	0.236	0.942	0.233	0.933	0.25
Diesel Engine	0.217	0.412	0.158	0.365	0.156	0.363
Diesel Engine, if Heavy-Duty	0.848	0.359	0.819	0.385	0.828	0.378
Miles/Gallon	15.958	2.105	15.743	1.909	15.65	1.948

[a] Ownership A indicates that I do not observe a vehicle transaction which requires an odometer reading to be taken. Option 1 indicates that between 366 days and 730 days elapsed between a vehicle's original purchase date and my observed odometer reading. Option 2 indicates that greater than 731 days elapsed between a vehicle's original purchase date and my observed odometer reading.

[b] Purchase date gives elapsed time, in years after January 1, 2012, at the time of vehicle purchase. Elapsed years gives number of years the vehicle was driven before observing an odometer reading. Odometer Reading Time gives elapsed time, in years after January 1, 2012, at the time of the odometer reading.

[c] Demographics are specified at the zip-code level, while weather data are specified at the county-level. Demographic data are only available for the bulk of zip-codes if ACS five-year averages are used, hence these data are drawn from 2010-2014 and as a result, unemployment figures are relatively high.

Table 3.7: Logit: Odometer Reading Dummy

Variable	**Specification 1**	**Specification 2**
Purchase Date	-1.076 (0.015)	-1.086 (0.015)
Rain, ft./year	-0.015 (0.012)	-0.011 (0.012)
Snow, ft./year	0.031 (0.023)	0.031 (0.023)
Commute Length, Minutes	0.114 (0.026)	0.109 (0.026)
Median Household Income	-0.037 (0.008)	-0.038 (0.008)
Rural Population Share	-0.232 (0.052)	-0.267 (0.052)
Unemployment Rate	0.006 (0.004)	0.006 (0.004)
Sport Utility Truck	-0.233 (0.100)	-
Half Ton	-0.041 (0.053)	-
Three-Quarter Ton	-0.244 (0.115)	-
One Ton	-0.484 (0.151)	-
Crew Cab	-0.007 (0.031)	-
Four Wheel Drive	-0.040 (0.066)	-
Diesel Engine	-0.061 (0.091)	-
Miles/Gallon	-0.045 (0.015)	-
Intercept	0.069 (0.315)	-0.770 (0.088)

Table 3.8: OLS Regressions of EPA Combined Fuel Economy on Self-Reported Measures

Measure	**Sample**	**Observations**	**Slope (β_1)**	**Intercept (β_0)**	R^2
SRP	$N_j > 0$	47	0.911 (0.064)	1.26 (1.038)	0.782
SRP	$N_j > 10$	46	0.919 (0.064)	1.097 (1.035)	0.788
SRP	$N_j > 20$	40	0.946 (0.082)	0.573 (1.350)	0.750
SRP	$VMT_j > 100K$	45	0.922 (0.068)	1.059 (1.096)	0.780
SRP	$VMT_j > 250K$	38	0.938 (0.081)	0.759 (1.328)	0.750
$\widetilde{SRP}$	$N_j > 0$	71	0.951 (0.052)	0.602 (0.872)	0.816
$\widetilde{SRP}$	$N_j > 10$	70	0.958 (0.052)	0.459 (0.872)	0.820
$\widetilde{SRP}$	$N_j > 20$	61	0.975 (0.067)	0.088 (1.132)	0.787
$\widetilde{SRP}$	$VMT_j > 100K$	68	0.955 (0.055)	0.511 (0.924)	0.812
$\widetilde{SRP}$	$VMT_j > 250K$	58	0.960 (0.066)	0.415 (1.113)	0.787

Figure 3.1: Lognormal Fit of Diesel VMT Distribution

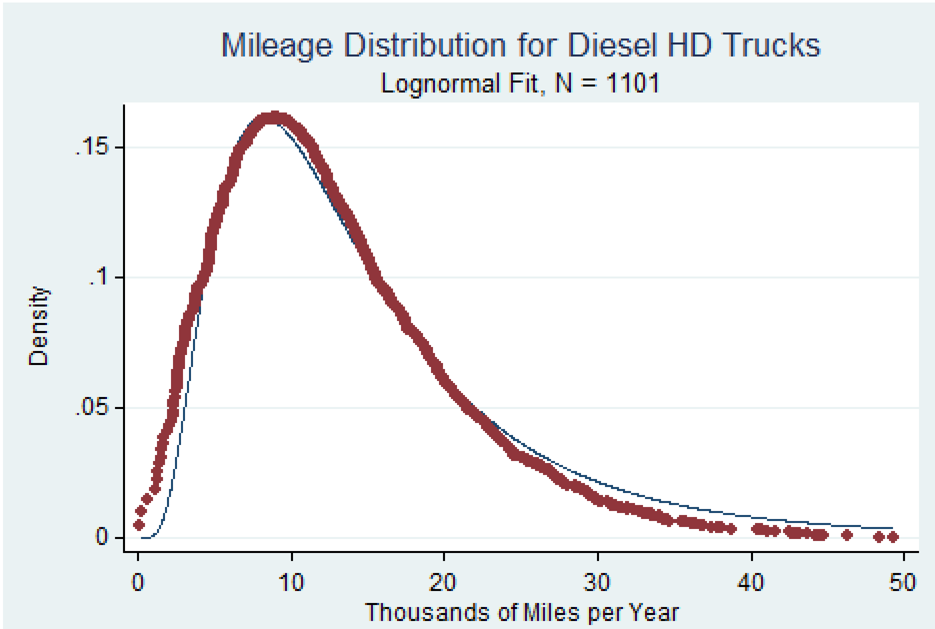


Figure 3.2: Comparison of Gasoline and Diesel VMT Distributions

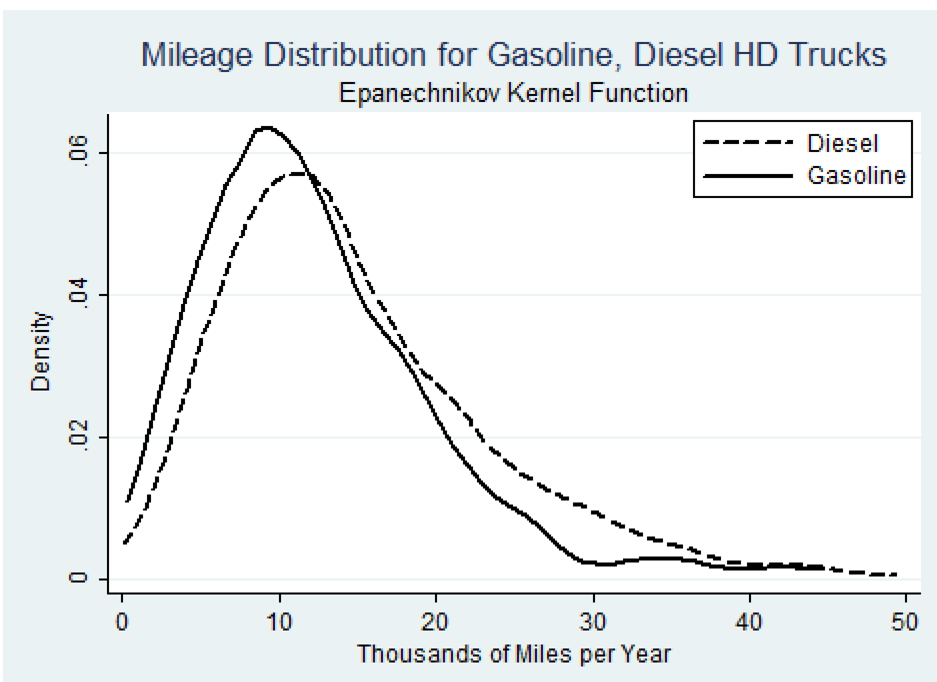
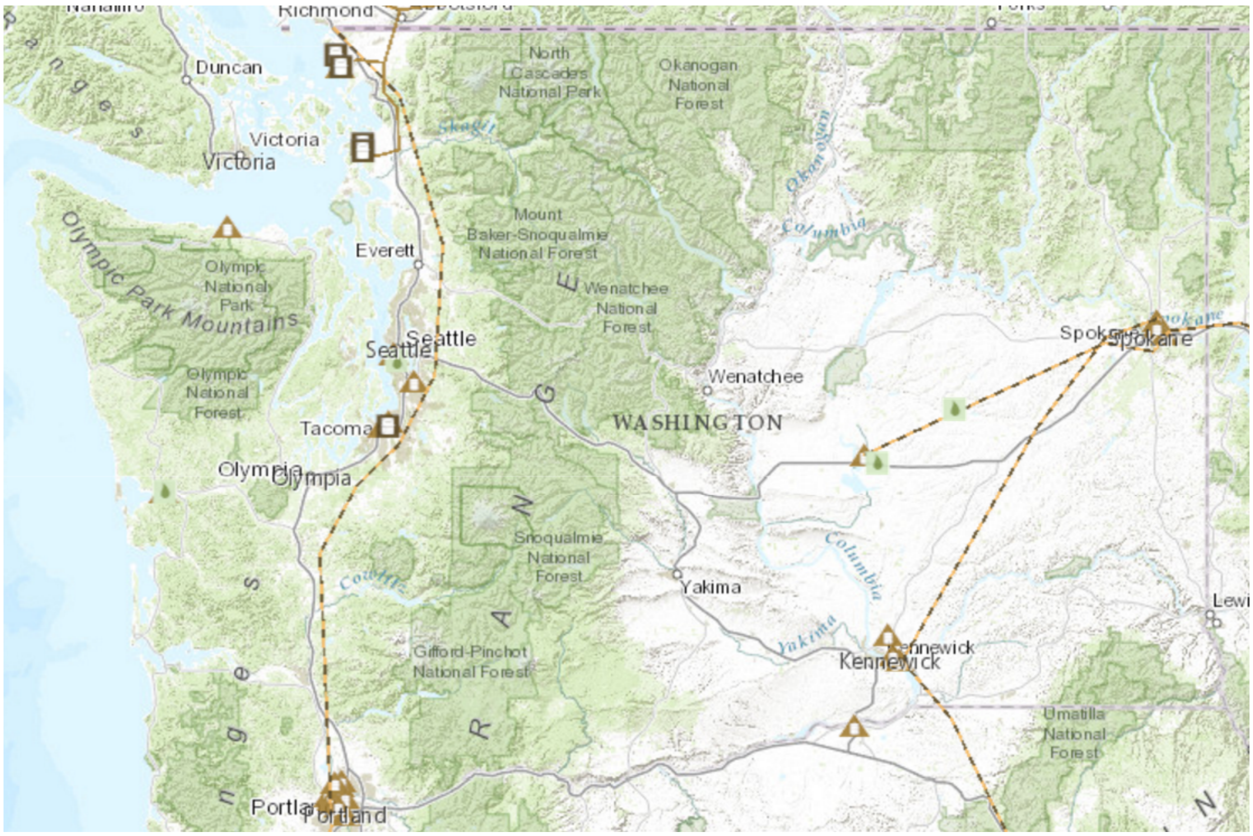


Figure 3.3: Topographical Map of Washington with Oil Distribution Network



Source: Energy Information Administration. Dashed pipelines are used to distribute gasoline, diesel, and other refined products, while the solid pipeline in the Northwest of Washington is used to distribute crude oil. The square 'oil-barrel' figures denote oil refineries, the triangle 'oil-barrel' figures denote petroleum product terminals, and the square 'droplet' figures denote bio-diesel plants.

Figure 3.4: Price Difference between Gasoline and Diesel, 01/01/2012-12/31/2015

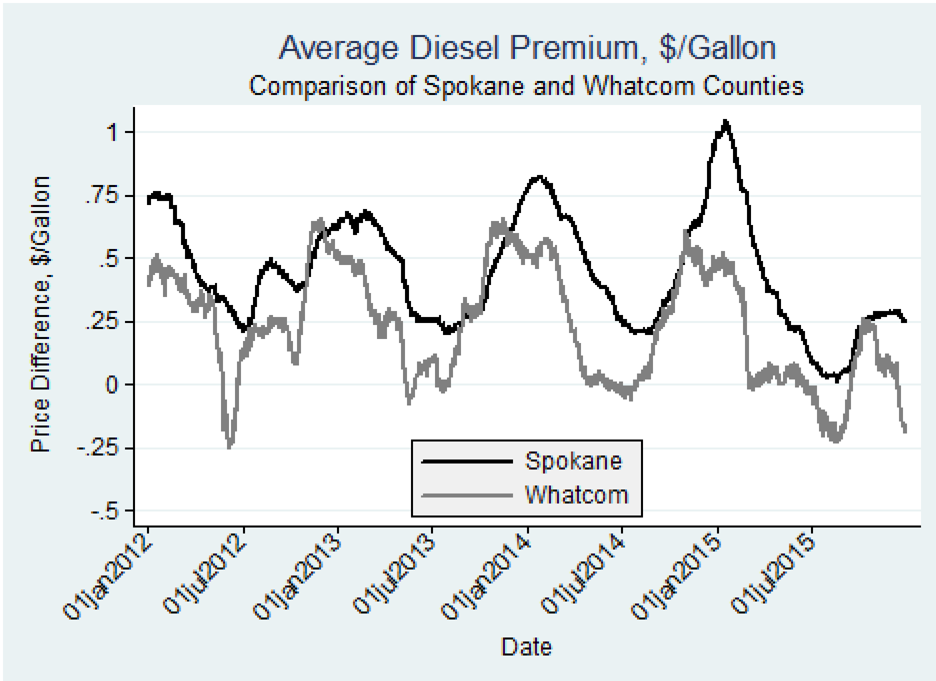


Figure 3.5: Gasoline Price, 01/01/2012-12/31/2015

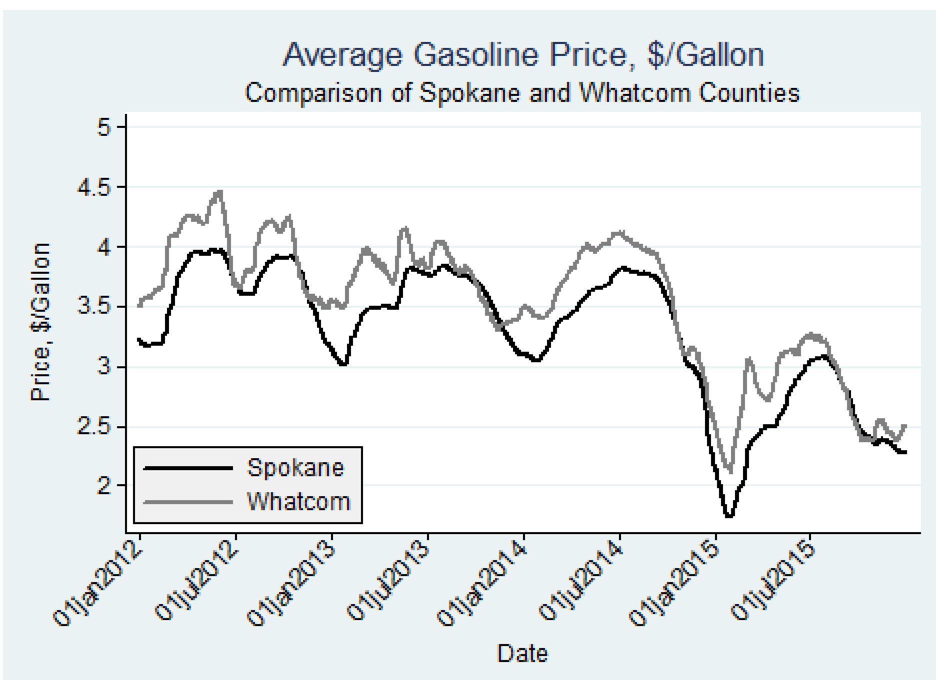


Figure 3.6: Diesel Price, 01/01/2012-12/31/2015

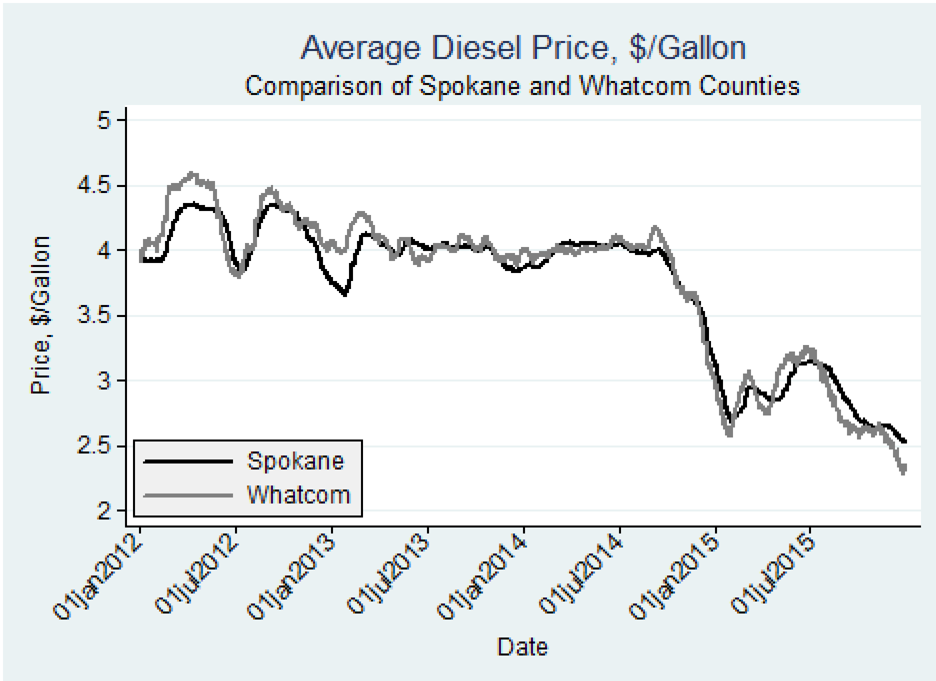


Figure 3.7: Example of Savings from Diesel Engine Upgrade

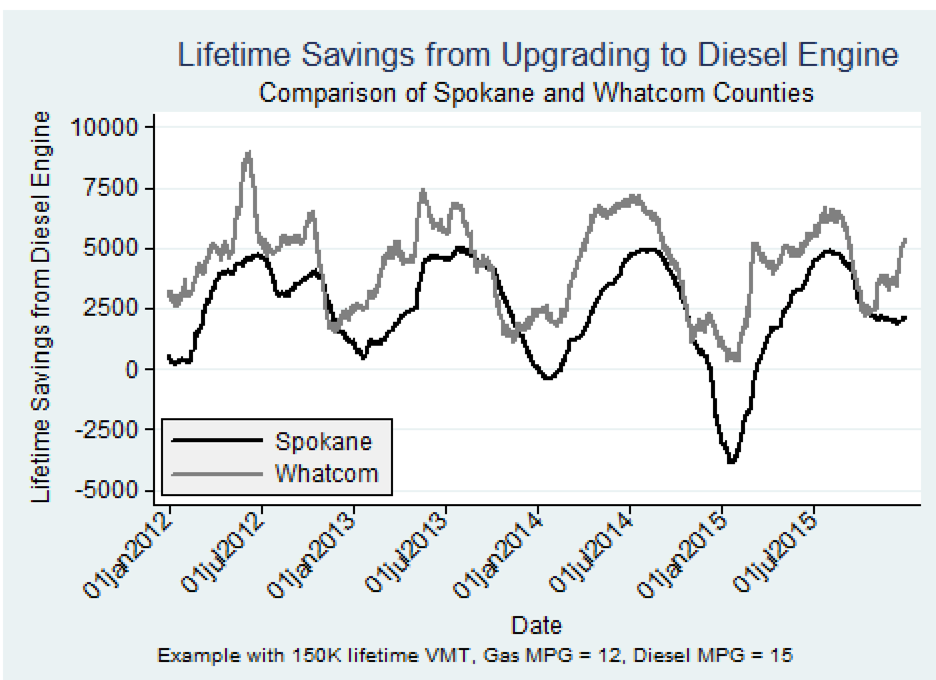


Figure 3.8: Screenshot of Fuelly Interface

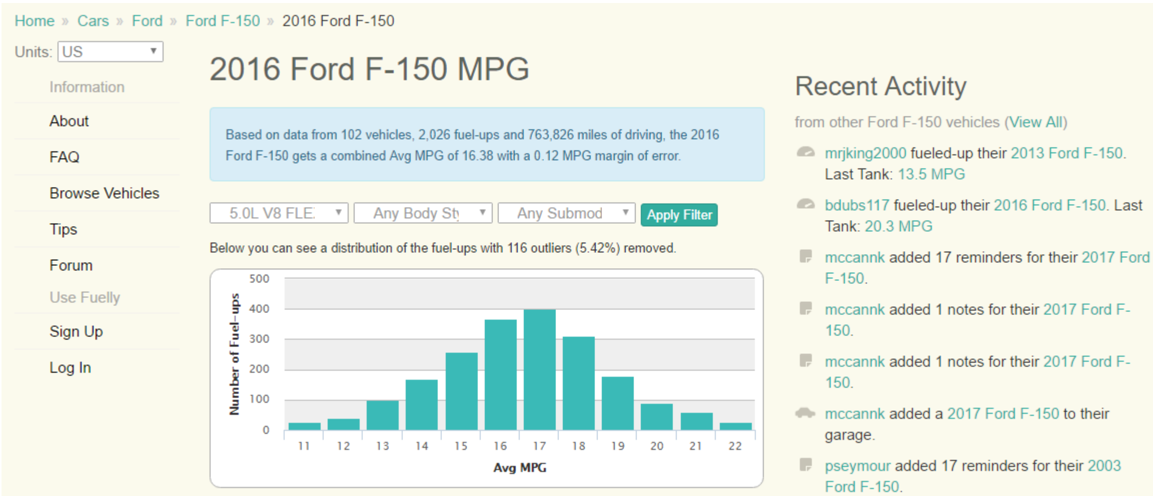


Figure 3.9: Differences between Self-Reported and EPA Fuel Economy Estimates

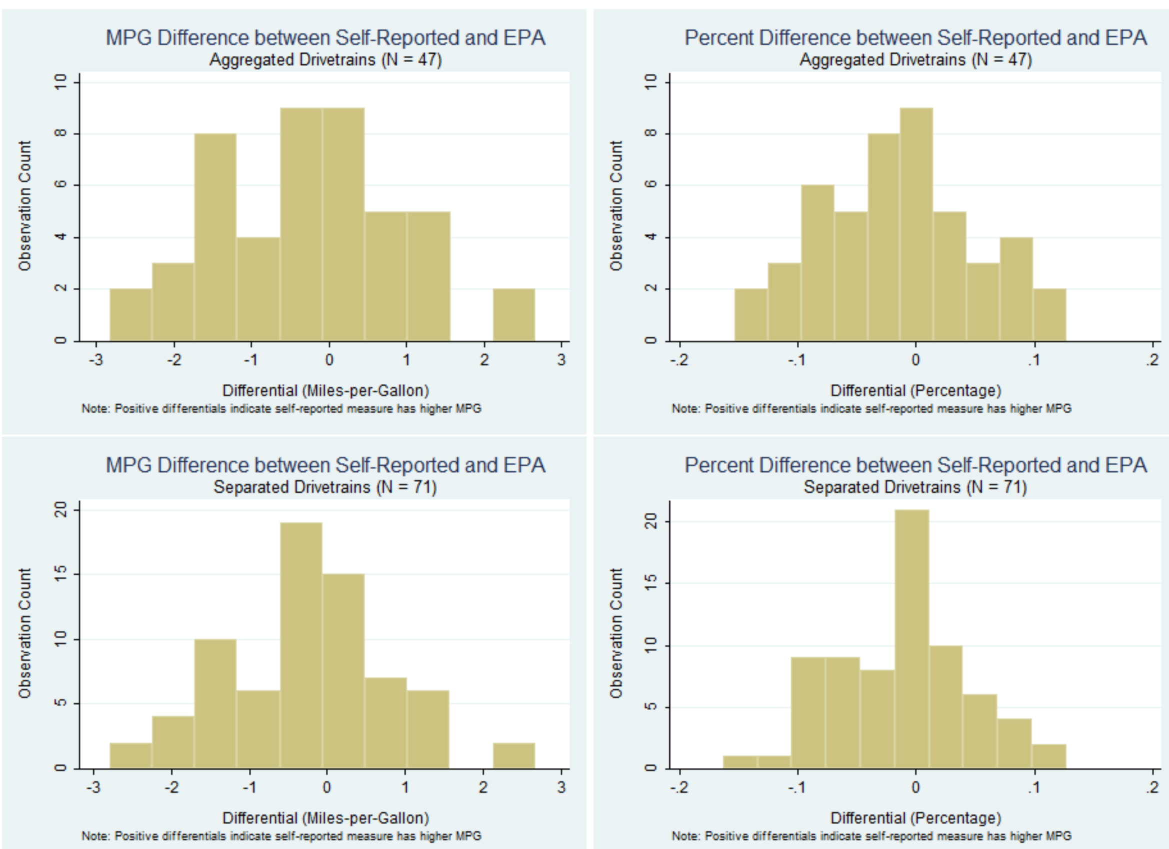


Figure 3.10: Scatterplot of Self-Reported and EPA Measures (A)

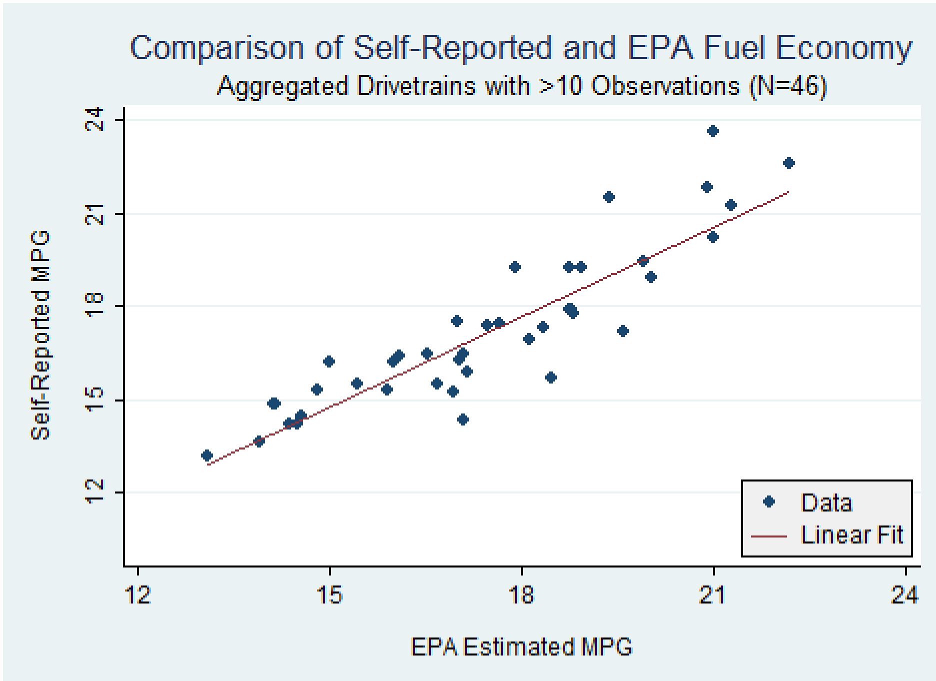
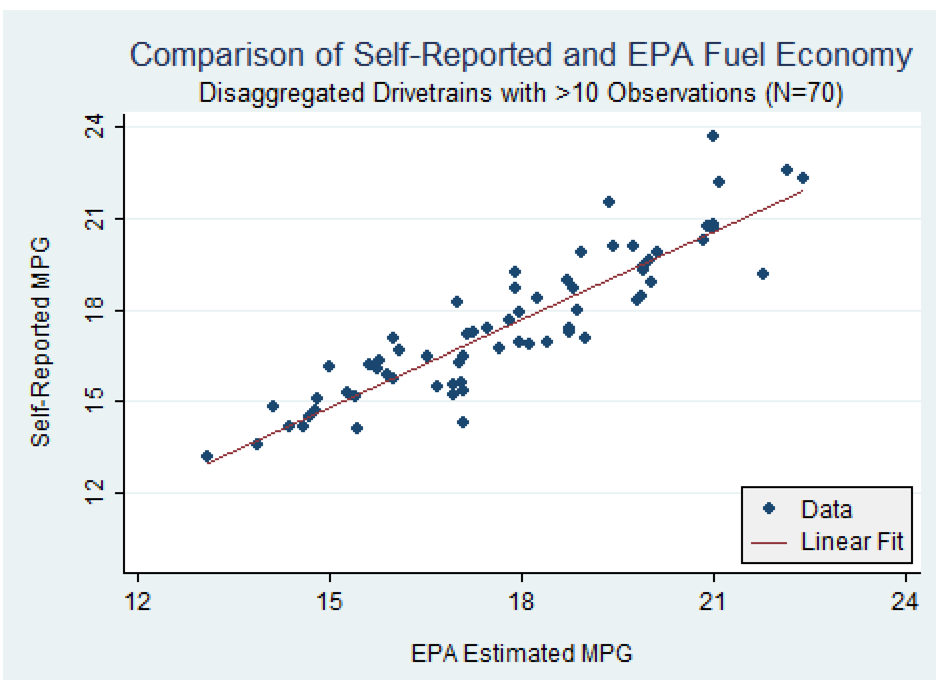


Figure 3.11: Scatterplot of Self-Reported and EPA Measures (B)



SECTION 4

DESCRIPTIVE ANALYSIS

In this section I present descriptive evidence on the determinants of fuel choice and subsequent usage intensity. This subsection is not intended to present or discuss any causal effects, but merely to illustrate patterns present in my data as a means of motivating my structural model, which is presented in §5.

4.1 Descriptive Evidence: Fuel Choice

As discussed in the introduction to this paper, the limited adoption of vehicles operating on any fuel other than gasoline has posed challenges for researchers studying fuel choice among U.S. automobile purchasers. The market for pickup trucks has a significantly higher adoption rate of non-gasoline powertrains, as diesel engines have been available in heavy-duty pickup trucks for several decades. In this subsection, I present a descriptive analysis of the determinants of fuel choice among purchasers of heavy-duty trucks. This analysis considers the determinants of fuel choice *conditional on purchasing a heavy-duty truck*, and as this explicitly ignores substitution between other classes of pickup truck and heavy-duty trucks, the usual caveats apply.[1] I estimate a linear-in-parameters binary logit model of the following form:

$$Pr(Diesel|Heavy - Duty) = \frac{exp(\beta_0 + \beta_1 Z_i + \beta_2 P_i)}{1 + exp(\beta_0 + \beta_1 Z_i + \beta_2 P_i)}$$

where Z_i refers to a vector of local taste-shifters and P_i refers to a vector of fuel price data. Results pertaining to the sample which consists of all registration types ($N = 31,166$) are presented in Table 4.1, while results pertaining to the sample which consists of only personal registrations ($N = 21,118$) are presented in Table 4.2. In the appendix, Table B.1 gives results from the sample

[1]The structural model presented in the next chapter allows for substitution between pickup truck classes.

of business registrations ($N = 7,147$). The primary parameters of interest are those attached to fuel prices. In specification (1) for each sample, I include the county-average gasoline price ($p^g_{C,t}$) and county-average diesel price differential ($p^d_{C,t} - p^g_{C,t}$).[2] In both samples with this specification, the marginal effects corresponding to the fuel price parameters are modest; in the sample of personal registrations the estimated marginal effect of increasing the price of both fuels by $1.00 (hence, leaving the difference unchanged) is a 2.2 percentage point increase in diesel engine uptake. Given that increasing the price of both fuels increases the savings associated with adopting the diesel engine, this marginal effect is intuitively plausible. The point estimate attached the price differential is small but statistically insignificant, the marginal effect is likewise small but statistically insignificant. Recalling the mean-reverting nature of the diesel premium over time, the small and imprecise point estimate is consistent with the possibility of forward-looking individuals not adjusting their beliefs about the diesel premium in response to transient variation in the premium.

In a second set of regressions, I construct fuel prices in an alternative manner to allow for different responses to transient sources of diesel price variation as opposed to more persistent sources of diesel price variation. In §3.2, I discuss the nature of the diesel price premium, and note that whereas time-series variation within a particular location is transitory, geographic variation across counties is reasonably persistent owing to the pipeline network in Washington. I decompose the current diesel price in a given location as

$$p^d_{C,t} = \underbrace{(p^g_{C,t} + \tau_C)}_{\equiv \bar{p}^d_{C,t}} + \tilde{p}^d_{C,t}$$

where for each county, I define $\tau_C = \frac{1}{T}\sum_{t=1}^{T}(p^d_{C,t} - p^g_{C,t})$ as the average price diesel premium within each county over the course of my sample. By construction, $\tilde{p}^d_{C,t}$ is mean-zero for each county, but varies over time. In specification (2) in Table 4.1 and 4.2, I include each component of the decomposed diesel price in the regression. Focusing on the marginal effects in the personal

[2] I define fuel prices on the date I observe the registration, which may lag the purchase date by a few days. Further, this could lag the date when the consumer made her purchase decision by more than a week. Given the rate at which fuel prices move, I don't believe this generates an overly concerning level of measurement error.

registration results from Table 4.2, the importance of allowing for different responses to different sources of fuel price variation is apparent. By construction, a \$1.00 increase in both the gasoline and diesel prices raises $p^g_{C,t}$ and $\bar{p}^d_{C,t}$ by \$1.00. The marginal effects corresponding to each parameter estimate indicate a similar response to such a price change as we saw in specification (1), with a 2.9 percentage point increase in diesel engine adoption. Fixing the gasoline price, the coefficient and marginal effect attached to $\bar{p}^d_{C,t}$ allows us to consider the relationship between τ_C and diesel engine uptake. This estimated marginal effect is large and statistically significant, and suggests that a \$1.00 increase in a county's average diesel premium is associated with a 25.0 percentage point decrease in diesel engine uptake. The point estimate attached to $\tilde{p}^d_{C,t}$ curiously has a positive sign; however, the marginal effect is statistically insignificant.

Across both specifications and samples, point estimates attached to local characteristics have the anticipated signs. Counties with a higher share of rural residents have significantly higher adoption rates of diesel engines, all else held constant. However, the point estimate and estimated marginal effect of the impact of rural population share on the probability of selecting a diesel engine are significantly lower among personal registrants than business registrants (see Table B.1 in the appendix for the results pertaining to business registrations). Given that the diesel engine is a costly upgrade on the order of \$6,000 - \$9,000, the positive point estimate on the local median income and negative point estimate on the local unemployment rate are generally reasonable.

While the analysis contained in this subsection is intended to be read as descriptive rather than causal, the results are crucial for motivating my construction of price beliefs in the structural model presented in the next chapter. To the extent that consumers purchasing a heavy-duty truck recognize that the time-series variation in the diesel premium is transitory, one would expect that forward-looking individuals would be largely irresponsive to such price variation. On the other hand, given the more persistent nature of geographic variation in the diesel price premium, forward-looking individuals ought to respond to such price variation when choosing between the two fuels.

4.2 Descriptive Evidence: Vehicle Miles Traveled

I now consider the determinants of usage intensity among pickup truck drivers. As mentioned earlier in the data section, I observe a non-random sample of 9,822 odometer readings attached to my new vehicle purchases, of which 7,071 correspond to personal registrations and 1,749 correspond to business registrations. For the descriptive results presented here, I do not account for selection into reporting an odometer reading and treat odometer readings as if they are missing-at-random. In the next chapter, I address the possibility of non-random entry with a selection correction. However, the reader should note that whereas the previous exercise largely serves to motivate my treatment of price beliefs in the next chapter, the exercises presented here are suggestive of the importance of accounting for selection on anticipated usage in the next chapter. For each observation in my sample, I define the usage intensity as the average yearly vehicle miles traveled between the two odometer readings (VMT_{ij}). I estimate an ordinary least squares regression of the form

$$\ln VMT_{ij} = \beta_0 + \beta_1 Z_i + \beta_2 X_j + \beta_3 \ln \overline{OC}_{ij} + \epsilon_{ij}$$

where Z_i represents a vector of local taste-shifters, X_j represents a vector of vehicle characteristics, and $\overline{OC}_{ij}$ represents individual i's average per-mile fuel cost of operating vehicle j.[3] The results of these regressions are presented in Table 4.3 for the full registration sample and Table 4.4 for the personal registration sample. In specifications (1) and (2) of Table 4.3 the reader can note that in absence of conditioning on vehicle characteristics, there is a positive association between operating costs and yearly mileage. This result is suggestive of the possibility that individuals with a high anticipated usage level self-select into larger vehicles with higher levels of performance, a feature which I account for in my structural model. Once controlling for a variety of vehicle characteristics, the anticipated negative coefficient attached to operating costs is observed. While this is a crude measure of the elasticity of driving to operating costs owing to the lack of controls for selection into vehicle type and selection into reporting an odometer reading, the estimates

[3] Hence, I am conditioning on individual i having purchased vehicle j.

contained in columns (3) and (4) lie within the range of estimates in the literature discussed in §2.3.

Table 4.1: Logistic Regression of Fuel Choice in Heavy-Duty Trucks: All Registrations

Variable	**(1)**	**(2)**
$p^g_{C,t}$	0.260 (0.047)	3.044 (0.272)
$p^d_{C,t} - p^g_{C,t}$	0.026 (0.076)	
$\bar{p}^d_{C,t}$		-2.778 (0.276)
$\tilde{p}^d_{C,t}$		0.205 (0.078)
Rain, ft./year	-0.004 (0.001)	-0.009 (0.001)
Snow, ft./year	-0.000 (0.002)	-0.001 (0.002)
Rural Population Percentage	0.926 (0.049)	0.837 (0.049)
Unemployment Rate	-0.025 (0.005)	-0.029 (0.005)
Median Income, $10K	0.117 (0.010)	0.119 (0.010)
Years Elapsed from 01/01/2012	0.112 (0.023)	0.114 (0.023)
Constant	-6.573 (1.420)	-5.597 (1.429)
Marginal Effects	**(1)**	**(2)**
$p^g_{C,t}$	0.052 (0.009)	0.611 (0.054)
$p^d_{C,t} - p^g_{C,t}$	0.005 (0.015)	
$\bar{p}^d_{C,t}$		-0.557 (0.055)
$\tilde{p}^d_{C,t}$		0.041 (0.016)
Rain, ft./year	0.001 (0.000)	-0.002 (0.000)
Snow, ft./year	0.000 (0.000)	0.000 (0.000)
Rural Population Percentage	0.186 (0.010)	0.017 (0.010)
Unemployment Rate	-0.005 (0.001)	-0.006 (0.001)
Median Income, $10K	0.024 (0.002)	0.024 (0.002)
Years Elapsed from 01/01/2012	0.022 (0.005)	0.023 (0.005)
(Negative) Pseudo-LL	18215	18157

Table 4.2: Logistic Regression of Fuel Choice in Heavy-Duty Trucks: Personal Registrations

Variable	**(1)**	**(2)**
$p^g_{C,t}$	0.173 (0.090)	2.162 (0.400)
$p^d_{C,t} - p^g_{C,t}$	0.046 (0.113)	
$\bar{p}^d_{C,t}$		-1.935 (0.406)
$\tilde{p}^d_{C,t}$		0.179 (0.116)
Rain, ft./year	-0.004 (0.001)	-0.007 (0.002)
Snow, ft./year	-0.002 (0.003)	-0.001 (0.003)
Rural Population Percentage	0.193 (0.066)	0.142 (0.066)
Unemployment Rate	-0.006 (0.007)	-0.010 (0.007)
Median Income, \$10K	0.053 (0.015)	0.055 (0.015)
Years Elapsed from 01/01/2012	0.070 (0.035)	0.074 (0.035)
Constant	-2.999 (2.141)	-2.484 (2.149)
Marginal Effects	**(1)**	**(2)**
$p^g_{C,t}$	0.022 (0.012)	0.279 (0.051)
$p^d_{C,t} - p^g_{C,t}$	0.006 (0.015)	
$\bar{p}^d_{C,t}$		-0.250 (0.052)
$\tilde{p}^d_{C,t}$		0.023 (0.015)
Rain, ft./year	0.000 (0.000)	-0.001 (0.000)
Snow, ft./year	0.000 (0.000)	0.000 (0.000)
Rural Population Percentage	0.025 (0.009)	0.018 (0.009)
Unemployment Rate	-0.001 (0.001)	-0.001 (0.001)
Median Income, \$10K	0.007 (0.002)	0.007 (0.002)
Years Elapsed from 01/01/2012	0.009 (0.005)	0.010 (0.005)
(Negative) Pseudo-LL	9027	9014

Table 4.3: OLS Regression of Log Yearly Mileage: Full Registration Sample

Variable	**(1)**	**(2)**	**(3)**	**(4)**
Log Operating Cost	0.329 (0.085)	0.303 (0.046)	-0.152 (0.067)	-0.180 (0.068)
Rain, ft./year	-0.002 (0.000)	0.019 (0.007)	-0.001 (0.000)	0.020 (0.007)
Snow, ft./year	-0.002 (0.001)	-0.028 (0.004)	-0.002 (0.001)	-0.028 (0.004)
Rural Population Percentage	0.162 (0.026)	0.162 (0.029)	0.147 (0.026)	0.156 (0.029)
Unemployment Rate	0.004 (0.003)	0.004 (0.003)	0.003 (0.003)	0.003 (0.003)
Median Income, \$10K	-0.009 (0.004)	-0.007 (0.005)	-0.009 (0.004)	-0.009 (0.005)
Extended Cab			0.111 (0.058)	0.102 (0.058)
Crew Cab			0.143 (0.058)	0.138 (0.058)
Four Wheel Drive			0.018 (0.038)	0.025 (0.038)
Dual Rear Wheels			-0.245 (0.106)	-0.249 (0.107)
Torque, 100 lb. ft.			0.000 (0.000)	0.000 (0.000)
Work Truck Trim			0.000 (0.023)	0.003 (0.023)
Length, ft.			0.025 (0.019)	0.028 (0.019)
Width, ft.			0.108 (0.047)	0.120 (0.047)
Height, ft.			0.065 (0.063)	0.042 (0.062)
Constant	3.044 (0.085)	3.383 (0.121)	0.187 (0.344)	0.553 (0.354)
County FEs	No	Yes	No	Yes
R^2	0.015	0.037	0.031	0.052

Table 4.4: OLS Regression of Log Yearly Mileage: Personal Registration Sample

Variable	**(1)**	**(2)**	**(3)**	**(4)**
Log Operating Cost	0.269 (0.052)	0.269 (0.052)	-0.178 (0.076)	-0.139 (0.077)
Rain, ft./year	-0.001 (0.001)	0.013 (0.008)	-0.001 (0.001)	0.012 (0.007)
Snow, ft./year	-0.001 (0.001)	-0.013 (0.007)	-0.001 (0.001)	-0.013 (0.007)
Rural Population Percentage	0.250 (0.028)	0.232 (0.033)	0.242 (0.028)	0.234 (0.032)
Unemployment Rate	0.004 (0.003)	0.003 (0.003)	0.004 (0.003)	0.003 (0.003)
Median Income, $10K	0.012 (0.005)	0.019 (0.006)	0.011 (0.005)	0.014 (0.006)
Extended Cab			0.225 (0.069)	0.226 (0.069)
Crew Cab			0.299 (0.068)	0.297 (0.068)
Four Wheel Drive			0.102 (0.043)	0.103 (0.044)
Dual Rear Wheels			-0.307 (0.130)	-0.304 (0.131)
Torque, 100 lb. ft.			0.000 (0.000)	0.000 (0.000)
Work Truck Trim			0.028 (0.024)	0.030 (0.024)
Length, ft.			0.023 (0.020)	0.019 (0.020)
Width, ft.			0.219 (0.057)	0.218 (0.058)
Height, ft.			-0..178 (0.076)	-0.174 (0.068)
Constant	2.691 (0.101)	2.818 (0.173)	0.529 (0.395)	0.890 (0.420)
County FEs	No	Yes	No	Yes
R^2	0.018	0.031	0.042	0.053

SECTION 5

MODEL

In this section, I present a two-period model of vehicle adoption and subsequent usage as a means to model price expectations in a concise manner and account for selection on anticipated usage. I then proceed by presenting an empirical analogue to my theoretical model which is estimated using maximum simulated likelihood (MSLE). The structure of the model is motivated by Gillingham (2013), who applies a model based on that of Einav et al. (2013) to automobile adoption.

5.1 Theoretical Model

I assume that individuals exogenously enter the market to purchase one new pickup truck. As a result of this assumption, estimates and counterfactuals presented in the next section should be viewed as conditional on selection into the new truck purchasing market. Based on the aggregation process described in Section 2, I use the subscript 'j' to refer to a vehicle's combination of model-series-model year-body-drivetrain-wheel configuration-engine. The timing of the model is as follows:

1. First Period: Consumers choose which truck to purchase based on:

   - Purchase price, denoted $MSRP_j$.
   - Vehicle attributes, denoted X_j. Fuel economy is not included in X_j, as I assume that individuals do not directly derive utility from operating a more efficient vehicle, but only through the fuel savings provided by a more efficient vehicle. Heterogeneity in consumers' preferences over vehicle characteristics can be introduced with random coefficients attached to X_j.

- Expected utility from driving the vehicle during the second period. Price expectations and fuel economy enter here, as they determine consumers' expected per-mile operating costs in the second period.

2. Second Period: Given the vehicle choice from the first period, consumers choose how many miles to drive after observing operating costs.

I proceed by discussing consumers' optimal second-period usage decisions conditional on owning a particular truck. Given consumers' second-period usage decisions, I then derive an expression for consumers' first-period expected utility under general fuel price expectations. I discuss the structure placed on price expectations prior to outlining my empirical analogue to the theoretical model.

5.1.1 Second Period: VMT Decision

The realization of the fuel price for the chosen vehicle occurs at the beginning of the second period, at which point consumers decide how many miles to drive. I parameterize second-period utility from using vehicle j, net of fuel costs, as

$$u_2 = \gamma_j \left(VMT_{ij} - \frac{1}{2\eta_i} VMT_{ij}^2 \right) - \alpha_i \tilde{p}_{ij} VMT_{ij} \tag{5.1}$$

The vector $\gamma_j \equiv \beta_X X_j$ acts to shift the marginal utility of driving vehicle j based on a vector of characteristics of vehicle j denoted by X_j, while the rate at which the marginal utility of driving diminishes is scaled by η_i. Fuel costs are specified as the per-mile price faced by individual i to drive vehicle j: $\tilde{p}_{ij} = \frac{p_{ij}}{MPG_j}$, where p_{ij} is the per-gallon fuel price faced by individual i.[1] Abstracting from corner solutions, the second-period utility function with respect to VMT yields the second-period demand for driving, conditional on owning vehicle j:

$$VMT_{ij}^* = \eta_i \left[1 - \frac{\alpha_i \tilde{p}_{ij}}{\gamma_j} \right] \tag{5.2}$$

[1] The per-gallon fuel price varies by fuel type, but does not vary for individual i across vehicles of the same fuel type.

The solution for VMT exhibits a couple of important properties. First, the individual-specific η_i represents the individual's level of driving in absence of fuel costs, and usage is multiplicative in η_i. This feature allows me to derive a tractable likelihood function for second-period usage decisions with a lognormally distributed usage shock. Second, driving is linear in fuel costs. In estimation, I only observe a single odometer reading attached to a subset of vehicles; the linearity of the demand curve allows me to model mileage as a function of the average fuel price during the second period.

5.1.2 First Period: Vehicle Choice

In the first period, I model consumers who have exogenously entered the market to purchase a new pickup truck and decide to choose the truck which maximizes their expected utility. I specify first-period expected utility as

$$E_i[u_1] = \delta E_i[u_2] - \alpha_i MSRP_j + \epsilon_{ij} \tag{5.3}$$

The individual's expected utility in the first period consists of two pieces. She incurs the cost of purchasing the vehicle in the first period and faces an expected utility flow from operating the vehicle in the second period. I abstain from modeling the resale market and instead assume that the second-period utility flows can be freely transferred to another individual. Given an individual's optimal driving level, expected second period utility can be written as[2]

$$E_i[u_2] = \delta E[\eta_i] \left[\frac{\gamma_j}{2} + \frac{\alpha_i^2 (Var(p_i) + E[p_i]^2)}{2\gamma_j MPG_j^2} - \frac{\alpha_i E[p_i]}{MPG_j} \right] \tag{5.4}$$

In the first period, individuals must take expectations over future realizations of fuel prices and the usage parameter η_i. Given that the empirical distribution of VMT can be approximated as a lognormal distribution, I specify $\eta_i = \exp\{\beta_Z Z_i + \sigma_k \epsilon_i^k + \sigma_u \epsilon_i^u\}$, where Z_i is a vector of exogenous taste shifters, $\epsilon_i^k \sim N(0,1)$ is a shifter of η_i which is known by the vehicle purchaser in the first period, and $\epsilon_i^u \sim N(0,1)$ is a second-period shock to η_i which is independent of Z_i and ϵ_i^k. Therefore, consumers' expected value of η_i at the time of vehicle purchase is given by

[2] A derivation of this result is given in §C.1.

$E_i[\eta_i] = exp\{\beta_Z Z_i + \sigma_k \epsilon_i^k + \frac{\sigma_u^2}{2}\}$. For the time being, I do not impose additional structure on price expectations, as I cover this at length when discussing my empirical implementation. Plugging this expression into $E_i[u_2]$ in (5.4) gives the individual's first period expected utility from purchasing a vehicle $j \in J$:

$$E_i[u_1] = \delta \exp\{\beta_Z Z_i + \sigma_k \epsilon_i^k + \frac{\sigma_u^2}{2}\} \left[\frac{\gamma_j}{2} + \frac{\alpha_i^2 (Var(p_i) + E[p_i]^2)}{2\gamma_j MPG_j^2} - \frac{\alpha_i E[p_i]}{MPG_j} \right] - \alpha_i MSRP_j + \epsilon_{ij} \tag{5.5}$$

This expected utility specification clarifies the role of anticipated usage on engine selection. Consider, for the sake of simplicity, an individual who is making a decision between a gasoline (G) and diesel (D) engine in an otherwise identical truck. If the diesel engine improves performance capabilities ($\gamma_D > \gamma_G$) and fuel economy ($\frac{E[p_D]}{MPG_D} < \frac{E[p_G]}{MPG_G}$), then for any value of η_i the second period expected utility is greater for the diesel engine. The difference in second-period utility associated with the diesel engine is multiplicative in η_i. The decision of whether to upgrade to the diesel engine depends on whether the difference in expected second period utility outweighs the difference in the purchase price ($\alpha_i(MSRP_D - MSRP_G)$). In this example, individuals with a sufficiently high level of $E[\eta_i]$ will select the diesel engine while those with a sufficiently low level of $E[\eta_i]$ will select the gasoline engine. By estimating the distribution of η_i, I am able to account for the role of selection on anticipated usage in fuel choice.[3]

5.2 Empirical Implementation

I model choices at the level of an individual deciding to purchase some vehicle j in her choice set J which maximizes her expected utility at the time of purchase. As discussed in §3.1, I define vehicles at the level of model-series-model year-cab type-drivetrain-engine-work trim-wheel configuration. For all vehicles of a particular model, series, and model year, I define the first date of availability as the first observed date of purchase. In a similar fashion, I define the last date of availability as the earlier of (i) the last date a vehicle of the same model, series, and model year is purchased or (ii) the first date a vehicle of the same model and series but subsequent model year

[3] More generally, I account for the role of selection on anticipated usage in truck choice.

is purchased. At this point, all vehicles of a particular model-series effectively 'switch' to the new model year.[4] The dates of availability typically apply to all different vehicles of a given model, series, and model year.[5]

In addition to estimating an empirical analogue of (5.5), I use my subsample with odometer readings attached to estimate a model of the second period usage decisions. For those observations with an odometer reading attached, the likelihood contributions are represented by the joint probability of purchasing vehicle j and conditional on purchasing vehicle j, choosing to drive VMT_{ij} miles. This strengthens identification of key parameters in the model, which I further discuss towards the end of this section. A significant shortcoming of my data on odometer readings is that I observe an odometer reading attached to just under 10% of my observations. I partition my data into two subsets of buyers; those individuals in subset N_1 owned the vehicle from the date of purchase through the end of 2015, and as such I do not observe an odometer reading for individuals in N_1. Purchasers in subset N_2 either transferred the vehicle by the end of 2015 or had an odometer reading recorded for another reason. I first consider the likelihood of observing outcomes recorded in each of these datasets separately, and then combine them to generate my likelihood function. Following the derivation of the likelihood function under the assumption that odometer readings are missing-at-random, I present a selection correction and derive the likelihood function which accounts for the non-random nature of odometer reading entries.

In estimation, I deviate from my theoretical model by specifying that the utility of operating the vehicle is accrued over repeated one year periods. As such, I attach a different coefficient on operating cost than is attached to the vehicle's purchase price, as the operating cost is incurred

[4] Within generation but across model years, there are minor differences in vehicle characteristics and MSRPs. However, this step reduces choice sets by around 50% to 60% and in turn vastly decreases the computational burden of estimation.

[5] When a new model year is released, most or all permutations of the truck are usually available. Suppose, for example, that I observe the first purchase of a four-wheel drive 2015 F-150 on October 3, 2014, while I do not observe an otherwise identical two-wheel drive 2015 F-150 until October 25,2014. Most likely, both vehicles were available in both choice situations, but the latter was simply not selected by a consumer until the later date. In one important instance, the 2014 Ram 1500 Diesel was not released until a few months after the other 2014 Ram 1500s. As such, I specify the first date of availability for all 2014 Ram 1500 Diesels as the first date a 2014 Ram 1500 with the diesel engine was purchased.

for several years. Given estimates of (i) consumers' responsiveness to operating costs and (ii) consumers' responsiveness to purchasing price, the discount rate is implicit given an assumption regarding the number of years a vehicle is operated (Hausman, 1979).

5.2.1 Specifying Price Beliefs

To estimate the model presented in the previous subsection, I must make an assumption regarding consumers' fuel price beliefs. I refer to three possible sets of price beliefs throughout the remainder of this paper. First, under a 'no-change' forecast, the expected price of both gasoline and diesel fuel at any future date t is given by the current price of each respective fuel $f \in \{d, g\} : E[p_t^f] = p_0^f$, where 0 denotes the date when the consumer forms her price beliefs. I follow much of the literature and use a 'no-change' forecast for gasoline price beliefs, but I consider an alternative possibility for diesel price expectations which is motivated by the results from my logistic regressions on fuel choice in the previous section. Under a 'decomposed' forecast, I decompose the current diesel price as $p_0^d = p_0^g + \tau_C + \tilde{p}_0^d$. I then specify the 'decomposed' forecast as $E[p_t^d] = p_0^g + \tau_C + \phi\tilde{p}_0^d$, where ϕ is a parameter to be estimated which represents the weight placed on the time-series variation in the diesel premium. By construction, the decomposed forecast is equivalent to the no-change forecast when $\phi = 1.00$. This specification allows for the possibility that consumers respond differently to different types of price variation. In this environment, the decomposition is motivated by the fact that under a random-walk process for gasoline prices as suggested by the prior literature, a forward-looking consumer would update their beliefs one-for-one when the prices of both fuels rise. Likewise, given the persistence of variation in the diesel premium across counties, a forward-looking individual in a county with a high average diesel premium (τ_C) should expect the diesel premium to remain higher than an individual in a county with a low average diesel premium over the life of her vehicle. On the other hand, a forward-looking consumer may not respond to time-series variation in the diesel premium ($\tilde{p}_0^d$) owing to its transitory nature. The extent to which consumers respond to time-series variation in the diesel premium relative to the aforementioned sources of price variation is captured by ϕ. I proceed by estimating the model under a decomposed forecast and a no-change forecast, and contrast the results in the next section.[6]

5.2.2 Likelihood Contributions

As discussed earlier in this chapter, I observe odometer readings attached to a potentially non-random subset of observations. In order to derive my likelihood function which allows for the possibility of non-random selection into reporting an odometer reading, I split my sample into two subsets based on whether an odometer reading is observed. For the remainder of the paper, I let N_1 denote the sample of observations where I observe a vehicle purchase but not an odometer reading, N_2 denote the sample of observations where I observe a vehicle purchase as well as an odometer reading, and θ denote the full set of parameters to be estimated in the model.

There is a variety of reasons one might suspect my odometer readings appear for a non-random subset of observations given the data entry process described earlier. For example, those individuals who have a high 'known' taste for driving, $\sigma_k \epsilon_i^k$, could be more likely than those with a low $\sigma_k \epsilon_i^k$ to wear out their vehicles and subsequently be less likely to transfer their vehicles owing to lower resale values. Alternatively, these individuals could have a stronger preference upgrading to newer, more luxurious vehicles given the greater amount of time they spend in their vehicles and exhibit a higher probability of recording an odometer reading. While neither of these are explicitly modeled, my selection correction aims to account for either possible relationship between $\sigma_k \epsilon_i^k$ and odometer reading reporting. One could similarly suspect that those individuals who receive an extreme draw of σ^u after purchasing their vehicle may be more likely to transfer their vehicle and subsequently receive an odometer entry. If an individual initially purchased a large, inefficient vehicle and then switched jobs to one which required an additional 20 miles of daily commuting (high σ^u), we might expect her to replace this vehicle with a more efficient one. For the sake of exposition, I begin by presenting the likelihood contributions under the assumption that odometer readings are entered randomly. I then continue to present a selection correction and discuss its implementation.

[6] I define the variance in each respective fuel price as the observed daily price variance in an individual's county throughout the sample. For the decomposed forecast, the variance is approximated by $Var(p_i^g + \tau_C) + \phi^2 Var(\tilde{p}_i^d)$. Because $Cov(p_i^g + \tau_C, \tilde{p}_i^d) \approx 0$, I drop the covariance term and consider my measure an approximation.

5.2.3 Likelihood Contributions: Random Odometer Entry

Because I do not observe an odometer reading for individuals in subsample N_1, the likelihood contributions for these individuals only considers their first-period vehicle choices. I assume that the idiosyncratic match value ϵ_{ij} follows an extreme value distribution which is independent and identically distributed across individuals and vehicles. Up to each individual's draw of ϵ_{ij}, the expected utility of individual i from adopting vehicle j is given by

$$V_{ij} = \exp\{\mu_k + \beta_Z Z_i + \sigma_k \epsilon_i^k + \frac{\sigma_u^2}{2}\} \left[\frac{\gamma_j}{2} + \frac{\alpha_f^2 (Var(p_i) + E[p_i]^2)}{2\gamma_j MPG_j^2} - \frac{\alpha_f E[p_i]}{MPG_j} \right] - \alpha_p MSRP_j \quad (5.6)$$

where notation is identical to that used in the theoretical model, with the exception that α_f is attached to yearly operating costs while α_p is attached to the purchase price. Given this specification, the discount rate is implied based on the vehicle's expected life and the values of α_f and α_p. I specify $\epsilon_i^k \sim N(0,1)$ where σ_k is a parameter to be identified. Given a particular draw of ϵ_i^k and the distributional assumption on ϵ_{ij}, the probability of an individual i choosing some vehicle $j \in J$ is given by

$$Pr_i(j|\theta) = \frac{e^{V_{ij}}}{\sum_{h \in J} e^{V_{ih}}} \quad (5.7)$$

This logit choice probability represents likelihood the contributions of individuals who do not report an odometer reading, conditional on a draw of ϵ_i^k. Integrating out over the distribution of ϵ_i^k, the likelihood contribution of an observation in N_1 given the full parameter vector θ is given by

$$L_i(\theta) = \int Pr_i(j|\theta) f(\epsilon^k) d\epsilon^k \quad (5.8)$$

In specifications which include random coefficient(s) attached to vehicle characteristics, likelihood contributions must also integrate over the distribution of these random coefficients. In §6, I estimate a specification which includes a random coefficient attached to torque which specifies $\beta_{iT} = exp\{\beta_T + \sigma_T \epsilon_{iT}\}$ where T denotes torque and $\epsilon_{iT} \sim N(0,1)$ is independent of ϵ_i^k and ϵ_i^u. For the sake of notational simplicity, I suppress this integral throughout this section.

For observations with an odometer reading, the likelihood contributions are given by the joint probability of purchasing vehicle j in the first period, and then conditional on this purchase, choosing to drive VMT_{ij} miles in the second period. I specify $\epsilon_i^u \sim N(0,1)$ and assume the shock is independent across i, j, ϵ_i^k. Given the realization of ϵ_i^u, the individual's optimal VMT is given by

$$VMT_{ij}^* = \exp\{\beta_Z Z_i + \sigma_k \epsilon_i^k + \sigma_u \epsilon_i^u\} \left[1 - \frac{\alpha_f \tilde{p}_{ij}}{\gamma_j}\right] \tag{5.9}$$

Taking the natural log of VMT_{ij}^* and letting $\mu_{ij} \equiv \left[1 - \frac{\alpha_f \tilde{p}_{ij}}{\gamma_j}\right]$, the density of $\ln(VMT_{ij})$ conditional on a draw of ϵ_i^k and owning vehicle j is given by

$$f(\ln(VMT_{ij}^*)|j,\theta) = \frac{1}{\sigma_u\sqrt{2\pi}} * \exp\left[\frac{-(\ln(VMT_{ij}) - \beta_Z Z_i - \sigma_k \epsilon_i^k - \ln(\mu_{ij}))^2}{2\sigma_u^2}\right] \tag{5.10}$$

Given a draw of ϵ_i^k, the likelihood contribution of an individual in N_2 is given by $Pr_i(j) * f(\ln(VMT_{ij}^*)|j)$. Integrating out over the distribution of ϵ_i^k, the likelihood contribution of an observation in N_2 is given by

$$L_i(\theta) = \int \left(Pr_i(j|\theta) * f(\ln(VMT_{ij}^*|j,\theta))\right) f(\epsilon^k) d\epsilon^k \tag{5.11}$$

5.2.4 Selection Correction

I now present an adjustment to the likelihood function which allows for the possibility of non-random entry of odometer readings. Recall that for all individuals sample N_1 and sample N_2, I observe a vehicle purchase, but only observations in N_2 have an odometer reading. I specify y_i as a latent variable which is determined by the following selection equation:

$$y_i = K_i \delta + \zeta \sigma_k \epsilon_i^k + v_i$$

I assume that the outcome of the selection equation determines whether I observe an odometer reading:

$$\ln VMT_i = \begin{cases} \beta_Z Z_i + \sigma_k \epsilon_i^k - \ln(\mu_{ij}) + \sigma_u \epsilon_i^u & \text{if } y_i > 0 \\ \text{unobserved} & \text{if } y_i \leq 0 \end{cases} \tag{5.12}$$

Following Heckman (1979), I specify the following joint distribution of the error term in the selection equation and the second-period shock to driving:

$$\begin{pmatrix} \sigma_u \epsilon^u \\ v \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \rho\sigma_u \\ \rho\sigma_u & 1 \end{pmatrix} \right]$$

where $\epsilon_i^u \sim N(0,1)$ and the variance of the error term in the selection equation is normalized to $\sigma_v^2 = 1$. Under these assumptions regarding the distribution of disturbances, the likelihood contribution of an observation in sample N_1 is the joint probability of (1) an individual choosing vehicle $j \in J$ and (2) not reporting an odometer reading:

$$L_i(\theta) = \int_{\epsilon^k} (Pr(y_i \leq 0|\theta) * Pr(j|\theta)) f(\epsilon^k) d\epsilon^k$$

where $Pr_i(j|\theta)$ is given by (5.7). Given our assumption that $v_i \sim N(0,1)$, note that $Pr(y_i \leq 0|\theta) = Pr(v_i \leq -K_i\delta - \zeta\sigma_k\epsilon_i^k) = 1 - \Phi(K_i\delta + \zeta\sigma_k\epsilon_i^k)$. Our likelihood contributions for observations in N_1 are thus given by

$$L_i(\theta) = \int_{\epsilon^k} (1 - \Phi(K_i\delta + \zeta\sigma_k\epsilon_i^k)) * \left[\frac{e^{V_{ij}}}{\sum_{h \in J} e^{V_{ih}}} \right] f(\epsilon^k) d\epsilon^k \tag{5.13}$$

For the subset of observations which carry an odometer reading (N_2), the likelihood contributions after making the selection correction are now given by the joint probability of (1) an individual choosing vehicle $j \in J$ and (2) the joint probability of driving a particular number of miles *and* reporting the odometer reading:

$$L_i(\theta) = \int_{\epsilon^k} (Pr(\ln VMT_{ij}, y_i > 0|j, \theta) * Pr(j|\theta)) f(\epsilon^k) d\epsilon^k$$

Using Bayes' rule, $Pr(\ln VMT_{ij}, y_i > 0|j, \theta) = f(\ln VMT_{ij}|j, \theta) * Pr(y_i > 0| \ln VMT_{ij}, j, \theta)$.

Given that $y_i = K_i\delta + \zeta\sigma_k\epsilon_i^k + v_i$, the probability of observing an odometer reading is given by

$$Pr(v_i > -K_i\delta - \zeta\sigma_k\epsilon_i^k | \ln VMT_{ij}, j, \theta) = \int_{-K_i\delta - \zeta\sigma_k\epsilon_i^k}^{\infty} f(v_i|\epsilon_i^u)dv_i$$

Using the properties of the conditional distribution function of the normal distribution,

$$\int_{-K_i\delta-\zeta\sigma_k\epsilon_i^k}^{\infty} f(v_i|\epsilon_i^u)dv_i = \int_{-K_i\delta-\zeta\sigma_k\epsilon_i^k}^{\infty} \phi\left(\frac{v_i - \frac{\rho}{\sigma_u}(\ln(VMT) - \beta_Z Z_i - \sigma_k\epsilon_i^k - \ln(\mu_{ij}))}{\sqrt{1-\rho^2}}\right) dv_i$$

$$= 1 - \Phi\left(\frac{-K_i\delta - \zeta\sigma_k\epsilon_i^k - \frac{\rho}{\sigma_u}(\ln(VMT) - \beta_Z Z_i - \sigma_k\epsilon_i^k - \ln(\mu_{ij}))}{\sqrt{1-\rho^2}}\right)$$

$$= \Phi\left(\frac{K_i\delta + \zeta\sigma_k\epsilon_i^k + \frac{\rho}{\sigma_u}(\ln VMT - \beta_Z Z_i - \sigma_k\epsilon_i^k - \ln(\mu_{ij}))}{\sqrt{1-\rho^2}}\right)$$

where the last step follows by the symmetry of the normal distribution. This gives us the likelihood contributions for individuals in sample N_2:

$$\begin{aligned} L_i(\theta) = & \\ & \int_{\epsilon^k} \left[\frac{e^{V_{ij}}}{\sum_{h\in J} e^{V_{ih}}}\right] * \left[\frac{1}{\sigma_u\sqrt{2\pi}} * exp\left\{\frac{-(\ln(VMT) - \beta_Z Z_i - \sigma_k\epsilon_i^k - \ln(\mu_{ij}))^2}{2\sigma_u^2}\right\}\right] * \\ & \Phi\left(\frac{K_i\delta + \zeta\sigma_k\epsilon_i^k + \frac{\rho}{\sigma_u}(\ln VMT - \beta_Z Z_i - \sigma_k\epsilon_i^k - \ln(\mu_{ij}))}{\sqrt{1-\rho^2}}\right) f(\epsilon^k)d\epsilon^k \end{aligned} \quad (5.14)$$

I approximate the integrals in (5.13) and (5.14) using simulation, as my model is estimated with MSLE. For each individual in the sample, I take 200 draws of ϵ_i^k using antithetic acceleration.[7] Given an individual's draws, her simulated likelihood contribution is given by the average of the likelihood contributions corresponding to each of her $R = 200$ draws. For individuals in each

[7] In specifications that include random coefficient(s) on vehicle characteristics, the same process is used. Train (2009) provides a thorough discussion of this procedure.

respective sample, the simulated likelihood contributions are given by

$$SL_i(\theta) = \begin{cases} \frac{1}{R}\sum_{r=1}^{R}\left[Pr(j|\theta) * Pr(y_i \leq 0|j,\theta)\right] & \text{if } i \in N_1 \\ \frac{1}{R}\sum_{r=1}^{R}\left[Pr(j|\theta) * f(\ln(VMT_{ij})|j,\theta) * Pr(y_i > 0|\ln VMT_{ij}, j, \theta)\right] & \text{if } i \in N_2 \end{cases} \tag{5.15}$$

Letting $SL_1(\theta) = \Pi_{i \in N_1} SL_i(\theta)$ and $SL_2(\theta) = \Pi_{i \in N_2} SL_i(\theta)$, the joint probability of observing the data in sample N_1 and sample N_2 is given by $SL(\theta) = SL_1(\theta) * SL_2(\theta)$. Taking logs gives the simulated log likelihood function which I maximize in estimation:

$$SLL(\theta) = \sum_{i \in N_1} \ln(SL_i(\theta)) + \sum_{i \in N_2} \ln(SL_i(\theta)) \tag{5.16}$$

5.2.5 Identification

The set of parameters to identify includes $\sigma_k, \sigma_u, \alpha_f, \alpha_p, \phi, \beta_Z$, and β_X. The key to the identification of fuel price parameters in this paper lies in the richness of my fuel price data and the breadth of my registration data. I observe significant cross-sectional and time-series variation in the fuel prices which consumers face. Within my decomposed model of fuel price expectations, I must identify two parameters on fuel prices. Consumers' disutility (α_f) associated with a one dollar increase in operating costs is identified through exogenous variation in fuel prices at the time and location of vehicle purchase, and strengthened by the second-period vehicle usage equation which I estimate. Both cross-sectional and time-series variation in fuel prices are especially useful in this setting where consumers have two different fuel options as they generate significant variation in the anticipated fuel savings associated with adopting the diesel engine. This variation strengthens my identification of consumers' valuation of fuel savings from the diesel engine separately from their valuation of the performance attributes, which would otherwise rely on very limited time-series variation in diesel engine offerings.[8] As discussed in §3.2, counties in the east of the state have

[8] All manufacturers always offer one diesel engine option in both their three-quarter ton and one-ton trucks, and engine characteristics did not significantly change over the course of the sample. Identification of consumer valuation of vehicle characteristics relies on exogenous variation in the choice set; however, there is significantly greater variation

persistently low gasoline prices and a high average diesel premium. Counties in western Washington have persistently high gasoline prices and a low average diesel premium. As shown in Figure 5.1, this difference can result in thousands of dollars of variation in the fuel savings associated with adopting a diesel engine. Given identical choice sets, the cross-sectional variation in incentives to adopt the diesel engine allows me to separately identify consumers' willingness to pay for fuel efficiency from their tastes for diesel engine attributes.

Given the identification of α_f, I use time-series variation in the diesel premium at the time of purchase to separately identify ϕ. The intuition behind this identification argument lies in Figure 5.2, which gives the lifetime expected fuel savings associated with adopting a diesel engine in one particular truck over the course of my sample. Given α_f, the extent to which diesel engine upgrades respond to time-series variation in the diesel premium identifies ϕ. As shown in Figure 5.2, under no-change fuel price expectations ($\phi = 1.00$), time-series variation in fuel prices generates massive variation in the incentive to adopt the diesel engine. Under this price forecast, within-year variation in the the diesel premium would result in upwards of ten thousand dollars of variation in the savings associated with adopting the diesel engine. Under a forecast where $\phi = 0$, the time-series variation in the incentives to adopt a diesel engine is modest. As such, the extent to which sales respond to the time-series variation in the diesel premium, relative to how they respond to all other sources of price variation, identifies ϕ.[9]

A shortcoming of the data used in this paper is that transaction prices are unavailable, and as such, I use MSRPs as a proxy for vehicle prices. As documented by Busse et al. (2013) and Langer and Miller (2013) using MSRPs as a proxy for transaction prices induces measurement error which is correlated with fuel price fluctuations. Such measurement error can bias estimates of consumers' willingness to pay for fuel economy downwards. There are two factors which may help mitigate the magnitude of the bias in this paper. First, I rely heavily on cross-sectional fuel

in gasoline engine offerings than diesel engine offerings.

[9]While I've primarily discussed the differing types of variation in the diesel premium for the sake of exposition, the other sources of price variation referred to here includes both cross-sectional variation in the diesel premium and variation in gasoline prices, based on the construction of decomposed price beliefs which is illustrated in §5.2.1.

price variation within a relatively small geographic area, while many of the incentive programs which are used by manufacturers in response to fuel price fluctuations may be set regionally or nationally. Further, in the case of discounts at the dealer-level, it is feasible for consumers make a one-time vehicle purchase at a dealership a couple of hours from their domicile. If dealers east of the Cascades began offering a discount of $1,000 on diesel powered trucks, shoppers from the west of the Cascades may also be able to take advantage of these savings.[10] Thus, the extent to which such geographical differentials in vehicle prices can persist may be mitigated by consumer search behavior.[11] Second, the extent to which dealers offset changes in operating costs over the life of vehicles in response to changing fuel prices is estimated by Langer and Miller (2013) to be smaller for trucks than smaller vehicles.[12]

As alluded to earlier, the parameters attached to vehicle characteristics are of interest in this project, but are identified with standard arguments from the vehicle demand estimation literature. The assumption that model characteristics evolve exogenously from the standpoint of the consumer is used to identify these parameters. This exogenous variation in choice sets allows me to observe how consumers substitute between products with varying attributes as their choice sets evolve over time, which in turn allows me to identify consumer preferences towards vehicle attributes.[13]

The role of both observed and unobserved heterogeneity in individuals' demand for VMT is

---

[10]As shown in Figure 5.1, this could represent a 25%-50% offset in the fuel savings associated with the diesel engine between the counties.

[11]Search costs are undoubtedly high. However, consumer search would presumably place a bound on persistent vehicle price differentials which could arise due to persistent fuel price variation across counties.

[12]They estimate a mean offset of 61% for cars, 30% for SUVs, and only 18% for trucks. Used vehicle prices also appear to be significantly more responsive to fuel price fluctuations than new vehicle prices. Busse et al. and Langer and Miller (2013) each estimate that a $1 change in gas prices is associated with around a $300 change in relative prices between a new vehicle in the top quartile of fuel economy and a new vehicle in the bottom quartile of fuel economy. Busse et al. (2013) estimate this figure is around $2,000 for used vehicles.

[13]Petrin (2002) provides a nice discussion of the intuition behind this identification argument, as well as the identification of random parameters attached to vehicle characteristics. The reader should note that I ignore the possibility of strategic delay in vehicle adoption due to the static nature of my model. However, because there is only modest technological advancement during my panel, I do not view this as a significant concern.

captured by the distribution of η_i, which is parameterized as $\exp\{\beta_Z Z_i + \sigma_k \epsilon_i^k + \sigma_u \epsilon_i^u\}$. Taste-shifters β_Z are identified by the variation in second-period usage decisions which results from exogenous variation in Z_i. The timing of the model and the observed second-period usage decisions allow me to separately identify the parameters of the known component (μ_k, σ_k) of unobserved heterogeneity and the unknown component of unobserved heterogeneity (σ_u). Given the assumption of a mean-zero second period usage shock, the coefficient attached to the constant in Z_i, which I denote as μ_k from now on, is identified based on the observed mean level of usage where $Z_i = \vec{0}$.[14] This identification argument is best illustrated through the choice of whether to adopt a gasoline or diesel engine in a heavy-duty truck. Recall Figure 3.2, which plots the distributions of usage for gasoline and diesel heavy-duty truck drivers. Given consumers' tastes for the performance characteristics of each engine contained in β_X and her fuel price expectations, there is a threshold of anticipated usage above which it is optimal to purchase the diesel engine and below which it is optimal to purchase the gasoline engine. This threshold varies across time and across locations as fuel prices vary. Variation in adoption rates of the diesel engine across time and counties, given an individuals' set of price expectations, allows me to trace out the distribution of $\sigma_k \epsilon_i^k$. The remaining parameter σ_u is then identified through observed variation in second-period usage decisions, given the estimated values of β_Z, μ_k, and σ_k.

[14]Recall that η_i represents the intercept of the VMT demand function, where $VMT_{ij}^* = \eta_i \left[1 - \frac{\alpha_f \tilde{p}_{ij}}{\gamma_j}\right]$. Given α_f, β_X, we can identify the mean of the intercept of the demand curve where $Z_i = 0$.

Figure 5.1: Cross-Sectional Identifying Variation with Kernel Density Estimate

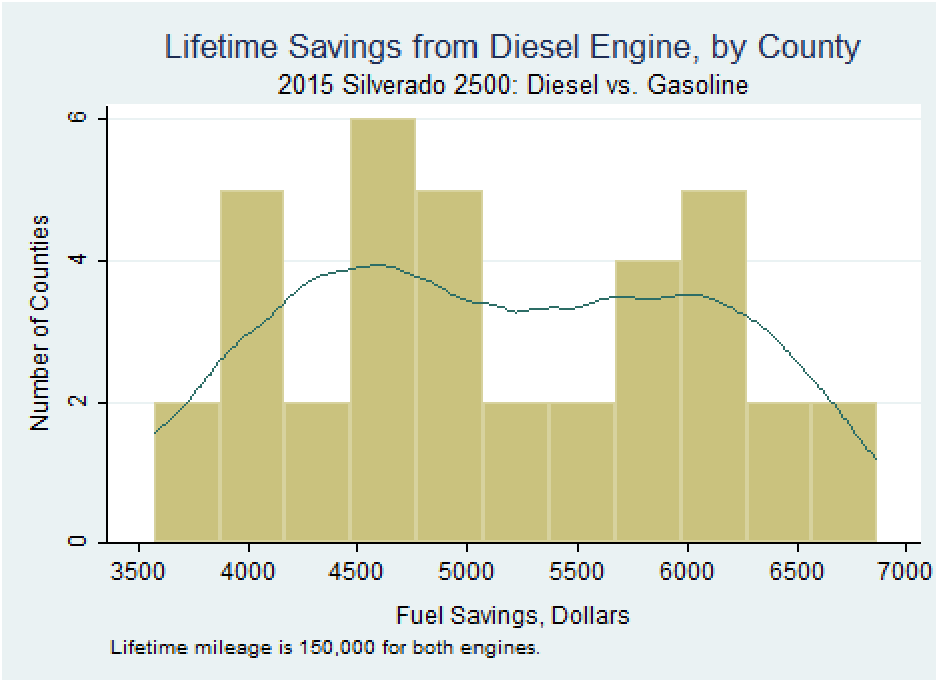
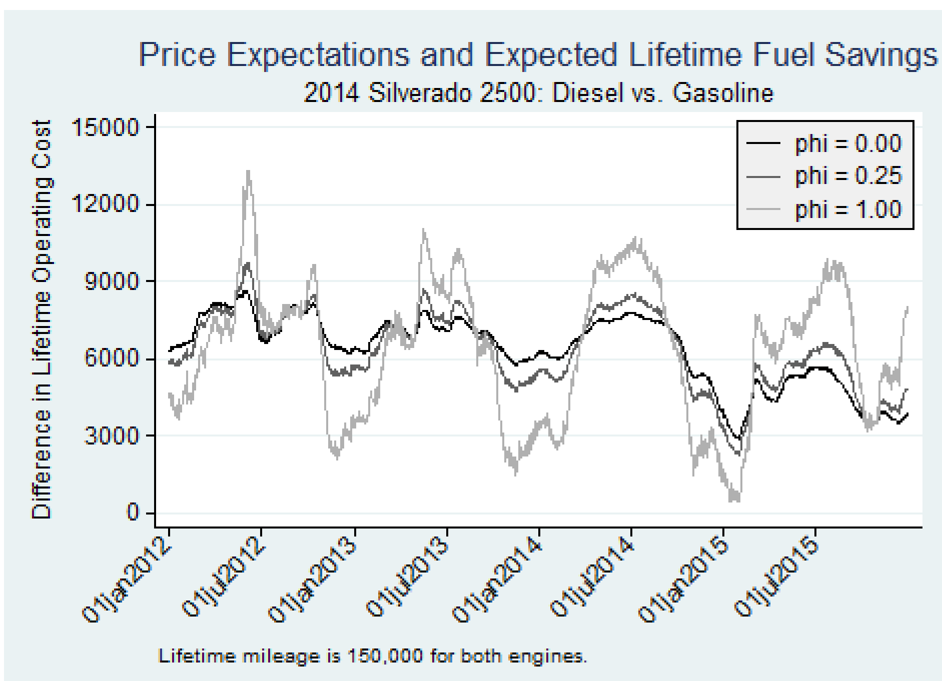


Figure 5.2: Time-Series Identifying Variation



SECTION 6

RESULTS

In this section, I present my parameter estimates and counterfactual simulations. My estimation routine simulates over 200 draws of ϵ_i^k which are drawn using antithetic acceleration. In specifications where a random coefficient is attached to torque, an identical procedure is used to take the draws. My estimation sample consists of all 7,040 of the observations with an attached odometer reading ($N_1 = 7,040$) and a randomly drawn sample of 20% of observations ($N_2 = 16,388$) which do not have an attached odometer reading. I present results from three different specifications of the model, each with two different sets of price beliefs. In Table 6.1, I first present my results from estimating the model without the selection correction discussed in §5.2.4. In Table 6.2, I present the results with the selection correction from §5.2.4. The final specification shown in Table 6.3 includes a random coefficient attached to torque, where I simulate over 200 draws of this coefficient once again using antithetic acceleration.[1] In all specifications, VMT is defined in thousands of miles per year, whereas vehicle prices are scaled in terms of thousands of dollars and operating costs are scaled in terms of thousands of dollars per thousand miles. All standard errors are computed by taking the square root of the elements along the diagonal of the inverse of the Hessian.[2]

6.1 Parameter Estimates

There are four sets of parameters which I present in my tables: price parameters which include α_f and α_p as well as ϕ when decomposed price beliefs are specified, the parameters of the usage

[1] In addition to these specifications, Table D.1 in §D presents results from a sample consisting only of those observations which have an odometer reading attached. Hence, the selection adjustment is not made for this specification.

[2] In §D, I also present the results from a linear-in-parameters multinomial logit model of vehicle choice which illustrates the importance of allowing consumers to respond differently to separate sources of fuel price variation in a more familiar discrete choice model.

distribution ($\mu_k, \beta_Z, \sigma_k, \sigma_u$), the parameters attached to vehicle characteristics (β_X), and finally those in the selection equation when relevant (δ, ζ, ρ). Each of the three tables includes two different forms of price beliefs. In the left-hand column beliefs are specified as a no-change forecast, whereas in the right-hand column I specify beliefs using the decomposed forecast. In addition the the parameter values, I report the implicit discount rate which is computed using α_f, α_p, and the median lifespan of a light truck as estimated by NHTSA (2006).

Recall that from the theoretical model, the parameters attached to vehicle characteristics represent the marginal utility of operating the vehicle for the first thousand miles in a year, and that the marginal utility of driving diminishes as VMT increases. Almost all estimated parameters in β_X have the expected sign. The lone exception is the parameter attached to the diesel engine; however, this is explainable as conditional on a particular level of torque (the sole measure of engine power used here), diesel engines have lower levels of horsepower than gasoline engines.[3] Operating a vehicle with an engine which produces more torque increases the marginal utility of operation, as more powerful engines by this measure improve vehicles' acceleration and smoothness of operation when towing or carrying heavy payload. In specifications which include a random coefficient on torque, I estimate little heterogeneity in the marginal utility of torque.[4] The omitted categories for cab, drivetrain, trim, and wheel configuration are regular cab, two wheel drive, standard trim, and single rear wheels. Upgrading to a more comfortable extended cab increases the marginal utility of driving the truck, but to a lesser extent than upgrading to the even larger crew cab. Likewise, the four-wheel-drive option, which entails a significant cost but is nonetheless selected by

[3] I limit the number of vehicle characteristics used here merely as a means to decrease the computational burden. I use torque as the sole measure of engine power due to its association with towing capability, and I do not include horsepower because it is highly collinear with torque conditional on the value of the diesel engine dummy variable.

[4] Two additional specifications of the random coefficient on torque were estimated. In one additional specification, I specified a normally distributed coefficient on torque. I also estimated a specification where a randomly distributed valuation of torque entered consumer utility linearly as a first-period payoff. Intuitively, this specification is consistent with consumers who value torque due to its impact on usage modes such as towing and hauling which are used only during a small percentage of driving situations for many consumers. In all cases, I estimate a low level of heterogeneity and find that including the random coefficient slightly decreases the objective function. In addition to this, I ran the estimation routine while fixing the shape parameter at higher levels than those estimated, but in each case this lead to a poorer fit and higher value of the objective function being minimized. In the appendix to this section, I include estimates from counterfactual simulations where I fix the shape parameter on torque to 0.25. This exercise has little effect on the relevant elasticities which I estimate using simulation.

nearly 95% of truck purchasers, increases the marginal utility of driving the vehicle by a similar magnitude as the crew cab upgrade.[5] The negative parameter attached to the dual rear wheel configuration is consistent with the possibility that this configuration is a nuisance to the majority of drivers, as driving a heavy-duty truck with this configuration makes it challenging to drive on small streets and park in a single spot. The work truck trim level represents a downgrade from the omitted trim level and decreases the marginal utility of driving the vehicle.[6] Truck class dummies account for unobservable vehicle characteristics which vary across truck classes. The negative coefficient on three-quarter ton trucks (e.g. Ford F-250) may reflect a lower likelihood of consumers selecting unobservable 'comfort' upgrades. In this market, the scale of unobservable product quality differences within a particular truck class is modest, so I use these class dummies rather than model-level fixed effects as a means to lower the computational burden.[7]

The distribution of the usage intensity parameter η_i is summarized by the estimated values of μ_k, σ_k, β_Z, and σ_u. As a result of the functional form of the second-period utility function, recall that η_i represents the intercept of the demand curve for VMT. The estimates of this distribution do not change significantly when price expectations are altered; to illustrate the properties of this distribution I consider the estimates from the model with decomposed price beliefs in Table 6.2. As a baseline, consider an individual who has the mean known component of usage and lives in a zip code which is entirely urbanized with a median household income of $60,000. For this individual, $E[\eta_i] \approx 14.824$, indicating that such an individual would expect to drive 14,824 miles per year in absence of fuel costs. Were this same individual living in a fully rural zip code, her expected value of η_i would indicate an expected usage level of 18,620 miles per year. The estimate of σ_k illustrates a significant level of heterogeneity in expected usage. Considering the individual with

[5]Recall that the crew cab upgrade is costlier than the four-wheel-drive upgrade, but selected by only about two-thirds of truck purchasers.

[6]It is common to aggregate up to a single trim level if it is not possible to uniquely identify vehicle trim levels. I leave the work trim option because certain configurations are not offered with upgraded trim levels (for example, two-wheel drive, regular cab trucks are sometimes only offered with the base trim package). The 'work trim' represents a base trim on trucks, while the omitted category represents a one level upgrade from the base trim.

[7]Beresteanu and Li (2011) employ model-level fixed effects as a means to account for the correlation between unobservable vehicle attributes and vehicle prices.

$E[\eta_i] = 14.824$, an otherwise identical individual with a draw of $\sigma_k \epsilon_i^k$ one standard deviation above the mean would expect to drive 22,137 miles per year in absence of fuel costs. I find that across all specifications, $\sigma_k < \sigma_u$, although the scale of the parameters is qualitatively similar. This indicates that the variance in second-period usage shocks is greater than that of the first-period unobservable heterogeneity. While perhaps counterintuitive, this estimate is not surprising given the observed usage distributions. The empirical usage distribution of gasoline heavy-duty truck purchasers shown in Figure 3.2 includes many individuals who drive over 20,000 miles per year. Fuel savings associated with the diesel engine for such individuals typically outweigh the upgrade cost even at a relatively high discount rate; accounting for the fact that the performance attributes of the diesel engine are greater than the gasoline engine, her choice is rationalized by the model with a large, positive second-period usage shock.

In the specifications in 6.2 and 6.3 where I include the selection correction, those parameters are presented in the bottom block of the table. This specification of the selection equation includes the same local taste-shifters and the VMT distribution as well as a variable capturing the initial purchase date of the vehicle, which is excluded from the VMT distribution. Given the assumption that individuals exogenously enter the market to purchase a new pickup truck, this exclusion restriction comes naturally; however, it should be noted that this places additional weight on an assumption which may be invalid. The large negative estimate of the coefficient attached to the vehicle purchase date (defined as years elapsed since the beginning of the sample in January 2012) is consistent with the observed pattern that individuals who purchased a vehicle earlier in the data sampling period were more likely to transfer their vehicle and have an odometer reading entered. The parameters attached to local income and local rurality are small by comparison, and have little impact on the likelihood that I observe an odometer reading. The negative estimates of ζ and ρ suggest that individuals with a higher (unobservable) first-period known usage level or second-period usage shock are less likely to report an odometer reading. Parameter estimates attached to fuel prices and vehicle characteristics scale downwards by roughly ten percent in magnitude once the selection correction is included. I further discuss the implications of controlling for selection into reporting an odometer reading in the next subsection where I conduct counterfactual simulations.

Of utmost interest in this project is the estimated responsiveness to fuel prices and the extent to which fuel price expectations are adjusted in response to different sources of fuel price variation. The former is captured by α_f, which represents the decrease in first-period expected utility associated with a one dollar increase in yearly fuel costs. The latter is captured by ϕ, which can be interpreted as the weight consumers place on time-series variation in the diesel premium relative to the weight that is placed on all other sources of fuel price variation when forming price expectations. In the three specifications which use the decomposed forecast, the estimate of ϕ lies between 0.110 and 0.117 and is not statistically significantly different from zero in any specification. Taking the highest estimate of ϕ for the sake of example, the interpretation of this point estimate is that individuals' price expectations adjust to \$1.00 of time-series variation in the diesel premium in the same manner as they adjust to a \$0.117 increase in the remaining sources of price variation. The importance of allowing individuals' price expectations to adjust differently to different types of price variation is illustrated by comparing the estimates of α_f across different models of fuel price expectations. The difference in α_f across each specification of price beliefs is similar in each of the three specifications, and ranges from 16.6% higher in the specification with the selection correction (and a random coefficient) to 17.3% higher in the specification with neither.

The bottom row of the table lists the implied discount rate under the assumption that fuel costs are incurred once yearly, and vehicle survival is given by the median light truck survival length of 14 years which is reported by NHTSA (2006). The level of these estimates should be taken with caution given the documented measurement error which is the result of using MSRPs rather than transaction prices. Across all specifications and models of price beliefs, estimated discount rates are higher than the average interest rate on loans for new automobiles during this time.[8] However, the comparison of estimated discount rates between different specifications of price expectations is illustrative of the importance of allowing for price expectations to adjust differently to different

[8]For reference, the average interest rate on a 48 month auto loan in February 2011 was 5.85% (FRED, 2017). Insofar as the discount from MSRP is lower (greater) for vehicles with high (low) fuel efficiency when fuel prices are high, the estimated discount rates with transaction price data would be lower. If dealers' prices for diesel and gasoline engines respond to time-series variation in the diesel premium, one could view ϕ as the lower bound of how consumers' price expectations respond to time-series variation in the diesel premium. Alternatively, the parameter can take a reduced-form interpretation and be thought of as how vehicle sales respond to such price variation.

types of fuel price variation. In each of the three main specifications, estimating the model with decomposed beliefs results in an implicit discount rate which is roughly four percentage points below the implicit discount rate which results when no-change beliefs are specified. Like the difference in the point estimates of α_f, this indicates that estimating price beliefs in the more flexible manner in this environment is important to adequately capture the consumer response to fuel price fluctuations.

6.2 Counterfactuals

To further illustrate the importance of allowing consumers to respond differently to different types of price variation, I conduct counterfactual simulations of consumers' first and second-period decisions in response to increased fuel prices. While simulations are conducted using the estimates of all three specifications each with two different sets of price beliefs, I focus on estimates from the specification which includes the selection adjustment but does not include a random coefficient on torque, as shown in Table 6.2. When relevant, I discuss the similarities and differences between the three specifications, but for the sake of brevity I leave the results from counterfactuals drawn from estimates in Table 6.1 and Table 6.2 in §D. All simulations are conducted over 20 draws of ϵ_{ij}, ϵ_{ik}, ϵ_{iu}, and when relevant, the random coefficient attached to torque. The reader should note a few caveats with regards to the interpretation of the counterfactuals and their place in the broader literature on the consumer response to price fluctuations. First and foremost, recall that all estimates presented here are conditional on the decision to purchase a new pickup truck. As such, the elasticities presented here are only applicable to the *new pickup truck fleet* rather than the broader fleet of passenger vehicles. Second, my price variation in each respective fuel covers a range of roughly \$2.50 per gallon to roughly \$4.50 per gallon. As such, these simulations are best applied to the scale of price variation which is used to identify my model. Choice sets are held fixed in my simulations; given that manufacturers may offer vehicles with significantly better fuel economy in response to a 50% to 100% increase in fuel prices, specifying a model which would allow firms to adjust product offerings in response to fuel price fluctuations may lead to a higher estimated elasticity of fleet fuel economy. However, while these restrictions should be taken seriously with regards to the scale of my estimates, they carry less importance when comparing estimates within this paper across different price beliefs. As discussed throughout this manuscript, modeling price beliefs in a more flexible manner carries importance when forward-looking consumers respond differently to different types of price variation; as such, I give special attention to the difference in counterfactual outcomes across price beliefs.

My simulations provide estimates of three relevant measures of the consumer response to price fluctuations: the elasticity of the vehicle fleet's average fuel economy, the elasticity of vehicle miles traveled, and the elasticity of fuel consumption. In Table 6.4, I present approximated elasticities resulting from a 5% change in fuel prices for each of the three specifications across each set of price beliefs. Comparing estimates using the no-change price beliefs to those using the decomposed price beliefs, those which are produced using the more flexible beliefs are roughly 15-20% larger in magnitude across each of the three elasticities. The intuition behind the difference in the outcomes lies in the nature of time-series variation in the diesel price premium. As illustrated in Figure 5.2, specifying no-change price beliefs ($\phi = 1.00$) results in a massive amount of variation in the savings associated with upgrading to a diesel engine. Under the no-change forecast, expected savings can vary by upwards of \$10,000 in a given year. If we observe similar diesel uptake regardless of the current level of expected savings under this forecast, the model must moderate α_f in order to rationalize these decisions. In essence, treating a source of price variation (time-series variation in the diesel premium) which forward-looking vehicle purchasers may not respond to in the same fashion as all other sources of fuel price variation depresses the estimated consumer responsiveness. At the end of this section, Figures 6.1-6.5 compare the consumer response across different specifications of price beliefs which illustrate the same bias as the discussed differences in elasticities across a wider range of simulated prices.

Compared to the literature, I estimate a relatively low elasticity of fleet fuel economy of 0.061 in the preferred specification. A highly inelastic estimated elasticity of the new vehicle fleet's fuel economy is typical within studies which do not specify a supply-side model where firms can respond to high fuel prices by increasing the fuel economy of their offerings. Gillingham (2013) similarly accounts for selection on anticipated usage and estimates an elasticity of 0.09, whereas Klier and Linn (2010) consider a significantly longer panel than I do and estimate an elasticity of

0.12. Studies which include a supply side model or consider the elasticity of the entire vehicle fleet's fuel economy generally produce higher estimates. While Li et al. (2009) include the effect on scrappage of used vehicles and estimate an elasticity of 0.20, Austin and Dinan (2005) include a supply-side model and produce a similar estimate of 0.22. My estimated elasticity of vehicle miles traveled of -0.090 lies on the low end of recent estimates. Because this estimate includes the rebound effect's impact on driving, it is directly comparable to Gillingham (2013) who estimates the elasticity of demand for VMT of -0.15. Accounting for the fact that my estimates include the rebound effect, fixing the individual's fuel efficiency would modestly increase the magnitude of the estimate. Comparing my estimates to those of Linn (2013), whose estimates of the elasticity of VMT are isolated from the rebound effect, they lie within his estimated range of -0.09 to -0.20.

Considering the difference between my estimates and those in the recent literature, it is worth considering a few reasons why this might occur in my market of interest. One possibility is that the elasticity of the new pickup truck's fleet fuel economy is relatively low because pickup trucks are more likely to be needed to haul heavy payload or tow cargo than other vehicles. The cost to a sedan purchaser from choosing a less powerful, more fuel efficient engine may be a reduction in acceleration, but it may not fundamentally alter how the individual can use the vehicle. A pickup truck purchaser who chooses a less powerful, more efficient engine which decreases payload capacity and towing capabilities may no longer be able to operate the vehicle in the manner which they desire.[9] The possibility that purchasers of pickup trucks are less responsive than purchasers of other vehicles is also noted by Langer and Miller (2013), who document that manufacturer incentives attached to trucks are less responsive to fuel prices than smaller vehicles. Within-household substitution between pickup trucks and smaller vehicles could explain why estimates of the elasticity of vehicle miles traveled do fall within the range of estimates produced by researchers studying car and light truck markets. While testing these possibilities is outside the scope of this paper, future research may consider the reasons pickup truck purchasers appear to be less responsive to

[9] For example, a purchaser of either a truck or a car would find it more difficult to pass another vehicle on a two-lane expressway with a less powerful engine. For a truck purchaser, however, the decrease in power may prevent them from towing their boat, hauling gravel for a home-improvement project or require them to risk an engine malfunction.

fuel price fluctuations in greater detail.

The selection correction has a negligent impact of the relevant elasticities, but it does have a noteworthy impact on the simulated distributions of usage. In Table 6.5, I present percentiles of the distribution of vehicle miles traveled at the observed level of fuel prices for the true usage distribution as well as the simulated distribution. Comparing the simulated distribution in N_1 to the observed odometer readings reveals a reasonably good fit between the 5th and 95th percentiles of the distribution. The quality of fit is poorer in the tails of the distribution, as the lognormal specification of η_i fits the distribution with a fatter right-tail than I observe in the data. This can also be seen in Figure 6.6, which plots the observed and simulated distributions of usage for N_1. The impact of higher fuel prices on the simulated usage distribution is presented in Figure 6.7 for the simulation corresponding to decomposed price beliefs. Consistent with the estimated elasticity of vehicle miles traveled, this distributions shifts to the left as fuel prices increase. The primary impact of the selection adjustment on my results lies in the difference in the distribution of usage between individuals in N_1 versus those in N_2. As a result of the estimated negative correlation coefficient between the error term in the selection equation and σ_u, the simulated usage levels of those individuals who did not report an odometer reading are higher than those who did report an odometer reading. However, this difference small due to the magnitude of ρ, with the median level of simulated usage only 2.4% higher for sample N_2. As illustrated in Figure 6.8, the selection adjustment results in higher mean vehicle miles traveled than the model without the selection adjustment, but it does not have a significant impact on the scale of the consumer response to increased fuel prices.

Table 6.1: Estimates without Selection Correction

Parameters	No-Change Beliefs		Decomposed Beliefs	
Price Parameters:				
α_f	0.856	(0.033)	1.004	(0.037)
α_p	0.126	(0.003)	0.127	(0.003)
ϕ			0.117	(0.071)
Usage Distribution:				
μ_k	2.351	(0.021)	2.365	(0.021)
Income (x $10K)	0.016	(0.003)	0.017	(0.003)
Rural Share	0.224	(0.018)	0.223	(0.018)
σ_k	0.372	(0.009)	0.372	(0.009)
σ_u	0.526	(0.007)	0.527	(0.007)
Vehicle Characteristics:				
Extended Cab	0.286	(0.009)	0.286	(0.009)
Crew Cab	0.436	(0.009)	0.436	(0.009)
Four Wheel Drive	0.381	(0.007)	0.379	(0.007)
Dual Rear Wheels	-0.174	(0.010)	-0.173	(0.010)
Work Truck Trim	-0.194	(0.004)	-0.192	(0.004)
Length (ft.)	-0.017	(0.003)	-0.019	(0.003)
Width (ft.)	0.050	(0.007)	0.055	(0.007)
Height (ft.)	0.163	(0.010)	0.170	(0.010)
Torque (100 lb. ft.)	0.111	(0.002)	0.114	(0.002)
Diesel Engine	-0.163	(0.009)	-0.180	(0.009)
Sport-Utility Truck	-0.031	(0.009)	-0.030	(0.009)
Full-Size Truck	-0.052	(0.007)	-0.051	(0.007)
Heavy-Duty: $\frac{3}{4}$ Ton Truck	-0.209	(0.011)	-0.195	(0.011)
Heavy-Duty: One Ton Truck	-0.191	(0.012)	-0.169	(0.012)
Log Likelihood	108,967		108,911	
Implicit Discount Rate	14.18%		10.53%	

Table 6.2: Estimates with Selection Correction

Parameters	No-Change Beliefs		Decomposed Beliefs	
Price Parameters				
α_f	0.772	(0.038)	0.901	(0.043)
α_p	0.126	(0.003)	0.127	(0.003)
ϕ			0.110	(0.070)
Usage Distribution				
μ_k	2.436	(0.035)	2.458	(0.036)
Income (x $10K)	0.017	(0.003)	0.017	(0.003)
Rural Share	0.228	(0.018)	0.228	(0.018)
σ_k	0.401	(0.010)	0.401	(0.010)
σ_u	0.521	(0.008)	0.522	(0.008)
Vehicle Characteristics:				
Extended Cab	0.269	(0.011)	0.267	(0.011)
Crew Cab	0.407	(0.014)	0.403	(0.014)
Four Wheel Drive	0.356	(0.012)	0.351	(0.012)
Dual Rear Wheels	-0.154	(0.010)	-0.152	(0.011)
Work Truck Trim	-0.178	(0.006)	-0.175	(0.006)
Length (ft.)	-0.016	(0.002)	-0.018	(0.002)
Width (ft.)	0.044	(0.006)	0.048	(0.006)
Height (ft.)	0.152	(0.010)	0.157	(0.010)
Torque (100 lb. ft.)	0.101	(0.004)	0.102	(0.004)
Diesel Engine	-0.145	(0.009)	-0.160	(0.010)
Sport-Utility Truck	-0.028	(0.008)	-0.027	(0.008)
Full-Size Truck	-0.047	(0.006)	-0.045	(0.006)
Heavy-Duty: $\frac{3}{4}$ Ton Truck	-0.192	(0.012)	-0.177	(0.012)
Heavy-Duty: One Ton Truck	-0.175	(0.012)	-0.154	(0.012)
Selection Parameters:				
Constant	1.210	(0.079)	1.222	(0.079)
Income (x 10K)	-0.011	(0.005)	-0.011	(0.005)
Rural Share	-0.082	(0.034)	-0.082	(0.034)
Purchase Date	-0.708	(0.011)	-0.708	(0.011)
ζ	-0.489	(0.060)	-0.500	(0.060)
ρ	-0.062	(0.055)	-0.077	(0.056)
Log Likelihood	120,213		120,155	
Implicit Discount Rate	16.88%		12.99%	

Table 6.3: Estimates with Selection Correction and Random Coefficient

Parameters	No-Change Beliefs		Decomposed Beliefs	
Price Parameters: α_f	0.771	(0.037)	0.899	(0.042)
α_p	0.126	(0.003)	0.126	(0.003)
ϕ			0.111	(0.071)
Usage Distribution:				
μ_k	2.439	(0.033)	2.461	(0.034)
Income (x $10K)	0.017	(0.003)	0.017	(0.003)
Rural Share	0.226	(0.018)	0.225	(0.018)
σ_k	0.402	(0.010)	0.402	(0.010)
σ_u	0.520	(0.007)	0.521	(0.008)
Vehicle Characteristics:				
Extended Cab	0.269	(0.011)	0.267	(0.011)
Crew Cab	0.406	(0.014)	0.403	(0.014)
Four Wheel Drive	0.355	(0.011)	0.351	(0.011)
Dual Rear Wheels	-0.154	(0.010)	-0.152	(0.010)
Work Truck Trim	-0.178	(0.006)	-0.175	(0.006)
Length (ft.)	-0.016	(0.002)	-0.017	(0.002)
Width (ft.)	0.044	(0.006)	0.048	(0.006)
Height (ft.)	0.152	(0.010)	0.157	(0.010)
Torque: Mean (100 lb. ft.)	0.101	(0.036)	0.102	(0.036)
Torque: Shape Parameter	0.006	(0.032)	0.001	(0.020)
Diesel Engine	-0.145	(0.009)	-0.160	(0.009)
Sport-Utility Truck	-0.028	(0.008)	-0.027	(0.008)
Full-Size Truck	-0.047	(0.006)	-0.046	(0.006)
Heavy-Duty: $\frac{3}{4}$ Ton Truck	-0.193	(0.011)	-0.178	(0.011)
Heavy-Duty: One Ton Truck	-0.176	(0.012)	-0.155	(0.012)
Selection Parameters:				
Constant	1.222	(0.079)	1.233	(0.078)
Income (x 10K)	-0.011	(0.005)	-0.011	(0.005)
Rural Share	-0.081	(0.034)	-0.081	(0.035)
Purchase Date	-0.709	(0.011)	-0.709	(0.011)
ζ	-0.500	(0.060)	-0.510	(0.059)
ρ	-0.062	(0.051)	-0.077	(0.052)
Log Likelihood	120,211		120,153	
Implicit Discount Rate	16.94%		13.03%	

Table 6.4: Approximated Elasticities from 5% Fuel Price Change

Specification	Price Beliefs	Fleet Fuel Economy	VMT	Fuel Consumption
No Selection Adjustment	No-Change	0.052	-0.080	-0.136
Selection Adjustment	No-Change	0.052	-0.078	-0.135
Random Coef. on Torque	No-Change	0.053	-0.078	-0.135
No Selection Adjustment	Decomposed	0.062	-0.092	-0.160
Selection Adjustment	Decomposed	0.061	-0.090	-0.158
Random Coef. on Torque	Decomposed	0.062	-0.090	-0.158

Table 6.5: Observed vs. Simulated VMT Percentiles

Percentile	**Observed:** N_1	**Simulated:** N_1	**Simulated:** N_2
1%	1,434	2,590	2,587
5%	3,650	4,049	4,025
10%	5,160	5,056	5,128
25%	8,068	7,598	7,703
50%	11,758	11,738	12,020
75%	16,595	18,482	18,668
90%	22,534	26,733	27,956
95%	27,015	33,770	36,002
99%	37,916	52,365	55,125

Figure 6.1: Counterfactual Fleet Fuel Economy

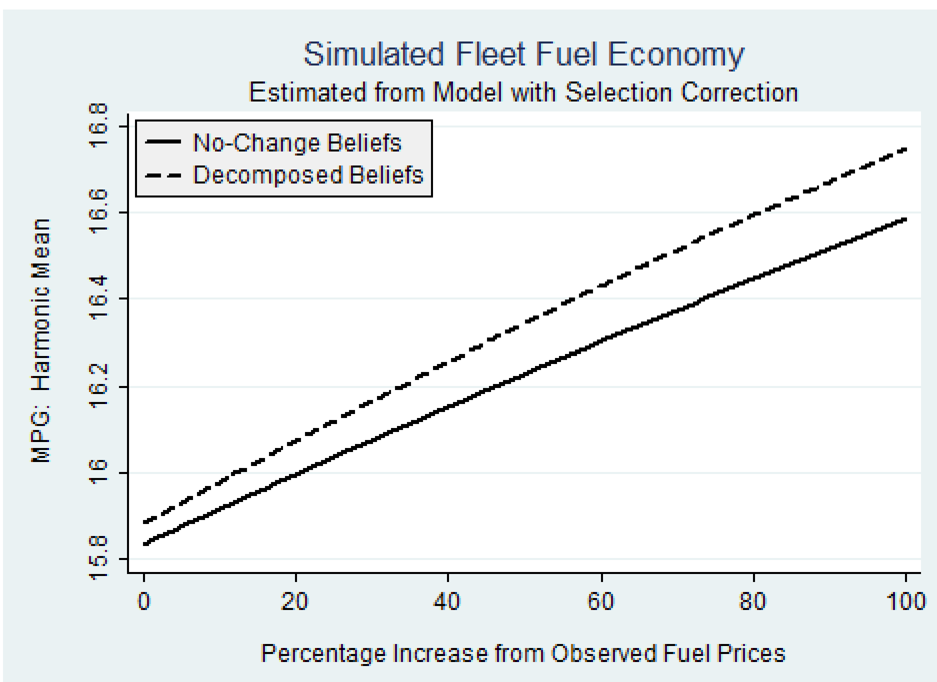


Figure 6.2: Counterfactual Change in Fleet Fuel Economy

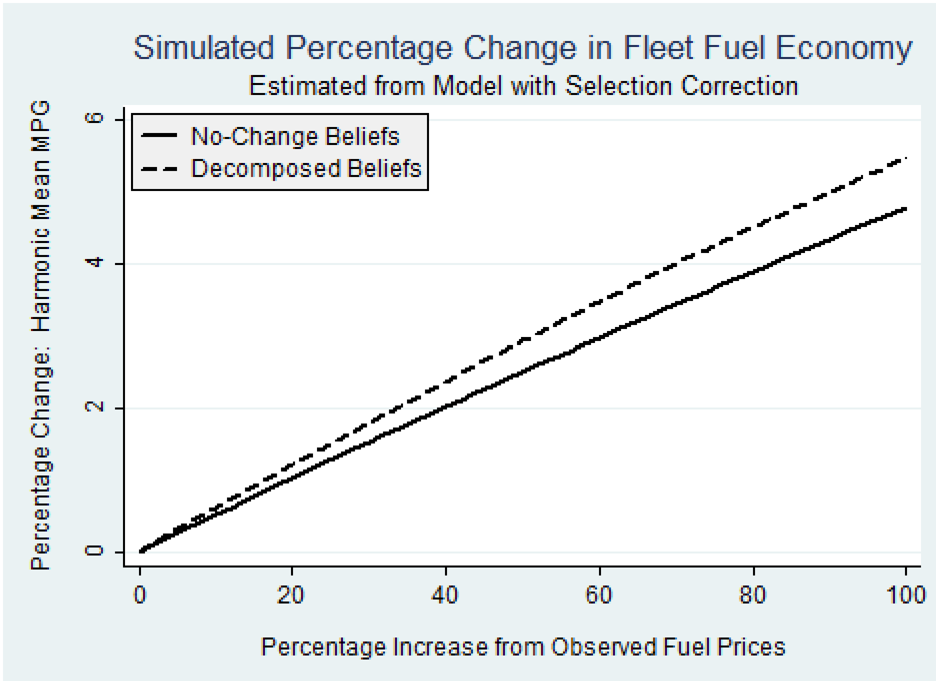


Figure 6.3: Counterfactual Change in Mean VMT

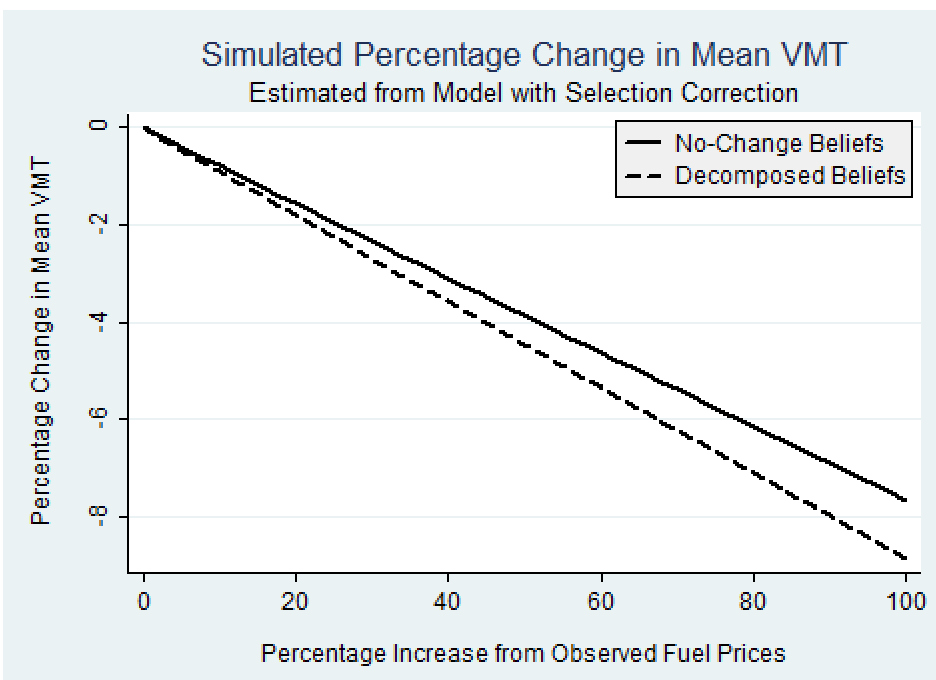


Figure 6.4: Counterfactual Average Fuel Consumption

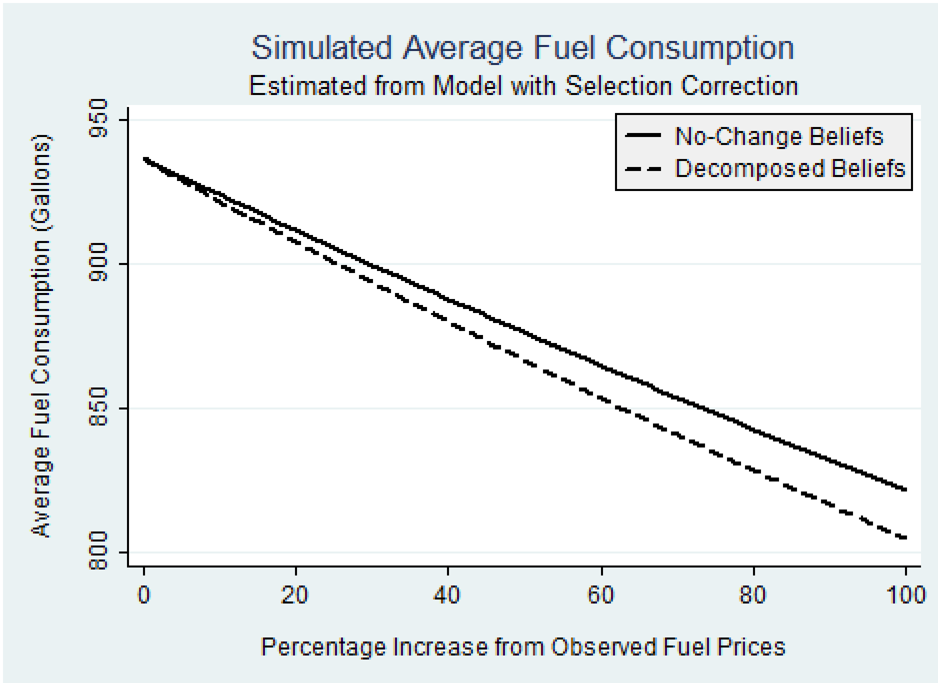


Figure 6.5: Counterfactual Change in Fuel Consumption

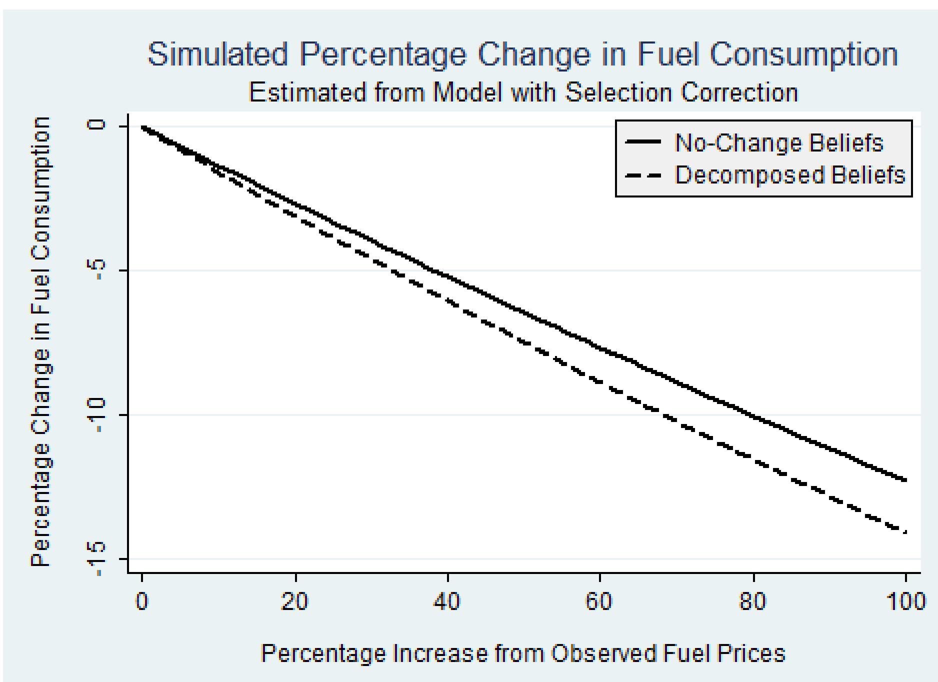


Figure 6.6: Observed and Simulated VMT Distribution for Sample N_1

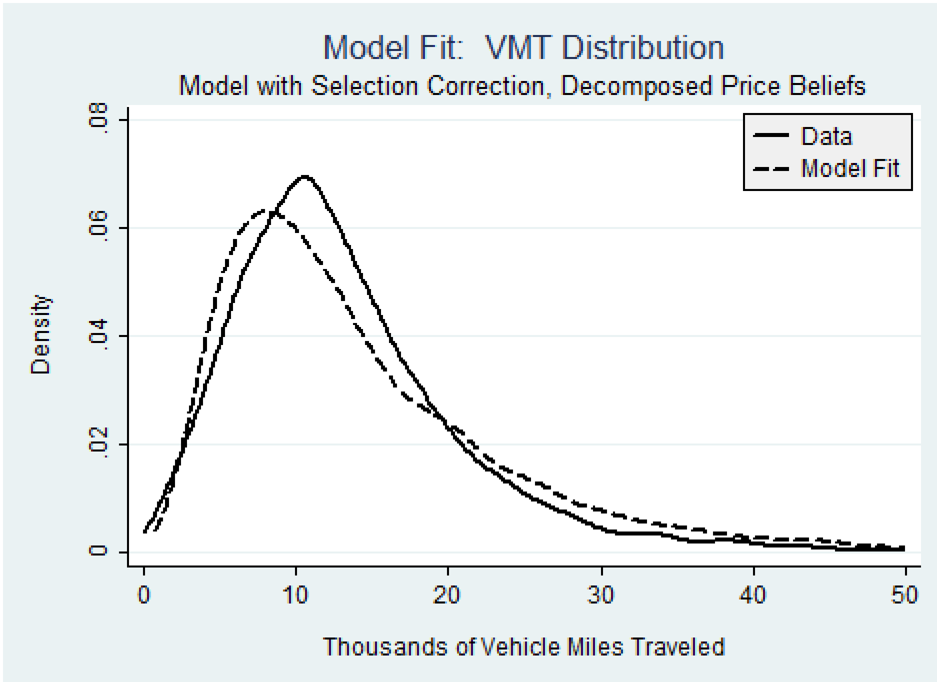


Figure 6.7: Counterfactual VMT Distribution, Full Sample

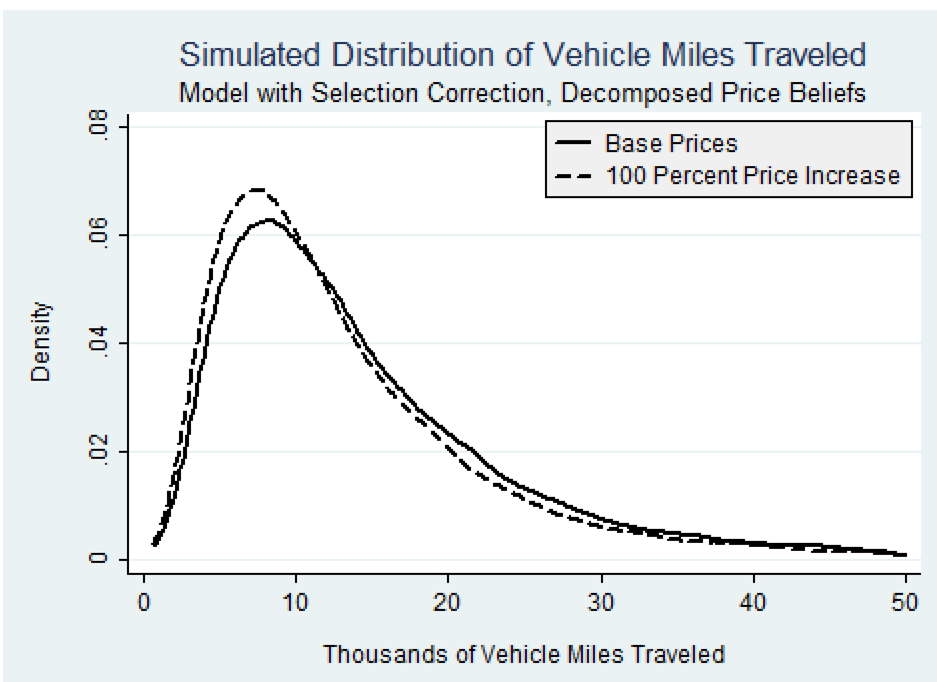
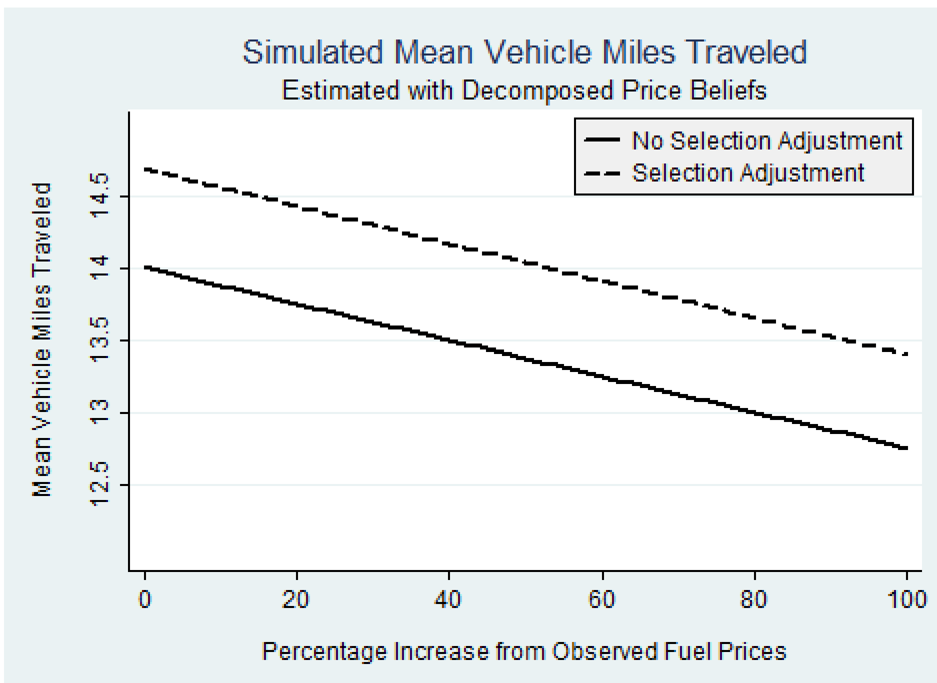


Figure 6.8: Counterfactual Mean VMT With/Without Selection Adjustment



SECTION 7

CONCLUSION

For the past century the U.S. market for automobiles has been dominated by gasoline-powered vehicles. Because gas prices have been notoriously difficult to forecast, modeling price expectations with a no-change forecast when gasoline is the sole fuel considered can be a reasonable approximation to consumer beliefs. When a fuel other than gasoline is offered on vehicles in consumers' choice sets, using a no-change forecast for each fuel may not approximate consumers' beliefs. The proposed approach here, where price expectations are tailored to account for the transient nature of time-series variation in the diesel premium, allows for the possibility that such variation results in smaller changes in price expectations than is assumed under a no-change forecast.[1] My results indicate that price expectations are considerably less responsive to time-series variation in the diesel premium, and in turn, a no-change forecast results in underestimation of consumers' responsiveness to fuel price fluctuations. The magnitude of the bias is significant despite the modest scale of time-series variation in the diesel premium, as parameter estimates and the relevant elasticities change by about 15-20% when the decomposed forecast is specified.

Future research on consumers' responsiveness to fuel price fluctuations will need to give increasing consideration to adoption of non-gasoline powered vehicles as their market shares rise. Just as forward-looking pickup truck purchasers should form price beliefs which account for the mean-reverting nature of time-series variation in the diesel premium, forward-looking purchasers of alternative-fuel vehicles ought to consider the evolution of the operating costs when forming

[1]The challenge of specifying beliefs over product attributes which are consistent with those generating the data is by no means limited to automobile adoption. In many cases, the quality of product attributes (in this case, fuel economy) evolve according to some process wherein the expected attribute quality over the life of the product differs from its value at purchase. For example, Bishop and Murphy (2016) consider the willingness to pay for neighborhood safety among home purchasers, which also has a mean-reverting tendency, and derive a bias adjustment to incorporate the possibility of forward-looking behavior in a static model.

price beliefs. In the case of electric vehicles, forward-looking consumers should account for the time-trend in electric vehicle charging infrastructure. Given the increasing quality and quantity of EV chargers, consumers' expected refueling cost should decrease over the life of the vehicle.[2] Further, with rapid change in both refueling costs and EV quality over the past decade, this would be a natural application to consider the role of price beliefs within a dynamic model of vehicle adoption. I leave the prospect of modeling price beliefs in a dynamic model of vehicle adoption for later work.

A second area of future research to be considered is the adoption and usage of large vehicles. Medium-duty and heavy-duty vehicles are covered under a different regulatory umbrella than light-duty vehicles in the United States. While such vehicles represent only about 5% of all vehicles in operation in the U.S., they represent roughly one fifth of greenhouse gas emissions associated with the transportation sector (EPA 2011). Under the most recent standard developed jointly by the EPA and NHTSA, these vehicles will have to decrease emissions by roughly 25% over the next decade, some of which must result from increased fuel economy. Despite the renewed policy interest placed on trucks, the rich literature on consumers' responsiveness to fuel price fluctuations is largely focused on vehicles covered by CAFE regulations. While Parry (2008) considers optimal taxation for heavy-duty trucks and Leard et al. (2016) estimate a rebound effect for drivers of such vehicles, additional research into the costs of taxation versus the newly developed emissions regulation is warranted. Such research poses several challenges, including the difficulty of modeling the decisions of fleet managers rather than individual consumers, measuring fuel prices when consumers purchase fuel in many different locales, and obtaining high-quality data on fuel economy of heavy-duty trucks. For the aforementioned policy importance and the intrigue of modeling vehicle choice in this context, I leave this avenue open for future research.

[2] In the case of EVs, this can refer to either the nuisance cost of recharging the vehicle or the monetary cost.

APPENDIX A

APPENDIX: CHAPTER 3

A.1 Institutional Details

The United States automobile market has historically seen a high proliferation of personally owned pickup trucks compared to most other markets. These vehicles have a sizable market share, with the 'big three' pickup models (Ford F-Series, Chevrolet Silverado, and Dodge Ram) often ranking first, second, and third among all vehicle models in U.S. sales. It is well documented that the light truck market segment, which includes pickup trucks, SUV's, and vans, has been highly important to U.S. manufacturers over the past quarter-century. While the market share of domestic automobiles fell steadily over much of the past half-century, its stabilization from the early 1980s through mid 1990's is largely attributed to sales of light trucks (Klier, 2009). Looking more closely at the sub-classes of vehicles within the 'light truck' category further reveals the importance of the pickup truck sub-class. Whereas domestic manufacturers captured only a 34% market share of cars from the 2015 model year and a 49% market share of SUV's and vans, they enjoyed a market share of better than 85% of pickup trucks, including a 100% market share among heavy-duty pickup trucks. Although the entry of Japanese automakers such as Toyota, Honda, and Nissan into the U.S. pickup truck market has lead to some erosion in the domestic market share, this decline is comparatively small to that observed in other vehicle classes.

Pickup trucks in the U.S. typically fall into one of five classes which constitute the vast majority of sales: mid-size (e.g. Chevrolet Colorado), sport-utility truck (e.g. Honda Ridgeline), half ton (e.g. Chevy Silverado 1500), three-quarter ton (e.g. Chevy Silverado 2500), and one ton (Chevy Silverado 3500).[1] Vehicles vary significantly across classes based on their performance characteristics. Across Chevrolet's line of model year 2016 (MY2016) trucks, maximum towing (payload) capacities are reported as 7,700 (1,580) pounds for the Colorado, 12,000 (2,260) pounds

[1] Half-ton trucks are also commonly referred to as full-size trucks, while three-quarter-ton and one-ton trucks are commonly referred to as heavy-duty trucks. Compact trucks such as the Chevrolet S-10 and Ford Ranger enjoyed large market shares throughout the late 20th century, but no pickup truck currently offered on the market fits into the compact category. The Ranger is expected to return to the U.S. market for MY2019 with a diesel engine available to pose as a competitor to the Chevrolet Colorado.

for a Silverado 1500, 18,000 (3,534) pounds for the Silverado 2500, and 23,200 (7,153) pounds for the Silverado 3500 with dual rear wheels. The reader should note the dissonance between the class definitions of 'half', 'three quarter', and 'one' ton. Historically, these figure were loosely tied to the maximum amount of payload which could be hauled in a truck of each respective class. Just as the shift towards larger, more powerful vehicles in the light-duty segment is documented by Knittel (2013) among others, this trend is also readily apparent among large pickup trucks.

Owing to the significant growth in these vehicles' size and performance capabilities, many of the pickup trucks produced today are not regulated under the same set of policies as cars, SUV's, and most vans. Under the first and second phases of the Greenhouse Gas Emissions Standards and Fuel Efficiency Standards set forth jointly by the Environmental Protection Agency (EPA) and National Highway Traffic Safety Administration (NHTSA), pickup trucks with a gross vehicle weight rating (GVWR) under 8,500 pounds are regulated under Corporate Average Fuel Economy (CAFE) standards as light-duty vehicles whereas pickup trucks with a GVWR over 8,500 pounds are regulated under a separate standard for medium and heavy-duty vehicles.[2] When CAFE standards were initially developed in the late 1970's, all vehicles under a GVWR of 6,000 pounds fell under this regulatory umbrella; this threshold was raised to its current level of 8,500 pounds for pickup trucks shortly thereafter. The predecessors to many of the trucks discussed in this paper which now fall above the 8,500 pound threshold fell below the threshold when it was put in place some thirty-five years ago. For example, the 1982 the Chevrolet C/K 20 ranges in GVWR from 6,400 to 8,600 pounds while its 2013 analogue, the Chevrolet Silverado 2500, ranges in GVWR from 9,300 to 10,000 pounds. The 8,500 GVWR threshold has garnered criticism from some groups (see Mackenzie et al. 2005), and was increased to 10,000 pounds for SUV's and vans, but it has not been adjusted for pickup trucks. Today, pickup trucks over the 8,500 GVWR threshold remain classified as medium-duty vehicles. Although recent regulations require medium and heavy-duty vehicles to significantly reduce emissions levels over the next decade, manufacturers which are constrained by CAFE standards may retain an incentive to offer or report vehicles with GVWR's

[2] GVWR refers to the maximum weight of the vehicle when it is fully loaded with passengers, cargo, payload, fuel, etc.

above the 8,500 pound threshold as a means of improving their CAFE.[3]

The diesel engine option in pickup trucks is typically a costly upgrade which historically was available only on heavy-duty models as a means to improve towing capability as well as decrease fuel costs. In recent years, however, many manufacturers have given serious consideration to developing and offering diesel engines in smaller trucks as a means of reaching more stringent CAFE requirements and differentiating truck offerings from competitors' offerings. Dodge became the first manufacturer to offer a diesel engine option in a half-ton truck when it released a RAM 1500 equipped with a diesel engine late in MY2014. More recently, General Motors began offering a diesel engine option in both the Chevrolet Colorado and the alternatively-badged GMC Canyon, while Nissan has taken aim at the U.S. manufacturers' market share of heavy-duty trucks with the release of their so-called 'five-eighths ton' Nissan Titan XD, which has been offered with a 5.0L Cummins diesel engine since late in the 2016 model year. While the petroleum savings associated with adoption of diesel engines rather than conventional gasoline engines are modest compared to the savings associated with hybrid electric vehicles (e.g. Toyota Prius) or plug-in electric vehicles (e.g. Nissan Leaf), the latter technologies have failed to gain any significant traction in the U.S. pickup truck market. However, given the size and corresponding fuel efficiency of pickup trucks, modest improvements in fuel efficiency are associated with comparatively large reductions in petroleum usage and carbon dioxide emissions. As a simple comparison, note that driving 12,000 miles in a 2015 Toyota Camry equipped with the available 4-cylinder hybrid engine rather than the 4-cylinder gasoline engine decreases carbon emissions by roughly one ton whereas driving 12,000 miles in a 2015 RAM 1500 equipped with the available 6-cylinder diesel engine rather than the 8-cylinder gasoline engine decreases carbon emissions by roughly 1.2 tons.[4] Although a

[3] As shown in Mackenzie et al. (2005), ten of the eighteen vehicles considered which crossed the 8,500 pound threshold did so by under 200 pounds. Selling vehicles above the 8,500 pound threshold which satisfy emissions requirements for medium-duty vehicles rather than selling vehicles just below the threshold can result in improved CAFE. For example, selling one additional truck with a GVWR of 8,400 pounds with fuel economy of 15 miles-per-gallon would decrease any major automaker's CAFE. However, selling one additional truck with a GVWR of 8,500 pounds with fuel economy of 14 miles-per-gallon would not.

[4] Calculations based on EIA estimates of CO2 released resulting from burning each respective fuel (18.95 pounds per-gallon of E10 gasoline, 22.3 pounds per-gallon of B5 diesel) and user-generated fuel economy data from fuelly.com which is used throughout the rest of this paper. RAM 1500's equipped with the V8 engine do enjoy a modestly higher

few manufacturers have been experimenting with hybrid engine offerings, it remains to be whether others will follow suit and when these engines will ultimately be released.[5]

A.2 Data Cleaning

A.2.1 Sample Extraction

In this subsection, I describe the process I undertook to extract new pickup truck purchases from the registration files provided to me by the DOL. The DOL recorded each registered vehicle's model in the registration database, but a significant proportion of these entries either did not map to a single model (e.g. "PICKUP", "F-SERIES") or contained an error (e.g. a vehicle with make listed as "HOND" and model listed as "COROLLA"). Vehicles with a recorded model year before 2000 were also dropped from the sample. In order to identify any vehicle which could likely be a pickup truck, I extracted all VINs where the model field may indicate the vehicle is a pickup truck. When it was unclear whether the VIN represented a pickup truck, it was included. From this set of VINs, I kept one VIN for each unique combination of the first ten digits of the VIN. This process produced a list of 10,238 unique VIN Stubs containing the first ten digits of a VIN. For each VIN Stub, I retained one VIN which I then fed through my online decoder. I was able to decode a total of 8,954 out of the 10,238 VIN Stubs, but I was unable to decode a nontrivial proportion of the VIN stubs. However, while the proportion of VIN Stubs which I was unable to decode is somewhat high (12.54%), these observations represent a relatively small proportion of vehicles in my data. Many of the observations which I failed to decoded appear to be associated with an error in data entry, whereas others are associated with larger commercial vehicles than those considered in this project. Out of a total of 7,464,165 observations in my registration files, a total of 7,417,400 (99.37%) are associated with a VIN Stub which was decoded. Out of the remaining 8,954 VIN Stubs which correspond to 7,464,165 observations, I proceeded to drop any VIN Stub which did

---

level of performance, but the point of this calculation is to note that substitution towards more efficient engines in trucks can yield surprisingly large decreases in carbon emissions. This point is also made by Larrick and Soll (2008).

[5] General Motors offered a hybrid engine from MY2009-MY2013 in the Chevy Silverado and GMC Sierra, but uptake was low and the option was abandoned. Ford and Toyota developed a partnership in 2011 in hopes of developing a hybrid pickup truck offering, but Ford pulled out of the project in 2013. Both firms have stated publicly that they plan to release hybrid pickup trucks in the future, but official release dates have not been set.

not represent a pickup truck; these observations represented a total of 1,267 VIN Stubs which are associated with a total of 2,158,578 registration observations.[6]

Given the timing of my fuel price data, the project is restricted to new vehicle purchases between January 1, 2012 through December 31, 2015. In order for an observation to be included in the estimation sample, it must satisfy the four conditions. First, I must observe the original transcation for a vehicle some time between January 1, 2012 and December 31, 2015; any vehicle which does not carry an entry coded as an 'original transaction' by the DOL within these dates was not purchased as a new vehicle during those dates. Second, vehicles must carry a model year of 2010 or later. This is done so as to limit choice sets; a subset of vehicles purchased after January 1, 2012 do have a model year before 2010, but the vast majority of new vehicle sales correspond to a model year no more than one year lower than the calendar year. Third, I retain only those vehicles with an odometer reading below 1,000 miles at the time of their initial registration. This distinction is made to eliminate vehicles which were likely purchased outside of Washington and only registered in the state at a later date. [7] Fourth, pickup trucks must be either at or below the 'one-ton' pickup truck size class.[8] A total of 128,174 vehicles satisfy all four conditions, and this figure should be taken as the starting value with regards to sample construction. I drop 395 observations which correspond to MY 2010, are listed with a specialty trim which could not be aggregated, or have a body type which is not some sort of pickup truck.[9] In order to limit choice sets in the third chapter of

[6]As discussed earlier, any vehicle with a model name (as recorded by the DOL) which could have reasonably represented a pickup truck was decoded. A large share of these observations which were dropped are associated with the Ford Explorer, a popular full-size SUV. From 2000 through 2010, Ford produced the Explorer Sport Trac, a pickup truck based on the Explorer. For either of these vehicles, it was common for the DOL to record the model as 'EXPL', and a total of 411 of the decoded VIN Stubs which were not pickup trucks were from some variant of the Ford Explorer.

[7]For example, if an individual purchased a new F-150 in Oregon and later moved to Washington, this entry would be coded as an 'original transaction'. However, the odometer reading threshold would eliminate this vehicle from the sample of 'new vehicle purchases'.

[8]For reference, the Ford F-350 is a 'one-ton' pickup truck and is included in the sample, whereas the larger Ford F-450 is dropped from the sample.

[9]These body types include 'Cab and Chassis' trucks and a few commercial vans.

this project, I drop 2,399 observations which correspond to products (defined at the level of Make-Model-Cab-Drivetrain-Dual Rear Wheels-Engine-Work Truck Trim-Generation) which makes up less than 0.02% of the sample. Due to minor limitations in fuel price data for two sparsely populated counties, I drop 846 observations from San Juan County and Garfield County, as well as those with an unknown registration county. Dropping an additional 44 observations which are listed with an owner type of "In care of" results in a final sample of 124,310 observations.

A.2.2 Aggregation Process

As mentioned in the body of this paper, reducing choice sets to a point where estimation is computationally feasible requires that the researcher aggregate vehicles up to certain characteristics; I describe my aggregation process in detail here.

1. For vehicles with identical Model-Model Year-Series-Body-Drivetrain-Wheel Configuration-Engine-Trim-Wheelbase Option-Doors, but different MSRPs, all values are specified to the values of the vehicle with the lower MSRP. Vehicles which were identical according to the aforementioned characteristics but a higher MSRP typically include some small upgrade, e.g. a chrome wheel package, which carries an upgrade cost on the order of a couple hundred dollars.

2. For vehicles with identical Model-Model Year-Series-Body-Drivetrain-Wheel Configuration-Engine-Trim-Wheelbase Option, but a different number of doors, all values are specified to the value of the vehicle with four doors. This is only pertinent to a few trucks which were offered with both a 2-door and 4-door extended cab.

3. For vehicles with identical Model-Model Year-Body-Drivetrain-Wheel Configuration-Engine-Trim, but a different wheelbase, all values are specified to the value of the vehicle with the standard wheelbase (in most cases, this gives a vehicle with similar interior dimensions but a smaller pickup box). Upgrading to a longer pickup box generally costs somewhere around a few hundred dollars.

I aggregate trim to a maximum of two trim levels per vehicle.[10] Trim levels are entered as follows:

- For vehicles where the trim level cannot be determined from the VIN (this occurs for around half of my VINs, primarily Ford F-Series trucks), I enter a trim which is one level up from the 'base' or 'work' specification. For example, I enter the trim for all F-Series as 'XLT'.

- For vehicles where the trim level can be determined from the VIN, I initially enter the trim implied by the VIN. All trim levels above the 'base' or 'work' specification are then aggregated down a level comparable to Ford's 'XLT' line.

- The exception to the prior rule applies to two models. For the Chevrolet Avalanche and Honda Ridgeline, an even higher trim was chosen by the majority of purchasers. For these vehicles, which make up a small percentage of my sample, I aggregate to these higher trim levels.[11]

For many vehicles, this leaves a single trim option. I leave the 'work truck' trim level as an option where applicable, because in some instances this is the only trim level available for particular combinations of body-type and drivetrain (typically regular cab, 2-wheel-drive configurations).[12] In order to further pare-down choice sets to decrease computation time, I aggregate to the level of generation rather than Model Year. Like cars, pickup trucks do not undergo significant changes in style and characteristics on a yearly basis. Rather, truck models are overhauled somewhere in the neighborhood of once every five years. [13] I aggregate characteristics to those of the newest model

[10]I define the Ford F-150 SVT Raptor, a high-performance model with a 6.2L engine typically utilized in their heavy-duty trucks, as a separate series of vehicle.

[11]There are Model Years where for these vehicles, I do not observe a purchase below the specified trim level. Thus, some fairly laborious manual data entry would be necessary to aggregate at a lower trim level. These are the only 'Sport Utility Trucks' available; specifications which include a dummy for truck class should account for the different trim entry.

[12]This stipulation applies to the Nissan Frontier as well as most RAM and GM trucks.

[13]For example, the 12th generation of the Ford F-150 was released beginning on Model Year 2009, and the 13th generation was released beginning in Model Year 2015.

year that has been purchased up to a given point in time in the data; however, within a vehicle generation, changes in characteristics tend to be minor.[14]

A.3 Fuel Economy Drivetrain Adjustment

Because equipping a vehicle with four-wheel drive typically decreases fuel economy (relative to the two-wheel drive counterpart), I adjust fuel economy based on the vehicle's drivetrain in order to properly estimate the willingness to pay for four-wheel drive. While this process undoubtedly leads to some measurement error in the fuel economy variable, I believe it to be of a fairly small magnitude, and significantly decreases measurement error relative to a process where I make no such adjustment. To account for differences in fuel economy between four-wheel-drive and two-wheel-drive versions of otherwise identical vehicles, I use the EPA's fuel economy datafiles to compute the estimated percentage change in fuel economy which results from a vehicle having four-wheel-drive. At the level of Model-Series-Generation-Engine (the same level as the fuelly data is entered), I compute the proportion of vehicles which are equipped with two-wheel-drive and four-wheel-drive, respectively. I assume that the vehicles which generated my fuelly data had this proportion of two-wheel-drive and four-wheel-drive vehicles.[15] Thus, the observed fuel economy pulled from fuelly is given by:

$$FE = \frac{Pr(4WD) + Pr(2WD)}{\frac{Pr(4WD)}{MPG_2 * P} + \frac{Pr(2WD)}{MPG_2}}$$

where MPG_2 denotes the fuel economy of the two-wheel drive version, P denotes the fuel economy of the four-wheel-drive version as a percentage of the two-wheel drive version (hence, $MPG_4 = MPG_2 * P$), and $Pr(4WD), Pr(2WD)$ give the probabilities of a particular vehicle

[14]The most significant change within generation is generally the availability of a new engine (a tune-up of the previous year's engine is common in diesel heavy-duty pickup trucks, where the manufacturer will find a way to yield a few extra horsepower or pound-feet of torque out of the same base engine), or in some instances a new drivetrain or cab type. However, given that I leave these disaggregated, aggregating to the level of generation rather than model year will not aggregate vehicles with the same generation, a different Model Year, and significantly different characteristics.

[15]A potential problem here is that distributions of drivetrains may be different across different geographies, and while my data are confined to Washington, the fuelly data may be entered from across the country. For example, regions with heavy (light) snowfall likely have greater (lower) rates of four-wheel-drive adoption. However, given that the fuel economy penalty from four-wheel-drive is generally in the neighborhood of five percent, such differences would not significantly alter estimated fuel economy.

having four-wheel drive or two-wheel drive, respectively. Solving this for MPG_2 yields the estimated fuel economy of the two-wheel drive version, and multiplying this estimate by P gives the estimated fuel economy of the four-wheel drive version.

$$MPG_2 = FE * \frac{Pr(4WD) + P * Pr(2WD)}{P}$$

The fuel economy penalty for purchasing a four-wheel drive truck, rather than an otherwise identical two-wheel drive version, is typically on the order of five percent. Because heavy-duty trucks are not tested by the EPA, I do not observe any measure of the difference between two-wheel drive and four-wheel drive versions of these vehicles. In order to adjust fuel economy of heavy-duty trucks, I must impute the percentage difference between fuel economy of the two-wheel drive and four-wheel drive versions. To do so, I regress the four-wheel drive penalty for vehicles tested by the EPA on the vehicle's wheelbase, engine cylinders, engine displacement, torque, and a dummy for whether the engine uses diesel.[16] I then impute values as predicted by this regression as estimated fuel economy penalties for heavy-duty trucks with four-wheel drive.

A.4 Fuel Prices

While I briefly discuss OPIS data collection process in §3.2, it is worthwhile noting a few more details about this process as it pertains to my sample of retailers. First, the quantity of retailers which sell diesel is substantially higher than the ratio of daily diesel reports to daily gasoline reports (for reference, the mean ratio is approximately 51.88%, and a time-series of price reports by fuel type is shown in Figure A.1). In Figure A.2 and Figure A.3, I present a histogram of the count of daily price reports by retailer for diesel and gasoline, respectively. Out of the 2,649 retailers which report a gasoline price at least once in my sample, only 102 (3.85%) never report a diesel price, whereas 472 retailers (17.82%) report a diesel price on fewer than one dozen days throughout the sample. Although my data do not allow me to see when retailers are offering diesel fuel, the pattern of diesel price reports suggests that at any given point in time, the vast majority of

[16] As with earlier, this is computed by comparing the EPA-estimated unrounded combined fuel economy for vehicles which are identical up to Model-Series-Model Year-Engine, but have a different drivetrain.

retailers in the state sell diesel.

As mentioned in §3.2, a few counties in Washington have very low population levels and few filling stations. Thus, even when aggregating at the level of day-county, my data contain missing observations in several counties. To impute observations, I run a simple algorithm. In the first round, the missing daily county-average fuel price is coded as the county-average price from the next day; if the price is still missing, it is coded as the county-average price from the previous day. In the second round, any remaining missing daily county-average fuel price is coded as the county-average price from two days later; if the price is still missing, it is coded as the county-average price from two days prior. For all but two counties, one iteration is sufficient to impute all missing gasoline prices, while four iterations is sufficient to impute all missing diesel prices. The two remaining counties, Garfield and San Juan, have significantly more missing observations than any other county and as such, require dozens of iterations to fill in all of the relevant prices.[17] For this reason, I drop observations which are registered in either of these counties. The outcomes of this data entry process are shown in Table A.1.

A.5 Stationarity of the Diesel Premium

To verify the stationarity of the diesel premium, I gathered monthly average gasoline and diesel prices in the United States from the Energy Information Administration (EIA). Beginning in March 1994, I compute the monthly average price difference between diesel and gasoline for the country as a whole as well as five regions within the United States, and deflate all prices using the Consumer Price Index as provided by the Bureau of Labor Statistics. As shown in Table A.3, the diesel premium in the West Coast Petroleum Administration for Defense District (PADD) has been noticeably lower than the diesel premium in the four other PADDs. The time-series of diesel premium over this period is illustrated in A.4 for the United States, A.5 for the Rocky Mountain PADD, and A.6 for the West Coast PADD. The reader should note a shift in the diesel premium during this period as Ultra-low-sulfur-diesel was phased in throughout the United States during the

[17] San Juan County is a set of islands which are not accessible from the mainland by road with a population of 15,769. Garfield County is the least densely populated and least populated county in Washington with just 2,266 residents (Census 2010).

2000's. The phase-in was complete by the beginning of the data period for all other sections of this manuscript, which begin in January 2012.

To consider the stationarity of the diesel premium, I conduct an augmented Dickey-Fuller test on the time series of diesel premium data in each of the three aforementioned regions from 1994 through 2016. Letting the price difference as $\tilde{p}_t$, this test runs an OLS regression on

$$\Delta\tilde{p}_t = \alpha + \rho\tilde{p}_{t-1} + \delta t + u_t$$

Under the null hypothesis, there is a unit root and $\rho = 1$. Insofar as the seasonality in fuel prices (among other factors) could result in serial correlation in u_t, I also run an augmented Dickey-Fuller test with three lags of the form

$$\Delta\tilde{p}_t = \alpha + \beta\tilde{p}_{t-1} + \delta t + \zeta_1\Delta\tilde{p}_{t-1} + \zeta_2\Delta\tilde{p}_{t-2} + \zeta_3\Delta\tilde{p}_{t-3} + \epsilon_t$$

where we seek to test the null hypothesis that there is a unit root ($\beta = 0$). Results for the United States, the Rocky Mountain region, and the West Coast are presented in Tables A.4-A.6. As noted by the Mackinnon approximate p-values in the bottom row of the table, in all scenarios we reject the null hypothesis of a unit root at a 1% significance level.

Table A.1: Missing Observations by Stage of Data Entry Algorithm by County

Diesel	Zero	One	Two	Three	Four	**Gasoline**	Zero	One	Two	Three	Four
Adams	1					Adams					
Asotin	357	48	10	2		Asotin	12				
Benton						Benton					
Chelan						Chelan					
Clallam	1					Clallam					
Clark						Clark					
Columbia	313	36	4	1		Columbia	18				
Cowlitz						Cowlitz					
Douglas	6					Douglas					
Ferry	291	23	4			Ferry	15				
Franklin						Franklin					
Garfield	1271	987	786	636	549	Garfield	341	76	34	19	12
Grant						Grant					
Grays Harbor						Grays Harbor					
Island	1					Island					
Jefferson	24	1				Jefferson					
King						King					
Kitsap						Kitsap					
Kittitas						Kittitas					
Klickitat	22					Klickitat					
Lewis						Lewis					
Lincoln	87	2				Lincoln					
Mason	1					Mason					
Okanogan	5					Okanogan					
Pacific						Pacific					
Pend Oreille	21					Pend Oreille	2				
Pierce						Pierce					
San Juan	1152	811	608	474	384	San Juan	125	46	39	33	28
Skagit						Skagit					
Skamania	163	9				Skamania	3				
Snohomish						Snohomish					
Spokane						Spokane					
Stevens	7					Stevens					
Thurston						Thurston					
Wahkiakum	98	9				Wahkiakum	18				
Walla Walla						Walla Walla					
Whatcom						Whatcom					
Whitman						Whitman					
Yakima	7					Yakima					

Table A.2: Fuel Price Summary Statistics, by County

County	$\bar{p}_D$	p_D^{Min}	p_D^{Max}	$\bar{p}_G$	p_G^{Min}	p_G^{Max}	Δ	Δ^{Min}	Δ^{Max}
Adams	3.78	2.35	4.50	3.43	2.00	4.24	0.35	-0.36	0.97
Asotin	3.68	2.24	4.50	3.29	1.81	4.10	0.39	-0.43	1.12
Benton	3.73	2.38	4.46	3.31	1.80	4.16	0.42	-0.05	0.92
Chelan	3.83	2.48	4.58	3.43	2.12	4.27	0.39	-0.25	0.91
Clallan	3.75	2.40	4.62	3.50	2.13	4.41	0.25	-0.23	0.77
Clark	3.72	2.51	4.47	3.44	2.11	4.31	0.28	-0.08	0.69
Columbia	3.85	2.30	4.53	3.44	1.99	4.29	0.41	-0.13	1.20
Cowlitz	3.71	2.43	4.46	3.47	2.21	4.32	0.23	-0.27	0.70
Douglas	3.78	2.35	4.61	3.37	2.05	4.25	0.42	-0.17	1.04
Ferry	3.74	2.40	4.51	3.40	1.86	4.27	0.34	-0.42	1.37
Franklin	3.77	2.46	4.50	3.33	1.82	4.16	0.45	0.01	0.97
Garfield*	3.68	2.50	4.47	3.41	2.02	4.13	0.41	-0.28	1.01
Grant	3.77	2.29	4.57	3.40	1.96	4.25	0.37	-0.26	0.78
Grays Harbor	3.80	2.56	4.65	3.53	2.22	4.40	0.27	-0.13	0.77
Island	3.73	2.32	4.55	3.53	2.30	4.37	0.20	-0.24	0.57
Jefferson	3.78	2.45	4.60	3.48	2.15	4.31	0.30	-0.19	0.94
King	3.86	2.70	4.59	3.52	2.29	4.36	0.34	-0.03	0.73
Kitsap	3.79	2.52	4.61	3.44	2.15	4.29	0.35	-0.09	0.81
Kittitas	3.68	2.18	4.55	3.42	1.90	4.34	0.26	-0.45	0.74
Klickitat	3.74	2.35	4.58	3.50	2.24	4.38	0.25	-0.32	0.71
Lewis	3.72	2.43	4.53	3.46	2.16	4.35	0.26	-0.23	0.70
Lincoln	3.80	2.21	4.52	3.37	1.89	4.19	0.42	-0.22	1.05
Mason	3.70	2.13	4.54	3.45	2.10	4.32	0.26	-0.37	0.84
Okanogan	3.85	2.39	4.63	3.46	1.98	4.31	0.39	-0.09	1.01
Pacific	3.79	2.25	4.63	3.57	2.32	4.37	0.22	-0.40	0.73
Pend Oreille	3.82	2.54	4.52	3.39	1.88	4.10	0.43	-0.14	1.01
Pierce	3.79	2.61	4.56	3.45	2.17	4.33	0.34	-0.06	0.77
San Juan*	4.16	2.88	5.08	4.05	2.78	5.08	0.13	-0.49	0.72
Skagit	3.71	2.27	4.57	3.46	2.13	4.38	0.24	-0.25	0.65
Skamania	3.76	2.50	4.60	3.54	2.32	4.44	0.23	-0.26	0.70
Snohomish	3.80	2.56	4.55	3.48	2.21	4.32	0.32	-0.04	0.72
Spokane	3.72	2.52	4.36	3.27	1.74	3.99	0.45	0.01	1.05
Stevens	3.72	2.12	4.55	3.35	1.75	4.17	0.37	-0.38	0.86
Thurston	3.81	2.64	4.58	3.46	2.19	4.34	0.35	-0.05	0.81
Wahkiakum	3.82	2.77	4.57	3.56	2.22	4.40	0.26	-0.20	0.69
Whatcom	3.76	2.29	4.61	3.52	2.11	4.46	0.23	-0.25	0.66
Whitman	3.76	2.45	4.60	3.37	1.89	4.15	0.39	-0.19	0.84
Yakima	3.74	2.46	4.51	3.38	1.96	4.24	0.37	-0.09	0.76

Table A.3: United States Gasoline and Diesel Prices, by PADD

Region	$\bar{p}_D$ 1994-2016	$\bar{p}_G$ 1994-2016	$(\bar{p}_D - \bar{p}_G)$ 1994-2016	$(\bar{p}_D - \bar{p}_G)$ 1994-2000	$(\bar{p}_D - \bar{p}_G)$ 2010-2016
United States	2.69	2.54	0.15	0.02	0.32
East Coast	2.71	2.52	0.19	0.04	0.38
Gulf Coast	2.61	2.41	0.21	0.01	0.44
Midwest	2.65	2.50	0.16	0.02	0.34
Rocky Mountain	2.73	2.54	0.19	0.02	0.39
West Coast	2.86	2.79	0.06	-0.01	0.16

Table A.4: Dickey-Fuller Test Results: United States

Specification	**(1)**	**(2)**	**(3)**
ρ	0.8938 (0.0272)	0.8413 (0.0325)	
β			-0.2173 (0.0375)
ζ_1			0.2667 (0.0601)
ζ_2			0.0300 (0.0612)
ζ_3			0.0528 (0.0612)
δ		0.0003 (0.0001)	0.0004 (0.0001)
α	0.0162 (0.0076)	-0.0138 (0.0129)	-0.0192 (0.0130)
Approximate P-Value	0.0020	0.0003	0.0000

Table A.5: Dickey-Fuller Test Results: Rocky Mountain Region

Specification	**(1)**	**(2)**	**(3)**
ρ	0.9046 (0.0258)	0.8547 (0.0351)	
β			-0.2114 (0.0351)
ζ_1			0.4093 (0.0588)
ζ_2			-0.0272 (0.0622)
ζ_3			0.0616 (0.0611)
δ		0.0003 (0.0001)	0.0004 (0.0001)
α	0.0193 (0.0084)	-0.0096 (0.0136)	-0.0149 (0.0130)
Approximate P-Value	0.0041	0.0010	0.0000

Table A.6: Dickey-Fuller Test Results: West Coast

Specification	**(1)**	**(2)**	**(3)**
ρ	0.8414 (0.0327)	0.8231 (0.0342)	
β			-0.2314 (0.0401)
ζ_1			0.1849 (0.0610)
ζ_2			0.0562 (0.0611)
ζ_3			0.0571 (0.0612)
δ		0.0002 (0.0001)	0.0002 (0.0001)
α	0.0097 (0.0078)	0.0130 (0.0152)	-0.0167 (0.0155)
Approximate P-Value	0.0000	0.0001	0.0000

Figure A.1: Count of Retailers Reporting Fuel Prices

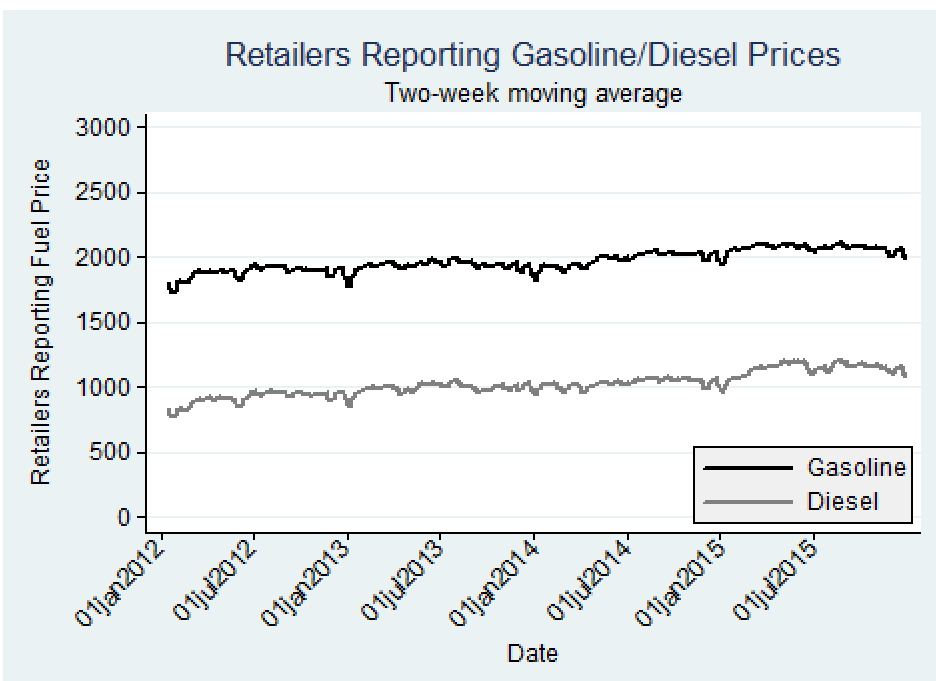


Figure A.2: Count of Daily Diesel Price Reports, by Filling Station

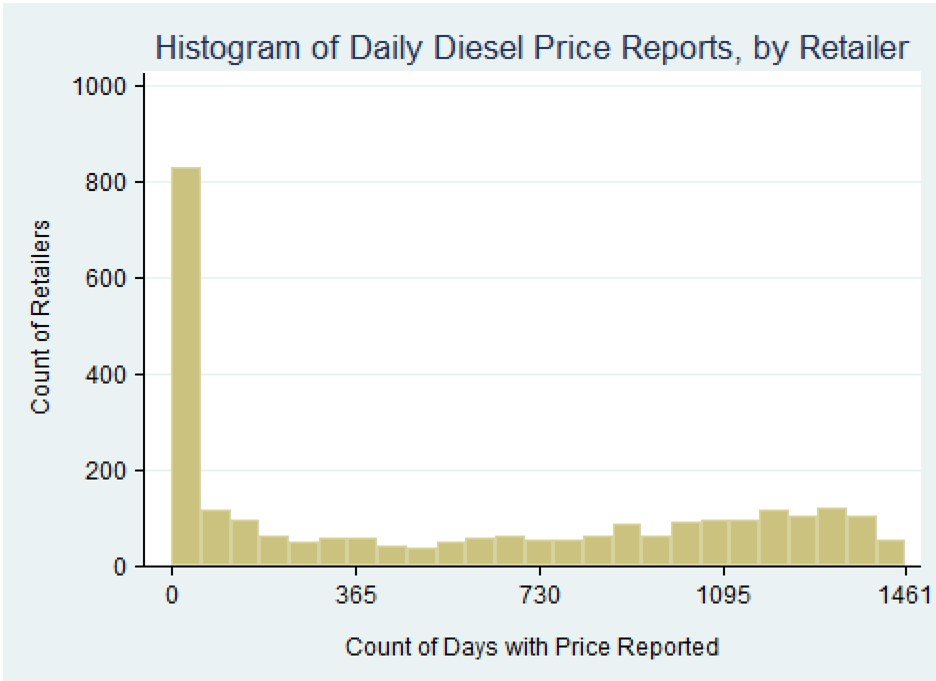


Figure A.3: Count of Daily Gasoline Price Reports, by Filling Station

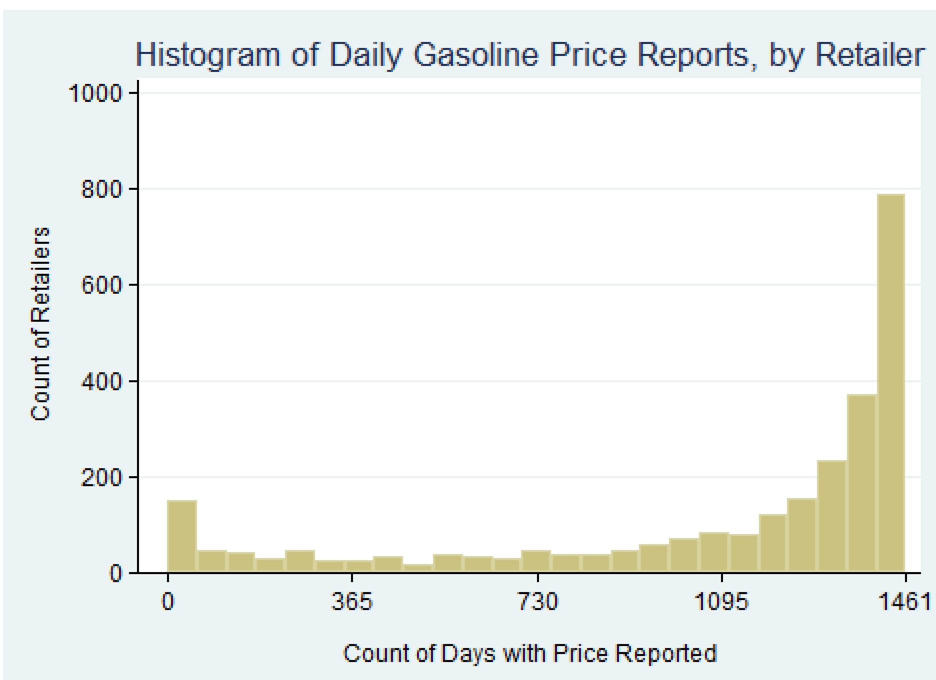


Figure A.4: U.S. Diesel Premium, 1994-2016

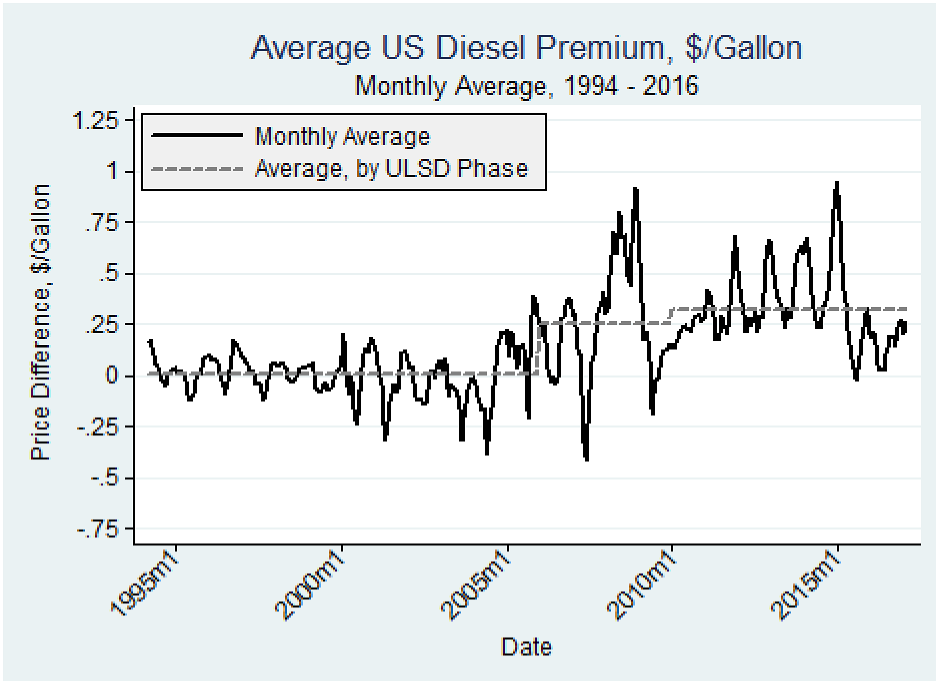


Figure A.5: Rocky Mountain Region Diesel Premium, 1994-2016

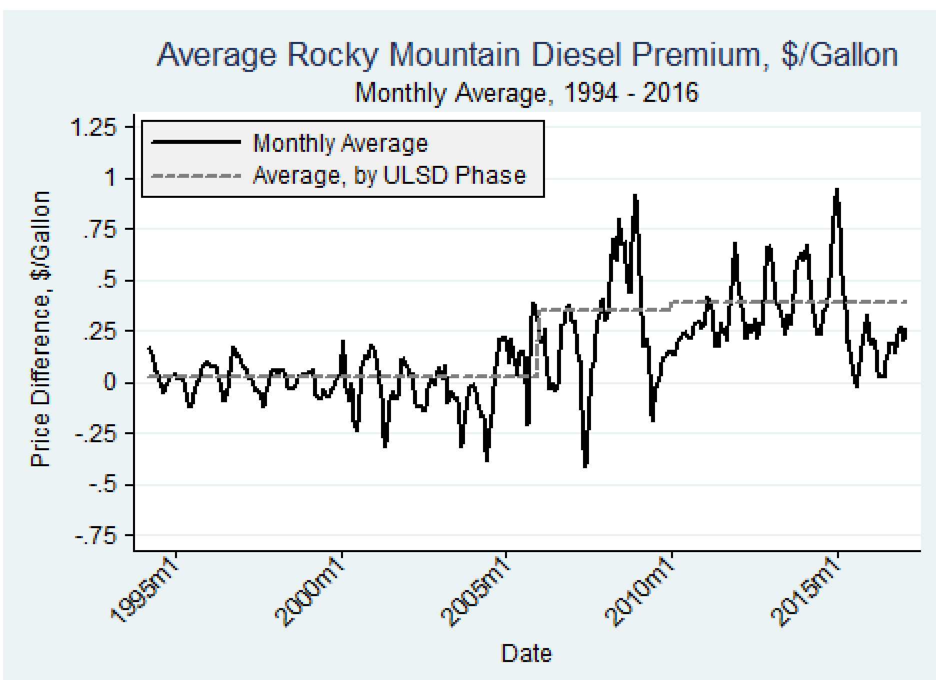
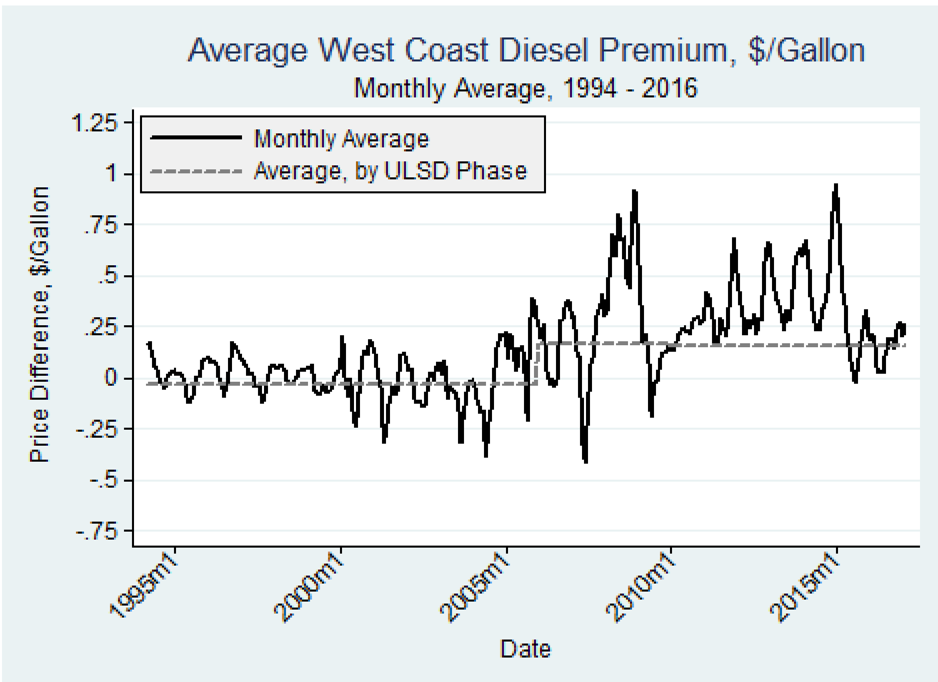


Figure A.6: West Coast Region Diesel Premium, 1994-2016



APPENDIX B

APPENDIX: CHAPTER 4

Table B.1: OLS Regression of Log Yearly Mileage: Business Registration Sample

Variable	**(1)**	**(2)**	**(3)**	**(4)**
Log Operating Cost	0.104 (0.094)	-0.022 (0.101)	-0.516 (0.145)	-0.656 (0.156)
Rain, ft./year	-0.004 (0.001)	0.002 (0.015)	-0.003 (0.002)	0.003 (0.016)
Snow, ft./year	0.000 (0.003)	-0.022 (0.005)	-0.001 (0.003)	-0.022 (0.005)
Rural Population Percentage	0.274 (0.065)	0.164 (0.086)	0.214 (0.065)	0.132 (0.089)
Unemployment Rate	0.001 (0.007)	-0.006 (0.008)	0.000 (0.007)	-0.005 (0.008)
Median Income, $10K	-0.016 (0.010)	-0.014 (0.010)	-0.012 (0.011)	-0.009 (0.012)
Extended Cab			0.159 (0.117)	0.126 (0.116)
Crew Cab			0.164 (0.125)	0.128 (0.123)
Four Wheel Drive			0.076 (0.089)	0.095 (0.098)
Dual Rear Wheels			0.029 (0.150)	0.066 (0.155)
Torque, 100 lb. ft.			0.000 (0.000)	0.000 (0.000)
Work Truck Trim			-0.098 (0.074)	-0.110 (0.071)
Length, ft.			-0.114 (0.055)	-0.091 (0.050)
Width, ft.			0.007 (0.091)	0.024 (0.087)
Height, ft.			0.715 (0.165)	0.652 (0.150)
Constant	3.113 (0.192)	3.469 (0.246)	-0.966 (0.767)	-0.736 (0.156)
County FEs	No	Yes	No	Yes
R^2	0.028	0.117	0.074	0.157

Table B.2: Logistic Regression of Fuel Choice in Heavy-Duty Trucks: Business Registrations

Variable	**(1)**	**(2)**
$p^g_{C,t}$	0.290 (0.088)	2.872 (0.529)
$p^d_{C,t} - p^g_{C,t}$	0.365 (0.143)	
$\bar{p}^d_{C,t}$		-2.574 (0.536)
$\tilde{p}^d_{C,t}$		0.551 (0.147)
Rain, ft./year	-0.003 (0.002)	-0.011 (0.002)
Snow, ft./year	0.002 (0.004)	0.000 (0.004)
Rural Population Percentage	0.881 (0.092)	0.795 (0.093)
Unemployment Rate	-0.019 (0.008)	-0.022 (0.009)
Median Income, $10K	0.038 (0.018)	0.040 (0.018)
Years Elapsed from 01/01/2012	0.129 (0.044)	0.134 (0.044)
Constant	-8.064 (2.655)	-7.103 (2.671)
Marginal Effects	**(1)**	**(2)**
$p^g_{C,t}$	0.072 (0.022)	0.713 (0.131)
$p^d_{C,t} - p^g_{C,t}$	0.091 (0.035)	
$\bar{p}^d_{C,t}$		-0.639 (0.133)
$\tilde{p}^d_{C,t}$		0.137 (0.036)
Rain, ft./year	-0.001 (0.000)	-0.003 (0.001)
Snow, ft./year	0.000 (0.001)	0.000 (0.001)
Rural Population Percentage	0.219 (0.023)	0.197 (0.023)
Unemployment Rate	-0.005 (0.002)	-0.005 (0.002)
Median Income, $10K	0.010 (0.004)	0.010 (0.004)
Years Elapsed from 01/01/2012	0.032 (0.011)	0.033 (0.011)
(Negative) Pseudo-LL	4856	4840

APPENDIX C

APPENDIX: CHAPTER 5

C.1 Theoretical Model

I now show how I derive (5.4) starting with the first-period expected utility expression given in (5.3). Recall that the solution to the individual's second-period VMT decision is given by $VMT^*_{ij} = \eta_i \left[1 - \frac{\alpha_i \tilde{p}_{ij}}{\gamma_j}\right]$. First, recall that the expected second-period utility can be written as

$$E_i[u_2] = E_i\left[\gamma_j\left(VMT_{ij} - \frac{1}{2\eta_i}VMT^2_{ij}\right) - \alpha_i \tilde{p}_{ij} VMT_{ij}\right]$$

First, consider the terms within the expectations operator, which I will now simplify after substituting in my derived expression for VMT^*_{ij}. Concerning the term on the right-hand side of the previous equation, this can be expanded as

$$\alpha_i \tilde{p}_{ij} VMT_{ij} = \eta_i \alpha_i \tilde{p}_{ij} - \frac{\eta_i \alpha_i \tilde{p}_{ij}}{\gamma_j} \tag{C.1}$$

Squaring the expression for VMT^*_{ij} gives the following:

$$VMT^{*2}_{ij} = \eta_i^2 - \frac{2\eta_i^2 \alpha_i \tilde{p}_{ij}}{\gamma_j} + \frac{\eta_i^2 \alpha_i^2 \tilde{p}^2_{ij}}{\gamma_j^2} \tag{C.2}$$

The second-period utility associated with driving the vehicle (without considering the fuel costs) is given by $\gamma_j\left(VMT^*_{ij} - \frac{1}{2\eta_i}VMT^{2*}_{ij}\right)$. Plugging (C.2) and VMT^*_{ij} to this expression yields

$$\gamma_j\left(VMT^*_{ij} - \frac{1}{2\eta_i}VMT^{2*}_{ij}\right) = \left(\frac{\gamma_j \eta_i}{2} - \frac{\gamma_j \eta_i \alpha_i^2 \tilde{p}^2_{ij}}{2\gamma_j^2}\right) \tag{C.3}$$

Subtracting (C.1) from (C.3) and recalling that $\tilde{p}_{ij} \equiv \frac{p_i}{MPG_j}$, the term inside the expectation operator is given by

$$\eta_i\left[\frac{\gamma_j}{2} + \frac{\alpha_i^2 p_i^2}{2\gamma_j MPG_j^2} - \frac{\alpha_i p_i}{MPG_j}\right] \tag{C.4}$$

Taking expectations, noting (i) the assumption that fuel prices evolve exogenously with respect to the second-period shocks ϵ_i^u and rewriting $E[p_{ij}^2] = Var(p_i) + E[p_i]^2$ leads to the following expression:

$$E_i[u_2] = E_i[\eta_i] \left[\frac{\gamma_j}{2} + \frac{\alpha_i^2 [Var(p_i) + E[p_i]^2]}{2\gamma_j MPG_j^2} - \frac{\alpha E[p_i]}{MPG_j} \right] \tag{C.5}$$

Given the assumption that $\epsilon_i^u \sim N(0,1)$ is independent of ϵ_i^k and Z_i, the expected value of η_i can be written as $E_i[\eta_i] = \exp\left\{ \beta_Z Z_i + \sigma_k \epsilon_i^k + \mu_k + \frac{\sigma_u^2}{2} \right\}$. Substituting this expression into $E_i[\eta_i]$ in (C.5) gives the second-period expected utility expression. Plugging this into (5.3) result in (5.4).

APPENDIX D

APPENDIX: CHAPTER 6

In addition to the structural model presented in §5, I estimate a linear-in-parameters parameters multinomial logit model of vehicle choice. This specification notably ignores the importance of selection on anticipated usage in vehicle choice. However, it is instructive as a means to illustrate the importance of allowing consumers to respond differently to separate sources of fuel price variation in a common discrete choice model. I specify

$$E[u_{ij}] = \beta_X X_j - \alpha_p MSRP_j - \alpha_f E[OC_{ij}] + \epsilon_{ij}$$

Notation is consistent with the two-period model presented in §5 wherever possible. Likewise, I assume that the idiosyncratic match value ϵ_{ij} follows an extreme value distribution which is independent and identically distributed across individuals and vehicles. Expected yearly operating costs are denoted by $E[OC_{ij}]$. This specification does not model second-period driving behavior or allow for selection on anticipated usage, so expected operating costs are defined using the mean usage level of pickup trucks observed in my sample ($\overline{VMT} = 13,072$ miles per year). Using this value, I specify $E[OC_{ij}] \equiv \frac{\overline{VMT} * E[p_i]}{MPG_j}$. I follow the same construction of price beliefs as §5, where the 'no-change' forecast specifies that the expected price of both gasoline and diesel at any future date t is given by the current price of each respective fuel $f \in \{d, g\} : E[p_t^f] = p_0^f$. Under the 'decomposed' forecast, I specify $E[p_t^d] = p_0^g + \tau_C + \phi \tilde{p}_0^d$ and retain the no-change forecast for gasoline prices. The interpretation of ϕ is identical to its interpretation in §5, as it represents the weight attached to variation in $\tilde{p}_0^d$ relative to all other sources of price variation. For the sake of notational simplicity, I replace the t subscript with i and omit this subscript when possible, taking i to denote the choice situation of i, including the date of purchase.

Comparing across the two columns in Table D.1 gives results from each specification of price beliefs. I find that consumers do not exhibit a statistically significant response to variation in $\tilde{p}_0^d$, and the point estimate of ϕ from the linear-in-parameters logit specification is qualitatively similar to the estimate from the two-period structural model. Likewise, the difference in α_f is similar across the specifications of price beliefs to §5, with decomposed beliefs resulting in an estimate roughly 15.75% higher than no-change beliefs. Unlike the two-period model from §5, this approach is limited by its inability to estimate the distribution of consumers' driving behavior after making a vehicle purchase. As shown in Figure 6.6, the two-period model estimates a usage distribution which accounts for selection on anticipated usage and fits the empirical distribution closely. The second-period usage model which is used to fit this distribution is used to simulate counterfactual VMT and fuel consumption. In absence of a model of second-period driving behavior, the linear-in-parameters logit model is limited in this regard.

Table D.1: Multinomial Logit Estimates

Parameters	**No-Change Beliefs**		**Decomposed Beliefs**	
Price Parameters:				
α_f	0.889	(0.034)	1.029	(0.037)
α_p	0.112	(0.003)	0.113	(0.003)
ϕ			0.122	(0.083)
Vehicle Characteristics:				
Extended Cab	1.568	(0.045)	1.585	(0.045)
Crew Cab	2.616	(0.048)	2.641	(0.048)
Four Wheel Drive	2.200	(0.032)	2.227	(0.032)
Dual Rear Wheels	-1.319	(0.072)	-1.331	(0.073)
Length (ft.)	-0.146	(0.019)	-0.156	(0.019)
Width (ft.)	0.353	(0.049)	0.383	(0.049)
Height (ft.)	1.107	(0.069)	1.144	(0.069)
Torque (100 lb. ft.)	0.807	(0.016)	0.836	(0.016)
Diesel Engine	-1.293	(0.064)	-1.422	(0.067)
Sport-Utility Truck	-0.369	(0.063)	-0.348	(0.063)
Full-Size Truck	-0.429	(0.049)	-0.408	(0.049)
Heavy-Duty: $\frac{3}{4}$ Ton Truck	-1.480	(0.076)	-1.371	(0.077)
Heavy-Duty: One Ton Truck	-1.353	(0.085)	-1.184	(0.087)
Log Likelihood	102,739		102,691	
Implicit Discount Rate	10.39%		7.55%	

Table D.2: Estimates from Sample Limited to N_1

Parameters	No-Change Beliefs		Decomposed Beliefs	
Price Parameters:				
α_f	0.986	(0.065)	1.106	(0.070)
α_p	0.139	(0.005)	0.139	(0.005)
ϕ			0.121	(0.148)
Usage Distribution:				
μ_k	2.367	(0.028)	2.379	(0.028)
Income (x $10K)	0.014	(0.004)	0.014	(0.004)
Rural Share	0.221	(0.023)	0.220	(0.023)
σ_k	0.305	(0.011)	0.304	(0.011)
σ_u	0.550	(0.007)	0.551	(0.007)
Vehicle Characteristics:				
Extended Cab	0.260	(0.014)	0.260	(0.014)
Crew Cab	0.395	(0.014)	0.394	(0.014)
Four Wheel Drive	0.373	(0.011)	0.373	(0.011)
Dual Rear Wheels	-0.247	(0.022)	-0.165	(0.022)
Work Truck Trim	-0.166	(0.006)	-0.165	(0.006)
Length (ft.)	0.003	(0.005)	0.002	(0.005)
Width (ft.)	0.072	(0.011)	0.076	(0.011)
Height (ft.)	0.111	(0.017)	0.113	(0.017)
Torque (100 lb. ft.)	0.130	(0.005)	0.133	(0.005)
Diesel Engine	-0.230	(0.021)	-0.253	(0.022)
Sport-Utility Truck	-0.005	(0.015)	-0.004	(0.015)
Full-Size Truck	-0.051	(0.013)	-0.050	(0.013)
Heavy-Duty: $\frac{3}{4}$ Ton Truck	-0.225	(0.022)	-0.212	(0.022)
Heavy-Duty: One Ton Truck	-0.167	(0.025)	-0.147	(0.025)
Log Likelihood	37,024.8		37,011.4	
Implicit Discount Rate	13.12%		10.49%	

Figure D.1: Simulated VMT Distribution, by Subsample

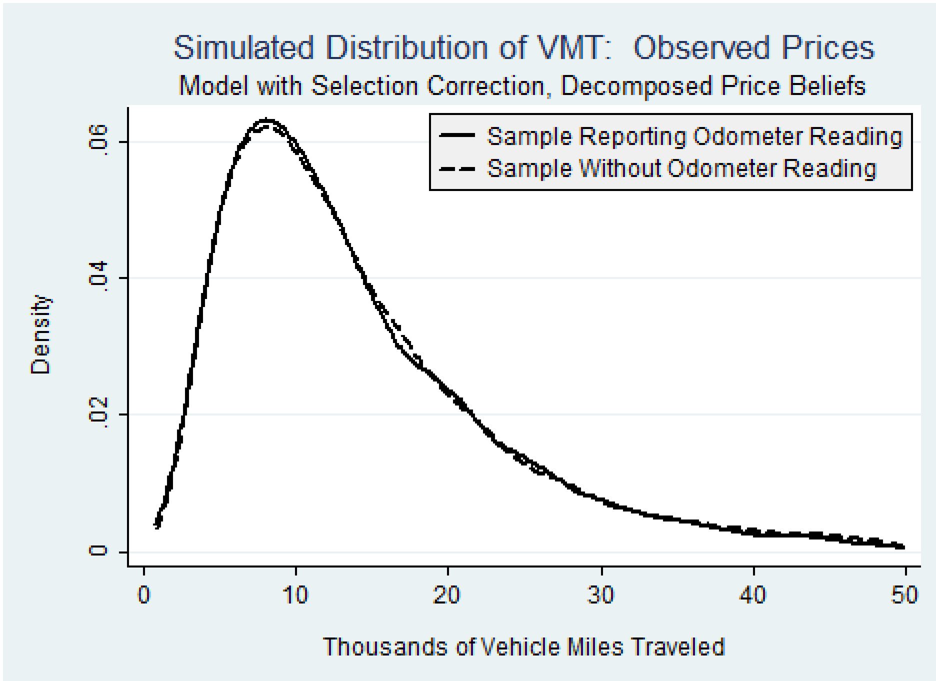


Figure D.2: Counterfactual VMT Distribution, by Subsample

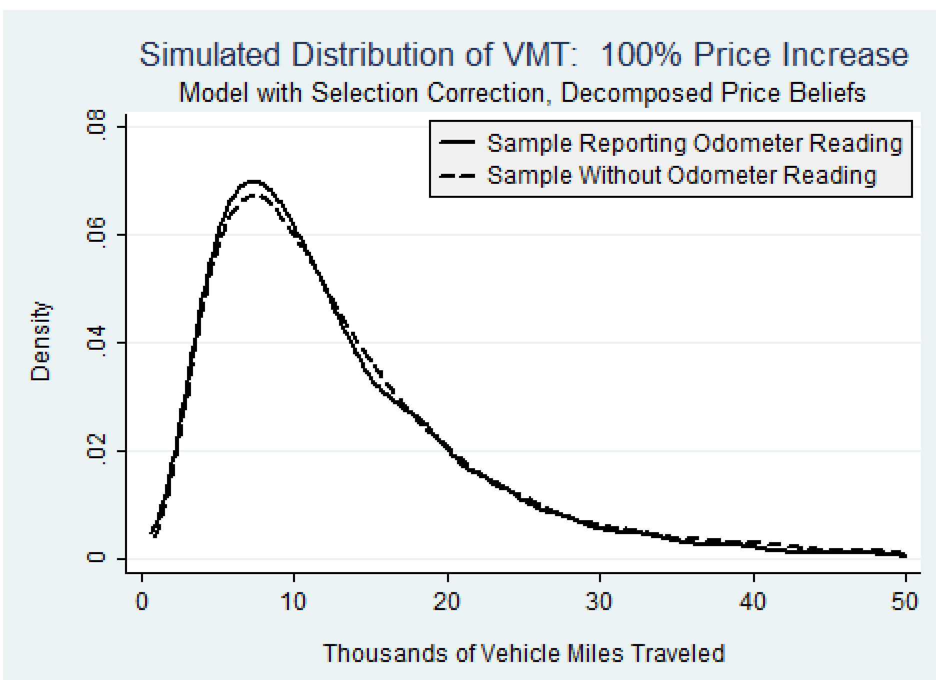


Figure D.3: Comparison of Results Across Specifications: Fleet Fuel Economy

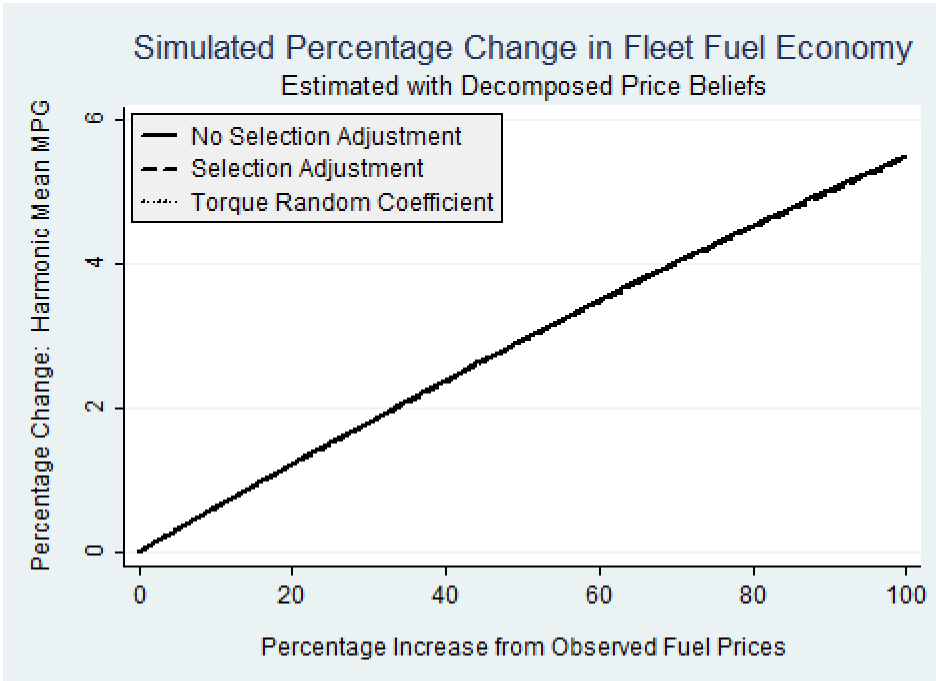


Figure D.4: Comparison of Results Across Specifications: Vehicle Miles Traveled

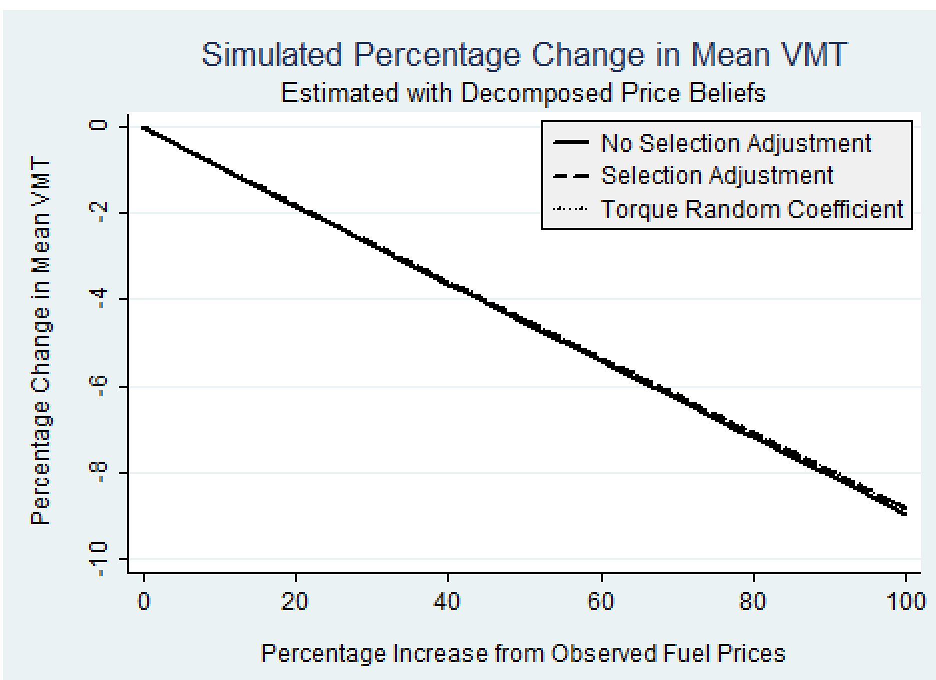


Figure D.5: Comparison of Results Across Specifications: Fuel Consumption

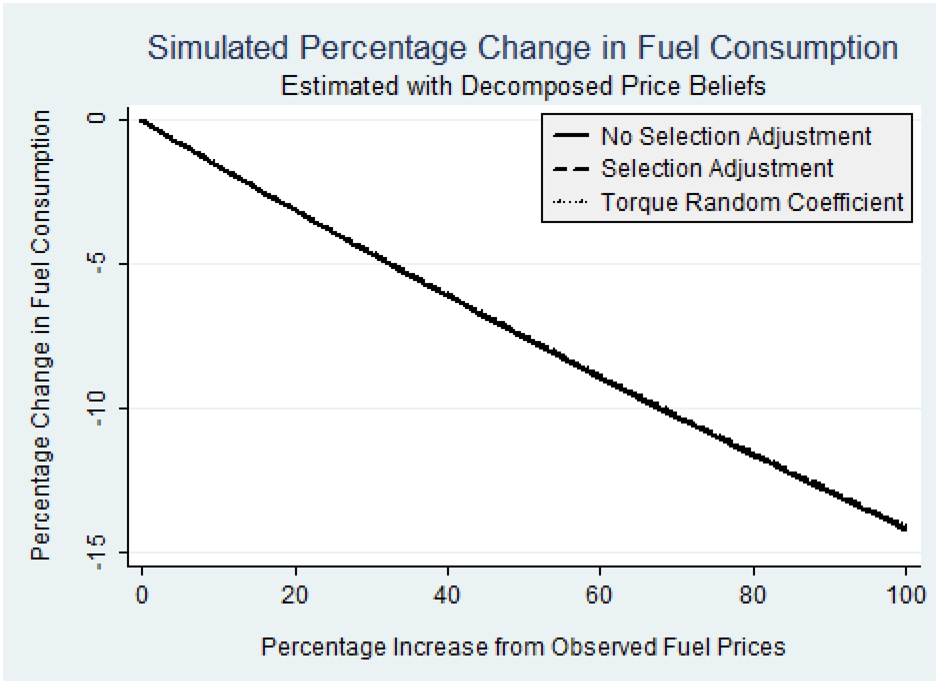


Figure D.6: Counterfactual Fleet Fuel Economy (A)

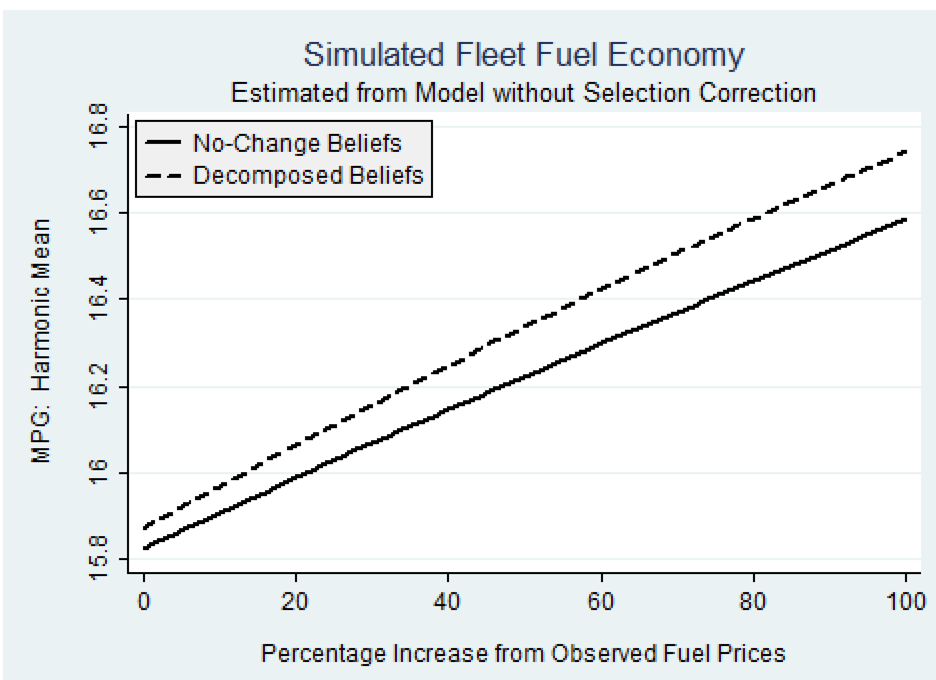


Figure D.7: Counterfactual Change in Fleet Fuel Economy (A)

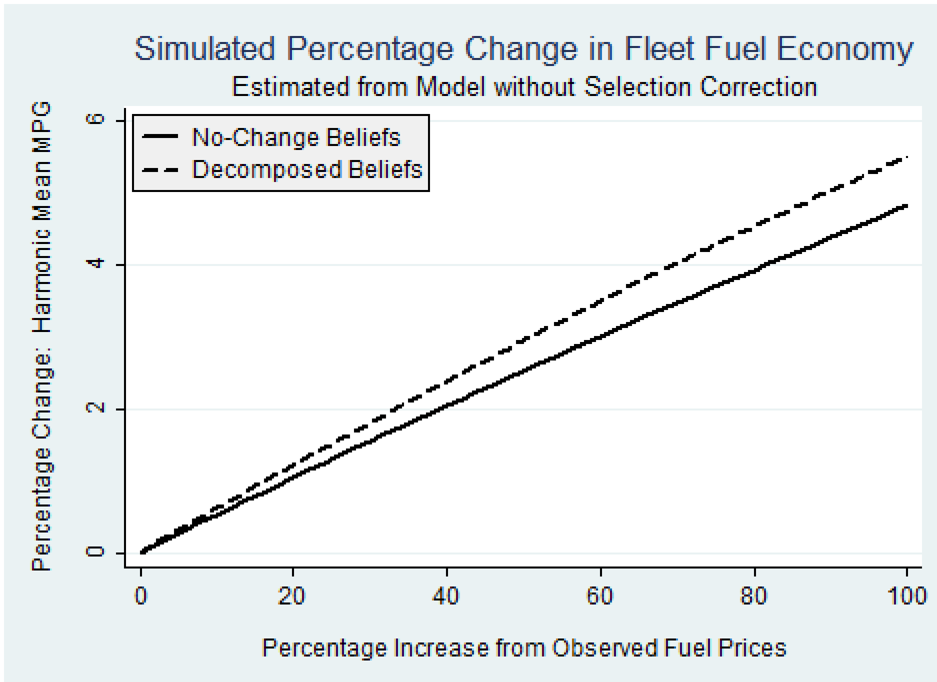


Figure D.8: Counterfactual Change in Mean VMT (A)

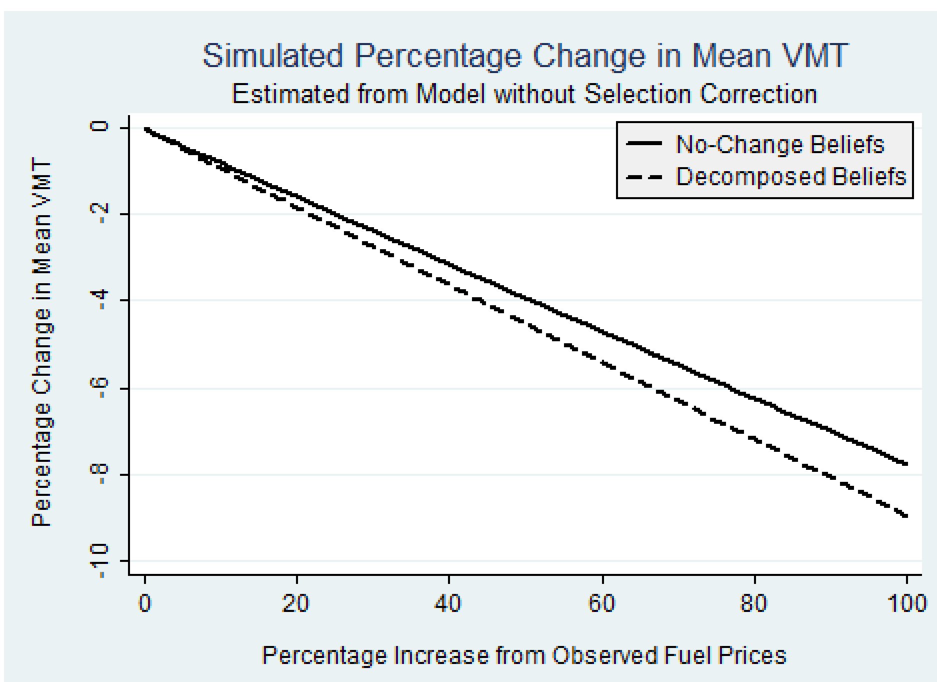


Figure D.9: Counterfactual Average Fuel Consumption (A)

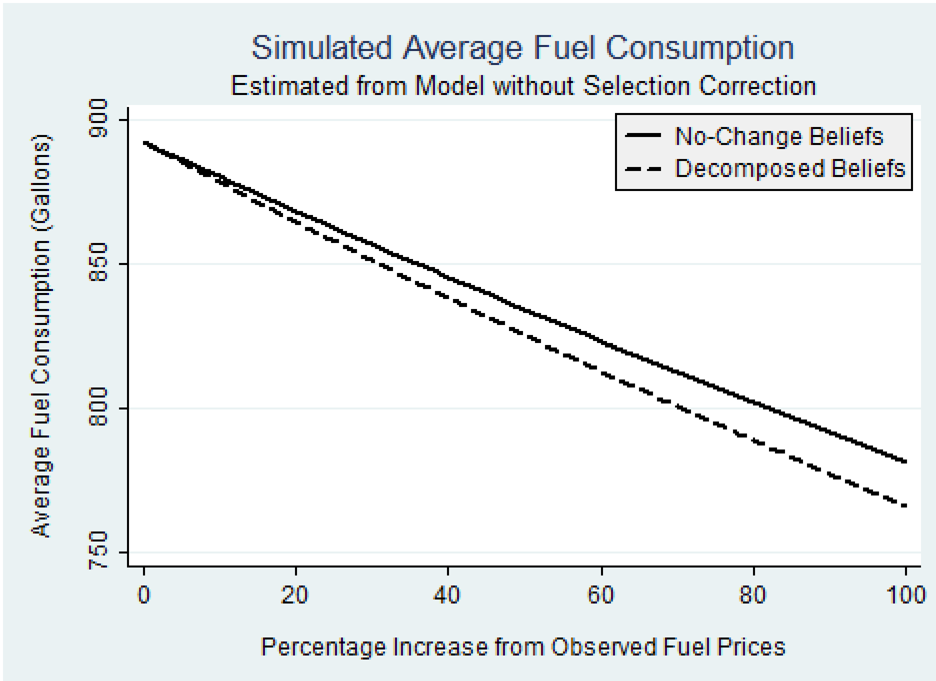


Figure D.10: Counterfactual Change in Fuel Consumption (A)

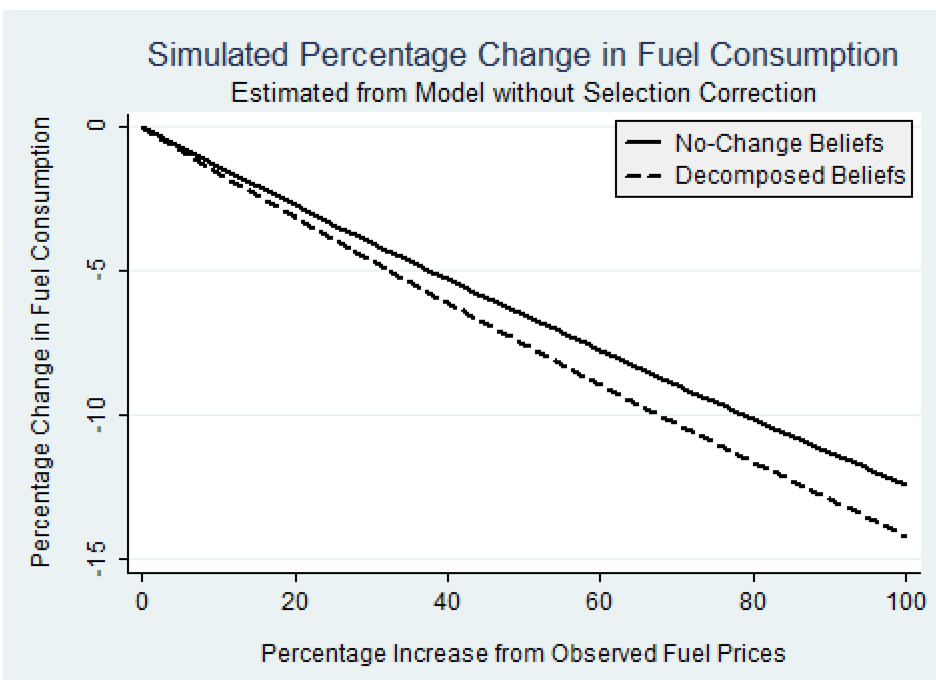


Figure D.11: Counterfactual Fleet Fuel Economy (B)

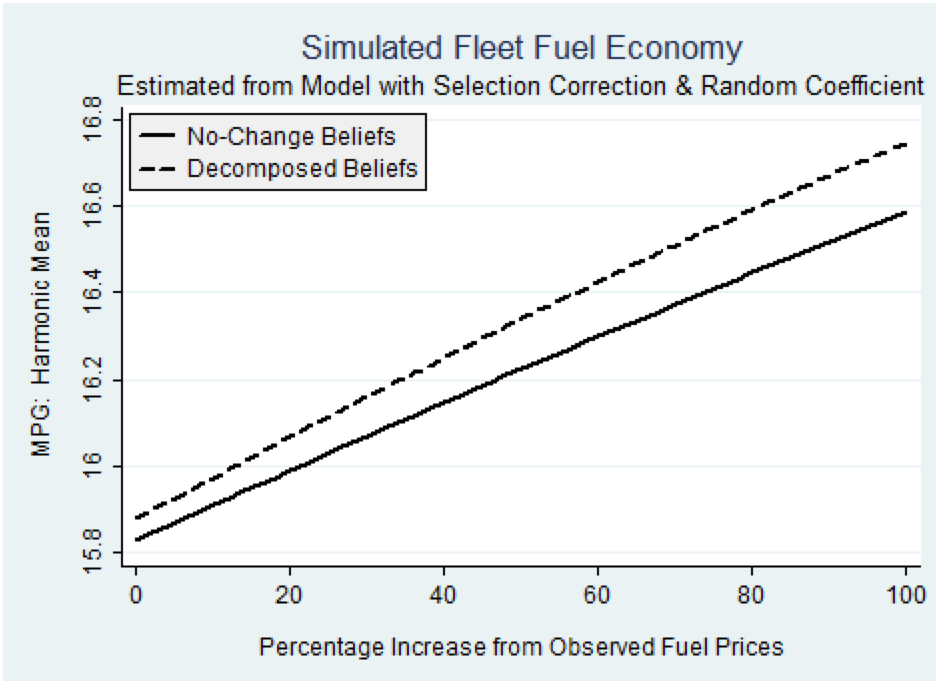


Figure D.12: Counterfactual Change in Fleet Fuel Economy (B)

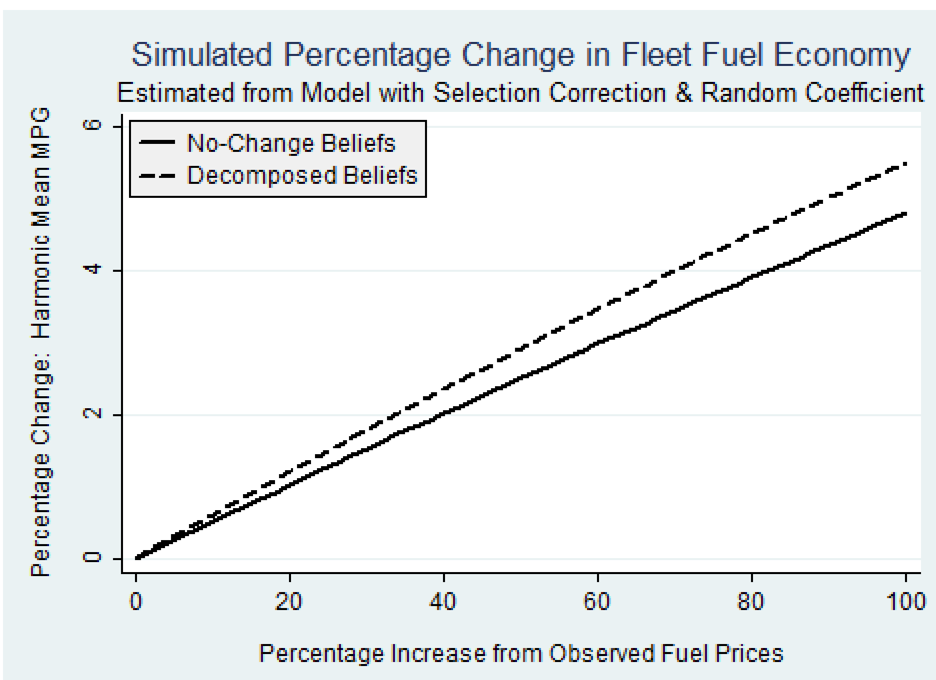


Figure D.13: Counterfactual Change in Mean VMT (B)

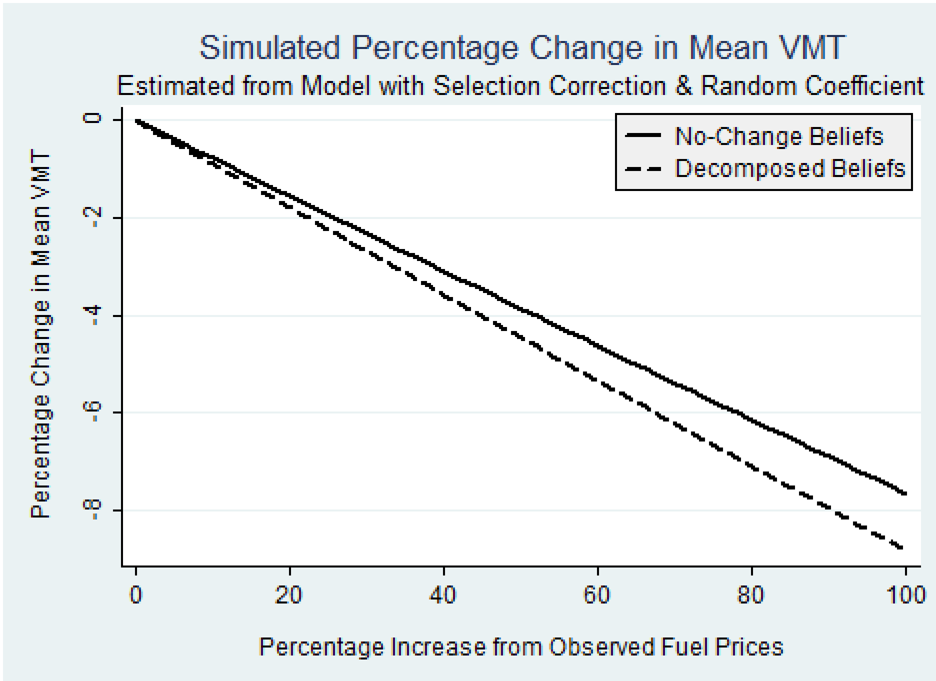


Figure D.14: Counterfactual Average Fuel Consumption (B)

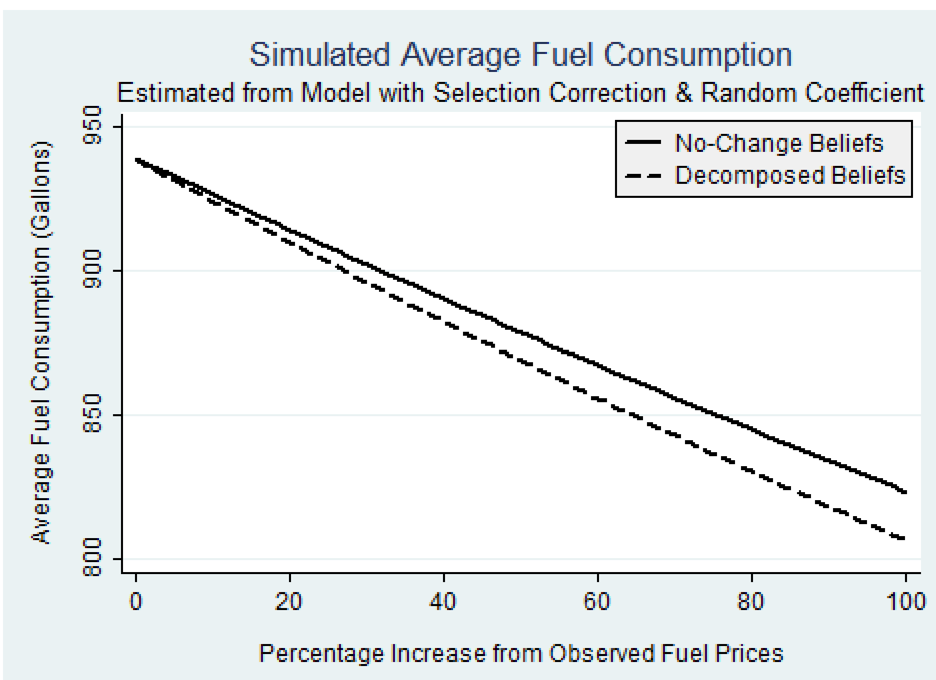
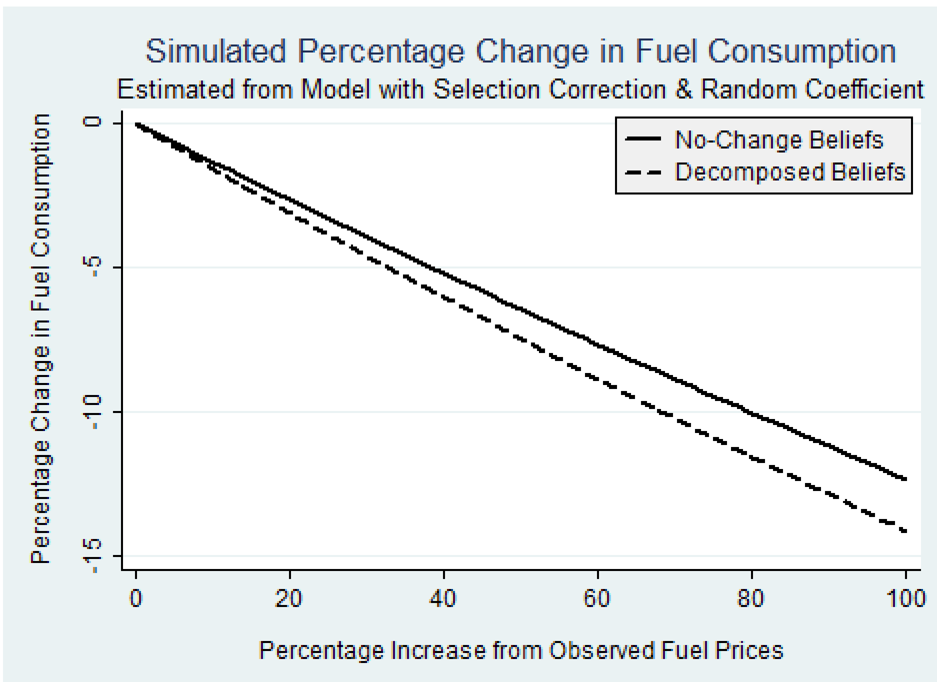


Figure D.15: Counterfactual Change in Fuel Consumption (B)



BIBLIOGRAPHY

Allcott, H. and Wozny, N. (2014). Gasoline Prices, Fuel Economy, and the Energy Paradox. *Review of Economics and Statistics* 96(5): 779-795.

Alquest, R., Kilian, Lutz, and Vigfusson, R. J. (2013). Forecasting the Price of Oil. *Handbook of Economic Forecasting*: 427-507.

Anderson, S. (2012). The Demand for Ethanol as a Gasoline Substitute. *Journal of Environmental Economics and Management* 63: 151-168.

Anderson, S., Kellogg, R., and Sallee, J. M. (2013). What do Consumers Believe about Future Gasoline Prices? *Journal of Environmental Economics and Management* 66(3): 151-168.

Anderson, S. and Sallee, J. M. (2011). Using Loopholes to Reveal the Marginal Cost of Regulation: The Case of Fuel-Economy Standards. *American Economic Review* 101(3): 1375-1409.

Austin, D. and Dinan, T. (2005). Clearing the Air: The Costs and Consequences of Higher CAFE Standards and Increased Gasoline Taxes. *Journal of Environmental Economics and Management*, 50: 562-582.

Bento, A. M., Li, S., and Roth, K. (2012). Is there an Energy Paradox in Fuel Economy? A Note on the Role of Consumer Heterogeneity and Sorting Bias. *Economic Letters* 115: 44-48.

Beresteanu, A. and Li, S. (2011). Gasoline Prices, Government Support, and the Demand for Hybrid Vehicles in the United States. *International Economic Review* 52: 161-182.

Bishop, K. C. and Murphy, A. D. (2016). Valuing Time-Varying Attributes using the Hedonic Model: When is a Dynamic Approach Necessary? *Working Paper*.

Brownstone, D., Bunch, D. S., Golob, T. F., and Ren, W. (1996). A Transactions Choice Model for Forecasting Demand for Alternative-Fuel Vehicles. *Research in Transportation Economics* 4: 87-129.

Brownstone, D., Bunch, D. S., and Train, K. (2000). Joint Mixed Logit Models of Stated and Revealed Preferences. *Transportation Research Part B* 34: 315-338.

Brownstone, D. and Train, K. (2000). Forecasting New Product Penetration with Flexible Substitution Patterns. *Journal of Econometrics* 89: 109-129.

Busse, M. R., Knittel, C. R., and Zettelmeyer, F. (2013). Are Consumers Myopic? Evidence from New and Used Car Purchases. *American Economic Review* 103(1): 220-256.

Corts, K. (2010). Building out Alternative Fuel Retail Infrastructure: Government Fleet Spillovers in E85. *Journal of Environmental Economics and Management* 59: 219-234.

Einav, L., Finkelstein, A., Ryan, S. P., Schrimpf, P., and Cullen, M. R. (2013). Selection on Moral Hazard in Health Insurance. *American Economic Review* 103(1): 178-219.

Espey, M. (1998). Gasoline Demand Revisited: An International Meta-Analysis of Elasticities. *Energy Economics*, 20: 273-295.

Gillingham, K. (2013). Selection on Anticipated Driving and the Consumer Reponse to Changing Gasoline Prices. *Working Paper*.

Gillingham, K. (2014). Identifying the Elasticity of Driving: Evidence from a Gasoline Price Shock in California. *Regional Science & Urban Economics*, 47(4): 13-24.

Gramlich, J. (2010). Gas Prices, Fuel Efficiency, and Endogenous Product Choice in the U.S. Automobile Industry. *Working Paper*.

Green, D. L. and Lin, Z. (2011). Predicting Individual Fuel Economy. *Society of Automotive Engineers International Journal of Fuels and Lubricants* 4(1): 84-95.

Griffin, C. E. (1928). Recent Literature on the Automotive Industry. *Quarterly Journal of Economics*, 43(1): 142-153.

Hausman, J. A. (1979). Individual Discount Rates and the Purchase of Energy-Using Durables. *The Bell Journal of Economics* 10(1): 33-54.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica* 47(1): 153-161.

Helfand, G. and Wolverton, A. (2009). Evaluating the Consumer Reponse to Fuel Economy: A Review of the Literature. *National Center for Environmental Economics Working Paper*.

Holland, S. P., Mansur, E. T., Muller, N. Z., and Yates, A. J. (2016). Are There Environmental Benefits from Driving Electric Vehicles? The Importance of Local Factors. *American Economic Review* 106(12): 3700-3729.

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

Kahn, J. A. (1986). Gasoline Prices and the Used Automobile Market: A Rational Expectations Asset Price Approach. *Quarterly Journal of Economics*, 101(2): 323-339.

Klemick, H., Kopits, E., Sargent, K., and Wolverton, A. (2014). Heavy-Duty Trucking and the Energy Efficiency Paradox. *National Center for Environmental Economics Working Paper*.

Klier, T. (2009). From Tail Fins to Hybrids: How Detroit Lost its Dominance of the U.S. Auto Market. *Economic Perspectives* 33(2), 2-17.

Klier, T. and Linn, J. (2010). The Price of Gasoline and New Vehicle Fuel Economy: Evidence from Monthly Sales Data. *American Economic Journal: Economic Policy* 2: 134-153.

Knittel, C. (2012). Automobiles on Steroids: Product Attribute Trade-Offs and Technological Progress in the Automobile Sector. *American Economic Review* 101(5): 3368-3399.

Langer, A. and Miller, N. H. (2013). Automakers' Short-Run Responses to Changing Gasoline Prices. *Review of Economics and Statistics* 95: 1198-1211.

Larrick, R. P. and Soll, J. B. (2008). The MPG Illusion. *Science* 320: 1593-1594.

Leard, B., Linn, J., McConnell, V., and Raich, W. (2015). Fuel Costs, Economic Ativity, an the Rebound Effect for Heavy-Duty Trucks. *Resources for the Future Working Paper*.

Li, S., von Haefen, R., and Timmins, C. (2009). How do Gasoline Prices affect Fleet Fuel Economy? *American Economic Journal: Economic Policy* 1(2): 113-137.

Li, S., Linn, J., and Muehlegger, E. (2014). Gasoline Taxes and Consumer Behavior. *American Economic Journal: Economic Policy* 6(4): 302-342.

Li, S., Tong, L., Xing, J., and Zhou, Y. (2015). The Market for Electric Vehicles: Indirect Network Effects and Policy Design. *Working Paper*.

Linn, J. (2013). The Rebound Effect for Passenger Vehicles. *Resources for the Future Working Paper*.

Linn, J. (2014). Explaining the Adoption of Diesel Passenger Cars in Europe. *Resources for the Future Working Paper*.

Luechinger, S. and Roth, F. (2016). Effects of a Mileage Tax for Trucks. *Journal of Urban Economics* 92: 1-15.

Mackenzie, D., Bedsworth, L., and Friedman, D. (2005). Fuel Economy Fraud: Closing Loopholes that Increase U.S. Oil Dependence. *Union of Concerned Scientists Report*.

National Highway Traffic Safety Administration. (2006). Vehicle Survivability and Travel Mileage Schedules. *NHTSA Technical Report No. 809 952*.

Parry, I. (2008). How Should Heavy-Duty Trucks be Taxed? *Journal of Urban Economics* 63(2): 651-668.

Petrin, A. (2002). Quantifying the Benefits of New Products: The Case of the Minivan. *Journal of Political Economy* 110(4): 705-729.

Sallee, J. M., West, S., and Fan, W. (2016). Do Consumers Recognize the Value of Fuel Economy? Evidence from Used Car Prices and Gasoline Price Fluctuations. *Journal of Public Economics* 135: 705-729.

Shriver, S. (2015). Network Effects in Alternative Fuel Adoption: Empirical Analysis of the Market for Ethanol. *Marketing Science* 34(1): 78-97.

Small, K. and Van Dender, K. (2007). Fuel Efficiency and Motor Vehicle Travel: The Declining Rebound Effect. *Energy Journal*, 28(1): 25.51.

Train, K. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Verboven, F. (2002). Quality-Based Price Discrimination and Tax Incidence: Evidence from Gasoline and Diesel Cars. *RAND Journal of Economics* 33: 275-297.

Yeh, S. (2007). An Empirical Analysis on the Adoption of Alternative Fuel Vehicles: The Case of Natural Gas Vehicles. *Energy Policy* 35: 5865-5875.