Title
Essays in Transportation Economics and Industrial Organization

Author
Makuch, Kim

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Essays in Transportation Economics and Industrial Organization

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Economics

by

Kim Makuch

Dissertation Committee:
Assistant Professor Kevin Roth, Chair
Professor Linda R. Cohen
Professor Jan K. Brueckner

2016

DEDICATION

To my parents.

TABLE OF CONTENTS

Page

LIST OF FIGURES **v**

LIST OF TABLES **vi**

ACKNOWLEDGMENTS **vii**

CURRICULUM VITAE **viii**

ABSTRACT OF THE DISSERTATION **ix**

1 Pricing Patterns in Competition between Two Dimensional Products **1**
1.1 Introduction 1
1.2 A Two Dimensional Product Differentiation Model 4
1.3 Empirical Framework 6
1.3.1 Setting 6
1.3.2 Data 8
1.3.3 Model Specification 9
1.4 Results 11
1.5 Discussion 14
1.6 Conclusion 16

2 The External Congestion Costs of Light Trucks **17**
2.1 Introduction 17
2.2 Theory 20
2.2.1 Model Assumptions 21
2.2.2 Individual Optimization 22
2.2.3 Effect of Vehicle Size on Following Distance Choice 23
2.3 Data 25
2.3.1 Study Area and Available Data 26
2.3.2 Unit of Observation 26
2.3.3 Summary Statistics and Characteristics of Travel 27
2.4 Specification and Results 30
2.4.1 Effect of Vehicle Type on Following Distance 31
2.4.2 Effects of Other Variables on Following Distance 34
2.4.3 Following Behavior at Different Congestion Levels 35

2.5 Discussion of Driver Selection 36
2.5.1 Driver Selection into Vehicle Type 38
2.5.2 Driver Selection into Lead Vehicle Type 40
2.6 Implications for Congestion Costs 43
2.7 Conclusion 47

3 Driver (Mis)Perception of Speed in Congested Traffic 49
3.1 Introduction 49
3.2 Theoretical Background 52
3.2.1 Vehicle Count Heuristic 53
3.2.2 Interval Count Heuristic 55
3.3 Data 59
3.4 Analysis 62
3.4.1 Heuristic Performance in Heavy Congestion 67
3.5 Conclusion 67

Bibliography 68

LIST OF FIGURES

Page

2.1 Primary Lane Distribution by Vehicle Type 39
2.2 Switch Subsample 42
2.3 Traffic Speed and Flow 44
2.4 Lane 3 Speed at 4:00pm 45

3.1 Vehicle Count Heuristic 53
3.2 Constant Vehicle Spacing 57
3.3 Increasing Vehicle Spacing 58
3.4 Performance of the Heuristics 63
3.5 Comparison of the Heuristics 64
3.6 Kernel Density Estimate 65

LIST OF TABLES

Page

1.1 Dunkin' Donuts Franchises 10
1.2 Regressions of Coffee and Latte Price 11
1.3 Latte Subsample Regressions 13

2.1 Following Distance Choice 25
2.2 Following Events Sample 29
2.3 Vehicle Characteristics 30
2.4 Effect of Vehicle Type on Following Distance 32
2.5 Effect of Light Truck as Following and Lead Vehicle 33
2.6 Predicted Following Distance 35
2.7 Predicted Headway 36
2.8 Average Headway 37
2.9 Lane Change Behavior by Vehicle Type 40
2.10 Switch and Merge Subsamples: Effect of Vehicle Type on Following Distance 41

3.1 Possible Outcomes for Net Vehicles Passed and Speed Difference 55
3.2 Traffic Flow and Speed 60
3.3 Driver Comparison Sample 61
3.4 Regression: Vehicle Ratio on Interval Ratio 66

ACKNOWLEDGMENTS

I am grateful to my dissertation committee, Kevin Roth, Linda Cohen, and Jan Brueckner, for invaluable guidance on my dissertation. Specifically, I would like to thank Kevin Roth for pushing me to clarify my thinking and to improve my writing. I also thank Linda Cohen for her stimulating questions and honest advice. I thank Jan Brueckner for working diligently with me to improve my writing. I am also grateful to Ami Glazer for his suggestions and tough questions, which improved the quality of my work.

I also thank my family, Mark, Betsy, Eddie, Frank, Joe, and Annie, and friend, Matt Kaminsky, for help collecting data for the first chapter of this dissertation. Beyond that, I am grateful to my parents and my brothers and my sister for grounding support and providing perspective at difficult times.

I thank Arjun Ravikumar and Kara Dimitruk for many insightful conversations, boundless understanding, and bringing me joy.

Funding provided by the Department of Economics and School of Social Sciences, University of California, Irvine is acknowledged.

CURRICULUM VITAE

Kim Makuch

EDUCATION

Doctor of Philosophy in Economics **2016**
University of California, Irvine

Master of Arts in Economics **2015**
University of California, Irvine

Bachelor of Arts in Economics and German **2009**
College of the Holy Cross

AWARDS AND FELLOWSHIPS

Summer Research Fellowship **2013**
Department of Economics, University of California, Irvine

Phi Beta Kappa **2009**
College of the Holy Cross

Edna Dwyer Grzebien Prize **2009**
Modern Languages and Literatures, College of the Holy Cross

EXPERIENCE

Teaching Assistant **2011-2016**
Department of Economics, University of California, Irvine

- Introductory Microeconomics
- Introductory Macroeconomics
- Intermediate Microeconomics
- Intermediate Macroeconomics
- Industrial Organization
- Probability and Statistics
- Applied Econometrics

Research Assistant to Assistant Professor Kevin Roth **2014**
University of California, Irvine

Research Assistant to Professor Katherine Kiel **2007**
College of the Holy Cross

ABSTRACT OF THE DISSERTATION

Essays in Transportation Economics and Industrial Organization

By

Kim Makuch

Doctor of Philosophy in Economics

University of California, Irvine, 2016

Assistant Professor Kevin Roth, Chair

This dissertation consists of three chapters: one chapter in industrial organization and two chapters in transportation economics. The first chapter studies competition between geographically differentiated firms selling differentiated products. Using an original data set of prices at a franchised retail coffee bar in New York City, this chapter examines how a retailer's price varies with its degree of geographic differentiation from competitors. The evidence shows that the relationship between price and geographic differentiation depends on the extent to which a product is differentiated from competitors' products in other characteristics. If products are otherwise sufficiently differentiated, prices are higher when competing retailers are located geographically close together. Products that are not sufficiently differentiated in other characteristics, by contrast, face lower prices when geographic differentiation is smaller. The findings appear to be consistent with multidimensional product differentiation models.

The second and third chapters study driver behavior in congested traffic. The second chapter determines the size of the congestion externality generated by different vehicle types. Analysis of detailed vehicle trajectory data provides evidence that light trucks (pickups, sport utility vehicles, and vans) use 4-6% more road space than cars because drivers follow light trucks at a greater distance and light truck drivers follow other vehicles at a greater distance

than car drivers. Using a bottleneck congestion model, I estimate that the average delay on a California freeway could be reduced by 26% or $2.2 million annually if all light trucks were replaced with cars.

The third chapter explores the reliability of heuristics drivers may use to judge relative speed and make lane change decisions. Using vehicle trajectory data, I find that a driver may underestimate her speed relative to other lanes. She is also more likely to do so in heavily congested traffic. Given the external costs of lane changing and the evidence presented in this chapter, many lane changes are likely neither socially nor privately optimal.

Chapter 1

Pricing Patterns in Competition between Two Dimensional Products

1.1 Introduction

This chapter empirically studies the effect of geographic proximity to a competitor on price when firms sell differentiated products. Consider a market in which two stores each sell a variant of a product. The product variant will be called the brand. Sellers make three strategic choices: the store's location, the brand to sell and the price to charge. Consumers are distributed uniformly in both location and in preference for a particular brand. A consumer buys the product that delivers the greatest utility.

What will this market look like? Will the stores choose nearby locations at the market center or distant locations? Will they sell similar brands or will each store cater to a particular subset of consumers? If a retailer's location and brand are considered the two horizontal characteristics of its product, two dimensional product differentiation models offer insight into the equilibrium outcome. These models find that firms choose maximal differentiation in one characteristic and minimal differentiation in the other (Neven and Thisse, 1990; Irmen and Thisse, 1998; Ansari et al., 1998). Equilibrium profits are highest under this configuration because differentiation in one characteristic weakens price competition in the other characteristic. Therefore firms choose similar, broadly appealing to consumers positions in the second characteristic and, because differentiation in the other characteristic softens competition, they are able to charge higher prices.

The model further implies that firms choose maximal differentiation in the "dominant" characteristic, the one that matters most to consumers. Assuming that brand is dominant and geographic location is the dominated characteristic, the model then implies that firms will locate adjacent to one another in geographic space (at a point near the market center) while choosing highly differentiated brands. In reality, of course, location choices will exhibit considerable "noise," with competitors not located immediately adjacent to one another but instead found at a range of distances. This fact can be exploited along with another prediction of the model to carry out an empirical test. In particular, the analysis of Irmen and Thisse (1998) shows that, in the second-stage price game, prices rise as firms become closer to one another and the market center in the dominated characteristic (geographic location). Empirically, the implication is then that competitors' prices rise as the physical distance between them falls.

This chapter offers a test of this prediction using a unique and detailed data set of prices at a franchised retail coffee bar, Dunkin' Donuts, in the borough of Manhattan in New York City. In this market, competition is duopolistic and occurs between Dunkin' Donuts and Starbucks. The dimensions of product differentiation are geographic location and brand, where brand refers to the taste and advertising appeal of the product. The market is well-suited for analysis because Dunkin' Donuts are owned and operated as independent franchises. Therefore prices are set independently but the products offered at each Dunkin' Donuts franchise are the same.

If brand is the dominant characteristic, a two dimensional product differentiation model predicts that prices at Dunkin' Donuts will be higher at Dunkin' Donuts stores closer to Starbucks and the market center. The results show evidence of this pricing pattern in regular coffee prices but show an opposite pattern in latte prices. When Dunkin' Donuts and Starbucks are located closer together geographically, the price of regular coffee at Dunkin' Donuts is higher but the price of lattes is lower.

The results may be consistent with a two dimensional product differentiation model under certain assumptions. If each characteristic retains the same relative weight in the consumer utility function for coffee and latte choice, the dominance of a particular characteristic depends on the extent of differentiation between the retailers in each characteristic. For regular coffee and lattes, geographic differentiation between Dunkin' Donuts and Starbucks is the same but brand differentiation may differ. Closer examination reveals that the difference in caffeine content and therefore taste between the retailers is much larger for regular coffee than for lattes. In addition, advertising may effectively differentiate Dunkin' Donuts regular coffee from Starbucks regular coffee while failing to do so for Dunkin' Donuts lattes. Given this information, the premise that brand differentiation is larger for regular coffee than for lattes is a reasonable possibility. Under this case, brand would be the dominant characteristic for coffee and geographic location would be the dominant characteristic for lattes. Then the two dimensional product differentiation model predicts that Dunkin' Donuts closer to Starbucks have higher coffee prices but lower latte prices. In light of this argument, the empirical findings are consistent with multidimensional product differentiation models although a different view could be argued.

The empirical literature on product differentiation in multiple characteristics is fairly sparse. Netz and Taylor (2002) and Iyer and Seetharaman (2008) study geographic and service level differentiation of retail gasoline stations. Netz and Taylor find that as geographic differentiation increases, retailers become more differentiated in service offerings. By contrast, Iyer and

Seetharaman find that retailers are more likely to provide differentiated service quality levels when geographic differentiation is limited. Netz and Taylor's findings provide evidence of a product-location equilibrium where firms maximally differentiate themselves in all product characteristics. Iyer and Seetharaman's findings, on the other hand, are consistent with multidimensional product differentiation models; firms maximally differentiate themselves in one dimension (service quality) and minimally differentiate in the other dimension (geographic location). Watson (2009), in an analysis of eyeglass retailers, also finds results consistent with theoretical models. His evidence suggests that retailers maximally differentiate themselves in product style and minimally differentiate themselves in geographic location.

In contrast to the literature above, this chapter examines the relationship between price and the extent of geographic differentiation. Because other mechanisms could generate the product-location outcomes found, studying price and its relationship to product differentiation provides a better test of the workings of the two dimensional product differentiation model.

The chapter is structured as follows. Section II lays out a two dimensional product differentiation model, highlighting its testable predictions. The empirical setting, data and econometric specification are introduced in Section III. Section IV presents the results and Section V discusses the empirical findings in the model's context. Finally, Section VI concludes.

1.2 A Two Dimensional Product Differentiation Model

The theoretical literature has developed a multitude of product differentiation models in one and more recently, multiple product dimensions.[1] These models explore both horizontal

[1] See Hotelling (1929); D'Aspremont et al. (1979); Gabszewicz and Thisse (1979); de Palma et al. (1985); Ben-Akiva et al. (1989), and Wauthy (1996) among others for one dimensional models. For multidimensional models, see Neven and Thisse (1990); Tabuchi (1994); Irmen and Thisse (1998), and Ansari et al. (1998).

and vertical product characteristics.[2] Most relevant for the empirical application are models in which products are differentiated in multiple horizontal characteristics as in a model by Irmen and Thisse (1998). Their findings are summarized here for the case of two horizontal product characteristics. The product characteristics are referred to as geographic location and brand to match those studied in the empirical analysis.

In the model, two firms sell two dimensional products, a and b. Each product has a particular geographic location and brand that can be represented by its position on a rectangle where each dimension represents one characteristic. Consumers are distributed uniformly over the rectangle in geographic location and preferred brand and have unit demands. A consumer with geographic location, x, and preferred brand, y, who buys a has the conditional indirect utility, $V_a(x, y) = Y - P_a - w_1(x - a_1)^2 - w_2(y - a_2)^2$ where a_1 and a_2 are geographic location and brand for a and w_1 and w_2 are utility weights (Y is income and P_a is a's price). The utility from b, $V_b(x, y)$, is given by an analogous expression.

Brand is said to dominate geographic location when $w_2(b_2 - a_2) > w_1(b_1 - a_1)$ or when the weighted extent of brand differentiation between products is larger than the weighted extent of geographic differentiation.[3] Intuitively, under dominance of brand, the difference in the marginal utility of brand between a and b is larger than the difference in the marginal utility of geographic location.

Firms choose product characteristics and price to maximize profit in a two-stage game. Geographic location and brand are chosen simultaneously in the first stage and in the second stage, price is chosen. Solving the game by backward induction yields the sub-game perfect product-location equilibrium of maximal differentiation in the dominant characteristic and

[2] Horizontal product characteristics are characteristics over which consumers do not share the same preferences (for example, color and location). Vertical characteristics, on the other hand, are characteristics over which consumers share the same preferences. Quality is a vertical characteristic because all consumers prefer higher quality products.

[3] Without loss of generality assume $b_2 > a_2$ and $b_1 > a_1$.

minimal differentiation in the dominated characteristic. In other words, firms are geographically adjacent but choose maximally differentiated brands.

The comparative statics of the price sub-game equilibrium highlight the forces that drive the product-location equilibrium. Somewhat surprisingly, when a product moves closer to its competitor in the dominated characteristic (geographic location), its price will rise provided this movement also makes the product closer to the market center. The reason is that the positive effect of a more central geographic location outweighs the negative effect of a geographically closer competitor, a consequence of the dominance of brand, which weakens the effect of competition in geographic location. In the dominant dimension (brand), however, the negative effect of a closer competitor outweighs the centrality effect, so that moving closer to a competing product in brand leads to a lower price.

1.3 Empirical Framework

1.3.1 Setting

The empirical analysis studies competition between two retail coffee bars, Dunkin' Donuts and Starbucks, in the borough of Manhattan in New York City. In 2012, there were 201 Starbucks stores and 122 Dunkin' Donuts stores in Manhattan, making the chains the largest retail coffee bars by far (Laney, 2012).[4] Therefore, competition in the market for quick service coffee is characterized as duopolistic.

Further, Starbucks and Dunkin' Donuts are differentiated in two dimensions - geographic location and brand. A walk through the streets shows that each retailer's stores are located throughout the city. The second characteristic, brand, refers to the appeal of each brand

[4] McDonalds, Tim Hortons and Peet's Coffee and Tea had 69, 15 and 0 outlets respectively in Manhattan in 2012 (Laney, 2012).

created through coffee taste and advertising. Starbucks is known for strong, bitter coffee with an upscale flair and its emphasis on customer service. Dunkin' Donuts cultivates an image of unfussy coffee for everyday, hardworking people. For example, in 2007, Dunkin' Donuts ran television commercials featuring the voice of actor John Goodman and the slogan "America runs on Dunkin'." Dunkin' Donuts also made fun of beverages with foreign sounding, hard to pronounce names at other retailers in a 2006 ad (Champagne and Iezzi, 2014).

This competitive setting is well-suited for study of the interesting implication of the model, namely, that price may be higher when a product is more similar to a competitor in the dominated characteristic. If brand is the dominant characteristic and geographic location is the dominated characteristic, then, according to the model, there will be an inverse relationship between price and distance to Starbucks; in other words, Dunkin' Donuts stores far from Starbucks will have lower prices.

Although competition between Starbucks and Dunkin' Donuts in Manhattan fits many of the assumptions of the two dimensional product differentiation model there is one important difference between the model and the empirical application. In the model, the market center of each characteristic is well defined. The location of the market center is important because equilibrium price increases as a product becomes more similar to its competitor only if the product also becomes closer to the market center. According to the model, Dunkin' Donuts' price will be higher at Dunkin' Donuts stores closer to Starbucks only if Starbucks lies in the direction of the market center. This assumption cannot be directly verified, but, because Starbucks stores are located to maximize joint profits and Dunkin' Donuts stores are located sequentially by individual profit maximizing franchisees, it is plausible that Starbucks stores generally have more central locations than Dunkin' Donuts stores.

1.3.2 Data

The data set contains prices at all Dunkin' Donuts stores and the locations of all Dunkin' Donuts and Starbucks stores in the area of Manhattan south of 110th Street (the northern border of Central Park).[5]

Starbucks and Dunkin' Donuts each sell two main coffee products, regular brewed coffee and espresso drinks such as lattes and cappuccinos. Hereafter the former is referred to as coffee and the latter as lattes. Unfortunately, the original study design did not include an analysis of latte prices and consequently, they were not purposely collected. However, because prices were recorded by photographing the large price boards at each location and latte prices are frequently listed near coffee prices, latte prices were recorded at most Dunkin' Donuts stores (72%). A comparison of the summary statistics for the full sample and the latte subsample suggests that the latte subsample is a random sample from the population of Dunkin' Donuts stores in the sample area.

Aside from prices at Dunkin' Donuts and the location of each Dunkin' Donuts and Starbucks store, the data set also includes information on franchise ownership and market characteristics of the area surrounding each Dunkin' Donuts store. Using franchise ownership data from the NYC Department of Health, franchises can be identified as jointly or independently owned. This information is useful in understanding the competitive environment of each Dunkin' Donuts store.

[5] Prices at 94 Dunkin' Donuts stores in Manhattan were collected on one day in December 2012. In December 2012, there were 98 Dunkin' Donuts stores in the sample area. Two were excluded from the sample because they are located within hotels; one was excluded because it is one of two Pennsylvania Station outlets with identical location, price and ownership and the fourth was excluded because its owner refused to allow data collection both in person and over the phone. There were 190 Starbucks stores not located in other stores, hotels or malls in December 2012. The excluded stores are not expected to compete with Dunkin' Donuts in the same way as standalone stores. In total, ten Starbucks stores in the sample area were excluded.

Variables to identify consumer locations and characteristics are also included in the data set. The number of fares paid at all subway entrances throughout the city during the week of price data collection was obtained from the Metropolitan Transit Authority. In addition, the data set contains information on worker and resident densities by census tract and the locations of top tourist attractions in Manhattan.[6] To measure consumer characteristics, median household income by census tract as well as property value data from the NYC Department of Finance for buildings adjacent to each Dunkin' Donuts is included.

Summary statistics for all variables are found in Table 1.1 for the full sample and the latte subsample. Lattes are about six cents more expensive per ounce than coffee at Dunkin' Donuts on average. Turning to location, Dunkin' Donuts stores are located about 0.12 miles on average from the nearest Starbucks store and about 0.23 miles from the nearest independent Dunkin' Donuts store on average. Most Dunkin' Donuts stores are jointly owned with only about 23% of stores in the sample area operating independently.

1.3.3 Model Specification

The equation below is estimated by OLS separately for coffee and latte price.

$$P_i = \alpha + \beta_1 \text{DistSB}_i + \beta_2 \text{DistSB}_i^2 + \boldsymbol{X_{1i}\delta_1} + \boldsymbol{X_{2i}\delta_2} + \epsilon_i$$

P_i is either coffee or latte price at Dunkin' Donuts store i, $DistSB_i$ is the distance to the nearest Starbucks from Dunkin' Donuts store i and $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ contain variables to control for other competitors (aside from Starbucks) and market characteristics. To control for other competitors, the number of independent Dunkin' Donuts stores in the immedi-

[6]Top tourist attractions come from a list published on Timeout.com. Included in the list are the Brooklyn Bridge, the Staten Island Ferry, the World Trade Center site, Times Square, the Empire State Building, the Chrysler Building, Rockefeller Center, the American Museum of Natural History, Battery Park, Grand Central Terminal, the Metropolitan Museum of Art, New York Public Library and theaters on Broadway.

Table 1.1: Dunkin' Donuts Franchises

	Full Sample		Latte Subsample	
	Mean	SD	Mean	SD
Coffee price (per oz.)	0.13	0.01	0.13	0.01
Latte price (per oz.)	.	.	0.19	0.02
Distance to SB (miles)	0.12	0.11	0.11	0.08
Number SB within 0.1 miles	0.86	0.97	0.90	1.01
Distance to independent DD (miles)	0.23	0.09	0.22	0.09
Number independent DD within 0.1 miles	0.05	0.23	0.04	0.21
Number other franchises owned	7.95	7.20	7.74	7.25
Multiple ownership dummy	0.77	0.43	0.74	0.44
Number subway passengers within 0.1 miles	165,107	235,407	187,256	244,392
Tourist attraction within 0.25 miles dummy	0.12	0.32	0.12	0.32
Worker density (workers/sq. mile)	226,966	208,383	234,991	201,088
Resident density (population/sq. mile)	84,774	51,983	84,492	48,177
Median household income	92,568	39,491	94,009	37,444
Property value ($/sq. ft)	305.42	231.33	294.01	175.67
Observations	94		68	

ate area is included. Subway ridership, worker and resident densities, median household income, proximity to tourist attractions and property value are used to control for market characteristics.

Controlling for other competitors and market characteristics is necessary because these factors affect price and could be correlated with distance to Starbucks. Including these variables should separate an area's general price level from the effect predicted by the model and captured by β_1 and β_2. Omitted variable bias, however, may affect the coefficient estimates if there are unobserved variables that affect price and are correlated with distance to Starbucks. To mitigate omitted variable bias, the number of Starbucks stores in the immediate area of each Dunkin' Donuts store is included in addition to the above variables. Because Starbucks chooses its locations with better knowledge of all factors that affect demand and price, it seems likely that Dunkin' Donuts stores near the same number of Starbucks stores face similar conditions.

Table 1.2: Regressions of Coffee and Latte Price

Dependent Variable	Coffee Price		Latte Price	
	(1)	(2)	(3)	(4)
Distance to SB	-0.0175	-0.0525**	0.0347	0.1062*
	(0.0232)	(0.0263)	(0.0609)	(0.0617)
Distance to SB^2	0.0206	0.0666†	-0.0945	-0.2039**
	(0.0384)	(0.0432)	(0.1006)	(0.0981)
SB within 0.1 miles		-0.0025*		0.0040
		(0.0015)		(0.0040)
Income ($100,000s)	0.0020	0.0022	0.0101	0.0092
	(0.0025)	(0.0023)	(0.0081)	(0.0084)
Subway passengers (1,000,000s)	-0.0066	-0.0045	-0.0212*	-0.0238†
within 0.1 miles	(0.0057)	(0.0054)	(0.0118)	(0.0146)
Property value ($10,000s/sq. ft)	0.0019	-0.0005	0.0466	0.0555
	(0.0331)	(0.0292)	(0.1266)	(0.1251)
Number other franchises owned	0.0002**	0.0003***	-0.0005**	-0.0006**
	(0.00008)	(0.00009)	(0.0003)	(0.0003)
Independent DD within 0.1 miles	-0.0003	0.0003	-0.0114†	-0.0128*
	(0.0032)	(0.0030)	(0.0071)	(0.0071)
Constant	0.1272***	0.1317***	0.1858***	0.1778***
	(0.0042)	(0.0043)	(0.0111)	(0.0116)
Observations	94	94	68	68
R-squared	0.099	0.139	0.149	0.164
Marginal effect of distance to SB	-0.0014	-0.0040	0.0010	0.0047
F-Statistic for marginal effect	0.66	4.51**	0.09	2.05†

Notes: Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$, † $p<0.2$

1.4 Results

The results are presented in Table 1.2. In the first two columns coffee price is the dependent variable and in the next two columns latte price is the dependent variable. Columns 2 and 4 include the number of Starbucks stores within 0.1 miles in the regression. The bottom two rows of Table 1.2 show the marginal effect of a one standard deviation increase in distance to Starbucks (from the mean) and the F-statistic for the test that this effect is zero.

Interestingly, the marginal effect of an increase in distance to Starbucks is negative for coffee price but positive for latte price. In other words, greater distance to Starbucks leads to lower coffee prices and higher latte prices. The effect is statistically significant for coffee price and marginally significant for latte price when the number of Starbucks stores within 0.1 miles is included. The results show that a one standard deviation increase in distance to Starbucks leads to a \$0.004 per ounce decrease in coffee price and a \$0.005 per ounce increase in latte price on average. These effects are small but have opposite signs.

When the number of Starbucks stores within 0.1 miles is added to the model, the effect of distance to Starbucks on coffee and latte price increases in magnitude. The effect, however, is tempered by the effect of an increase in the number of Starbucks stores within 0.1 miles. For coffee, a shorter distance to Starbucks leads to higher prices but having a greater number of Starbucks stores close by leads to lower prices. Although it seems odd that the presence of Starbucks has two opposite effects on coffee price, the effect captured by the number of Starbucks stores in the immediate area may be, instead, the effect of demand conditions on price. In addition, even if moving one standard deviation closer to Starbucks increases the number of Starbucks stores within 0.1 miles by one, the net effect of closer proximity to Starbucks remains positive for coffee price and negative for latte price.

The coefficient estimates for the other variables are largely as expected. Areas with higher median household income have higher prices though the effect is not statistically significant. Areas with a larger number of subway passengers have lower prices and the effect is statistically significant for latte price. This result could indicate that locations with more subway passengers have higher demand and, because of a larger number of unobserved competitors, lower prices. The coefficient estimate for property value is generally positive but standard errors are large in all specifications.

Dunkin' Donuts stores owned by an individual who owns multiple franchises have higher coffee prices and lower latte prices. For coffee price, the result is consistent with findings

Table 1.3: Latte Subsample Regressions

Dependent Variable	Coffee Price	
	(1)	(2)
Distance to SB	0.0037	-0.0205
	(0.0294)	(0.0268)
Distance to SB^2	-0.0317	0.0053
	(0.0462)	(0.0426)
SB within 0.1 miles		-0.0014
		(0.0017)
Income ($100,000s)	0.0029	0.0032
	(0.0030)	(0.0031)
Subway passengers (1,000,000s) within 0.1 miles	-0.0029	-0.0021
	(0.0060)	(0.0061)
Property value ($10,000s/sq. ft)	0.0716*	0.0686†
	(0.0424)	(0.0421)
Number other franchises owned	0.0002***	0.0003**
	(0.0001)	(0.0001)
Independent DD within 0.1 miles	-0.0051*	-0.0046†
	(0.0028)	(0.0032)
Constant	0.1214***	0.1241***
	(0.0052)	(0.0043)
Observations	68	68
R-squared	0.1462	0.1591
Marginal effect of distance to SB	-0.0005	-0.0021
F-Statistic for marginal effect	0.05	1.32

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1, † p<0.2

by Thomadsen (2005), who found that fast food franchises owned jointly by a single owner had higher prices than independent franchises. For latte price, the effect may be negative because more experienced owners recognize that lower prices encourage customers to try new products, like lattes.[7] Finally, the presence of other independent Dunkin' Donuts stores nearby is negatively related to price in all but one specification and is significant for latte price. This result is consistent with a model of one dimensional product differentiation; when a store is located near other independent Dunkin' Donuts stores price is lower due to price competition.

The variables measuring subway passengers, resident and worker densities and proximity to tourist attractions were found to be highly collinear. Therefore, only subway passengers are included in the main specification. The results using resident density, worker density, and proximity to tourist attractions instead of subway passengers are unchanged but somewhat weaker.

Table 1.3 shows the estimation results for each model estimated on the latte subsample and using coffee price as the dependent variable. As in the full sample there is a negative relationship between coffee price and distance to Starbucks. This finding further suggests that the latte subsample is a random sample from the population of Dunkin' Donuts stores in the sample area.

1.5 Discussion

The empirical analysis shows that the relationship between price and distance to Starbucks is different for coffee and lattes at Dunkin' Donuts. Coffee prices are lower and latte prices are higher when Dunkin' Donuts is located farther from Starbucks. In the language of the

[7] Espresso drinks are a relatively new product at Dunkin' Donuts. They were introduced in 2003.

model, Dunkin' Donuts stores that are more similar to Starbucks in geographic location, have higher coffee prices and lower latte prices.

These findings are consistent with the model if brand is the dominant product characteristic for coffee but geographic location is the dominant characteristic for lattes. For the above premise to be true and assuming that consumers have the same utility weights in coffee and latte choice for each characteristic, it must be that Dunkin' Donuts coffee is well differentiated in brand from Starbucks coffee but Dunkin' Donuts lattes are not well differentiated from Starbucks lattes in brand, making geographic location, instead, the dominant characteristic.

Although Starbucks is well known for its espresso offerings it may still be true that brand differentiation between the retailers is greater for coffee than lattes. This argument appears reasonable if product taste and appeal created by advertising determine brand. In particular, differences in caffeine content indicate that coffee is more differentiated than lattes by taste between the retailers. Coffee at Starbucks contains 21.4 mg caffeine per ounce on average while coffee at Dunkin' Donuts contains only 12.5 mg caffeine per ounce on average. For lattes, the difference in caffeine content is much smaller. Starbucks lattes contain 8.1 mg caffeine per ounce on average and Dunkin' Donuts lattes contain 7.3 mg caffeine per ounce on average.[8] Since caffeine content affects the taste of food and drinks, consumers may perceive a greater difference in coffee taste than in latte taste between Dunkin' Donuts and Starbucks.

Considering the appeal created by advertising, it seems likely that Dunkin' Donuts' strategy to market itself as coffee for everyday Americans is more effective for coffee than lattes. Espresso drinks are relatively new to the American market and creating an unpretentious image for them may be difficult. While consumers accept Dunkin' Donuts coffee as an unfussy beverage for everyday people, they are less likely to accept the same claim for lattes.

[8] Caffeine content information was obtained through the Starbucks website and Dunkin' Donuts customer service.

Using advertising, Dunkin' Donuts is able to further differentiate its coffee from Starbucks coffee but it may not be able to do so for lattes.

1.6 Conclusion

Estimation results presented above suggest that for two dimensional products, the degree of differentiation in each dimension is important in equilibrium price determination. For the retail coffee bars studied, the results show that if retailers are sufficiently differentiated in brand, prices are higher when they are located geographically close to each other and close to the market center. Without sufficient brand differentiation, prices are lower if geographic differentiation is smaller. These empirical results are consistent with models of two dimensional product differentiation.

The analysis suggests that the competitiveness of a market is determined by the degree of differentiation in the dominant product characteristic. Minimal differentiation between products in some characteristics does not necessarily indicate a high degree of competition in the market. In the empirical analysis, closer geographic proximity to competitors is not associated with lower coffee prices because coffee is highly differentiated in brand, the dominant characteristic. To effectively assess market competition, extensive knowledge of all product characteristics and consumer preferences is needed.

Chapter 2

The External Congestion Costs of Light Trucks

2.1 Introduction

The share of light trucks in the U.S. vehicle fleet has more than doubled over the past three decades.[1] Today, almost half of all registered vehicles are light trucks (Federal Highway Administration, 2015).[2] Because of the rise in popularity and the physical characteristics of light trucks, many researchers have studied the external accident costs of light trucks (White, 2004; Anderson, 2008; Li, 2012; Anderson and Auffhammer, 2014). The relative congestion externality of light trucks, however, has received little attention. This is somewhat surprising because external congestion costs are estimated to be larger than external accident costs (Small and Verhoef, 2007). Further, policy makers must consider differences in congestion costs between vehicle types when designing regulations that affect the composition of the

[1] Pickups, sport utility vehicles (SUVs), and vans are called light trucks.

[2] The share of light trucks in the vehicle fleet increased from 22% to 47% between 1980 and 2013 (Federal Highway Administration, 2015).

vehicle fleet. The structure of vehicle fuel economy standards, for example, may create incentives for auto manufacturers to produce certain vehicle types. With the Corporate Average Fuel Economy (CAFE) standards for model years 2022-2025 currently awaiting approval, this analysis provides timely, new evidence that can be used in assessing the unintended congestion consequences of the new standards.

The congestion externality of light trucks may be large in comparison to that of cars because drivers respond to the size of light trucks in ways that reduce traffic flow and increase delays. Because light trucks make it more difficult for following drivers to see the road ahead, drivers following light trucks are likely to choose greater following distances for safety. Further, light trucks have relatively poor braking, leading light truck drivers to also choose greater following distances for safety. Taking these behavioral adjustments and the length of light trucks into account, it is likely that light trucks use more road space than cars. As a result, fewer vehicles will be able to traverse a road segment in a given amount of time if a greater share of vehicles are light trucks and speed is held constant. The effect of light trucks on congestion costs may be substantial because total traffic congestion costs are very large. In 2014, the costs of wasted time and fuel in congested traffic were estimated to be $160 billion (Schrank et al., 2015).

To estimate the effect of light trucks on traffic congestion, I first estimate the effect of light trucks on following distance using an engineering data set of observed trajectories for all vehicles traveling in afternoon peak period traffic on Interstate 80 in Emeryville, CA. I then model the effect of road space usage on traffic congestion using a bottleneck congestion model calibrated to the observed traffic conditions. Using this model, I simulate the delay reduction possible if all light trucks were replaced with cars.

Previous analyses of the effect of light trucks on following distance provide mixed results. Sayer et al. (2003) and Brackstone et al. (2009) find that drivers follow all light trucks (pickups, SUVs, and vans) more closely than cars while Cosgrove (2011) finds that drivers follow pickups more closely than cars but SUVs at a greater distance. All three studies, however, rely on data from vehicles equipped with measurement devices to record detailed driving data. In this type of study, vehicles are driven by study participants for an extended period. These studies face important limitations that may affect their conclusions. First, because participants know that their driving behavior is recorded and later analyzed, the Hawthorne effect, the tendency to improve behavior when being observed, may affect the results. In addition, these studies provide no information about the following behavior of light truck drivers because the same vehicle, a sedan, was lent to every participant. Further, the expense of such studies limits the size of each sample, and it is unclear whether the sample represents the population of peak period drivers.[3] Finally, it is difficult to determine the implications of the following distance results for traffic congestion because broader traffic conditions are unknown.

By using trajectory observations for all vehicles traveling on the same freeway in the afternoon peak period, the analysis of this chapter improves on previous studies. Two key results emerge from the estimates: compared to cars, light trucks follow other vehicles at a greater distance and other vehicles follow light trucks at a greater distance. Including the additional space a light truck physically occupies, light trucks use 4-6% more space in congested traffic than cars. A simple bottleneck congestion model calibrated to observed traffic conditions shows that the average delay in the study area could be reduced by 26% if all light trucks were replaced with cars. The time savings from this replacement on the observed freeway are valued at $2.2 million or $333 per light truck replaced annually. The time savings associated with substituting cars for light trucks in congested traffic is of the same magnitude as the previously estimated safety benefit of such a policy (White, 2004).

Driver selection into vehicle type, however, could cause upward bias in the estimates if cautious drivers choose to drive light trucks. To mitigate this problem I include information

[3] In the studies above between 6 and 70 drivers were observed.

about lane changing behavior and lane choice that captures driver caution. Further, observable behavioral differences between drivers of different vehicle types are small and largely statistically insignificant. Driver selection is also plausible in *lead* vehicle type. Not only does a driver choose the vehicle she drives but she may also choose the vehicle she follows by maneuvering in traffic. Driver selection into lead vehicle type could bias the results if a certain type of driver chooses to follow a light truck. I examine this possibility by limiting the sample to observations in which drivers were effectively randomly assigned to follow a certain vehicle because of lane changes by other drivers or the need to merge with traffic. In the subsample, the effect of following a light truck on following distance confirms and, in one case, is larger than the full sample estimates. The analysis shows that the estimates are not upwardly biased due to driver selection into lead vehicle type.

This remainder of the chapter is organized as follows. Section 2 provides a model of driver decision making under vehicle size heterogeneity. Sections 3 and 4 describe the data and the results of the empirical analysis. Section 5 discusses the robustness of the results to selection effects. Section 6 estimates the effect of reducing the number of light trucks in traffic on congestion. Finally, Section 7 concludes.

2.2 Theory

A driver's choice of following distance may be affected by the size of her own vehicle and the size of the vehicle she follows (the lead vehicle) if accident costs depend on these factors. The following model adds vehicle size heterogeneity to a model developed by Rotemberg (1985). The model outlines the effect of following and lead vehicle size on following distance choice under various assumptions.

2.2.1 Model Assumptions

Suppose drivers travel on a one lane road and passing is not allowed. Every driver (except the first and the last) is both a following and lead driver. Drivers face time, impatience, and accident costs. Accident costs are paid by the following driver for accidents occurring with the lead vehicle.[4] The total cost each following driver faces in one mile of travel is given by

$$C(V, D, S, S_l) = \theta \frac{1}{V} + \gamma \frac{D}{V} + \rho A(V, D, S, S_l). \tag{2.1}$$

V and D represent the velocity (in miles per hour) and following distance (in miles) at which the following driver follows the lead vehicle.[5] S and S_l are the size of the driver's vehicle and size of the lead vehicle. The probability of an accident is given by $A(\ )$. Finally, θ and γ are values of travel time and frustrated time. The cost of an accident is given by ρ.

The first term in the cost function is the travel time cost. This cost is the product of the time it takes to travel one mile and the value of travel time. The next term is the impatience cost. The impatience cost represents the frustration or effort associated with maintaining a given following distance.[6] In a multi-lane context, drivers may prefer to follow closely to avoid being cut off by other drivers. The impatience cost increases with following distance and decreases with velocity because, at a greater velocity, a given distance can be traversed more quickly, leaving the driver less frustrated.

The accident cost is equal to the product of the cost of an accident and the probability that an accident occurs. The shape of the accident probability function with respect to V and D

[4] This assumption is justified because, in most cases, the driver of the following vehicle in a rear-end collision is legally at fault.

[5] Velocity is used interchangeably with speed.

[6] In Rotemberg (1985), this term is considered part of the time cost because a driver who chooses a larger following distance will arrive at her destination later.

follows Rotemberg (1985) and is given by the following inequalities:

$$\frac{\partial A}{\partial V} \geq 0,\ \frac{\partial A}{\partial D} \leq 0,\ \frac{\partial^2 A}{\partial V^2} \geq 0,\ \frac{\partial^2 A}{\partial D^2} \geq 0 \text{ and } \frac{\partial^2 A}{\partial V \partial D} \leq 0. \tag{2.2}$$

Both driving faster and following closely increase the probability of an accident. Next, a speed increase is marginally more dangerous at higher speeds because braking distance increases quadratically with velocity (Brown et al., 2002). Further, allowing more following distance is more effective in reducing accident risk at small following distances. Finally, as velocity increases, an increase in following distance is more effective in reducing the probability of an accident.

2.2.2 Individual Optimization

Drivers choose following distance, D, to minimize total cost. Velocity is not a choice variable and each driver follows the lead vehicle at the same speed. I make this assumption because in the empirical application, traffic is heavily congested and drivers cannot freely choose speed. Thus, the driver's optimization problem is

$$\begin{aligned} \underset{D}{\text{minimize}} \quad & \gamma \frac{D}{V} + \rho A(V, D, S, S_l) \\ \text{subject to} \quad & D \geq 0. \end{aligned} \tag{2.3}$$

The first order condition with respect to following distance is

$$\gamma \frac{1}{V} + \rho \frac{\partial A}{\partial D} = 0. \tag{2.4}$$

This condition states that a driver sets her marginal impatience cost equal to the marginal accident prevention benefit from increasing following distance. If vehicles were identical, all vehicles would choose the same following distance. In this model, however, vehicles are heterogeneous in size. Large vehicles are taller and heavier than small vehicles. Totally differentiating (2.4) yields

$$\rho\frac{\partial^2 A}{\partial D^2}dD + \rho\frac{\partial^2 A}{\partial D\partial S}dS + \rho\frac{\partial^2 A}{\partial D\partial V}dV + \rho\frac{\partial^2 A}{\partial D\partial S_l}dS_l = 0 \tag{2.5}$$

which rearranges to give the relationship between following distance and vehicle size,

$$\frac{dD}{dS} = -\frac{\frac{\partial^2 A}{\partial D\partial S}}{\frac{\partial^2 A}{\partial D^2}} - \frac{\frac{\partial^2 A}{\partial D\partial V}}{\frac{\partial^2 A}{\partial D^2}}\frac{dV}{dS} - \frac{\frac{\partial^2 A}{\partial D\partial S_l}}{\frac{\partial^2 A}{\partial D^2}}\frac{dS_l}{dS}. \tag{2.6}$$

The second and third terms above are zero because velocity and lead vehicle size are not chosen by drivers and do not vary with vehicle size. The sign of the first term is determined by the cross partial derivative, $\frac{\partial^2 A}{\partial D\partial S}$. Because the denominator is positive, the effect of vehicle size on following distance has the sign opposite of the sign of the cross partial derivative. Following similar reasoning, the relationship between lead vehicle size and following distance depends on the cross partial derivative, $\frac{\partial^2 A}{\partial D\partial S_l}$.

2.2.3 Effect of Vehicle Size on Following Distance Choice

Accidents occur when the following driver fails to slow her vehicle to avoid a collision with the lead vehicle. This failure can occur in two ways: (1) the following driver does not react to a speed change quickly enough to avoid a collision or (2) the following vehicle's braking system is incapable of avoiding a collision. The size (height or weight) of the following and lead vehicle affect both driver reaction time and vehicle braking capability. The dominant effect determines the relationship between vehicle size and following distance. I consider each case in turn.

First, suppose that driver reaction time is a larger determinant of accident probability than vehicle braking capability. Under this assumption, a tall following vehicle gives the driver a better view of the road ahead so she can better anticipate and react to speed changes. The probability of an accident falls as following vehicle size increases $\left(\frac{\partial A}{\partial S} \leq 0\right)$. Therefore, an increase in following distance is more effective in reducing accident probability for small following vehicles $\left(\frac{\partial^2 A}{\partial S \partial D} \geq 0\right)$. The sign of the cross partial derivative implies that chosen following distance decreases as following vehicle size increases and large vehicles follow other vehicles more closely.

Now, consider lead vehicle height. If driver reaction time remains dominant in determining whether an accident occurs, then, as lead vehicle height increases, the probability of an accident increases $\left(\frac{\partial A}{\partial S_l} \geq 0\right)$ because the following driver's view of traffic is obstructed by the tall lead vehicle. Therefore, an increase in following distance is more effective in preventing accidents if the lead vehicle is large $\left(\frac{\partial^2 A}{\partial S_l \partial D} \leq 0\right)$. In this case, chosen following distance increases as lead vehicle size increases and vehicles follow large lead vehicles at a greater distance.

Alternatively, suppose vehicle braking capability is the primary determinant of accident probability. Because heavier vehicles cannot decelerate as quickly as lighter vehicles, the probability of an accident increases as following vehicle size increases $\left(\frac{\partial A}{\partial S} \geq 0\right)$.[7] Therefore, an increase in following distance is more effective in reducing the probability of an accident for large following vehicles $\left(\frac{\partial^2 A}{\partial S \partial D} \leq 0\right)$. Chosen following distance increases as following vehicle size increases and large vehicles follow other vehicles at a greater distance.[8]

Turning to lead vehicle size, because a heavier lead vehicle takes more time to slow down, the following vehicle has more time to reduce its own speed to avoid a collision. In this case, the probability of an accident decreases as lead vehicle size increases $\left(\frac{\partial A}{\partial S_l} \leq 0\right)$. Therefore,

[7] The NHTSA tested ten 1999 model year vehicles in the development of a vehicle brake testing protocol. Difference of means tests show that all cars had braking distances statistically smaller than braking distances of light trucks in the sample (Schultz and Babinchak, 1999).

[8] In this model, the cost of an accident is the same for all vehicles. In reality, however, drivers of large vehicles may be better protected and suffer smaller losses in an accident. It is possible that drivers behave according to the Peltzman effect. Drivers of large vehicles may, because they are better protected, take on more risk in driving and follow other vehicles more closely.

Table 2.1: Following Distance Choice

	Reaction time effect dominant	Braking effect dominant
Following vehicle size (S)	$\frac{\partial A}{\partial S} \leq 0$, $\frac{\partial^2 A}{\partial D \partial S} \geq 0$, $\frac{dD}{dS} \leq 0$	$\frac{\partial A}{\partial S} \geq 0$, $\frac{\partial^2 A}{\partial D \partial S} \leq 0$, $\frac{dD}{dS} \geq 0$
Lead vehicle size (S_l)	$\frac{\partial A}{\partial S_l} \geq 0$, $\frac{\partial^2 A}{\partial D \partial S_l} \leq 0$, $\frac{dD}{dS_l} \geq 0$	$\frac{\partial A}{\partial S_l} \leq 0$, $\frac{\partial^2 A}{\partial D \partial S_l} \geq 0$, $\frac{dD}{dS_l} \leq 0$

increasing following distance is more effective in accident prevention when the lead vehicle is small $\left(\frac{\partial^2 A}{\partial S_l \partial D} \geq 0\right)$ because small vehicles may stop very quickly. Chosen following distance decreases as lead vehicle size increases and vehicles follow large lead vehicles more closely. Table 2.1 summarizes the signs of the relevant partial and cross partial derivatives under each assumption and the resulting effect of vehicle size on following distance choice. Given the ambiguity of the theoretical findings, empirical analysis is useful in determining which cases describe freeway following behavior.

2.3 Data

To study the effect of vehicle type on following distance, I use an engineering data set containing the trajectories of all vehicles traveling on a particular freeway during a given time interval. The data were collected as part of the Federal Highway Administration's Next Generation Simulation (NGSIM) program in which traffic is videotaped and vehicle trajectories are later transcribed from the video (Federal Highway Administration, 2008). In this section I describe the location of data collection and the available data, explain the unit of observation for the analysis, and highlight summary statistics for key variables.

2.3.1 Study Area and Available Data

The data were collected on Interstate 80 eastbound in Emeryville, California during the afternoon peak period (4:00-4:15 pm and 5:00-5:30 pm) on April 13, 2005. On this segment of I-80, the I-80 and I-580 freeways run concurrently. The two freeways join upstream of the study area at the MacArthur Maze, a freeway interchange near the east end of the San Francisco-Oakland Bay Bridge in Oakland. The freeways later split in Albany at the Hoffman Split, downstream of the study area.

The observed segment is 0.31 miles in length and has one high occupancy vehicle (HOV) lane, five mainline lanes, and one on-ramp. There is an off-ramp about 100 feet downstream of the study area.

The trajectory data include each vehicle's position, speed, and acceleration recorded every 0.1 seconds. In addition, the lane of travel and the length and width of each vehicle are observed. Therefore, lane changes can be identified, and following distance can be determined using vehicle position and length. To supplement this data, I also identified all vehicles by type from the video recording. Every vehicle was classified as a motorcycle, car, pickup, van/SUV, or heavy-duty vehicle. Unfortunately, vans and SUVs must be grouped together because they could not be distinguished in the video. Heavy-duty vehicles are defined as heavy-duty trucks, recreational vehicles, and buses following the Highway Capacity Manual's definition (Transportation Research Board, 2010). Pickups, vans, and SUVs are classified as light trucks.

2.3.2 Unit of Observation

Following Sayer et al. (2003) and Cosgrove (2011), the unit of observation is a *following event*. A following event is defined as a following vehicle continuously following a lead vehicle.

Because vehicles often change lanes, one vehicle may be observed in multiple following events as a following or lead vehicle. In addition, following events vary in duration.

Previous studies of following behavior have used criteria for following to exclude following events in which the following driver's behavior is likely not influenced by the lead vehicle. For example, a driver may not be engaged in following the lead vehicle if she is a large distance behind the lead vehicle or if she chooses a slower speed. Employing a following criterion reduces random noise in the sample but, the stricter the criterion, the more likely it is that the sample represents a selected group of following events and drivers. I take a conservative approach and exclude following events in which the absolute difference in speed between the following and lead vehicle has a mean value greater than 10 mph.[9]

Further, following events that take place on the on-ramp or in the HOV lane are excluded from the sample. Vehicles traveling on the on-ramp are accelerating to join the flow of freeway traffic, and because cars are able to accelerate more quickly than light trucks (Kockelman and Shabih, 2000), differences in following behavior may reflect vehicle performance differences rather than driver behavior differences. Traffic conditions in the HOV lane differ significantly from conditions in the mainline lanes. In addition, following behavior in the HOV lane does not affect the delay in the mainline lanes. Finally, following events that involve motorcycles are excluded. The final data set contains 10,747 following events and 4,325 unique drivers.

2.3.3 Summary Statistics and Characteristics of Travel

Table 2.2 presents summary statistics for the sample of following events during the time periods 4:00-4:15 pm, 5:00-5:15 pm, and 5:15-5:30 pm. Following distance is defined as the distance (in feet) between the rear bumper of the lead vehicle and the front bumper of the

[9]Sayer et al. (2003) and Cosgrove (2011) use stricter speed difference criteria of 1.5 m/s and 1 m/s respectively (3.4 mph and 2.2 mph). Imposing a following criterion of 10 mph excludes 126 following events from the sample.

following vehicle. Following distance does not include either vehicle's length, only the space between the vehicles. Both mean following distance and speed decrease substantially across the three time periods, indicating that traffic conditions are changing quickly. Average following distance decreases by 16% while average speed decreases by 21% from the first period to the second. From the second period to the third, average following distance decreases by 13% and average speed decreases by 12%.

Dummy variables indicate the following and lead vehicle (the following pair) in each following event. The variable, Car-LT, indicates that a car follows a light truck. The other following pair dummy variables are named analogously. In the full sample, approximately 31% of following events involve a car following a car, 21% involve a car following a light truck, 22% involve a light truck following a car, and 16% involve a light truck following a light truck. The remaining 10% of following events involve heavy-duty vehicles. Given the vehicle type proportions, the following pair proportions are nearly identical to the expected proportions under random assignment of vehicles to following pairs.

Lane change or merge events (events that were initiated by a lane change or merging maneuver by the following driver) make up about 21% of all following events. Across all following events, following drivers made, on average, 0.83 lane changes, and 47% of following events involve a following driver who made at least one lane change. Finally, about 25% of following events occur in the rightmost lane, and the proportion of following events occurring in a lane decreases when moving from right to left lanes.

Table 2.3 presents vehicle characteristics by vehicle type for all periods. Approximately 56% of vehicles in the sample are cars, 13% are pickups, 27% are vans/SUVs, and 4% are heavy-duty vehicles. Car and light truck drivers appear similar in observable driving behavior. Cars, vans, and SUVs are similar in length (approximately 14.6 feet) while pickup trucks are about two feet longer on average. This difference in length, however, is not included in the following distance variable.

Table 2.2: Following Events Sample

	4:00-4:15 pm	5:00-5:15 pm	5:15-5:30 pm
Mean following distance (ft)	56.94	47.91	41.74
	(42.24)	(36.10)	(28.17)
Mean speed (mph)	16.23	12.77	11.30
	(5.29)	(5.76)	(5.08)
		Following Pair Type	
Car-Car	0.29	0.31	0.34
	(0.46)	(0.46)	(0.47)
Car-LT	0.21	0.21	0.22
	(0.41)	(0.41)	(0.41)
LT-Car	0.22	0.21	0.22
	(0.41)	(0.41)	(0.42)
LT-LT	0.15	0.16	0.16
	(0.36)	(0.37)	(0.36)
Car or LT-Heavy	0.05	0.04	0.03
	(0.21)	(0.20)	(0.16)
Heavy-Car or LT	0.07	0.06	0.03
	(0.25)	(0.24)	(0.18)
Heavy-Heavy	0.01	0.00	0.00
	(0.10)	(0.06)	(0.04)
		Following Event Characteristics	
Lane change	0.18	0.16	0.16
	(0.38)	(0.37)	(0.37)
Merging maneuver	0.04	0.05	0.05
	(0.21)	(0.22)	(0.22)
Number lane changes	0.83	0.80	0.87
	(1.16)	(1.23)	(1.26)
Made lane change	0.48	0.46	0.48
	(0.50)	(0.50)	(0.50)
Lane 2 (left lane)	0.15	0.16	0.14
	(0.36)	(0.36)	(0.35)
Lane 3	0.17	0.16	0.15
	(0.38)	(0.36)	(0.36)
Lane 4	0.20	0.18	0.20
	(0.40)	(0.38)	(0.40)
Lane 5	0.22	0.24	0.24
	(0.42)	(0.43)	(0.42)
Lane 6 (right lane)	0.25	0.27	0.27
	(0.44)	(0.44)	(0.44)
Following Events	3942	3213	3319
Unique Following Vehicles	1663	1356	1306

Notes: Mean values are reported with standard deviations in parentheses.

Table 2.3: Vehicle Characteristics

	Car	Pickup	Van/SUV	LT	Heavy
Primary lane	4	3	4	4	3
Made lane change	0.41	0.37	0.39	0.38	0.28
Number lane changes	0.63	0.54	0.61	0.58	0.45
Vehicle length	14.56	16.68	14.71	15.35	44.47
N	2407	557	1159	1716	202

Notes: Mean values are reported for all variables except primary lane for which the median value is reported.

2.4 Specification and Results

The basic specification is a linear model of following distance choice in following events involving particular vehicle types. The estimating equation is

$$\text{Following Distance}_i = \beta_1\text{Car-LT}_i + \beta_2\text{LT-Car}_i + \beta_3\text{LT-LT}_i + \boldsymbol{X_{1i}}\delta_1 + \boldsymbol{X_{2i}}\delta_2 + \epsilon_i. \quad (2.7)$$

The key explanatory variables are dummy variables for the following pair, where the omitted category is Car-Car. The coefficients, β_1, β_2, and β_3, therefore, measure the difference between the following distance of the indicated pair and the following distance of a car following a car. Dummy variables for following pairs involving heavy-duty vehicles are included in $\boldsymbol{X_{1i}}$ and $\boldsymbol{X_{2i}}$ represents a set of characteristics of the following event.

The set of following event characteristics include speed, lane choice, and lane changing behavior of the following driver. Speed enters the estimating equation quadratically because braking distance increases quadratically with speed (Brown et al., 2002). I also include the number of lane changes made by the following driver and a dummy variable for following events initiated by a lane changing or merging maneuver by the following driver. These variables, along with lane indicator variables, are included to control for driver aggressiveness.

I estimate the model separately for the 4:00-4:15 pm, 5:00-5:15 pm, and 5:15-5:30 pm periods because traffic conditions are changing quickly. As traffic congestion worsens, vehicle density increases and speeds decline, changing accident risk and potentially, the relative following behavior of drivers. At low congestion levels, drivers may not differentiate following distance by vehicle type because drivers are able to choose sufficiently large following distances. At higher congestion levels, drivers are likely more fearful of accidents and the effect of vehicle type on following distance may be larger. Under peak congestion, however, vehicles are densely packed together and speeds are lowest. At this time, drivers may simply follow all vehicles very closely. Because the data come from these three periods, I estimate the model separately for each period.

2.4.1 Effect of Vehicle Type on Following Distance

Table 2.4 presents results from estimating the linear model on the three samples of following events by OLS. Standard errors are clustered at the following vehicle level to account for correlation between observations pertaining to the same following vehicle. Alternatively, clustering standard errors at the lead vehicle level does not affect the conclusions.

The first row of Table 2.4 shows the difference in following distance between a car following a light truck and a car following a car. The results indicate that in the 5:00-5:15 pm period, cars follow light trucks at a 2.8 feet larger distance. At lower and higher levels of congestion during the 4:00-4:15 pm and 5:15-5:30 pm periods, cars are not found to follow light trucks and cars at statistically different distances. The second and third rows of Table 2.4 show that the LT-Car and LT-LT coefficients are positive in all periods, indicating that following pairs that have a light truck as the following vehicle have greater following distances than Car-Car pairs. The LT-Car coefficient is statistically significant in the 5:00-5:15 pm period while the LT-LT coefficient is statistically significant in all periods.

Table 2.4: Effect of Vehicle Type on Following Distance

	Dependent Variable: Following Distance (ft)		
follower-leader	4:00-4:15 pm	5:00-5:15 pm	5:15-5:30 pm
Car-LT	-0.522	2.799^{*}	-0.206
	(1.725)	(1.516)	(1.218)
LT-Car	1.987	3.702^{**}	1.182
	(1.617)	(1.578)	(1.344)
LT-LT	3.731^{**}	4.727^{**}	2.636^{*}
	(1.836)	(2.170)	(1.379)
Car or LT-Heavy	-3.542	9.933^{**}	3.531
	(2.692)	(3.956)	(3.001)
Heavy-Car or LT	36.130^{***}	28.644^{***}	38.829^{***}
	(5.364)	(4.193)	(5.726)
Heavy-Heavy	38.164^{***}	17.484^{***}	45.759
	(13.368)	(6.284)	(34.390)
Speed (mph)	-2.645^{***}	-0.504	1.086
	(0.647)	(0.520)	(0.780)
Speed^2	0.166^{***}	0.117^{***}	0.046
	(0.022)	(0.019)	(0.034)
Lane Change/Merge	-19.325^{***}	-16.854^{***}	-14.527^{***}
	(1.374)	(1.334)	(1.040)
# Lane Changes	-0.421	-0.311	0.023
	(0.688)	(0.566)	(0.473)
Lane Fixed Effects	Yes	Yes	Yes
Observations	3,942	3,213	3,319
R-squared	0.736	0.748	0.769

Notes: Standard errors clustered at the following vehicle level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Further testing of the coefficients is required to determine whether light trucks follow other vehicles at a greater distance than cars do, whether other drivers follow light trucks at a greater distance than cars, or if both statements are true. Table 2.5 presents the relevant tests with each test's F-statistic. I explain each comparison below.

Table 2.5: Effect of Light Truck as Following and Lead Vehicle

Difference	4:00-4:15 pm	5:00-5:15 pm	5:15-5:30 pm
LT-Car − Car-Car	1.987	3.702**	1.182
	(1.51)	(5.5)	(0.77)
LT-LT − Car-LT	4.253**	1.928	2.842*
	(4.41)	(0.66)	(4.06)
Car-LT − Car-Car	-0.522	2.799*	-0.206
	(0.09)	(3.41)	(0.03)
LT-LT − LT-Car	1.744	1.025	1.454
	(1.14)	(0.22)	(1.28)

Notes: In each row, the difference between the estimated coefficients indicated is shown with the F-statistic for the test that the difference is equal to zero in parentheses. *** p<0.01, ** p<0.05, * p<0.1

First, to determine whether light trucks follow other vehicles at a greater distance than do cars, I test the equality of coefficients for following pairs that have the same lead vehicle type but a different following vehicle type. In the first row of Table 2.5, the lead vehicle is a car and I compare the following distance maintained by cars and light trucks. This comparison can be made simply by using the estimated LT-Car coefficient because Car-Car is the omitted category. Next, I examine the following distance of cars and light trucks following light trucks by taking the difference between the estimated LT-LT and Car-LT coefficients. When comparing the following distance of cars and light trucks following the same lead vehicle type, light trucks are always estimated to follow the lead vehicle at a greater distance than cars. In addition, the difference in following distance is statistically significant in three of six cases.

Now, to determine whether drivers follow light trucks at a greater distance than they follow cars, I test the equality of coefficients for following pairs that have the same following vehicle type but a different lead vehicle type. In the first comparison, I compare the following distance chosen by cars following cars and light trucks. This comparison can be made using the estimated Car-LT coefficient because Car-Car is the omitted category. Then, to compare following distance chosen by light trucks following cars and light trucks, I take the difference between the estimated LT-LT and LT-Car coefficients. In this set of comparisons, the only statistically significant result is that cars follow light trucks at a greater distance in the 5:00-5:15 pm period. Light trucks are always estimated to follow other light trucks at a greater distance than they follow cars, but the difference is never statistically significant. Although the evidence is weaker, the data suggest that drivers follow light trucks at a greater distance than they follow cars.

2.4.2 Effects of Other Variables on Following Distance

The signs of the other coefficient estimates are as expected. Heavy-duty vehicles follow all vehicles more distantly than a car follows a car, and cars and light trucks follow heavy-duty vehicles at a greater distance in the 5:00-5:15 pm sample. Consistent with the effect of speed on braking distance, as speed increases, following distance increases at an increasing rate. At the mean speed, a 10% increase in speed leads to a 7% increase in following distance. Following distance after a lane change or merge is much smaller than in other following events. These following events have, on average, following distances between 14 and 19 feet smaller than other following events. The coefficient on the number of lane changes made by a driver is negative in the 4:00-4:15 pm and 5:00-5:15 pm periods although not statistically significant.

Table 2.6: Predicted Following Distance

	4:00-4:15 pm		5:00-5:15 pm		5:15-5:30 pm	
	Dist (ft)	Ratio	Dist (ft)	Ratio	Dist (ft)	Ratio
Car-Car	53.03	1.00	43.40	1.00	41.55	1.00
Car-LT	52.51	0.99	46.19	1.06	41.27	0.99
LT-Car	55.02	1.04	47.10	1.09	42.66	1.03
LT-LT	56.76	1.07	48.12	1.11	44.12	1.06
Car or Light truck - Heavy	49.49	0.93	53.33	1.23	45.01	1.08
Heavy - Car or Light truck	89.16	1.68	72.04	1.66	80.31	1.93
Heavy - Heavy	91.19	1.72	60.88	1.40	87.24	2.10

Notes: Each reported following distance is the average of the predicted following distance in each lane, weighted by the flow of vehicles. To predict following distance in each lane, mean values of speed and number of lane changes made are used.

2.4.3 Following Behavior at Different Congestion Levels

Table 2.6 presents the predicted following distance for each following pair type and the ratio of the pair's following distance to the Car-Car following distance. Comparing the results across the three time periods makes clear that relative following behavior varies across congestion levels. Cars follow light trucks more distantly only in the 5:00-5:15 pm period, and the additional following distance allowed by light trucks, compared to the Car-Car pair is greatest during this period as well. It seems likely that the severity of congestion in the 5:00-5:15 pm period, compared to the 4:00-4:15 pm period, induces drivers to fear accidents and choose following distance based on vehicle size. In the 5:15-5:30 pm sample, traffic congestion is at its peak and differences in following distance are moderate. Under extreme congestion, drivers may follow all vehicles closely because low speeds make accidents less severe. In addition, drivers may want to avoid being passed by other drivers and therefore follow all vehicles closely.

Finally, I consider all freeway space used by each following pair. Table 2.7 shows predicted headway by following pair type. A pair's headway is the sum of the pair's predicted following distance (from Table 2.6) and the average length of the lead vehicle type. By using vehicle

Table 2.7: Predicted Headway

	4:00-4:15 pm		5:00-5:15 pm		5:15-5:30 pm	
follower-leader	HW (ft)	Share	HW (ft)	Share	HW (ft)	Share
Car - Car	67.59	0.29	57.96	0.31	56.11	0.34
Car - Light truck	67.86	0.21	61.54	0.21	56.62	0.22
Light truck - Car	69.58	0.22	61.66	0.21	57.22	0.22
Light truck - Light truck	72.11	0.15	63.47	0.16	59.47	0.16
Car or Light truck - Heavy	93.96	0.05	97.80	0.04	89.48	0.03
Heavy - Car or Light truck	104.05	0.07	86.93	0.06	95.20	0.03
Heavy - Heavy	135.66	0.01	105.35	<0.00	131.71	<0.00

Notes: Each reported headway is the sum of the predicted following distance for the indicated following pair type and the sample average length of the lead vehicle type.

headway, all freeway space used by vehicles is taken into account. Table 2.8 shows the average headway for cars and light trucks as the following and lead vehicle. I take the average of the two values because each vehicle is both a following and lead vehicle. On average, light trucks use 4-6% more freeway space than cars.

2.5 Discussion of Driver Selection

The results presented in Section 4 demonstrate that (1) drivers of light trucks follow other vehicles at a greater distance than drivers of cars and (2) other drivers follow light trucks at a greater distance than they follow cars. Although the relationship is robust to the inclusion of controls for speed, lane choice, and lane change behavior, bias in the estimated coefficients remains a possibility. If drivers choose a vehicle type to drive or to follow based on unobserved characteristics that also affect following distance choice, the findings cannot be interpreted as the causal effect of light trucks on following distance. In this section, I explore driver selection into vehicle and lead vehicle type.

Table 2.8: Average Headway

	4:00-4:15 pm		
	Car Headway (ft)	Light Truck Headway (ft)	Ratio
As Following Vehicle	69.93	73.23	1.05
As Lead Vehicle	72.60	75.05	1.03
Average	71.26	74.14	1.04
	5:00-5:15 pm		
	Car Headway (ft)	Light Truck Headway (ft)	Ratio
As Following Vehicle	62.28	66.09	1.06
As Lead Vehicle	62.23	65.75	1.06
Average	62.25	65.92	1.06
	5:15-5:30 pm		
	Car Headway (ft)	Light Truck Headway (ft)	Ratio
As Following Vehicle	57.84	60.22	1.04
As Lead Vehicle	58.72	60.86	1.04
Average	58.28	60.54	1.04

Notes: The average headway as the following or lead vehicle is the average of all headways in which the indicated vehicle is the following or lead vehicle, weighted by the share of each following pair type. The average headway is the average of headway as following vehicle and headway as lead vehicle.

2.5.1 Driver Selection into Vehicle Type

Driver selection into vehicle type on unobserved characteristics has been recognized as a potential source of bias in several studies of the effect of vehicle type on traffic safety (White, 2004; Anderson, 2008; Li, 2012; Anderson and Auffhammer, 2014). Jacobsen (2013) explicitly models and estimates unobserved driver riskiness by vehicle type. His estimates indicate that unobserved risk is largest for large pickup truck drivers and smallest for minivan drivers. Those estimates, however, represent drivers in urban and rural areas at all times of day and may not reflect the riskiness of drivers in the I-80 sample. Looking at the raw data from the Fatal Accident Reporting System, Jacobsen (2013) notes that compact cars have the greatest relative frequency of single-vehicle fatal accidents and the pair of a compact car and a large pickup has the greatest relative frequency of two-vehicle fatal accidents in high income urban areas, like the San Francisco metropolitan area, during the day. The accident frequency data do not provide a clear picture of driver riskiness by vehicle type in areas similar to the study area but they do not reject the hypothesis that unobserved riskiness is similar for drivers of cars and light trucks in the I-80 sample.

Using the data from the I-80 sample, I analyze observed driving behavior by vehicle type to explore the relationship between driver risk and vehicle type. Figure 2.1 shows the distribution of primary travel lane by vehicle type for each period. In a chi-square test, statistical independence of vehicle type and primary lane cannot be rejected at the 10% significance level in the 4:00-4:15 pm and 5:00-5:15 pm periods. The data do not provide evidence that drivers of cars and light trucks choose to travel in different lanes. In the 5:15-5:30 pm period, independence is rejected with a p-value of 0.096. The results of the test, however, do not necessarily indicate a difference in aggressiveness between car and light truck drivers. Cars are overrepresented in lanes 3 and 6 while light trucks are overrepresented in lanes 2 and 4.[10] Therefore I group lanes 2 and 3 together as left lanes and lanes 4,5, and 6 as right

[10] Lane 2 is the leftmost lane while lane 6 is the right most lane.

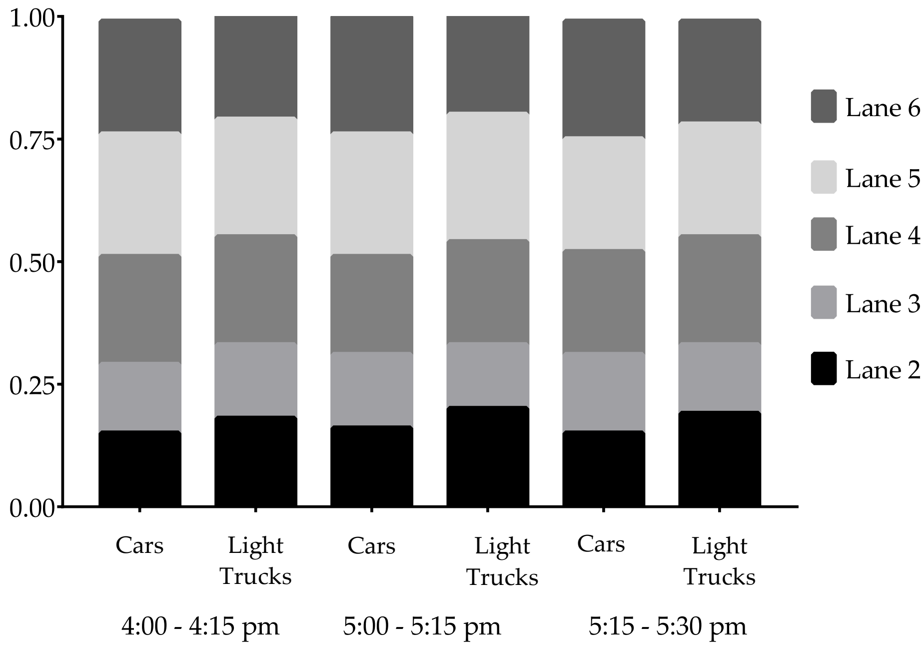


Figure 2.1: Primary Lane Distribution by Vehicle Type

lanes to test whether vehicle type is associated with choosing a left or a right lane. In a test of independence of vehicle type and primary lane (now grouped into left and right lanes) statistical independence cannot be rejected.

I also investigate differences in lane changing behavior across vehicle types. Table 2.9 shows the mean number of lane changes made and the proportion of drivers who made at least one lane change by vehicle type. The mean number of lane changes made is statistically different for car and light truck drivers in the 5:15-5:30 pm period, with car drivers making a greater number of lane changes. The proportion of drivers who made at least one lane change is statistically different for car and light truck drivers only in the 4:00-4:15 pm period with a larger proportion of car drivers making a lane change. Analysis of lane changing behavior shows some evidence that car drivers are more aggressive than light truck drivers in the 4:00-4:15 pm and 5:15-5:30 pm periods. This may indicate that the coefficients, LT-Car and LT-LT (following pairs in which a light truck is the following vehicle), are biased upward.

Table 2.9: Lane Change Behavior by Vehicle Type

	4:00 - 4:15 pm		5:00 - 5:15 pm		5:15 - 5:30 pm	
	Cars	LTs	Cars	LTs	Cars	LTs
Number Lane Changes	0.63	0.57	0.59	0.61	0.68	0.58
	(0.94)	(0.94)	(0.96)	(0.95)	(1.05)	(0.92)
Made Lane Change	0.41	0.37	0.39	0.40	0.43	0.39
	(0.49)	(0.48)	(0.49)	(0.49)	(0.50)	(0.49)

Notes: Mean values are reported with standard deviations in parentheses.

These differences between car and light truck drivers, however, appear small and are not found during the 5:00-5:15 pm period.

2.5.2 Driver Selection into Lead Vehicle Type

A second selection effect with the potential to bias the coefficient estimates is driver selection into lead vehicle type based on unobserved characteristics. This type of selection occurs if drivers choose to follow a certain vehicle type based on unobserved characteristics that also affect following distance choice. To examine driver selection into lead vehicle type, I estimate the main specification using subsamples of following events in which the following and lead vehicles were arguably randomly paired together. As shown in Figure 2.2, consider three vehicles traveling in a single lane: vehicle 1 leads followed by vehicle 2, and then vehicle 3. Now, for some reason, vehicle 2 changes lanes, leaving vehicle 3 to follow vehicle 1. The *Switch* subsample consists of following pairs formed in the way that the pair between vehicles 1 and 3 was formed. I also create a *Merge* subsample of following events, which includes only following events in which the following or lead vehicle has just merged with traffic. Drivers entering at the on-ramp have not yet been able to maneuver through traffic and have little choice of the vehicle to follow immediately upon leaving the on-ramp. Table 2.10 presents the results from estimating the linear model of following distance using the Switch and Merge subsamples. Although the subsamples are small and the estimation results are

Table 2.10: Switch and Merge Subsamples: Effect of Vehicle Type on Following Distance

follower-leader	4:00-4:15 pm		
Sample	*Full*	*Switch*	*Merge*
Car-LT	-0.522	8.045	8.217*
	(1.725)	(5.443)	(4.784)
LT-Car	1.987	-0.058	9.617***
	(1.617)	(4.630)	(3.074)
LT-LT	3.731**	5.839	8.980**
	(1.836)	(4.970)	(4.008)
	5:00-5:15 pm		
Car-LT	2.799*	10.714***	3.391
	(1.516)	(4.053)	(2.055)
LT-Car	3.702**	5.556	2.941
	(1.578)	(3.852)	(2.392)
LT-LT	4.727**	4.951	3.670
	(2.170)	(5.166)	(3.105)
	5:15-5:30 pm		
Car-LT	-0.206	-1.538	-3.241
	(1.218)	(3.110)	(2.323)
LT-Car	1.182	0.979	-1.191
	(1.344)	(3.797)	(2.590)
LT-LT	2.636*	0.450	2.893
	(1.379)	(3.199)	(2.838)

Notes: All variables used in main specification included. Standard errors clustered at the following vehicle level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

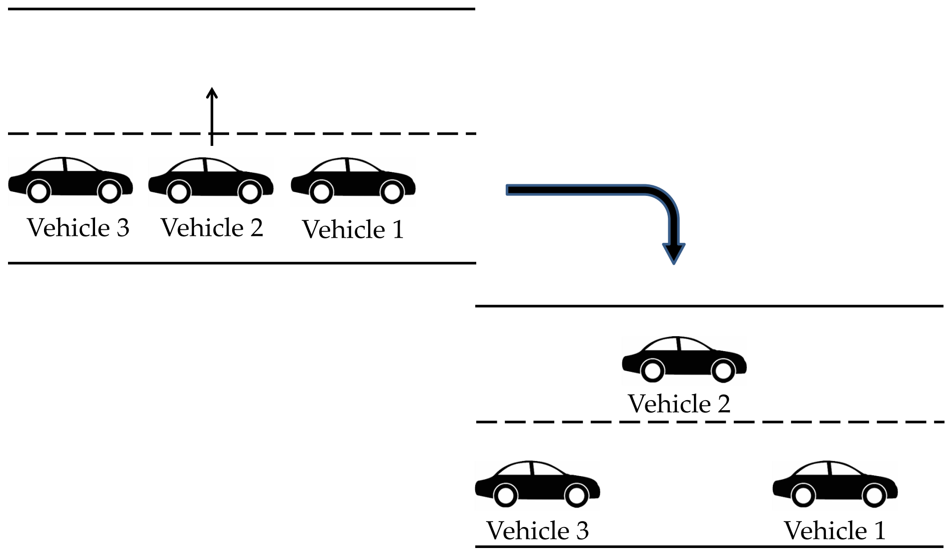


Figure 2.2: Switch Subsample

less precise than the full sample estimates, the subsample provides useful insight into the full sample results. In contrast to the full sample, the estimated coefficient, Car-LT, is positive and very large in both subsamples and statistically significant in the Merge subsample for the period 4:00-4:15 pm. This finding indicates downward bias in the full sample Car-LT coefficient due to driver selection into lead vehicle type. When traffic is less congested, as it is from 4:00-4:15 pm, drivers can more easily change lanes and cautious car drivers may avoid following light trucks perhaps because light trucks obstruct view of traffic. More aggressive car drivers are left to follow light trucks and the Car-LT coefficient is biased downward. In the 5:00-5:15 pm Switch and Merge subsamples, the Car-LT coefficient estimate retains its positive sign and is statistically significant in the Switch subsample. The robustness of this result in the subsamples suggests that driver selection into lead vehicle type is not a source of bias in the full sample Car-LT coefficient estimate. As in the full sample, the Car-LT coefficient is not statistically significant in either subsample in the 5:15-5:30 pm period.

The subsample results for drivers of light trucks are similar to the full sample results. The difference between the LT-LT and LT-Car coefficients is given in the third row from the bottom of Table 2.10. As in the full sample, the LT-Car and LT-LT coefficients are not statistically different in any period in either subsample but it is not always the case that the LT-LT coefficient is larger than the LT-Car coefficient. Still, the hypothesis that light trucks follow other light trucks at a greater distance than they follow cars is within the range of the estimates.

2.6 Implications for Congestion Costs

The econometric evidence demonstrates that relative to cars, light trucks follow other vehicles and are followed by other vehicles at a greater distance. When both following distance and vehicle length are taken into account, light trucks use 4-6% more freeway space than cars. This section examines the effect of light trucks on traffic congestion. I use a bottleneck congestion model to analyze traffic congestion because, as I show below, congestion in the study area best fits this model. Then, I consider the effect of a policy that replaces all light trucks with cars on congestion.

A traffic bottleneck occurs where there is a reduction in capacity so that the inflow of vehicles to a freeway segment exceeds its capacity. The segment of reduced capacity is called the traffic bottleneck. When the bottleneck's capacity is exceeded, drivers must wait to pass through the bottleneck. The implication of a change in freeway space usage by vehicles is straightforward. If vehicles use less freeway space, more vehicles will be able to pass through the bottleneck in a given amount of time. This increase in effective bottleneck capacity reduces delay for drivers. Below, I present evidence that traffic congestion in the study area is characterized by bottleneck congestion.

Traffic speed and flow data from freeway detectors in the area show that during the period of vehicle trajectory data collection, traffic is hypercongested. Figure 2.3 shows lane speed and

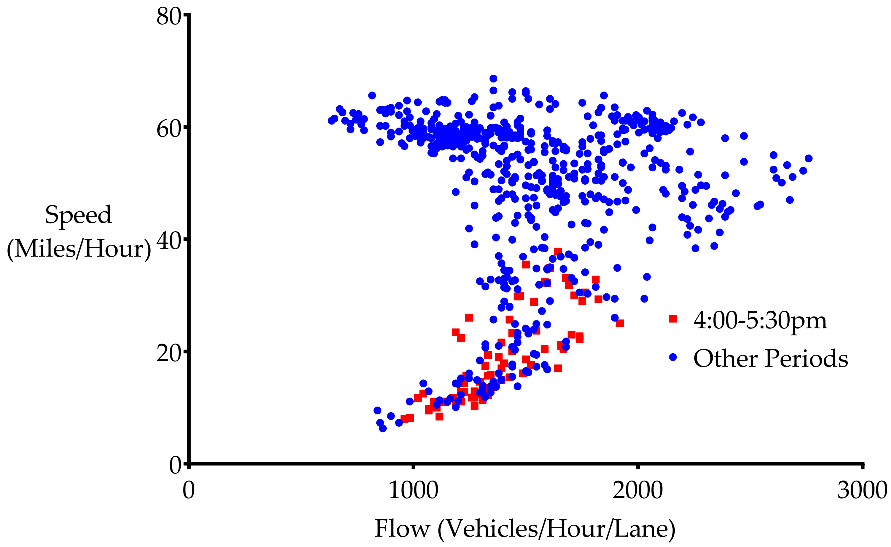


Figure 2.3: Traffic Speed and Flow

flow observations from the study day with the observations from the period of vehicle trajectory data collection, 4:00-5:30 pm, highlighted.[11] Consistent with the standard speed-flow curve, the observations form two distinct branches. On the upper, congested branch, flow increases as speed decreases because more vehicles are using the road. The highlighted observations fall on the lower, hypercongested branch where flow increases as speed increases.[12] Because hypercongested conditions often result from a traffic bottleneck (Small and Chu, 2003), I look for evidence that a traffic bottleneck exists.

Spatial analysis shows a large speed increase near the downstream I-80/I-580 interchange called the Hoffman Split. Figure 2.4 shows traffic speed in lane 3 at 4pm by detector post mile. The study area is closest to the detector at post mile 9.73 and the interchange is located near post mile 12. The graph shows that traffic speed hovers around 20 mph until the first traffic detector after the interchange. Beyond the interchange, speed increases to 54 mph. Such a large increase in speed indicates that a traffic bottleneck releases between the two detectors (California Department of Transportation, 2014). The slow moving vehicles

[11]These data come from the California Department of Transportation Performance Measurement System.
[12]In traffic engineering, the upper branch is called uncongested and the lower branch is called congested.

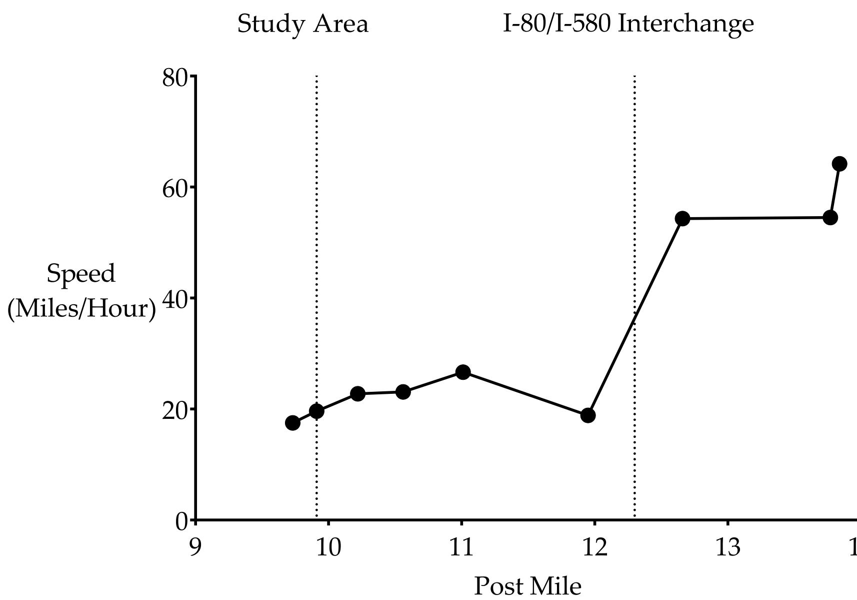


Figure 2.4: Lane 3 Speed at 4:00pm

at the first detector are traveling through the bottleneck while the vehicles at the second detector have exited the bottleneck and are traveling at nearly free-flow speeds.

Given this evidence, a bottleneck congestion model is appropriate. In the model, vehicles arrive at a bottleneck (with capacity V_K) at a constant rate, V_a, during the time interval $[t_p, t_{p'}]$.[13] If the arrival rate of vehicles exceeds the bottleneck capacity, vehicles wait in a queue to pass through the bottleneck. The average delay is given by $\frac{1}{2}\left(t_{p'} - t_p\right)\left(\frac{V_a}{V_K} - 1\right)$ if $V_a > V_K$. To apply the model, I find the average delay and the duration of the congested period from the traffic data and solve for the ratio, $\frac{V_a}{V_K}$. I then adjust this ratio to simulate the increase in capacity that would result from a replacement of light trucks with cars. I compare the resulting delay with the original delay to estimate the effect of light trucks on traffic congestion.

[13]The model described is called bottleneck congestion with a fixed peak period. Further information can be found in Small and Verhoef (2007).

The traffic data show that the bottleneck capacity is exceeded from 3-6pm and the average delay (relative to 65 mph) during that period is 6.8 minutes.[14] Therefore, the ratio, $\frac{V_a}{V_K}$, is 1.076, meaning the arrival rate of vehicles is 7.6% greater than the bottleneck capacity.

To determine the capacity change that would result from a replacement of light trucks with cars, I use the headway results from Table 2.7 along with the shares of each vehicle type. In the sample, light trucks make up 40% of vehicles and over the three periods use an average of 4.67% more space than cars. Therefore, a replacement of all light trucks with cars would increase the bottleneck capacity by 1.87% (0.40×0.0467). The ratio, $\frac{V_a}{V_K}$, would fall to 1.056 and the average delay would decrease by 26% to 5.03 minutes. In total, the delay would be reduced by 495 vehicle-hours per day and 126,389 vehicle-hours per year.[15] Using the 2014 value of travel time employed by the Texas Transportation Institute for the San Francisco metropolitan area ($17.67/hour), the annual time savings possible during the afternoon peak period on this freeway segment are valued at $2,233,290 or $333 per light truck replaced.

The congestion effect of light trucks appears to be economically significant. The estimated annual time savings benefit is of the same magnitude as a previously estimated safety benefit of replacing light trucks with cars. White (2004) studies the external accident costs imposed by light trucks and estimates the annual benefit of such a policy to be $520 per light truck replaced.[16] Because the time savings estimate only accounts for the afternoon peak period, the total value of time savings is likely very similar to the estimated safety benefit.

The estimates show that, in the study area, light trucks impose considerable travel time costs on other vehicles. It is important, however, to determine whether these external congestion costs are likely to be large on other freeways. In the study area, traffic is hypercongested because of a traffic bottleneck. Therefore, light trucks are expected to impose large conges-

[14]The delay is calculated for the freeway segment between the MacArthur Maze and the Hoffman Split.

[15]The traffic flow over this time period (3-6pm on the sample day) was 16,765 vehicles and I assume the delay affects 255 working days per year.

[16]This value is the estimate from the original paper ($447) adjusted for inflation to 2014 US dollars.

tion costs on other freeways with similar traffic conditions. In California, traffic bottlenecks are the cause of approximately 49% of total delay faced by drivers (California Department of Transportation, 2014). Traffic bottlenecks, which induce hypercongested conditions, are clearly a major source of delay. The results, therefore, are directly relevant not only for the particular freeway under study, but for other hypercongested freeways as well. Further study using data from less congested conditions is needed to determine whether vehicle size affects traffic flow at higher speeds.

2.7 Conclusion

The share of light trucks in the US vehicle fleet has greatly expanded over the past three decades. Light trucks now make up nearly half of vehicle registrations (Federal Highway Administration, 2015). Using detailed vehicle trajectories, I show that compared to cars, light trucks are followed by other vehicles and follow other vehicles more distantly. This finding is robust in tests of selection bias. On average, light trucks use 4-6% more road space than cars. Using a bottleneck congestion model, I find that the average delay on a California freeway in the afternoon peak period could be reduced by 26% if all light trucks were replaced with cars. Such a policy would have an annual benefit of $333 per light truck.

In light of the large external congestion costs imposed by light trucks, policies to internalize this negative externality should be considered. The first-best policy to internalize congestion costs is to impose congestion tolls which charge drivers for the marginal external time cost they impose while driving (Small and Verhoef, 2007). The evidence suggests that a differentiated congestion toll for cars and light trucks would be optimal because the external time costs imposed by each vehicle type differ. If congestion tolls are not feasible, second-best policies that reduce the number of light trucks purchased should be considered. Several European countries tax vehicle weight at the time of purchase or set vehicle registration fees based on vehicle weight (Greven, 2015). Compared to the United States, these countries also have higher gasoline taxes which discourage the purchase of vehicles with low fuel economies, often large vehicles. The evidence that light trucks impose greater external congestion costs adds to previous findings that light trucks impose larger accident costs and further justifies the use of these policies.

Finally, the results are important for assessing the consequences of any policy that affects the composition of the vehicle fleet. Whitefoot and Skerlos (2012) show that fuel economy standards based on a vehicle's footprint create an incentive for manufacturers to increase vehicle size, and this incentive is larger for light trucks than cars. The authors point out that these incentives undermine fuel economy gains and may increase traffic safety risks by increasing vehicle size heterogeneity. In addition to those concerns, this chapter shows that congestion costs are likely to increase as a result of the incentives created by this policy.

Chapter 3

Driver (Mis)Perception of Speed in Congested Traffic

3.1 Introduction

Changing lanes in congested traffic is a demanding and dangerous task for drivers. The driver must decide whether vehicle spacing is sufficient to allow a change, and during the change, the vehicle straddles two streams of traffic. Not surprisingly, lane changes often have negative consequences for surrounding drivers. A disproportionately large share of accidents involves a lane change, and lane changes also form and propagate traffic speed oscillations, which are better known as stop and go traffic (Sen et al., 2003; Ahn and Cassidy, 2007; Zheng, 2014). Because these costs are external, drivers do not consider them when changing lanes but weigh only the private costs and benefits of the decision. For many drivers, the lane change decision involves weighing accident costs against a speed gain. In heavily congested traffic, however, a driver may not be able to accurately judge her average speed relative to the average speed of an adjacent lane. This is particularly problematic if the driver underestimates her own speed relative to the speed in an adjacent lane. In this case, she overestimates the speed gain possible and may make a lane change that is neither privately nor socially optimal.

Consider two heuristic techniques drivers might use to judge relative speed. Under the first heuristic, called the vehicle count heuristic, the driver notes how many vehicles she passes and how many vehicles pass her in a comparison lane over some time period. If the number of vehicles she has passed is greater than the number of vehicles that have passed her vehicle, she concludes that her average speed is higher than the average speed in the comparison lane. This heuristic is similar to choosing a reference vehicle in the comparison lane and continuously making position comparisons. In fact, this heuristic fails only in certain special cases.

Under a second heuristic, called the interval count heuristic, a driver regularly glances at the comparison lane and notes whether she has passed or has been passed by any vehicle in the time interval since her last glance. This driver compares the number of intervals spent passing other vehicles to the number of intervals spent being passed by other vehicles. If the number of intervals she has spent passing is greater than the number of intervals she has spent being passed, she concludes that her average speed is higher than the average speed in the comparison lane.

Although drivers may assume that the vehicle count heuristic and the interval count heuristic are functionally equivalent, the conclusions drawn from each differ in a systematic way. Compared to the vehicle count heuristic, the interval count heuristic underestimates a driver's speed relative to the comparison lane.

The discrepancy between the two heuristics arises because of optimizing behavior by drivers in following distance choice. As speed increases, vehicle stopping distance also rises, and drivers choose greater following distances to reduce accident costs. As a result, when traveling at high speeds, vehicles are spread apart and when traveling slowly, vehicles are close together. If time is measured in discrete intervals, the time it takes for a slow vehicle to be passed by a certain number of fast vehicles is longer than the time it takes for a fast vehicle to pass the same number of slow vehicles. Therefore, it is possible for a driver to pass more vehicles than she is passed by but to spend more time (measures in intervals) being passed by other vehicles than passing other vehicles. Using the vehicle count heuristic, the driver would conclude she has a higher average speed than the other lane but using the interval count heuristic, she would conclude that her average speed is lower than the other lane. Because of the relationship between speed and following distance, the case described above, in which the driver underestimates her relative speed using the interval count heuristic is much more likely to occur than the case in which she overestimates her speed.

This chapter uses a data set of observed vehicle trajectories from congested freeways to build on work by Redelmeier and Tibshirani (2000) that first described and analyzed the above phenomenon using simulated vehicle trajectories. Using the vehicle trajectories, the hypothesized asymmetry is documented. Overall, the interval count heuristic underestimates relative speed compared to the vehicle count heuristic in 18% of cases while overestimating relative speed in only 1% of cases. The analysis also shows that the asymmetric discrepancy between the two heuristics is more pronounced in heavily congested traffic.

The findings suggest that a driver is likely to overestimate the speed gain from changing lanes if she relies on the interval count heuristic. She is also more likely to do so in heavily congested traffic. Because of the external costs associated with lane changes, many privately optimal lane changes are likely not socially optimal. But if drivers overestimate speed gains, many lane changes may not be privately optimal as either. Including information about the external costs of lane changing as well as reliable methods for judging relative speed in driver education programs could have safety benefits and reduce congestion.

In addition to building directly on Redelmeier and Tibshirani (2000), this work adds to a literature that examines the cognitive processes of drivers during the lane change decision. In this strand of the traffic engineering literature, duration models (Hamdar and Mahmassani, 2009) as well as fuzzy logic models (McDonald et al., 1997) of lane changing have been developed.

This chapter is organized as follows. Section 2 more rigorously explains the vehicle count and interval count heuristics. Section 3 describes the data used to evaluate the heuristics, and Section 4 presents the analysis. Section 5 concludes.

3.2 Theoretical Background

The vehicle count heuristic and interval count heuristic are two methods drivers might use to assess average speed relative to another lane in traffic. Using the vehicle count heuristic, a driver compares the number of vehicles she has passed to the number of vehicles she has been passed by in the other lane over some period of time. If the number of vehicles she has passed is greater than the number of vehicles she was passed by, the driver concludes she is traveling at a higher average speed than the other lane. If she has been passed by more vehicles than she has passed, she concludes her average speed is lower. Using the interval count heuristic, a driver applies the same logic in a comparison of the amount of time (measured in intervals) she has spent passing other vehicles and being passed by other vehicles over the time period. Although these heuristics appear similar, optimizing behavior by drivers in following distance choice causes their conclusions to diverge in a systematic way.

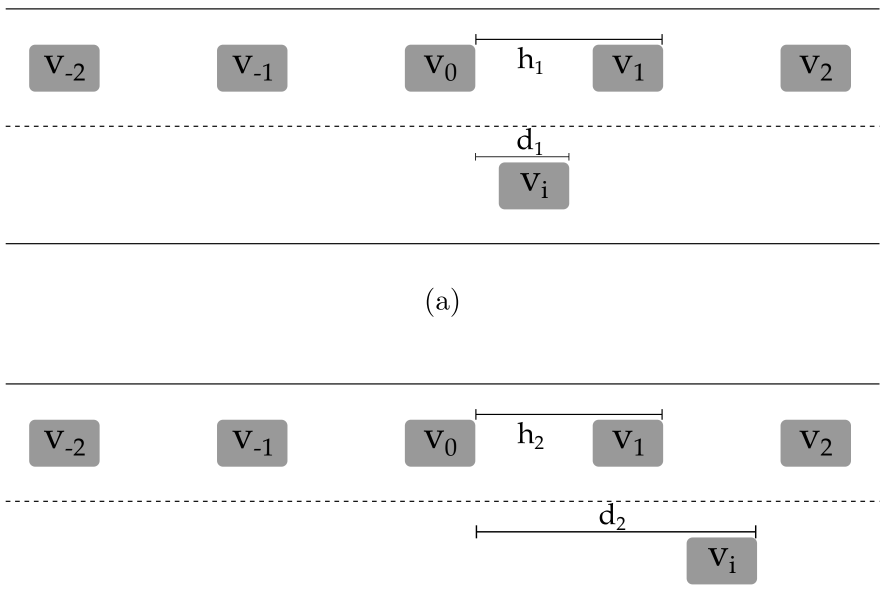


Figure 3.1: (a) Beginning of interval (b) End of interval

3.2.1 Vehicle Count Heuristic

Imagine a driver (called the index driver) who wants to compare her average speed (S_{index}) to the average speed of a comparison lane over a period of time. The top panel of Figure 3.1 shows the index driver's vehicle (v_i) and the position of vehicles in the comparison lane at the beginning of the time period. V_0 and V_1 are defined such that the index vehicle's longitudinal position is between the longitudinal positions of these vehicles at the start of the period. Vehicle numbering increases in the direction of traffic and decreases in the opposite direction. The distance between the index vehicle and V_0 at the start of the period is d_1. The space headway, the distance between two vehicles in the comparison lane, at the start of the period is h_1. The average speed of the other lane over the interval is estimated by the average speed of V_0 (S_0). S_0 is the average speed the index driver could achieve if she were traveling in the comparison lane.

The bottom panel of Figure 3.1 shows the index vehicle's position at the end of the time period. The distance between the index vehicle and V_0 is given by d_2. In this example, d_2 is positive because the index vehicle ends the period ahead of V_0. Unlike d_1, however, d_2 may be positive or negative depending on whether the index vehicle ends the period ahead or behind V_0. The space headway at the end of the period is h_2.

During the interval, the speed of each lane varies. For example, suppose the index vehicle is traveling faster than the vehicles in the comparison lane at the beginning of the period. The index vehicle passes three vehicles (V_1, V_2, and V_3). Later, the index vehicle travels slower than the vehicles in the comparison lane. At that time, V_3 and V_2 pass the index vehicle. Net vehicles passed by the index driver during the period (number of vehicles index driver passed − number of vehicles index driver was passed by) is 1. As you can see in the bottom panel of Figure 3.1, the index vehicle is just ahead of V_1. If there is no lane changing, net vehicles passed is always equal to the number of the vehicle just behind the index driver in the comparison lane at the end of the interval.[1]

The average speed of each vehicle is given by the distance the vehicle traveled divided by the length of the time period. Because the index vehicle and V_0 are compared over the same time period, the vehicle that traveled a greater distance has a higher average speed. Therefore, if the index vehicle ends the interval farther ahead of V_0 (if $d_2 > d_1$), the index vehicle traveled a greater distance and has a higher average speed over the period. If the index vehicle ends the interval not as far ahead of V_0 or behind V_0 (if $d_1 > d_2 > 0$ or $d_1 > 0 > d_2$), the index vehicle did not travel as far as V_0 and has a lower average speed.

Now, consider net vehicles passed. Net vehicles passed will be positive if $d_2 > h_2$, negative if $d_2 < 0$, and zero if $h_2 > d_2 > 0$. Table 3.1 shows all possible outcomes for which $d_1 \neq d_2$.

[1]It is important to note that net vehicles passed may not equal the number of the vehicle just behind the index driver if vehicles enter or exit the comparison lane. When this happens, the vehicle count heuristic may not give a reliable conclusion for relative speed.

Table 3.1: Possible Outcomes for Net Vehicles Passed and Speed Difference

	net vehicles passed >0 $d_2 > h_2$	net vehicles passed $= 0$ $h_2 > d_2 > 0$	net vehicles passed <0 $d_2 < 0$
index faster $d_2 > d_1$	$d_2 > h_2 > d_1$ $d_2 > d_1 > h_2$	$h_2 > d_2 > d_1$	
index slower $d_1 > d_2$	$d_1 > d_2 > h_2$	$d_1 > h_2 > d_2 > 0$ $h_2 > d_1 > d_2 > 0$	$d_1 > h_2 > 0 > d_2$ $h_2 > d_1 > 0 > d_2$

If net vehicles passed is positive, there are three possible orderings of d_1, d_2, and h_2. In the first two orderings, $d_2 > d_1$ indicating the index driver has a higher average speed than V_0. In the third ordering, $d_1 > d_2$ indicating V_0 has a higher average speed. In this case, the vehicle count heuristic would lead the driver to the wrong conclusion. If net vehicles passed is negative, there are only two possibilities, both of which indicate V_0 has a higher average speed than the index driver. Finally, when net vehicles passed is zero, there are three possible orderings. In two of the cases, V_0 has a higher average speed and in one the index driver has a higher average speed.

Overall, the conclusions drawn from this method are largely correct. If the driver has been passed by more vehicles than she has passed, she has a lower average speed over the time period. If the driver has passed more vehicles than she has been passed by she has a higher average speed in two out of three cases. If she has passed and been passed by an equal number of vehicles, she may have a higher or lower average speed but it is likely that her average speed is similar to V_0's average speed.

3.2.2 Interval Count Heuristic

A driver, intuitively grasping the logic of the vehicle count heuristic, may naturally jump to the interval count heuristic, assuming that if she has spent more time passing other vehicles than being passed by other vehicles she is traveling at a higher average speed than the comparison lane. The two heuristics, however, are only equivalent if, on average, the number of vehicles a driver passes during a passing interval and the number of vehicles a driver is passed by during an interval in which she is passed (a being passed interval) are equal. In vehicular traffic, this is not the case. A driver is likely to pass more vehicles in a passing interval than the number of vehicles she is passed by in a being passed interval. Therefore, if a driver passes and is passed by the same number of vehicles, she spends more time being passed by other vehicles than passing other vehicles. Drivers using the interval count heuristic systematically underestimate relative speed. This asymmetric discrepancy between the heuristics arises because of optimizing behavior by drivers in following distance choice. In trading off time and accident costs, drivers increase following distance as speed increases.

In this section, two cases are examined. In the first case, similar to train cars on two adjacent tracks, vehicle spacing is constant for all speeds. Under this case, the heuristics lead to the same conclusion. In the second case, similar to vehicles in adjacent lanes, vehicle spacing increases with speed. Under the second case, the interval count heuristic underestimates relative speed compared to the vehicle count heuristic.

Vehicle Spacing Constant for all Speeds

Suppose there are two lanes of traffic, moving at different speeds. The speed in the fast lane is S_{fast} and the speed in the slow lane is S_{slow}. Speed is measured in feet per second. Assume all vehicles have the same length and that drivers maintain the same following distance at every speed. Therefore, the space headway for vehicles in the fast lane is equal to the space headway for vehicles in the slow lane ($h_{fast} = h_{slow}$) and is measured in feet. Figures 3.2a and 3.2b show the positions of all vehicles at the start and end of a one second time interval. At the conclusion of the interval, each vehicle in the fast lane has passed one slow vehicle and each vehicle in the slow lane has been passed by one fast vehicle. The speed difference between the lanes determines how many slow vehicles each fast vehicle passes and how many

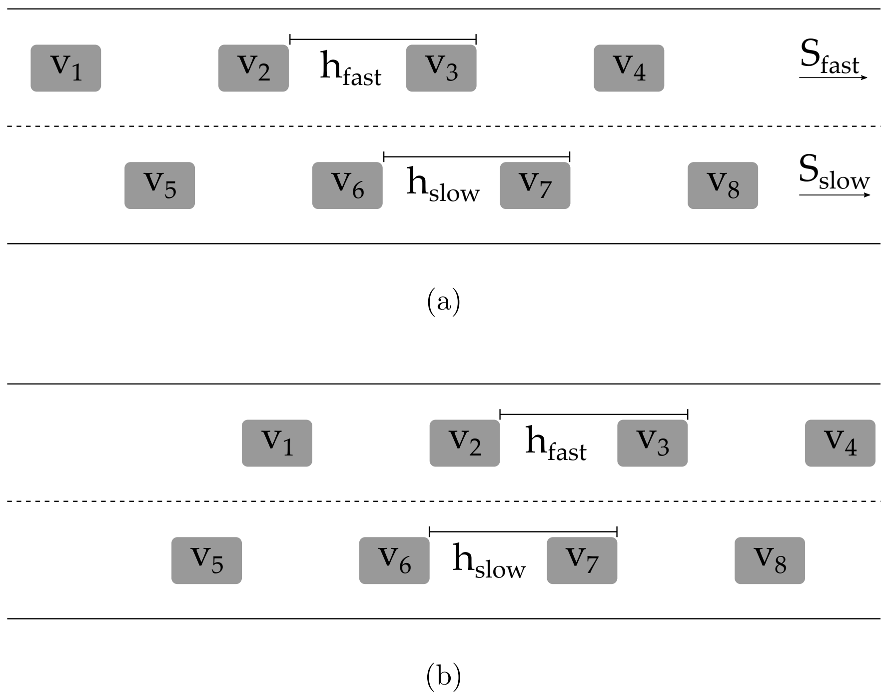
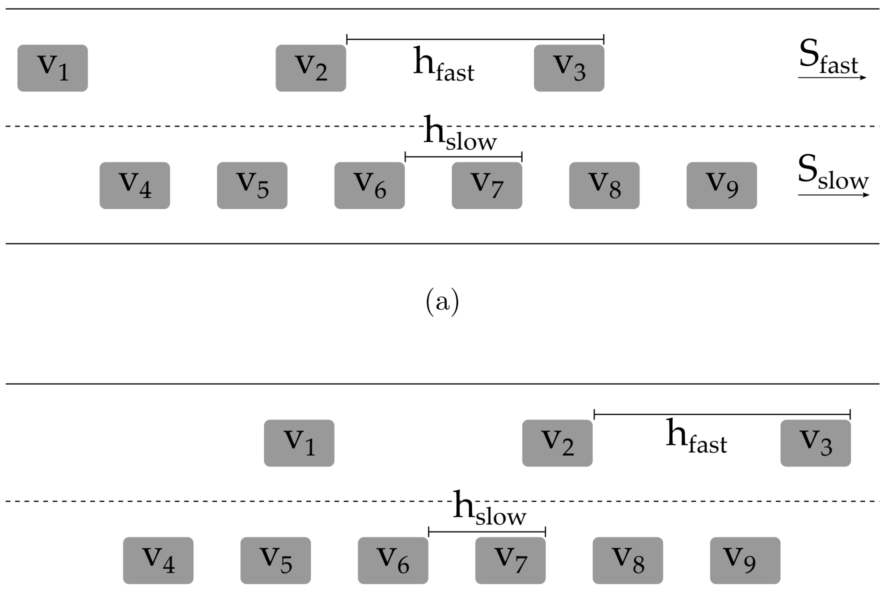


Figure 3.2: (a) Beginning of interval (b) End of interval

fast vehicles each slow vehicle is passed by but, because the space headway is the same at fast and slow speeds, the two quantities must be equal.

To see that this is the case, note that for a fast vehicle to pass k slow vehicles, $S_{fast} \geq S_{slow} + kh_{slow}$. To simplify, imagine $S_{slow} = 0$. If a fast vehicle travels a distance of kh_{slow} feet in a one second interval it will pass k slow vehicles and its speed is kh_{slow} feet/second by definition. Rearranging the equation above, a fast vehicle passes $\frac{S_{fast}-S_{slow}}{h_{slow}}$ vehicles each interval. Similarly, for a slow vehicle to be passed by k fast vehicles, $S_{fast} \geq S_{slow} + kh_{fast}$. Again, imagine $S_{slow} = 0$. A slow vehicle is passed by k fast vehicles in a one second interval if each fast vehicle travels a distance of kh_{fast}, meaning the speed of the fast vehicles is kh_{fast} feet/second. Rearranging, a slow vehicle is passed by $\frac{S_{fast}-S_{slow}}{h_{fast}}$ vehicles each interval. Because $h_{fast} = h_{slow}$, the number of slow vehicles a fast vehicle passes in one interval is always equal to the number of fast vehicles a slow vehicle is passed by in one interval.



(a)

(b)

Figure 3.3: (a) Beginning of interval (b) End of interval

Therefore, if vehicle spacing is constant for all speeds, a driver who has passed and has been passed by the same number of vehicles also spends the same number of intervals passing and being passed. Similarly, a driver who has passed more vehicles than she was passed by spends more intervals passing than being passed, and a driver who has been passed by more vehicles than she has passed spends more intervals being passed than passing other vehicles. If vehicle spacing is constant for all speeds, the vehicle count and interval count heuristics yield the same conclusion.

Vehicle Spacing Increases with Speed

Now, suppose that vehicles in the fast lane maintain a larger space headway than vehicles in the slow lane ($h_{fast} > h_{slow}$). Figures 3.3a and 3.3b show the positions of the vehicles at the start and end of a one second time interval. At the conclusion of the interval, each vehicle in the fast lane has passed two slow vehicles while each vehicle in the slow lane has only been passed by one fast vehicle. Because the space headway is larger in the fast lane than in the slow lane, each fast vehicle passes more slow vehicles than the number of fast vehicles each slow vehicle is passed by.

As in the case of constant spacing, the number of vehicles a fast vehicle passes in one interval is $\frac{S_{fast}-S_{slow}}{h_{slow}}$ and the number of vehicles a slow vehicle is passed by is $\frac{S_{fast}-S_{slow}}{h_{fast}}$. In this case, however, $h_{fast} > h_{slow}$ which means that for a given speed difference a fast vehicle passes more slow vehicles than the number of fast vehicles a slow vehicle is passed by.

Therefore, in vehicular traffic where vehicle spacing increases with speed, a driver who has passed and been passed by the same number of vehicles spends more intervals being passed than passing other vehicles. In addition, a driver who has passed more vehicles than she was passed by may also spend more time being passed by other vehicles than passing. In these two cases, a driver who relies on the interval count heuristic concludes she is traveling at a lower average speed than the comparison lane. In a third case, a driver who has been passed by more vehicles than she passed also spends more time being passed than passing other vehicles. This driver reaches the same conclusion using the vehicle count and interval count heuristics but she likely believes she is traveling even more slowly if she uses the interval count heuristic instead of the vehicle count heuristic. Using the interval count heuristic, these drivers underestimate speed relative to the comparison lane.

3.3 Data

The data set used to analyze each heuristic contains trajectories from vehicles traveling on two California freeways. The data were collected as part of the Federal Highway Administration's Next Generation Simulation (NGSIM) program. To collect the data, freeway traffic

Table 3.2: Traffic Flow and Speed

	Flow (vehicles/hour/lane)	Average vehicle speed (mph)			
		Mean	SD	Min	Max
I-80	1262	16.92	9.06	4.30	53.56
US-101	1356	22.54	6.31	10.98	43.64
Both Fwys	1309	19.83	8.25	4.30	53.56

was videotaped and later vehicle trajectories were transcribed using vehicle tracking software (Federal Highway Administration, 2008).

The trajectory data were collected on Interstate 80 in Emeryville, CA and US Route 101 in Los Angeles, CA. Data collection took place on I-80 eastbound during the afternoon peak period (4:00-4:15 pm and 5:00-5:30 pm) on April 13, 2005 and on US-101 southbound during the morning peak period (7:50 - 8:35 am) on June 15, 2005. The I-80 freeway segment is 1,650 feet (approximately 0.3 miles) in length and the US-101 segment is 2,100 feet (approximately 0.4 miles) in length. Each freeway has five mainline travel lanes on the observed segment.[2]

The trajectory data include each vehicle's position, speed, and lane of travel recorded every 0.1 seconds. Table 3.2 presents basic summary statistics for vehicles traveling on each freeway. Average vehicle speed is quite low on both freeways but the I-80 freeway is more severely congested than the US-101 freeway and has lower traffic flow and speed.

The vehicle count and interval count heuristics are applied to these data after the data are organized into comparisons drivers would make. First, each driver's trajectory is divided into episodes. An episode begins when the driver enters a lane and ends when the driver exits the lane. This driver is the index driver. Drivers who do not change lanes are involved in one episode while drivers who have traveled in multiple lanes are involved in multiple episodes. For each episode, the number of vehicles the index driver passes and the number of vehicles the index vehicle is passed by in each lane are recorded separately for each lane.

[2] I-80 also has one high-occupancy vehicle lane and one on-ramp and US-101 also has one on-ramp, one auxiliary lane, and one off-ramp on the observed segment.

Table 3.3: Driver Comparison Sample

	Mean	SD	Min	Max
Length of episode (seconds)	65.56	29.92	4	200
Index driver average speed (mph)	16.01	6.49	1.07	42.49
US-101 episode	0.52	0.50	0	1
Observations	26,651			

To apply the interval count heuristic, time is divided into two second intervals. The number of intervals the index driver spends passing other vehicles and the number of intervals she spends being passed by other vehicles in every lane are recorded.

To assess the accuracy of the heuristics, the average speed in each lane must be estimated. The driver's average speed and, for each lane, the average speed of the vehicle in that lane that begins the episode closest to the index driver is calculated.

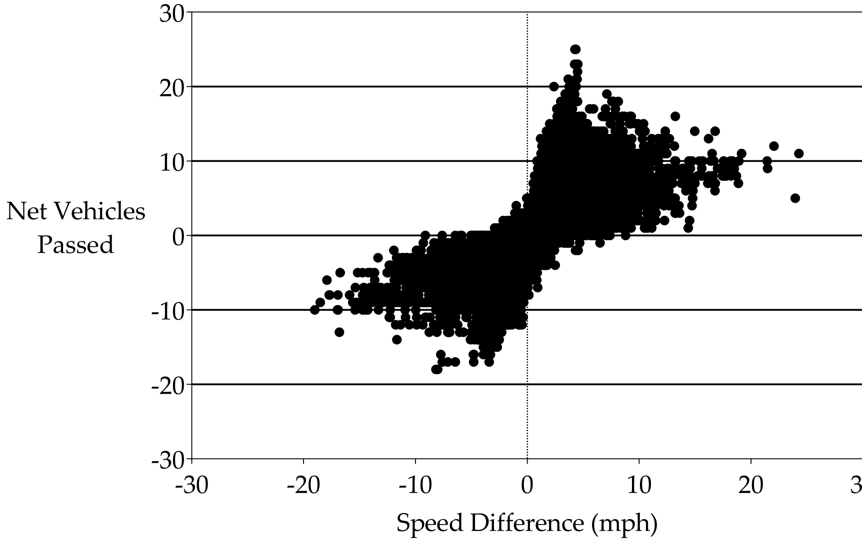
Because the unit of observation is a comparison and there are five lanes on each freeway, each driver is represented in at least four comparisons. In the sample, only driver comparisons in which the index driver was passed by another vehicle and also passed another vehicle are included. These drivers, because they experience both passing and being passed, are likely to question whether they could travel faster in another lane. Table 3.3 presents summary statistics for the sample of driver comparisons. The average speed of drivers in the comparison sample is somewhat lower than the average speed of drivers on the US-101 freeway but similar to the average speed of drivers on the I-80 freeway. Episodes lasted about 65 seconds on average. Approximately half of the driver comparisons are drawn from the US-101 freeway and half are drawn from the I-80 freeway.

3.4 Analysis

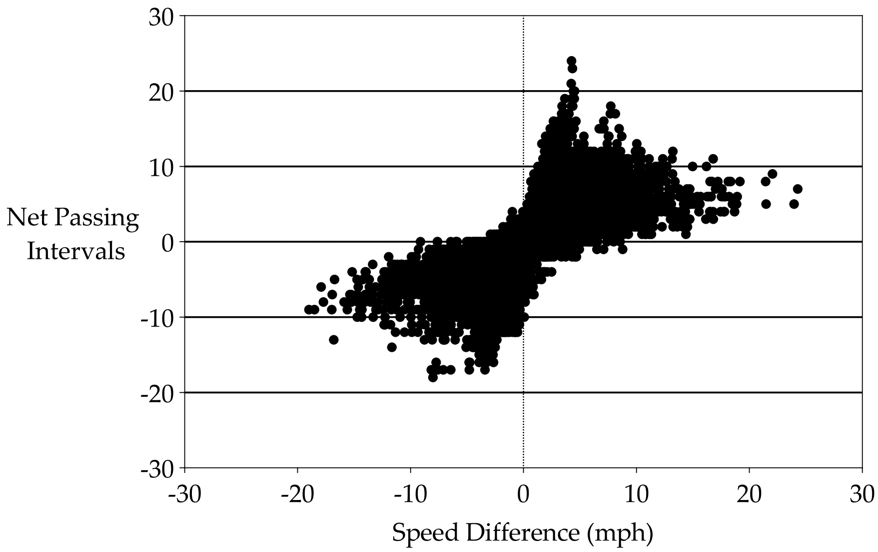
A starting point for comparing the vehicle count heuristic and the interval count heuristic is to examine how well each heuristic performs in judging relative speed. Figure 3.4a shows net vehicles passed on the vertical axis and the speed difference between the lanes on the horizontal axis. The speed difference is calculated as the difference between the index driver's speed and the comparison lane speed so that a positive difference indicates the index driver has a higher average speed. Figure 3.4b replaces net vehicles passed with net passing intervals. Both plots show clear positive trends with the great majority of observations falling in the lower left and upper right quadrants. For the drivers represented by these observations, the heuristic yields the correct judgment of relative speed.

Both heuristics, however, fail in a small number of cases. An observation in the upper left quadrant of either plot represents a driver who, using the chosen heuristic, concludes she is traveling at a higher average speed than the comparison lane but actually has a lower average speed. These drivers overestimate relative speed. An observation that falls in the lower right quadrant represents a driver who, using the heuristic, concludes she is traveling at a lower average speed than the comparison lane but actually has a higher average speed. These drivers underestimate relative speed. In total, the interval count heuristic fails in more cases than the vehicle count heuristic. While both heuristics more often underestimate relative speed, the asymmetry is more extreme for the interval count heuristic. The interval count heuristic underestimates speed in 68% of failed cases while the vehicle count heuristic underestimates speed in 57% of failed cases.

This finding is representative of a general trend. A comparison of the two plots shows that many of the observations in Figure 3.4a are shifted downward in Figure 3.4b. This downward shift further indicates that compared to the vehicle count heuristic, the interval count heuristic tends to underestimate relative speed.



(a)



(b)

Figure 3.4: (a) Vehicle Count Heuristic (b) Interval Count Heuristic

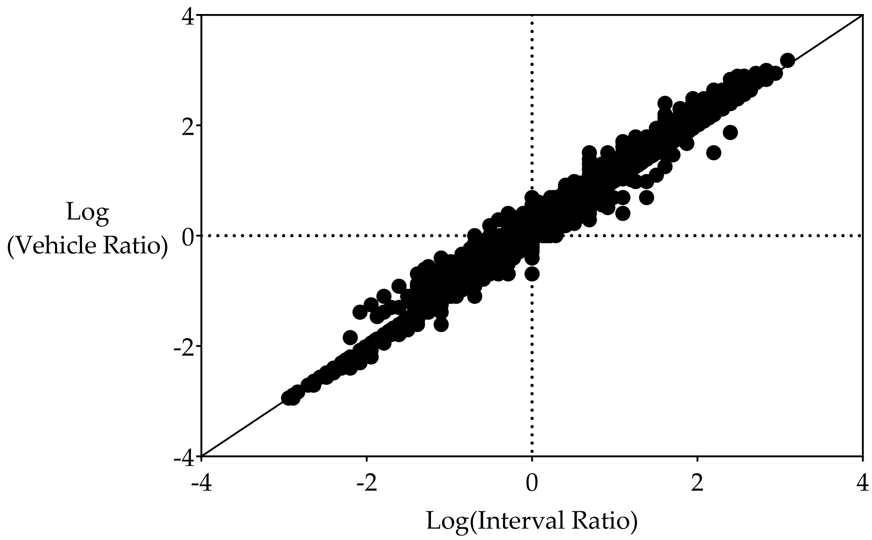


Figure 3.5: Comparison of the Heuristics

To see the differences between the two heuristics more clearly, it is useful to plot the heuristics together. Figure 3.5 shows the natural logarithm of the vehicle ratio on the vertical axis. On the horizontal axis the natural logarithm of the interval ratio is plotted. The vehicle ratio is defined as $\frac{\text{vehicles the index driver passed}}{\text{vehicles the index driver was passed by}}$ and the interval ratio is defined as $\frac{\text{passing intervals}}{\text{being passed intervals}}$.

Again, there is a clear positive trend and most observations fall in the upper right and lower left quadrants. This pattern indicates that the two heuristics yield the same main conclusion in most cases. For example, an observation in the upper right quadrant represents a driver who has passed more vehicles than she was passed by and has spent more time passing vehicles than being passed by other vehicles. With both heuristics, she concludes that she has a higher average speed than the comparison lane.

An observation that falls in the upper left quadrant, however, represents a driver who passed more vehicles than she was passed by yet spent more time being passed by other vehicles than passing other vehicles. This driver would conclude she has a lower average speed than the comparison lane using the interval count heuristic and a higher average speed than the comparison lane using the vehicle count heuristic. Conversely, an observation falling in the

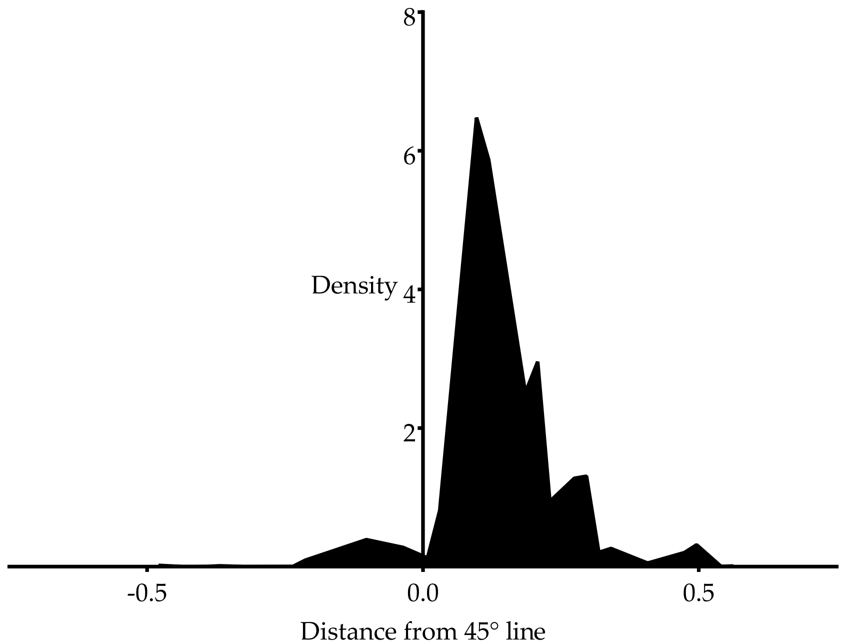


Figure 3.6: Kernel Density Estimate

lower right quadrant represents a driver who would conclude she has a higher average speed than the comparison lane using the interval count heuristic and a lower average speed than the comparison lane using the vehicle count heuristic. As shown in Figure 3.5, this scenario never arises in the data. There are no observations in the lower right quadrant but there are observations in the upper left quadrant. This pattern confirms that the interval count heuristic tends to underestimate relative speed compared to the vehicle count heuristic.

Even when the heuristics provide the same main conclusion for speed relative to a comparison lane, an asymmetric discrepancy between the heuristics is still present. Consider the 45 degree line. If an observation falls on this line, the two methods give exactly the same information. If an observation falls above the line, the interval ratio is smaller than the vehicle ratio. For example, a observation representing a driver who passed two vehicles and was passed by three vehicles but spent only one interval passing and three intervals being passed by other vehicles would fall above the 45 degree line. If drivers derive some information about the size of the speed difference from passing experience (in terms of both vehicles and

Table 3.4: Regression: Vehicle Ratio on Interval Ratio

Dependent Variable: Log(Vehicle Ratio)	(1)	(2)
Intercept	0.033	
	(0.001)	
I-80 Intercept		0.035
		(0.001)
US-101 Intercept		0.031
		(0.001)
Log(Interval Ratio)	1.016	1.016
	(0.001)	(0.001)
R-squared	0.991	0.991
Observations	26,651	26,651

All coefficients are statistically significant at the 1% level. Robust standard errors are reported in parentheses.

intervals), relative speed will be underestimated using the interval count heuristic compared to the vehicle count heuristic. For observations that fall below the line, the vehicle ratio is smaller than the interval ratio. These drivers overestimate relative speed using the interval count heuristic compared to the vehicle count heuristic.

The data show that the interval count heuristic is more likely to underestimate relative speed compared to the vehicle count heuristic. About 18% of observations fall above the line while only 1% of observations fall below the 45 degree line. Figure 3.6 shows the Epanechnikov kernel density estimate of the distribution of distance to the 45 degree line for the observations that do not fall on the line.

In addition, regressing the natural logarithm of the vehicle ratio on the natural logarithm of the interval ratio shows that the constant term in the regression is positive and significantly different from zero. On average, the vehicle ratio is larger than the interval ratio.

3.4.1 Heuristic Performance in Heavy Congestion

The data set also allows for differences between the heuristics to be examined at different levels of congestion. As noted earlier, the I-80 sample is more heavily congested than the US-101 sample. A first look at the observations shows that a larger share of observations fall above the 45 degree line in the I-80 sample than in the US-101 sample. Also, Table 3.4 shows that in a regression with separate constant terms for each sample, the I-80 constant term is statistically larger than the US-101 constant term. This finding indicates that drivers are more likely to underestimate relative speed in more heavily congested traffic using the interval count heuristic.

3.5 Conclusion

The data confirms there is an asymmetric difference between the vehicle count heuristic and the interval count heuristic. Although they appear similar, compared to the vehicle count heuristic, the interval count heuristic tends to underestimate relative speed because of optimizing behavior by drivers in following distance choice. In addition, drivers using the interval heuristic are more likely to underestimate relative speed in more heavily congested traffic. Given the external costs of lane changing and the evidence presented here that drivers may underestimate relative speed, many lane changes are likely socially and privately suboptimal. Educating drivers about the external costs of lane changing and illusions present in judging relative speed in freeway traffic could have safety benefits and reduce traffic congestion.

Bibliography

Ahn, S. and M. Cassidy (2007). Freeway traffic oscillations and vehicle lane-change maneuvers. *17th International Symposium ofTransportation and Traffic Theory*, 1–23.

Anderson, M. (2008). Safety for whom? The effects of light trucks on traffic fatalities. *Journal of Health Economics 27*(4), 973–989.

Anderson, M. L. and M. Auffhammer (2014). Pounds That Kill: The External Costs of Vehicle Weight. *The Review of Economic Studies 81*(2), 535–571.

Ansari, A., N. Economides, and J. Steckel (1998). The Max-Min-Min Principle of Product Differentiation. *Journal of Regional Science 38*(November 1996), 207–230.

Ben-Akiva, M., A. de Palma, and J.-F. Thisse (1989). Spatial Competition with Differentiated Products. *Regional Science Urban Economics 19*, 5–19.

Brackstone, M., B. Waterson, and M. McDonald (2009). Determinants of following headway in congested traffic. *Transportation Research Part F: Traffic Psychology and Behaviour 12*(2), 131–142.

Brown, J. F., K. S. Obenski, and T. R. Osborn (2002). *Forensic Engineering Reconstruction of Accidents*. Springfield: Charles C. Thomas Publisher LTD.

California Department of Transportation (2014). Caltrans PeMS.

Champagne, C. and T. Iezzi (2014). Dunkin' Donuts and Starbucks: A Tale of Two Marketing Giants.

Cosgrove, S. B. (2011). Passenger Car Equivalents of Light Duty Trucks and the Costs of Mixed Vehicle Traffic : Evidence from Michigan. *Journal of the Transportation Research Forum 50*(3), 63–75.

D'Aspremont, C., J. Gabszewicz, and J.-F. Thisse (1979). On Hotelling's 'Stability in Competition'. *Econometrica 47*(5).

de Palma, A., V. ictor Ginsburgh, Y. Y. Papageorgiou, and J.-F. Thisse (1985). The Principle of Minimum Differentiation Holds under Sufficient Heterogeneity. *Econometrica 53*(4), 767–781.

Federal Highway Administration (2008). Interstate 80 Freeway Data Set.

Federal Highway Administration (2015). Highway Statistics Series. Technical report, U.S. Department of Transportation.

Gabszewicz, J. J. and J. F. Thisse (1979). Price competition, quality and income disparities. *Journal of Economic Theory 20*(3), 340–359.

Greven, M. (2015). Tax Guide 2015. Technical report, European Automobile Manufacturers Association, Brussels.

Hamdar, S. H. and H. S. Mahmassani (2009). Life in the Fast Lane. *Transportation Research Record: Journal of the Transportation Research Board 2124*, 89–102.

Hotelling, H. (1929). Stability in Competition. *The Economic Journal 39*(153), 41–57.

Irmen, A. and J.-F. Thisse (1998). Competition in Multi-characteristics Spaces : Hotelling Was Almost Right. *Journal of Economic Theory 78*, 76–102.

Iyer, G. and P. B. Seetharaman (2008). Too close to be similar: Product and price competition in retail gasoline markets. *Quantitative Marketing and Economics 6*(3), 205–234.

Jacobsen, M. R. (2013). Fuel Economy and Safety: The Influences of Vehicle Class and Driver Behavior. *American Economic Journal: Applied Economics 5*(3), 1–26.

Kockelman, K. M. and R. a. Shabih (2000). Effect of Light-Duty Trucks on the Capacity of Signalized Intersections. *Journal of Transportation Engineering 126*(6), 506–512.

Laney, K. (2012). State of the Chains , 2012. Technical Report December, Center for an Urban Future.

Li, S. (2012). Traffic Safety and Vehicle Choice: Quantifying the Effects of the 'Arms Race' on American Roads. *Journal of Applied Econometrics 27*, 34–62.

McDonald, M., J. Wu, and M. Brackstone (1997). Development of a fuzzy logic based microscopic motorway simulation model. *Proceedings of the IEEE Conference on Intelligent Transportation System*, 82–87.

Netz, J. S. and B. a. Taylor (2002). Maximum or Minimum Differentiation? Location Patterns of Retail Outlets. *Review of Economics and Statistics 84*(1), 162–175.

Neven, D. and J.-F. Thisse (1990). On Quality and Variety Competition. In J. Gabszewicz, J. Richard, and L. Wolsey (Eds.), *Economic Decision-Making, Games, Econometrics and Optimisation*, pp. 175–199. Amsterdam: North Holland.

Redelmeier, D. A. and R. J. Tibshirani (2000). Are Those Other Drivers Really Going Faster? *Chance 13*(3), 8–14.

Rotemberg, J. J. (1985). The Efficiency of Equilibrium Traffic Flows. *Journal of Public Economics 26*, 191–205.

Sayer, J., M. Mefford, and R. Huang (2003). The effects of lead-vehicle size on driver following behavior: is ignorance truly bliss? In *Proceedings of the Second International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, pp. 221–225.

Schrank, D., B. Eisele, T. Lomax, and J. Bak (2015). Urban Mobility Scorecard. Technical Report August, Texas A&M Transportation Institute and INRIX, Inc.

Schultz, G. A. and M. J. Babinchak (1999). Final Report for the Methodology Study of the Consumer Braking Information Initiative. Technical report, National Highway Transportation Safety Administration.

Sen, B., J. D. Smith, and W. G. Najm (2003). Analysis of lane change crashes. Technical report, John A. Volpe National Transportation Systems Center, Cambridge, MA.

Small, K. A. and X. Chu (2003). Hypercongestion. *Journal of Transport Economics and Policy 37*(3), 319–352.

Small, K. A. and E. T. Verhoef (2007). *The Economics of Urban Transportation*. New York: Routledge.

Tabuchi, T. (1994). Two-stage two-dimensional spatial competition between two firms. *Regional Science and Urban Economics 24*(2), 207–227.

Transportation Research Board (2010). Highway Capacity Manual. Technical report, Washington, D.C.

Watson, R. (2009). Product variety and competition in the retail market for eyeglasses. *Journal of Industrial Economics 57*(2), 217–251.

Wauthy, X. (1996). Quality Choice in Models of Vertical Differentiation. *Journal of Industrial Economics 44*(3), 345–353.

White, M. J. (2004). The "Arms Race" on American Roads: The Effect of Sport Utility Vehicles and Pickup Trucks on Traffic Safety. *Journal of Law and Economics 47*(2), 333–355.

Whitefoot, K. S. and S. J. Skerlos (2012). Design incentives to increase vehicle size created from the U.S. footprint-based fuel economy standards. *Energy Policy 41*, 402–411.

Zheng, Z. (2014). Recent developments and research needs in modeling lane changing. *Transportation Research Part B: Methodological 60*, 16–32.