ABSTRACT

Title of dissertation: ESSAYS ON EMPIRICAL INDUSTRIAL

ORGANIZATION

Yong joon Paek

Doctor of Philosophy, 2018

Dissertation directed by: Professor Andrew Sweeting

Department of Economics

The dissertation focuses on two issues that are broadly related to the urban planning of Washington, DC. The first two chapters consider the burgeoning food truck industry in Washington, DC and the third chapter considers public transit ridership and the impact of large maintenance programs that cause temporary but large decreases in service quality.

In Chapter 1 I build and estimate a tractable model that captures some key characteristics of the food truck industry. Characteristics of the industry such as there being many small firms playing an entry game with various dimensions of heterogeneity (for example, cuisine genre and quality) render regulations and policies difficult to assess and design, leading to local regulators resorting to 'ad hoc' policies to regulate the industry. For example, in Washington, DC scarce parking spots at popular lunch locations are allocated through a random lottery. This highlights the importance of a tractable model that captures importance features

of the industry which can be estimated and used to consider counterfactual policies.

In Chapter 2 I consider two counterfactual scenarios. In the first counterfactual scenario I reduce the reach of the lottery and I find that the lottery allows for the survival of firms with lower quality and leads to higher prices. Expected utility for consumers are lower and firm profits are higher in current regime compared to a couterfactual regime where some locations are not included in the lottery. The net welfare effect for this counterfactual scenario is an increase in total daily welfare of $2.294.18. In my second counterfactual scenario where the non-lottery locations have their parking capacity increased by 2 spaces, I find a positive impact for both truck owners and consumers with a net increase in total daily welfare of $8,260.26.

Chapter 3 considers the impact of large public transit maintenance programs on long-run ridership. An agency that manages large transit systems must make investments to maintain a level of quality to sustain ridership. If consumers face switching costs when changing their mode of transport, the large and unavoidable disruptions to services resulting from a large maintenance program may provide a sufficient negative utility shock for riders to substitute to alternative modes of transport and not return after the repairs are completed. In this chapter I consider such indirect costs that a transportation agency may incur in the context of *Metrorail*, the subway system that stretches through the District of Columbia (DC), Maryland (MD), and Virginia (VA) operated by the Washington Metropolitan Area Transit Authority (WMATA). I find that there has been a persistent drop in ridership up to 10 months after the repairs on certain tracks have been completed.

Essays in Empirical Industrial Organization

by

Yong joon Paek

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Professor Andrew Sweeting, Chair
Professor Ginger G. Jin
Professor Dan Vincent
Professor Mary Zaki
Professor Joshua Linn

Dedication

To my parents E.H. Min, C.W. Paek, and grandparents G.S. Min, H.B. Paik, J.S. Lee, H.W. Kim.

Contents

List of Tables

List of Figures

Introduction

This dissertation covers two topics. Chapter 1 and 2 deals with competition and parking space allocation in the food truck industry in Washington, DC, and Chapter 3 deals with public transit ridership and the long-run impacts of a large-scale maintenance program.

An industry with many firms and various dimensions of heterogeneity is difficult to analyze and consequently poses challenges for the assessment and implementation of regulatory policies. One such example is a young but booming food truck industry. In this industry, hundreds of firms serve different types ('cuisine genre') of products with varying quality at many different locations that have capacity constraints (parking space is limited). Local regulators have responded to this emerging industry in various 'ad hoc' ways. To obtain a better understanding of the industry and assess counterfactual policies, in Chapter 1, I build a structural model of the food truck industry that accounts for important dimensions of heterogeneity and competition to assess the welfare effects of counterfactual scenarios. The model is kept tractable by borrowing ideas from the Oblivious Equilibrium [Benkard et al., 2008] literature. I apply this model to the food truck industry in Washington, DC and estimate its key parameters. I find that consumers preferences are only moderately correlated within cuisine genres and that trucks' own price elasticity's are on average about -1.61. Firms price-cost margin is on average 0.66, which closely resembles industry insiders rule of thumb for a truck that can operate into the foreseeable future. I find that the choice probabilities for lottery

locations are higher than that of the non-lottery locations given both are in a truck's choices set. In my model, this implies that given the lottery, entry costs for the lottery locations are lower than that of the non-lottery locations.

In Chapter 2, I employ the results from chapter one to shed light on the impacts of counterfactual regulatory regimes. Under the current regime the allocation of parking spaces for food trucks are determined by a lottery for some locations. My counterfactual simulations of reducing the reach of the lottery to include fewer locations suggest that the lottery is dampening competition and decreasing the costs that firms must incur to enter a market location. As a result, I find that it allows for the survival of firms with lower quality, higher prices, while lowering expected utility for consumers, and awarding higher profits to firms. The net welfare effect for this counterfactual scenario is an increase in total daily welfare of $2.294.18. I also consider a counterfactual scenario where the non-lottery locations increase their parking capacity by 2 spaces. I find that this is beneficial for both truck owners and consumers with a net increase in total daily welfare of $8,260.26.

Chapter 3 considers the impact of large public transit maintenance programs on long-run ridership. An agency that manages large transit systems must make investments to maintain a level of quality to sustain ridership. If consumers face switching costs when changing their mode of transport, the large and unavoidable disruptions to services resulting from a large maintenance program may provide a sufficient negative utility shock for riders to substitute to alternative modes of transport and not return after the repairs are completed. In this chapter I consider such indirect costs that a transportation agency will incur in the context of *Metrorail*, the subway system that stretches through the District of Columbia (DC),

Maryland (MD), and Virginia (VA) operated by the Washington Metropolitan Area Transit Authority (WMATA). I find that there has been a persistent drop in ridership of about 1.68% up to 10 months after the repairs on certain tracks have been completed. Also, commuters that are originating from VA seemed to have substituted more heavily away from using *Metrorail* compared to their MD counterparts.

Chapter 1 Location Choice and Price Competition with Differentiated Products and Many Firms: An Application to the Mobile Vending Industry in Washington, DC.

1.1 Introduction

In this chapter I build a structural model of the food truck industry that accounts for important dimensions of heterogeneity, market entry, and competition. Subsequently, I apply this model to the food truck industry in Washington, DC and estimate its key parameters. To begin thinking about this market as an economic issue, it is useful to observe the following: there are many small firms, who each truck chooses to serve a well-defined market based on geographical location (for example, street parking near a prominent park such as *Franklin Square*) and time of day (for example, in DC the "lunch break" shift). There are a limited number of parking spots at the most popular locations and different consumer demographics and preferences at each location.

Space is a scarce resource in a congested city and it is important for local regulators to allocate and manage such resources efficiently. For example, a regulator may want to know the welfare impacts of expanding street parking spaces for popular food truck lunch locations or they may want to consider how the current regulatory policy is impacting the industry relative to other policies. To be able to consider these counterfactual scenarios one needs a tractable but sufficiently

realistic model.

Many industries are comprised of a large number of small firms that are heterogeneous in various dimensions. These industries are difficult to model and analyze due to the rapid increase in computational burden for solving equilibria with many firms and because most of the theoretical literature considers stylized duopoly models. As a consequence most structural empirical industrial questions center around oligopolistic industries inhabited by a handful of firms.[1] Given these modelling challenges, the assessment and design of regulatory regimes in these industries are difficult. For example, in the US food truck industry within a city, there are hundreds of firms selling products of varying quality and type who must choose the location to enter and compete in every day. Also, given entry to a location, each type of firm may have different competitive effects on each other implying the existence of informational and allocative externalities depending on where a firm decides to operate in and the market structure that ensues from these decisions.

The difficulty in modelling this environment comes from the fact there are many heterogeneous firms. Outside of the food truck industry, any setting where a firm must secure the right to operate and then compete with other firms who have also secured this right will be similar. For example, the allocation of types of shops in a newly developed mall, the allocation of bundles of licenses. In this paper I build and estimate a model that attempts to account for these important features of the food truck industry while maintaining tractability. The question

---

[1]In a dynamic context, Berry and Pakes [1993] explore the curse of dimensionality and how quickly computational burden increases as the number of firms increases. My model is static; however, it captures other dimensions of heterogeneity which increases computational burden greatly if trying to solve for a standard Nash equilibrium.

of designing a parking location allocation mechanism for the industry is interesting. The multiple dimensions of firm and market heterogeneity, and the various externalities imply a simple auction may not be the ideal mechanism for this allocation problem. However, these interesting features are also the road blocks with respect to methodology and theory. With this in mind, the goal of this chapter is not to find and suggest some type of theoretically *optimal* mechanism, but to assess and quantify how competing regimes perform. One counterfactual regime I consider reduces the reach of the currently employed lottery system by 2 locations. The second counterfactual regime increases non-lottery locations' parking capacity.

My model of the food truck industry has buyers with nested logit utilities, where the nests are formed around cuisine type, which is a dimension of heterogeneity in the firms. This modelling choice allows seller's heterogeneity to impact demand in a meaningful and tractable way. The sellers (trucks) play a two-stage game where in the first stage given a conjecture about the state of each market location, they choose prices and choose a location to enter where they incur a entry cost that is specific to the location. Then in the second stage the trucks explicitly compete with other firms that have entered the particular market. This ordering of the seller's decision where he chooses price *then* location, is novel in the literature but is reasonable in my context. It is easily observed that the truck's prices do not vary over time but their locations change daily.[2] I find that removing a key food truck location (*L'Enfant* from the lottery increases daily total welfare by \$2,294.18 and that increasing the parking space capacity of current non-lottery locations by 2 spaces increases daily total welfare by \$8,260.26

---

[2]In Schmalensee [1992] the author briefly discusses motivations for the order in which firm choices are made. It is argued that the dynamic differences between price choice and product characteristic choice are not clear, and there is a tradition to think that maybe price is chosen before other attributes.

In Chapter 1, Section 1.2 will discuss where my work fits in the existing literature. Section 1.3 offers a description of institutional background for the food truck industry in Washington, DC and describes the data used to estimate the model. Section 1.4 will develop the structural model and provide an exposition of the equilibrium. Section 1.5 interprets and discusses the estimation stages and estimation results. Section 1.5 assesses the model's performance by comparing the model's equilibrium solution outcomes and their estimated counterparts.

1.2   Related Literature

There is a rich body of literature that considers the endogeneity of market structure. Classic examples include Berry [1992] and Bresnahan and Reiss [1990]. We learn from these seminal contributions that when a firm decides to enter a market it is a strategic decision. Favorable states of a market will increase profits, but also attract more entrants, hence a firm must evaluate these trade-offs before choosing to enter himself. When it comes to empirically estimating these types of models the researcher encounters a number of issues due to the large number of configurations that must be checked to find whether a distribution of players constitute an equilibrium. Such an example is Mazzeo [2002], where model of endogenous product type choice and market structure is considered under the assumption of complete-information. Mazzeo [2002] shows that even with two or three product types, the number of profit inequality constraints that must hold is large and is quite burdensome to estimate.

A paper that addresses these difficulties is Seim [2006]. Seim [2006] considers the product characteristic choice decision of firms and applies this framework to

7

the video rental industry. In this industry the product offered itself is relatively homogenous, but location is a major source of product differentiation and Seim considers this dimension as an endogenous choice by the firms. However, in Seim [2006] demand and explicit competition is not modeled, and a reduced form approach is taken to estimate a firm's entry probability. For me to estimate the dollar value of location entry costs and industry welfare I model entry and price competition. There is little work in the literature that models entry and also explicit competition after entry. For example, Suzuki [2010] models hotel chain entry decisions in a dynamic setting, but again doesn't model competition explicitly. The difficulties in modelling competition explicitly come from the lack of data and/or the computational burden in solving for the equilibrium with many firms.

I overcome these difficulties in modelling the firm's entry decision and the competition following entry to a location by using self-collected data and by using a simulation based iterative solution method. My model maintains tractability by assuming that the average of the realized distribution of firms in the market is equal to the firms' long-run conjecture of the state of the market, and I with this assumption can compute the equilibrium distribution of firm types, quality, and number of operational firms within the modelled locations. This key assumption is behaviorally plausible as it is reasonable to argue that individual small firms use their experiences and observations for a market realization to come up with a expected long run state of the market and is similar to the assumptions made for oblivious equilibrium in Benkard et al. [2008].[3]

One method of dealing with the burden of solving for a complete-information

---

[3]For more recent papers using moment-based equilibrium concepts see Xu [2008], Qi [2013], Saeedi [2014], Sweeting [2015].

Nash equilibrium is to assuming incomplete-information to effectively render all potential entrants to be homogenous as in Seim [2006]. This makes a model tractable, but in the process interesting economic insights related to firm heterogeneity and types may be lost. My model will allow the trucks to be of different types and qualities. Allowing for firm types, while considering the entry decisions for firms allows me to look at the equilibrium distribution of firm types that prevails in equilibrium. This is important because in many contexts, it is not just the number of competing firms that matter in an entry decision, but the number of firms that are more directly in competition to your type. This intuition is related to Wollmann et al. [2014] where the author considers the importance of product positioning in the context of commercial vehicles. Wollmann et al. [2014] finds that product entry has a dramatic impact on prices and purchases. In my context, this finding emphasizes the importance for researchers to allow the firms not to only choose whether to enter a market or not, but allowing them to strategically choose which market to enter given the state of each market as it appears to your own type.

The food truck industry in Washington, DC has been examined by Anenberg and Kung [2015]. The authors explore the impact of information technology on food truck growth and suggest that an advantage of food trucks over traditional brick and mortar establishments is that they can use mobility to capitalize on the consumer's taste-for-variety. Obtaining data from DC food trucks' Twitter feed for the trucks' location choice, a logit model is estimated by defining a reduced form profit function. Their results indicate that there is a negative impact of a truck visiting the same location in short succession. Relying on strong assumptions on industry total revenues, they estimate that the loss of choosing the same location two days in a row results in a $257 loss in the day's profit, which is about 38% of

average daily profits.

However, such reduced form analysis does not provide an opportunity for understanding the underlying parameters that dictate the consumer's and vendor's market behavior and hence is limited in scope when considering policy simulations and welfare analysis. Also, Anenberg and Kung [2015] does not deal with the cost arising from the fact that at each location, parking spots are scarce, which is an important factor when considering the industry from a regulation and policy perspective.

1.3 Industry Details and Data

1.3.1 The Food Truck Industry

Food trucks are very common in large cities and each truck requires a parking spot to operate. In these cities space is a scarce resource that needs to be carefully managed. With relatively low start-up costs, mobile vending as opposed to traditional brick-and-mortar setups buy the entrepreneurs an opportunity to be more experimental with their products. This coupled with an increase in the import and export of culinary culture in the last decade, has led to the rapid penetration of food truck food as standard meals and an increase in variety of the types of cuisine offered. Accurate numbers depicting the growth of culinary experimentation and the mobile vending industry are hard to come by, however a report by Mountain View-based financial software company, Intuit [2012], forecasts that the "rolling restaurants" are on track to be a $2.7 billion national industry in 2017 and its market share for meals to jump to 3 or 4 percentage points in the next five years.

This is particularly interesting when you consider the market research company, NPD Group's [2016] finding that for weekday casual dining and fast casual restaurants, food service lunch visits declined by 7% in the quarter ending June 2016 compared to same quarter the year before.

Simple Google trend searches support this phenomenon. In Figure 1.1 I show web search trends for three keywords; donburi, kimchi and food trucks. It can clearly be seen that the interests in foreign food items have increased noticeably in the last decade, particularly since 2010. With interests in food trucks following the pattern closely.[4] In particular, this effect is likely to be much stronger in large cities, where the growth of the food truck industries are concentrated. I can't pin down exactly why the mobile vending industry saw such an explosive growth and this paper does not attempt to explain this. I will be focusing on the issues that policy makers must consider now that this industry exists.

---

[4]The seasonal component of the interest in food trucks in panel (c) seem to be due to the fact that warmer temperatures that allow a customer to eat outside garner more interest for food trucks.
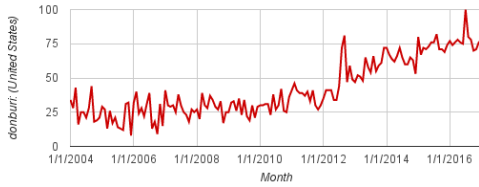
(a) Interest in Donburi over time.

(b) Interest in Kimchi over time.
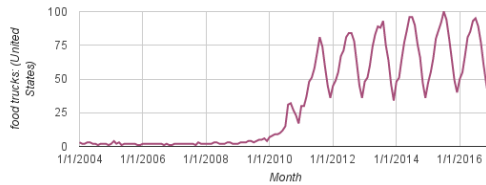
(c) Interest in Food Trucks over time.

Figure 1.1: Web Search Trends in America for Keywords Gauging the Advent of the Mobile Vending Industry.

These policy issues range from hygiene and quality regulation to parking, traffic and licensing regulations. Out of the broad range of issues, this project will specifically look at parking allocation at various locations where food trucks agglomerate at and the competition the food trucks engage in given a tendency for consumers to have preferences over different cuisine genres as well as the overall quality of the meals offered. Evidence of policy maker's concerns can be seen in DC's food truck industry. The DC Department of Consumer and Regulatory Affairs (DCRA) governs the licensing and regulation of mobile vendors in DC. Starting from December 2013 the DCRA has implemented a monthly lottery[5] system that randomly allocates vendors who register to enter the lottery to a predetermined list of popular locations. This system was the answer to the widespread problem of trucks showing up extremely early in the morning to secure the best spots, consequently congesting the area and inducing many parking violations and pedestrian safety

---

[5]The exact lottery mechanism is proprietary so I condition my analysis on the outcome of the monthly lotteries.

concerns. The officially designated lottery locations can be seen in Figure 1.2.[6]

Figure 1.2: The official DC Department of Consumer and Regulatory Affairs (DCRA) designated lottery locations and number of parking spots allotted.

Although a simple lottery is easy to implement and reduces the various entry

---

[6]The exact page can be found at **https://eservices.dcra.dc.gov/VendingLottery/MRVLocations.pdf**

costs that trucks need to incur to secure a popular spot, it may not be the ideal mechanism for allocating these scarce parking locations. For example on the 8th of November, 2016, at *Union Station*, there were 6 *Asian* trucks out of 11 trucks which may indicate a lack of variety. The lottery can generate situations where there is not an ideal level of variety at different locations, and at these locations trucks could trade spots for a Pareto improvement. An extreme example would be, assuming consumers like variety of cuisine genres, if there is no variety with only *Asian* trucks at *Franklin Square* and no variety with only *American* trucks at *L'Enfant*, it would be optimal to change the allocation of the trucks to reflect more variety at both locations. Also, the marginal value of variety maybe greater at *L'Enfant* where there are little outside options for lunch. Another dimension that must be considered in this allocation problem is truck quality, which opens up opportunities for more sophisticated mechanisms (for example, auctions) that may correctly price the value of a scarce spot for different trucks.

A simple motivating example can shed some light onto the subtleties involved in modelling competition and regulatory policies when consumers have preferences that are nested within a particular cuisine genre category (for example, *Asian* cuisine and *Mediterranean* cuisine). Consider a parking lot with 2 food trucks and no outside option for lunch. These two trucks are of different genres but have exactly the same quality and suppose that the consumer's preferences exhibit no correlations between similar genres. We can model this with a simple logit model and these two trucks will share the market evenly in equilibrium.

Under these circumstances, if we add a third truck to this location of the same quality, regardless of genre, the third truck will take equal market share from the

incumbent trucks ending up with a third of the market. However, if consumer's preferences are correlated within genres this will not be the case. Consider now a nested logit model, where consumers have preferences with nests formed around cuisine genres. Now the genre of the third truck impacts the incumbents and the entrant differently. In particular, both incumbent trucks would prefer the genre of the third truck to be different from their own, as then they will lose less market share compared to if the third truck was of the same genre. In other words, as the correlation within the nests become stronger, trucks and consumers want "more" variety in the market place. Depending on the correlation of preferences within cuisine genres, the optimal distribution of quality and types of trucks in the markets should be different. This simple example shows that when we are allocating scarce parking spaces to vendors, we must take into account these trade-offs to correctly measure and analyze welfare.

Allocative externalities where the value of entering a location for a truck depends on the identity of who is going to be at the location makes the setting theoretically intractable and characterizing and designing an optimal mechanism in this setting is not the scope of this paper. Knowing these effects exist in the market, I am attempting to build a model that captures these effects to assess the current policy and the impact of counterfactual scenarios that need not be theoretically optimal.

1.3.2 Data

The estimation of the model utilized data from multiple sources. Most of the demand data used to estimate the nested logit demand model is manually collected and processed in two stages. Firstly, to efficiently collect market share data I

investigated the proportions of consumer arrivals at a food truck location during different times of the consumer's lunch break. Secondly, using these time dependent proportions I scaled up the quantity count for each truck (number of consumers served) during the specific time interval by this proportion to obtain the "total lunch shift" market shares for each individual truck and the overall food truck segment market share by comparing it with the number of primary jobs in the area obtained from the publicly available Census Longitudinal Origin-Destination Employment Statistics (LODES) data set which is derived from the Longitudinal Employer-Household Dynamics (LEHD) data set.

In the process of collecting the arrival proportions I visited a total of 10 food truck locations (lottery designated and non-lottery) from 11am to 1 35pm and counted the stock of consumers in line at a truck at the location approximately every 10 minutes. Then assuming an average departure rate of 2 customers per minute per truck[7] I backed out the number of new arrivals and computed the proportion of arrivals that happened during each 10 minute interval. Effectively, this is a probability distribution of the consumers arriving to buy lunch during the aforementioned time frame. I conducted a Kolmogorov–Smirnov test on these distributions to assess if any statistical differences exist between every pair and find that any differences between all 10 of the distributions that I counted were statistically insignificant. There are limitations to this data, as it is manually collected and the sample size is limited, but it has helped me to collect market share data from multiple locations each day.

The resulting market share data consists of 30 location observations from 12

---

[7]This number was obtained from conversations with truck owners and I believe it is reasonable to believe that the rate at which a truck serves its customers are not time varying during the shift.

Figure 1.3: Map of Downtown Washington, DC and Surrounding Neighborhoods. Markers Indicate Food Truck Locations Revealed by Trucks' Twitter Feeds. The Triangles and Squares make up 86% of the Observations.

different locations, of which 9 are lottery and 3 are non-lottery and a total of 331 market share observations from 154 unique trucks (shown in Table 1.1). Location specific data such as the market size and mean earnings from the LODES data set[8] and truck specific data such as the number of Twitter Followers and their Yelp reviews are used to supplement the demand data. Figure 1.3 shows the food truck lunch locations that are observed in the Twitter data colored by lottery (all modelled), non-lottery but modelled, and not modelled.

I do not observe the average price of each order ticket for each truck, so we must determine what price to use when we estimate the demand model. For a subset of the data, I asked the truck owners at the end of the shift what the average price of their sales item on that day was. For these trucks I use these prices, for others I make an assumption. For example, I assume that for kabob trucks, that the menu item of interest is the average price of their pita sandwiches, for taco trucks I assume that a meal is buying 3 tacos, and for rice bowl trucks I use the median price of their entrees. If revenue data was available the measure of average price could be improved, but this was not available. I believe this method is a second best to thinking about the prices of each vendor. Tables 1.1 and 1.2 show some summary statistics from the demand data.

---

[8]Using the LODES OnTheMap web application (https://onthemap.ces.census.gov/) I drew a 0.15 mile radius circle and defined it as the "market".

Locations	Lottery	Mean Monthly Earnings (\$)	Utilized Market Size
19th & L	No	7,266.95	24,583
CNN (First St NE)	No	6,725.52	16,365
Farragut Square	Yes	7,439.35	23,779
Federal Center / Patriots Plaza	Yes	7,149.01	3,659
Franklin Square	Yes	7,996.54	14,317
Gallery Place / Chinatown	No	6,511.07	10,783
L'Enfant	Yes	6,974.15	16,901
Metro Center	Yes	7,817.77	23,020
Navy Yard	Yes	9,257.72	2,898
NoMa / New York Ave Metro	Yes	7,128.24	6,074
State Department	Yes	6,550.04	9,358
Union Station	Yes	6,512.05	14,124

Note: The Market Size is the total number of primary workers.

Table 1.1: Location Characteristics from the LODES data

Variable	Mean	Std. Dev.	Min.	Max.
Truck Characteristics				
Quantity	82.98	38.87	13	191
Price	9.48	1.47	7	16
Twitter Followers	1,385.20	2,228.03	0	14,000
Yelp Review	3.70	0.77	1	5
Market Share	0.0064	0.0062	0.0006	0.0614
Location Level Characteristics				
Total Consumers Served	1,035.55	346.445	272	1717
Potential Market	15,705.13	4,856.113	2,898	24,583
Outside Option Share	0.93	0.03	0.81	0.98
Mean Daily Earnings	237.74	19.25	217.04	308.59

Note: Outside Option Share is computed as $\frac{\text{no. of primary jobs} - \text{total consumers served}}{\text{no. of primary jobs}}$

Table 1.2: Summary Statistic from the Demand Data

Another data set that was compiled and used in the estimation of the model is Twitter data scraped from the DC food truck location aggregator *www.foodtruckfiesta.com*. This data set contains individual food trucks location choices that have been announced via the truck's Twitter account. In this data set I observe more locations

than the ones I have collected demand data from and have included in my model. The locations that I model account for approximately 86% of the Twitter data location choices. The locations I have modelled were chosen to include all of the lottery, and comparable key non-lottery locations. This means that the locations that are not modelled are fringe DC locations and also university-oriented locations such as George Washington University (Foggy Bottom) and Georgetown University (Georgetown). Due to the fact that the researcher does not observe exactly how the lottery is conducted it is impossible to simulate over different lottery outcomes, so my model takes the lottery outcomes as given (i.e. the trucks choice sets are fixed when the trucks are making their location choices). This implies that to complete this Twitter data, I must construct the choice set for each truck. I do this by matching the monthly DCRA lottery outcomes and the Twitter data over 4 months spanning July to October, 2017. This was a combination of a fuzzy string matching problem[9] and deducing the identity of trucks by matching where a truck has tweeted to be and where the lottery outcome suggests a truck to be. I have managed to match about 70% of all the trucks that appear in the Twitter data and these trucks supplemented with the location and truck characteristics from the above-mentioned data sources make up the final Twitter data set.

---

[9]This exercise involves scoring each pair of truck names in the Twitter data and the lottery outcomes based on similarity of their string names. Once I obtained these scores I investigated each Twitter data truck's top 3 similar lottery data truck names to find the correct match.

	June 2017 MRV Location Lottery Results					
Site Permit	Business Name	Monday	Tuesday	Wednesday	Thursday	Friday
VSP-01045	#222	Navy Yard/Capital River Front	OFF	OFF	Union Station	OFF
VSP-00170	A & J Food Enterprise L.L.C.	Union Station	OFF	OFF	Virginia Ave (State Dept)	Noma 2
VSP-01124	Abunai LLC	Noma 1	L'Enfant Plaza	OFF	Union Station	OFF
VSP-00019	Adil Ouinaksi	OFF	Union Station	OFF	Franklin Square 13th St	OFF
VSP-01097	Ali Baba	OFF	Franklin Square 13th St	OFF	Noma 2	OFF
VSP-00960	American & Mediterranean Food	OFF	Navy Yard/Capital River Front	OFF	OFF	L'Enfant Plaza
VSP-01266	American & Mediterranean Food	L'Enfant Plaza	OFF	Virginia Ave (State Dept)	OFF	Franklin Square 13th St
VSP-01207	Annie Mae's Cafe Express, LLC	OFF	OFF	OFF	Navy Yard/Capital River Front	OFF
VSP-01265	Arepa Station	OFF	Farragut Square 17th St	OFF	Union Station	OFF
VSP-00626	Arepa Zone LLC	OFF	Metro Center	OFF	Virginia Ave (State Dept)	OFF
VSP-00800	Astro Doughnuts & Fried Chicken	OFF	Patriots Plaza	OFF	Union Station	OFF
VSP-01217	Bahya Burger	Navy Yard/Capital River Front	OFF	OFF	Farragut Square 17th St	OFF
VSP-00336	Ball Or Nothing	OFF	Union Station	OFF	Franklin Square 13th St	OFF
VSP-00157	BBQ Bus	OFF	Metro Center	OFF	Noma 2	OFF
VSP-00986	Bel-Feast LLC	Virginia Ave (State Dept)	OFF	OFF	L'Enfant Plaza	OFF
VSP-01180	Best Kabob, Geno's And Pat Steak And Cheese	L'Enfant Plaza	OFF	Metro Center	OFF	Noma 2
VSP-01235	Best Kabob, Geno's And Pat Steak And Cheese	OFF	OFF	L'Enfant Plaza	OFF	Metro Center
VSP-00156	Bite 2 Go & Four Guys	L'Enfant Plaza	OFF	Virginia Ave (State Dept)	Navy Yard/Capital River Front	OFF
VSP-00877	Bite 2 Go & Four Guys	OFF	Virginia Ave (State Dept)	OFF	OFF	L'Enfant Plaza
VSP-01255	Bombay Cuisine, Downtown Chicken	OFF	OFF	L'Enfant Plaza	OFF	Union Station
VSP-01303	Bombay Cuisine, Downtown Chicken	Farragut Square 17th St	OFF	Franklin Square 13th St	OFF	Virginia Ave (State Dept)
VSP-00020	Borinquen Lunch Box	OFF	Union Station	OFF	Navy Yard/Capital River Front	OFF
VSP-01055	Bowtie Pastaria	Franklin Square 13th St	OFF	OFF	L'Enfant Plaza	Noma 2
VSP-01301	Cache Cuisine	Navy Yard/Capital River Front	OFF	OFF	Patriots Plaza	OFF
VSP-00979	Campfire	Franklin Square 13th St	Union Station	OFF	Metro Center	OFF
VSP-00338	Capital Chicken & Waffles	Farragut Square 17th St	Patriots Plaza	OFF	Metro Center	OFF
VSP-00227	Capmac LLC	OFF	Farragut Square 17th St	OFF	Franklin Square 13th St	OFF
VSP-00182	Captain Cookies & The Milk Man	L'Enfant Plaza	OFF	Farragut Square 17th St	OFF	OFF
VSP-00183	Captain Cookies & The Milk Man	OFF	Franklin Square 13th St	OFF	OFF	Virginia Ave (State Dept)
VSP-00830	Captain Cookies & The Milk Man	OFF	OFF	Metro Center	OFF	L'Enfant Plaza
VSP-00257	Carnivore BBQ	OFF	OFF	Farragut Square 17th St	OFF	Union Station
VSP-00259	Carnivore BBQ	OFF	Virginia Ave (State Dept)	OFF	OFF	Farragut Square 17th St
VSP-00287	Carnivore BBQ	OFF	OFF	L'Enfant Plaza	OFF	Franklin Square 13th St
VSP-00770	Carolina Q	OFF	Farragut Square 17th St	OFF	Metro Center	OFF
VSP-01257	Charming Nomad	Noma 2	L'Enfant Plaza	OFF	Farragut Square 17th St	OFF
VSP-00207	Chat Pat	Franklin Square 13th St	OFF	Noma 1	Metro Center	OFF
VSP-00047	Chef Alex/ Coco Loco	L'Enfant Plaza	OFF	Farragut Square 17th St	OFF	Franklin Square 13th St
VSP-01194	Chef Alex/ Coco Loco	OFF	Franklin Square 13th St	OFF	OFF	L'Enfant Plaza

Figure 1.4: The official DC Department of Consumer and Regulatory Affairs (DCRA) lottery outcomes.

The page from the actual lottery outcome of the month of June, 2017 available at DCRA's vending web page[10] is shown in Figure 1.4. It lists the site permit number, the name of the business, and the markets business is allowed to operate in on each day of the week. To register to enter and operate at these locations there is a total fixed cost of $175/month$ which is the sum of a $25 entry fee and a $150 location site permit fee [DCRA, 2013]. A truck can choose to not pay the permit fee and forgo its allocation, also trucks may trade their location on a given lottery draw with another truck if the DCRA approves the trade. This may explain some

---

[10]https://dcra.dc.gov/mrv

of the misfit of my model to the data as the rejection of the allocation post lottery entry and permit swaps are not modelled. My model assumes that the lottery outcomes we see are strictly adhered too and there is no convincing evidence that leads me to believe that a significant portion of the location allocations are traded or rejected. I observe most of the trucks that are supposed to be at the location inferring from the lottery to be at the location once I arrive to collect the data. From my demand data I can identify 68% of the trucks I actually observe are supposed to be there according to the lottery. Also, the non-adherence is typically not in the form of a different truck being at the location but a designated truck not showing up (i.e. location has fewer trucks than the lottery allocation suggests) implying that permit trades don't compose a significant share of my observations. There is a total of 139 designated lottery spots across the 9 locations and various measures of the total number of trucks operating in Washington, DC suggests there are many more trucks in total. The food truck location aggregator *www.foodtruckfiesta.com* lists 245 trucks listed as being permitted to operate in DC. In the lottery outcomes posted by the DCRA, I consistently count more than 220 individual site permits allocated through the week.

Given that I don't observe the whole industry (due to lack of man power and data collecting resources) I would want the Twitter data and the demand data to both be representative of the whole industry. Comparing distribution of cuisine genres which I have divided into 7 categories (dessert trucks have been dropped from the analysis), *American, Asian, Caribbean, Exotic, Indian, Latin American* and *Mediterranean*, the observations from the Twitter data and the demand data seem to represent quite a similar sample of the population as shown in Table 1.3. One exception is in the *Indian* genre which tends to have less online presence than

other genres, hence also Tweeting less.

	Demand Data		Twitter Data	
Genre	**Freq.**	**%**	**Freq.**	**%**
Asian	60	18.13	516	16.14
American	118	35.65	1,073	33.55
Mediterranean	64	19.34	678	21.20
Latin American	38	11.48	387	12.10
Indian	23	6.95	66	2.06
Caribbean	20	6.04	223	6.97
Exotic	8	2.42	67	2.10
Total	331		3,198	

Table 1.3: Comparison of the distribution of cuisine genres within each data set

Comparing the two data sets across the number of Twitter followers and Yelp reviews, the Twitter data seems to be capturing trucks that have on average more followers and higher Yelp reviews. This may be due to the fact that it is more likely that a truck with more online presence is more likely to be broadcasting their daily location through Twitter.

Truck heterogeneity and Quality

Trucks differ in both the quality of the product offered and their cuisine type. The paper focuses on taking into account these dimensions of heterogeneity in assessing regulatory policy. Figure 1.5 describes these features in the demand data. From Table 1.3 we can see that there are overwhelmingly more *American* genre trucks which are trucks that serve burgers, sandwiches, pizzas, Barbecue, etc. *Mediterranean* and *Asian* trucks are a distant second and third. Now considering this

23

together with the information in Figure 1.5, we can see that while there are an abundant number of *Mediterranean* trucks, they don't seem to be very active on social media (less Twitter followers on average) and also seem to have a wider distribution of Yelp reviews compared to the *Asian* trucks. This kind of heterogeneity will imply heterogeneous impacts given a change in regulatory regime. In particular, we could expect that the low quality metric genres like *Indian* and *Mediterranean* trucks along with abundant genre trucks like *American* trucks to be affected most heavily by a regime that does away with the lottery.
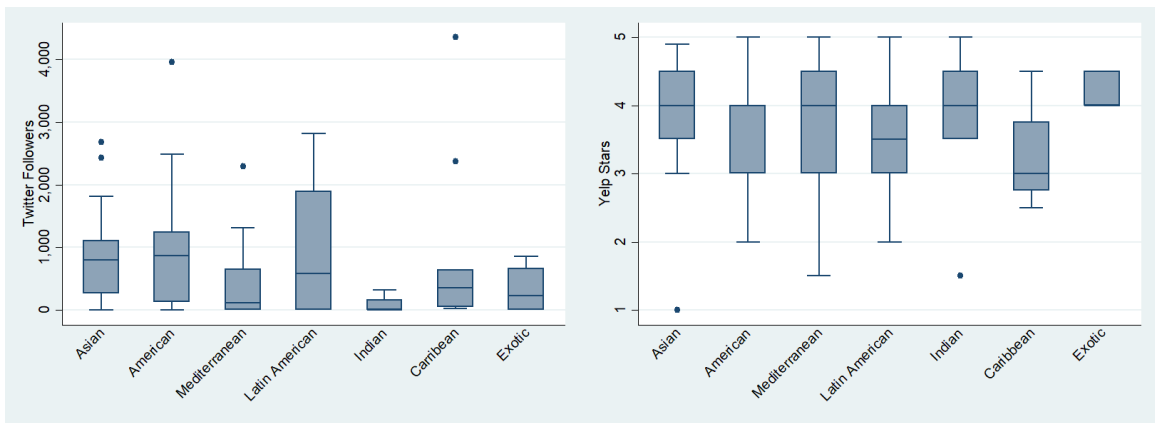
(a) Number of Twitter followers by Cuisine Genre
(Accounts with followers above 5000 excluded)

(b) Yelp Review Stars by Genre

Figure 1.5: Descriptive Truck Heterogeneity

1.4 Model

1.4.1 Demand

Suppose a location $l \in \{1, ..., L\}$ has entry cost c_l, market size M_l, capacity constraint μ_l, and mean income $\bar{Y}_l$. $g \in \{0, .., G\}$ denotes different cuisine types and $g = 0$ denotes the outside option. $j \in \{1, ..., J\}$ is an individual truck with quality to consumers v_j and marginal cost mc_j. Hence, a triplet $(v_j, g(j), mc_j)$ defines an individual vendor and these parameters are all exogenous. The model does not

consider cuisine genre choice of the food truck entrepreneur. Finally, let $\Gamma_{g(j)l}$ denote the set of all trucks at location l with the same genre as truck j.

Given the firms' prices and location outcomes (which will be described below), consumers at the location have nested logit demand with nests formed around the genre. This implies that the indirect utility of consumer i eating in location l, consuming truck j's product is:

$$U_{lij} = v_j + \alpha \ln(\bar{Y}_l - p_j) + \xi_{lj} + \zeta_{ig(j)} + (1 - \sigma)\epsilon_{ij} \tag{1.1}$$

and the indirect utility of consuming the outside good is:

$$U_{li0} = \alpha \ln(\bar{Y}_l) + \zeta_{i0(j)} + (1 - \sigma)\epsilon_{i0} \tag{1.2}$$

Where ϵ is i.i.d. Type I extreme value across products, $\zeta_{ig(j)}$ is the group specific taste of the consumer, and σ measures the relative weight of idiosyncratic and group preferences. Empirically, v_j will be parameterized by the truck's characteristics like Yelp reviews and the number of Twitter followers. Note that in this chapter including an income effect is important for two reasons. Firstly, it adds more variation in the variable such that I can estimate α. Secondly, buying lunch for consumers is a daily decision and is plausibly budgeted given the consumer's income.

The above setup implies location market shares:

$$
S_{jl} = \frac{\exp\left(\frac{v_j + \alpha \ln(\bar{Y}_l - p_j)}{1-\sigma}\right) \left(\sum_{k \in \Gamma_{g(j)l}} \exp\left(\frac{v_k + \alpha \ln(\bar{Y}_l - p_k)}{1-\sigma}\right)\right)^{-\sigma}}{\sum_{g=0}^{G} \left(\sum_{k \in \Gamma_{gl}} \exp\left(\frac{v_k + \alpha \ln(\bar{Y}_l - p_k)}{1-\sigma}\right)\right)^{1-\sigma}} \tag{1.3}
$$

1.4.2 Trucks

The Truck's expected profit[11] maximization is as follows. Firstly, he chooses a price, and then given this price and vendor's conjectures of the cuisine genre and location specific market shares, he selects a location to enter which I model as a standard logit discrete choice problem with zero profits if the choice is to not enter any locations. In other words, profits of vendor j at location l are:

$$
\Pi_{jl}\left(p_j, p_{-j}, v_j, mc_j, g(j), M_l, \bar{Y}_l\right) = (p_j - mc_j) S\left(p_j, p_{-j}, v_j, g(j), \bar{Y}_l, \sigma\right) M_l
$$
$$
- c_l + \lambda(\epsilon_{jl} - \epsilon_{j0}) \tag{1.4}
$$

The λ is the scaling parameter[12] and the ϵ are Type I Extreme Value. This ordering of events, where price choice happens before location choice is an innovation to the literature. An additional exogenous state of the model is the set of locations that each truck can choose from and the number of parking spots at each location. This location choice set is governed by the policy regime and the number of parking spots by physical space. In the data, this is the DCRA lottery outcome and respective parking space allotments as seen previously in Figure 1.2

---

[11]Expectations are taken over the truck's own location choice probabilities and rival's location probabilities through some equilibrium conjecture about cuisine genres and location specific market shares which will be introduced below.

[12]Since the "utility" of the truck is in actual dollars, the scale is important here

and the number of street parking spaces at the locations on non-lottery locations. Backwards induction implies the following maximization problem for truck j where L_j denotes truck j's location choice set:

$$\max_{p_j} \sum_{l \in L_j} \frac{\exp\left(\frac{\hat{\Pi}_{jl}\left(p_j, p_{-j}, v_j, mc_j, g(j), M_l, \bar{Y}_l, c_l, \sigma\right)}{\lambda}\right)}{1 + \sum_{l \in L_j} \exp\left(\frac{\hat{\Pi}_{jl}\left(p_j, p_{-j}, v_j, mc_j, g(j), M_l, \bar{Y}_l, c_l, \sigma\right)}{\lambda}\right)} \ldots$$

$$\ldots * \hat{\Pi}_{jl}\left(p_j, p_{-j}, v_j, mc_j, g(j), M_l, \bar{Y}_l, c_l, \sigma\right) \tag{1.5}$$

$$s.t. \quad E[n_l | c_1, ..., c_L] \le \mu_l \tag{1.6}$$

$E[n_l | c_1, ..., c_L]$ denotes the expected number of trucks at location l given the entry costs of the locations, note that the entry costs are the parameters which in equilibrium keep the number of trucks at a location feasible. More specifically, the entry costs affect the location choice probability and the average location choice probability determines the expected number of trucks at the location. An intuitive motivation for the entry costs is that it reflects the *value* of an additional spot added to a location. The exogenous μ_l is effectively the fixed supply of parking spots, while $E[n_l | c_1, ..., c_L]$ is the demand for parking spots in the market. The difference between these two variables can be thought of as the excess demand for parking at a location, and the vector of c_l i.e **c** will be such that the market in equilibrium clears. With this motivation, the interpretation of the entry costs become clear and practical. For example, if $c_1 = 100$, this would imply that a truck is willing to pay up to \$100 for an additional free parking spot at location 1. It is the fixed costs of securing a spot at a location relative to entering a fringe location where capacity doesn't matter. An alternative way to think about the

27

entry costs is to think of them as some kind of congestion cost arising out of excess demand. Too many trucks flock to one location inducing costs to the trucks if they are to secure a spot and do business. The entry costs in equilibrium will settle on a value such that a feasible number of trucks enter each location in expectation.

$\hat{\Pi}$ denotes expected profits at each location. To calculate this entity, you need to know the probability density for the numerous market configuration outcomes that can occur. This is intractable with over 200 vendors operating in the DC area, at many locations, several cuisine genres, and differing qualities. I assume that the exact price and location choices of the other truck vendors are not observed by the trucks, however an equilibrium conjecture of the 'inclusive values' provides a sufficient statistic that reflects the number of rival vendors and their genres, rival's prices and values. This assumption renders the problem tractable.

1.4.3   Equilibrium

Definition 1. An equilibrium is a $G \times L$ matrix of conjectures of I_{gl} denoted $\hat{I}_{gl}$ where:

$$I_{gl} = \ln \left(\sum_{k \in \Gamma_{gl}} \exp \left(\frac{v_k + \alpha \ln(\frac{\bar{Y}_l - p_k}{\bar{Y}_l})}{1 - \sigma} \right) \right)$$

and,

$$\hat{I}_{gl} = \frac{\sum_{s=1}^{S} I_{gls}}{S}$$

where S is the number of simulated outcomes given **a)** a set of prices $(J \times 1)$ such that:

$$\max_{p_j} \sum_{l \in L_j} \frac{\exp\left(\frac{\hat{\Pi}_{jl}\left(p_j, p_{-j}, v_j, mc_j, g(j), M_l, \bar{Y}_l, c_l, \sigma\right)}{\lambda}\right)}{1 + \sum_{l \in L_j} \exp\left(\frac{\hat{\Pi}_{jl}\left(p_j, p_{-j}, v_j, mc_j, g(j), M_l, \bar{Y}_l, c_l, \sigma\right)}{\lambda}\right)} \cdots$$

$$\cdots * \hat{\Pi}_{jl}\left(p_j, p_{-j}, v_j, mc_j, g(j), M_l, \bar{Y}_l, c_l, \sigma\right)$$

and is solved and **b)** entry costs $(L \times 1)$ such that location choice probabilities simulate outcomes which satisfy:

$$E[n_l | c_1, ..., c_L] \leq \mu_l$$

Specifically, the assumptions that I make is that the commonly perceived values of the inclusive values are formed as the average of actual outcomes of these inclusive values. Each individual firm does not perceive its own impact on the $\hat{I}_{gl}$ and the restriction of the equilibrium concept to focus on the average of the inclusive values means that the market share $(\hat{S}_{gl})$ is an approximation of the nested logit shares. This greatly simplifies computation. I forward simulate to compute each $\hat{I}_{gl}$.

The equilibrium is computed with the following iterative process:

On $iter1 = 1$ and $iter2 = 0$:

1. Guess a matrix $\hat{I}^{iter2=0}$ of size $G \times L$ and vector $c^{iter2=0}$ of size $L \times 1$. Set $p^{iter2=0}=0$ and set up a grid of v and g.

29

2. Draw S simulations from the random events of the model (i.e. Realizations of the trucks' logit location errors), and set tolerances ϵ_c, ϵ_I, and ϵ_p to assess convergence. Update $iter2 = 1$.

Then for $iter2 = m > 0$.

3. Set $iter1 = 1$, set $\hat{I}^{iter2=m-1} = \hat{I}^{iter1=1}$ and $c^{iter2=m-1} = c^{iter1=1}$. Now given $\hat{I}^{iter1=1}$ and $c^{iter1=1}$ solve for the profit maximizing prices ($p^{iter2=m}$) for the firm using $\hat{S}$, on a grid of genres (g), qualities (v), and marginal costs (mc).

Then for $iter1 = k > 0$

(a) Given $\hat{I}^{iter1=k}$ and $p^{iter2=m}$, solve the minimization problem $\min_c \sum_{l=1}^{L} \left(E[n_l|c_1, ..., c_L] - \mu_l \right)^2$. Forward simulate using the S simulation draws for obtaining $E[n_l|c_1, ..., c_L]$. Denote the solution $c^{iter1=k+1}$

(b) Given the solution $c^{iter1=k+1}$ compute $\hat{I}^{iter1=k+1}$ by forward simulating S times and taking the average over the simulated Is.

(c) Compute the distance.[13] between $\hat{I}^{iter1=k+1}$ and $\hat{I}^{iter1=k}$. Compute the distance between $c^{iter1=k+1}$ and $c^{iter1=k}$.

(d) If the distances in the previous step are less than ϵ_I and ϵ_c respectively, stop and save ($c^{iter1=k+1} = c^{iter2=m}$ and $\hat{I}^{iter1=k+1} = \hat{I}^{iter2=m}$). Otherwise return to the nested step a) with the updated inclusive values and costs ($c^{iter1=k} = c^{iter1=k+1}$ and $\hat{I}^{iter1=k} = \hat{I}^{iter1=k+1}$)

4. Compute the distance between $p^{iter2=m-1}$ and $p^{iter2=m}$, $c^{iter2=m-1}$ and $c^{iter2=m}$, and $\hat{I}^{iter1=m-1}$ and $\hat{I}^{iter1=m}$).

5. If the distances computed in 4. is less than ϵ_p, ϵ_c, ϵ_I respectively, stop.

---

[13]Note that one may choose different updating rules or distance measures for the iteration. I update the relevant variables in each iteration such that the latest iterations are a weighted average of the current iterations values and the past iterations values. I do this such that the model outcomes are not too volatile between iterations.

Otherwise update price vector $(p^{iter2=m} = p^{iter2=m-1})$. Update the index $iter2 = m + 1$ and return to step 3.

With 224 trucks, 7 cuisine genres, 13 (12 actual locations and 1 outside option) locations, tolerances set at 0.0001 for prices, entry costs, and the inclusive values solving the model takes approximately 17 minutes to simulate and solve.

1.5 Estimation

Estimation of the model will proceed in multiple stages:

1. Estimate demand parameters (α, β, σ) using demand data.

2. Estimate a reduced form conditional logit model of location choice using the Twitter data.

3. Using the estimates from Stage 1 and 2 and the trucks first order conditions back-out the marginal costs ($J \times 1$ vector **mc**) for a set of observed trucks and lottery outcomes.

4. Match model moments derived from estimates in stage 1 and stage 3 to moments generated from stage 2 to obtain estimates for entry costs and the scaling parameter ($L \times 1$ vector **c**, λ).

I will consider each stage in more detail in the following subsections.

1.5.1 Demand

I estimate demand using the realized observations after the firms have entered a location assuming that this is the equilibrium outcome given the entry game, more heterogeneity in the consumers was not modelled or estimated, for example, in the form of random coefficients because the data I am using to estimate the demand model doesn't allow me to observe the empirical distribution of consumer

characteristics.[14] Due to Berry [1994] we know there is a simple inversion for the nested logit model. Performing the inversion on the indirect utility formulation outlined in the previous section, the demand model that I estimate is the following:

$$\ln(s_{jl}) - \ln(s_{0l}) = X_j\beta - \alpha \ln(\frac{\bar{Y}_l - p_j}{\bar{Y}_l}) + L_l + G_j + \sigma \ln s_{j,g(j)} + \xi_{jl} \qquad (1.7)$$

X_j denotes truck characteristics such as Yelp reviews and the number of Twitter followers. $\bar{Y}_l$ denotes the location specific average daily earning and p_j is vendor j's price as before. L_l is a location dummy variable, G_j is a genre dummy variable and $s_{j,g(j)}$ denotes vendor j's share within his genre nest. The endogeneity of price is a general issue in demand estimation but on top of this I have the issue of there being very little price variation between and within food trucks. I alleviate these issues by including the difference between mean daily worker (consumer) earnings for a certain location and the truck's price in the utility function and to identify this *earnings-price* coefficient. The intuition is that I will compare the relative market shares of two trucks that are observed to compete in two different locations that have different mean earnings. This strategy adds variation and also somewhat deals with the endogeneity of price, as we are looking at the change in relative market share between two trucks and it is reasonable to assume that the quality differential between these two trucks shouldn't change with mean earnings. I try a variety of specifications but find that including truck specific fixed effects give me the most favorable estimates which I choose to use (specification (1) in Table 1.4).[15] The reason for this is as Nevo [2001] describes. I observe the same

---

[14]For example, income data from the LODES data set is top coded such that there is no meaningful variance data do be used.

[15]Robustness checks including day of the week dummies, controlling for the weather don't seem to make a significant difference on the key parameters of interest.

truck in multiple locations I can include these fixed effects and because the truck specific unobservables are not changing as the location of the truck changes the fixed effect accounts for these unobservables that may be correlated with price. Subsequently I project these fixed effect estimates on to the observed truck specific characteristics to obtain estimates for β.

In order to estimate the nesting parameter, I must instrument for the within group market share. I deal with this by using the DCRA lottery outcome. This source of randomization of vendors to locations gives me an exogenous number for how many vendors of a certain cuisine genre may be present at a given location. In my estimation, I have found this to be a good instrument with first stage $F-stat = 27.27$. Staiger and Stock [1997] suggest that instruments be declared weak if the first-stage $F-stat < 10$ and the first stage in my estimation clearly passes this rule of thumb. The demand estimates are shown in Table 1.4 and 1.5.

	(1) FE	(2) 2SLS	(3) 2SLS
$\ln s_{j,g(j)}$	0.241	0.158	0.195
	(0.178)	(0.127)	(0.124)
$\ln(\frac{\bar{Y}_l - p_j}{\bar{Y}_l})$	33.16	23.44***	21.87***
	(20.52)	(4.457)	(4.112)
Twitter followers		4.33e-05***	4.82e-05***
		(1.15e-05)	(1.08e-05)
Yelpstars			0.117***
			(0.0292)
Constant	-4.397***	-5.047***	-5.073***
	(0.729)	(0.323)	(0.252)
Genre FE	No	Yes	Yes
Yelp FE	No	Yes	No
Location FE	Yes	Yes	Yes
Truck FE	Yes	No	No
Observations	331	305	305
R-squared	0.865	0.673	0.681

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.4: Nested Logit Estimation Specifications

Table 1.5 shows the full specification including the location fixed effects. The location fixed effect coefficients suggest that *Navy Yard*, *Federal Center*, and *State Department* are the top three highest demand locations, which reflects the lack of outside options at these locations. Also, the fixed effects are driven by my assumption of the market size (i.e. using the number of primary jobs at each location for market size). For example, the high fixed effects in small markets such as *Navy Yard* may be due to the specified total market being very small, implying that the food truck segment of the market will be large. Table 1.6 shows the estimates of the projection of the fixed effect estimates on individual characteristics such as the number of Twitter followers, Yelp stars and genre. I will use these estimates in my

counterfactual simulations and in the process of backing out the marginal costs. The average price elasticity implied by the point estimates of α and σ is -1.61 (standard error $= 0.7351$). The nesting coefficient is estimated to be 0.241 and is feasible and consistent with utility maximization (i.e. between 0 and 1) however it is not significantly different from zero (*p-value=0.116*). I suspect with more observations (especially with the truck level fixed effects) these estimates will have tighter confidence intervals, but I proceed with these point estimates to the next stages of estimation. Other observations that can be made is that the *Asian* trucks tend to on average have larger market shares than any other genres. The signs on the number of Twitter followers and Yelp review stars are positive as expected.

	Preferred Specification
$\ln s_{j,g(j)}$	0.241
	(0.178)
$\ln(\frac{\bar{Y}_l - p_j}{\bar{Y}_l})$	33.16
	(20.52)
19th & L	1.029***
	(0.270)
CNN (First St NE)	0.949***
	(0.241)
Farragut Square	0.195
	(0.281)
Federal Center / Patriots Plaza	1.919***
	(0.506)
Franklin Square	0.711***
	(0.258)
Gallery Place / Chinatown	1.038***
	(0.307)
L'Enfant	0.891***
	(0.271)
Metro Center	-0.0350
	(0.310)
NOMA	1.625***
	(0.307)
Navy Yard	2.514***
	(0.426)
State Department (20th & Virginia Ave NW)	1.711***
	(0.266)
Union Station	1.043***
	(0.292)
Constant	-4.488***
	(0.729)
Observations	331
Number of Truck_id	154

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.5: Location F.E. from Consumer Demand Estimates

	Projection of Ind. F.E on Characteristics
Yelpstars=2	0.134
	(0.0866)
Yelpstars=3	0.376***
	(0.0629)
Yelpstars=4	0.392***
	(0.0629)
Yelpstars=5	0.450***
	(0.109)
American	-0.155***
	(0.0516)
Mediterranean	-0.624***
	(0.0582)
Latin American	-0.465***
	(0.0653)
Indian	-0.682***
	(0.0636)
Caribbean	-0.544***
	(0.0809)
Exotic	-0.607**
	(0.236)
Twitter followers (000s)	0.478***
	(9.87e-06)
Constant	-0.0414
	(0.0692)
Observations	305
R-squared	0.516

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1.6: Truck Fixed-Effects OLS Projection on Individual Characteristics.

1.5.2 Conditional Logit Location Entry Probabilities

Using with Twitter data, I fit a conditional logit model. Truck j's utility from entering location l is modeled as:

$$V_{jl} = L_l^{Stage2} + G_j^{Stage2} \times L_l^{Stage2} + X_j \beta_l^{Stage2} \times L_l^{Stage2} + \epsilon_{jl} \qquad (1.8)$$

ϵ_{ij} are Type I extreme value errors. Estimating the model above is simply a conditional logit regression with a dummy variable indicating the location choice made on the left-hand side with a full set of interactions between truck specific characteristics (cuisine genre, Twitter followers, and Yelp stars) and location dummies on the right-hand side. In the Twitter data, I see more locations tweeted by the trucks than I model. I model a total of 12 locations, and these locations account for approximately 86% of the locations visited by the trucks in the data. Locations that are not included are mostly in the fringes of central Washington, DC, one example of such a location is South Capitol (the most southern point in Figure 1.3). Other non-modeled locations include locations such as George Washington University (Foggy Bottom), Georgetown University (Georgetown) etc. I exclude these university-based locations as the markets that realized at these locations are different to the standard 'lunch shift' locations on many dimensions. In the empirical estimation, these non-modeled locations make up the outside option and net profits are normalized to be $500, this helps me pin down the scaling parameter as it provides a more stable outside option probability in the iterations where I am matching the moments for this parameter. In terms of interpretation, we can simply add this constant to the fixed entry costs once they have been estimated, to interpret the entry costs as a dollar value relative to the outside option being normalized to $0.

The specification that I proceed with has the full set of interactions as outlined

above but has the Yelp stars converted to a dummy variable for whether it is 'high' or 'low', where the category for 'high' is a rating of greater than 3.5 stars. Other specifications I've attempted to estimate faced issues in the estimation with non-convergence and non-convexity of the maximum-likelihood objective function. The model has an $Pseudo\ R^2$ of 0.1859. The coefficients of this estimation are not of interest directly and merely used to match with model moments in the next stages of estimation, hence I do not report the table of estimated coefficients. However, interesting patterns in the location choice probabilities can be observed to get a clear idea of the data moments that are being matched with the model moments for estimation of the location entry costs. Figure 1.6 shows a box plot of the estimated location entry choice probabilities. The figure excludes outliers to being out the relevant features of the estimates better. The vertical axis of Figure 1.6 shows the estimated choice probability of the location marked on the horizontal axis being chosen by a truck, given that the location is in his choice set. The variation in the mean of these estimates across locations will play a crucial role in the estimation of the location entry costs in the next estimation stage.
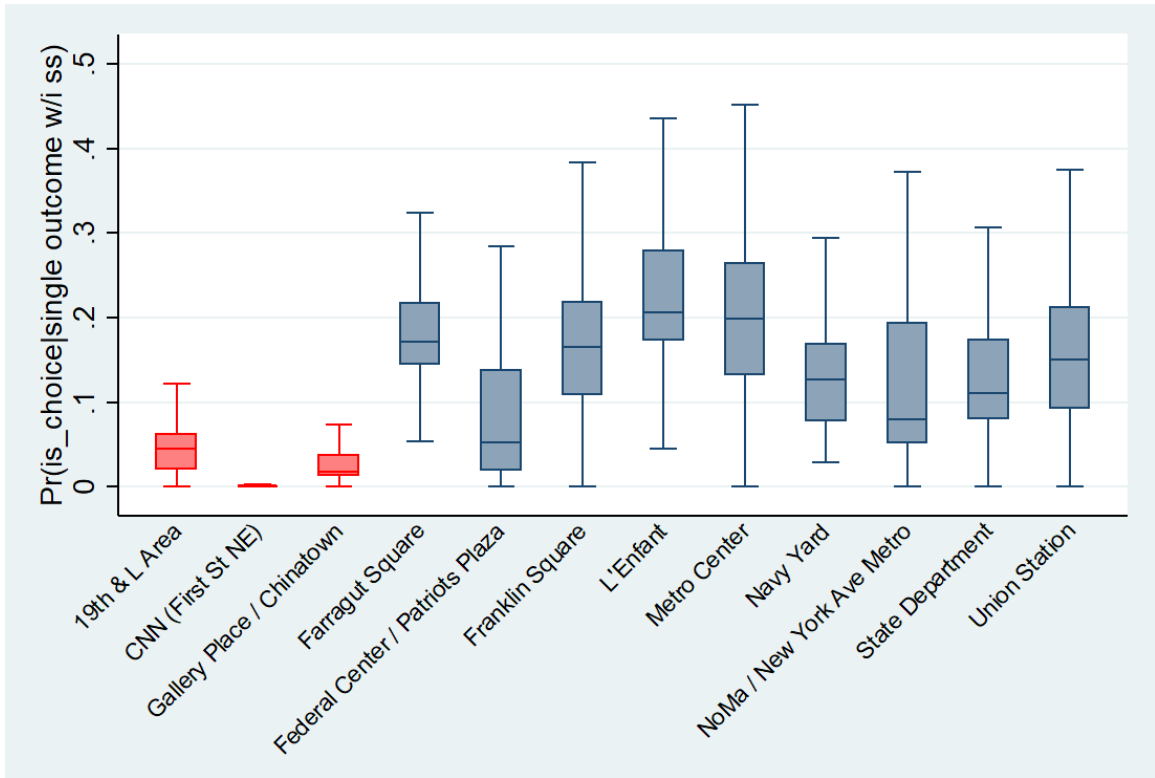
Figure 1.6: Variation in the Estimated Conditional Location Choice Probabilities by Locations.

The choice probabilities of the non-lottery locations (colored in red) are much lower than the lottery locations, this is because these locations appear in every truck's choice set but are not chosen as frequently as the locations that only appear in a truck's choice set if allocated by the lottery, which are chosen more frequently. In other words, trucks on a random day of the week, are more likely to be visiting a lottery allocated location over a non-lottery location given the lottery location is in the trucks' choice set. With respect to the structural model, these estimates imply that the implied entry costs will be lower in locations we observe with high choice probabilities and similarly, low choice probability locations will have high entry costs. This is discussed in more detail below.

1.5.3 Marginal Costs

In this stage I back out the constant marginal costs for a set of trucks. I model 224 individual trucks that I see consistently in the lottery outcome data. Given that www.foodtrukfiesta.com lists the total number of trucks permitted to operate in DC to be 245 trucks, my model contains the vast majority of the trucks that are operating consistently in Washington, DC.

Taking the first order conditions of the truck's maximization from above we get that the truck j's marginal costs are:

$$mc_j = p_j - \frac{\sum_l^L S_{lj} M_l Pr_j(l)}{\sum_l^L Pr_j(l) \frac{\partial S_{lj}}{\partial p_j} M_l} \tag{1.9}$$

Where M_l is the total number of primary jobs at location l and $Pr_j(l)$ is the location choice probability of truck j choosing to enter l which I estimate using a reduced form conditional logit model in Stage 2. Computing the marginal costs requires the derivatives:

$$\frac{\partial S_{lj}}{\partial p_j} = \frac{-\alpha S_{lj}(1 - \sigma S_{j|g} - (1-\sigma)S_{lj})}{(1-\sigma)(\bar{Y}_l - p_j)} \tag{1.10}$$

Which I simulate using the estimated conditional logit choice probabilities given lottery designated choice sets. The distribution of marginal costs and price-cost margins[16] in the DC food truck industry derived from my model are shown in

---

[16]price-cost margin $= \frac{p-c}{p}$

Figure 1.6. With these marginal costs, the average price-cost margin is calculated to be 0.6602. There is little past work on food truck competition and margins to compare this estimate with. However, to shed some light on its plausibility I interviewed an acclaimed industry leader. I learned that the general consensus rule of thumb in the industry was that for a truck truck to successfully operate in the long-run the food cost component of a menu item should be no higher than 1/3 of the price (i.e. a price-cost margin of 0.66). Interpreting the marginal costs in my model as the food costs required to make an additional order and assuming that labor costs are a longer-run decision,[17] the average price-cost margins that I find are strikingly similar to this number. In my estimates however, there is large variability in the price-cost margins but approximately 95% of the trucks have margins greater than 0.55.
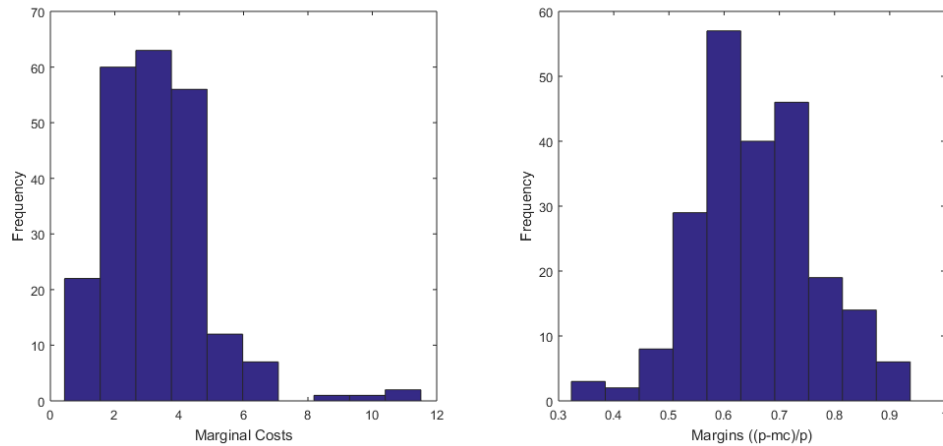
Figure 1.7: Marginal Costs ($) and Price Cost Margins. Recovered from Model First Order Conditions.

[17]It is plausible to think that the number of cooks on a truck are predetermined and they are paid by hours worked, not numbers of orders produced.

1.5.4 Location Entry Costs and Scale Parameter

All the estimates from Stages 1-3 are used to estimate the location specific entry costs and the scaling parameter. The scale parameter is usually unidentified solely with discrete choice data, however, in my model the profit function for the firm is not reduced form and the scale of the firm's utility is in dollars. Hence it is important to get the scale/variance of the conditional logit model correct.

The discrete choice model for the vendors can be written by re-scaling the location specific *profit - entry costs* (i.e. firm "utility") terms by λ. If we knew the entry costs, we could run a simple multinomial logit model on the actual firm utility and the coefficient on this variable will absorb the scale parameter. However, I don't observe the entry costs which also need to be scaled. I estimate the scaling parameter using the following definition and a fixed-point algorithm:

$$\lambda = \frac{\frac{1}{J} \sum_j E \left(\max_l \hat{\Pi}_{jl} - c_l \right)}{\frac{1}{J} \sum_j E \left(\max_l V_{jl} + \epsilon_{jl} - \epsilon_{j0} \right)} \tag{1.11}$$

This definition follows Anenberg and Kung [2015] and can be interpreted as defining the scale parameter as the ratio between the average expected maximum profit in dollar terms (the numerator), and the average expected maximum value in normalized utility terms (the denominator). More specifically, we this identity comes from the fact that the reduced form logit model (using reduced form profit functions) doesn't separately identify the scale parameter separately but estimates the product $\lambda\beta$. To identify λ we need to determine the scale of the reduced form profits.

Empirically, the expected maximum profit are the profits that we compute given the demand and marginal cost estimates. So, for some entry costs c_l, I will approximate the numerator as the average of estimated observed $profits - costs$ over the trucks I model. The denominator is obtained from calculating the truck's inclusive values from the estimation of the conditional logit model above.

Note that the algorithm will take the denominator as data, but in each iteration re-compute the numerator. The fixed-point algorithm for estimating the entry costs and the scaling parameter is as follows: Guess an initial value of $\lambda^{iter=0}$, set tolerances ϵ_c and ϵ_σ.

1. Scale each $\hat{\Pi}_{jl} - c_l$ by $\lambda^{iter=k}$.

2. Given the scaled $profit - costs$ (i.e. utility) solve the following minimization problem $\min_c \sum_{l=1}^{L} \left(Pr_l^{Model} - E[Pr_l] \right)^2$ to obtain a solutions vector of c which is a $L \times 1$ vector of c_ls.

3. Now given the solution vector of $\mathbf{c}$, compute the scaling parameter with the above definition for the scaling parameter and define as $\lambda^{iter=k+1}$.

4. Compute the distance between $\lambda^{iter=k+1}$ and $\lambda^{iter=k}$ check if it is less than the tolerance, if so, stop. otherwise update $\lambda^{iter=k} = \lambda^{iter=k+1}$ and repeat from 1.

Where:

$$Pr_l^{Model} = \frac{1}{J} \sum_{j=1}^{J} \frac{\exp\left(\frac{\hat{\Pi}_{jl}}{\lambda}\right)}{1 + \sum_{l \in L_j} \exp\left(\frac{\hat{\Pi}_{jl}}{\lambda}\right)}$$

This equation is a part of a system of $L + 1$ equations, one each for each modeled location and the outside option. Practically, the outside option that has the utility being normalized are the non-modelled locations. These include locations such as university car parks and more of the fringe DC locations marked on the

map shown in Figure 1.3. The total of 13 parameters to be estimated ($L = 12$ location specific entry costs and the scaling parameter) in this routine are the values of the parameters such that the model moments and the data moments are matched. Matching the model entry probabilities of firms and the observed probability of entry is similar to the strategy employed by Seim [2006] to obtain coefficient estimates on how the number of competitors within a distance band impact profitability. The difference is that in my model, I also need to consider the scaling parameter which requires an additional condition to pin down.

The estimates I obtain for the location specific entry costs are shown in Table 1.8. These estimates are taking the demand parameters as data, and have not been corrected for the imprecision in the demand estimation stage. We can observe the patterns in Figure 1.6, the lottery locations entry costs are in general lower than the non-lottery locations and also larger markets exhibit higher entry costs.

Locations	Lottery	Entry Costs ($)	GMM s.e.
19th & L Area	No	543.55	14.14
CNN (First St NE)	No	861.09	24.75
Farragut Square	Yes	122.64	6.96
Federal Center / Patriots Plaza	Yes	107.27	4.43
Franklin Square	Yes	142.02	7.27
Gallery Place / Chinatown	No	412.53	15.14
L'Enfant	Yes	76.23	7.53
Metro Center	Yes	76.40	6.48
Navy Yard	Yes	146.03	4.58
NoMa / New York Ave Metro	Yes	556.30	5.85
State Department	Yes	145.02	5.92
Union Station	Yes	159.95	6.54

Table 1.8: Estimated Entry Costs.

The estimates must be understood with careful consideration of the limitations

of the model. For example, for *CNN* the entry costs are estimated to be very high ($861.09), while this may be in line with the fact that this location is not part of the lottery, its market size is much smaller than *19th & L* which is also not part of the lottery but has a much lower entry cost. This implies that the Twitter data predicted choice probabilities for *CNN* are lower than the what the small market size in the model would imply. The reason seems to be that the model is not capturing the decisions of the trucks that are going to *CNN* well. In my Twitter data, I observe a few of the same trucks entering this location, maybe due to some type of dynamic reason such as customer base loyalty. Hence, the estimated choice probabilities for this location on average are quite low which then the model compensates by boosting the entry cost parameter. This is a issue that is present in all of the estimates but it seems to be more pronounced with the *CNN* entry cost estimate.

Another observation of note is that even among the lottery locations, there is considerable variation in the estimated entry costs. This suggests that despite the lottery in practice homogenizing the entry costs for designated locations (recall the $175 monthly total fee) these locations are still not valued equivalently to the trucks. There are events in reality where allocations awarded by the lottery are forgone and not entered, this behavior is unmodeled and inflates the entry cost parameter. The locations where the this seems to be an issue are the relatively less populous locations like *NoMa*.

1.6   Model Performance

Here I use the estimates discussed above and re-solve the model's equilibrium to assess how the model performs in explaining the data. Firstly, I consider consumer

46

utilities. Given the nested logit structure of my demand model, from [Train, 2009] we know that the expected utility for consumer i making a choice among alternatives in Γ_{gl} is:

$$I_{igl} = \ln \sum_{j \in \Gamma_{gl}} \exp(\frac{V_{ij}}{1 - \sigma}) \tag{1.12}$$

Figure 1.8 shows a scatter plot for quantity in equation (1.12) as predicted by the model and as implied by the location choice probabilities and prices observed in the data. The model captures the expected utilities quite well.
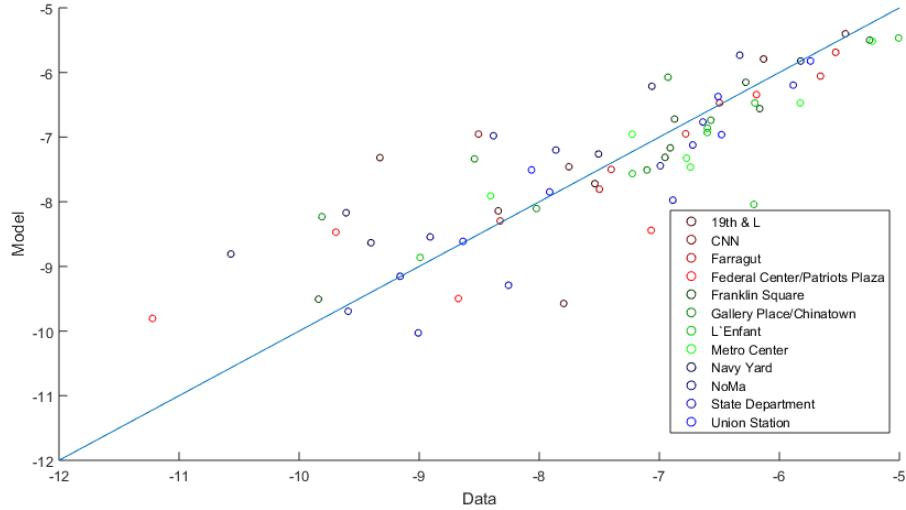
Figure 1.8: Model Fit of Location/Genre Specific Expected Utilities

Checking the model fit for prices (Figure 1.9 Panel (a)), we can see that the observed prices tend to be higher, also we see that the prices observed in the market are set in a almost discrete way (for example, 8.99, 9.99, etc.), and this bunching seen in reality throws the fit of the model off. It appears that the model is consistently predicting prices that are lower than we observe in the data. Given the bunching it is plausible that in reality the sellers round their price up to the

nearest 0.99 and this explains the consistent under-prediction. Given these features of the data, I believe the model fit is quite good. The market shares for the trucks at each location seem to be predicted well by the model too (Figure 1.9 Panel (b).
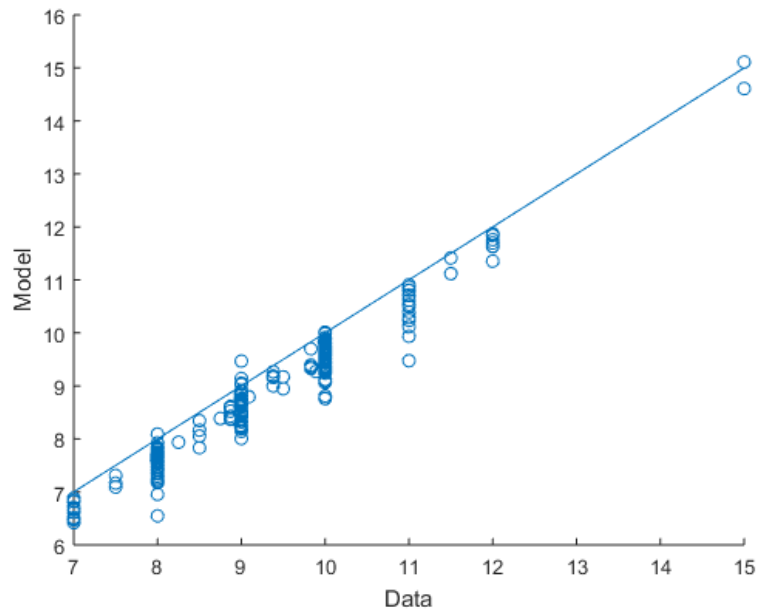
(a) Model Fit of Prices.
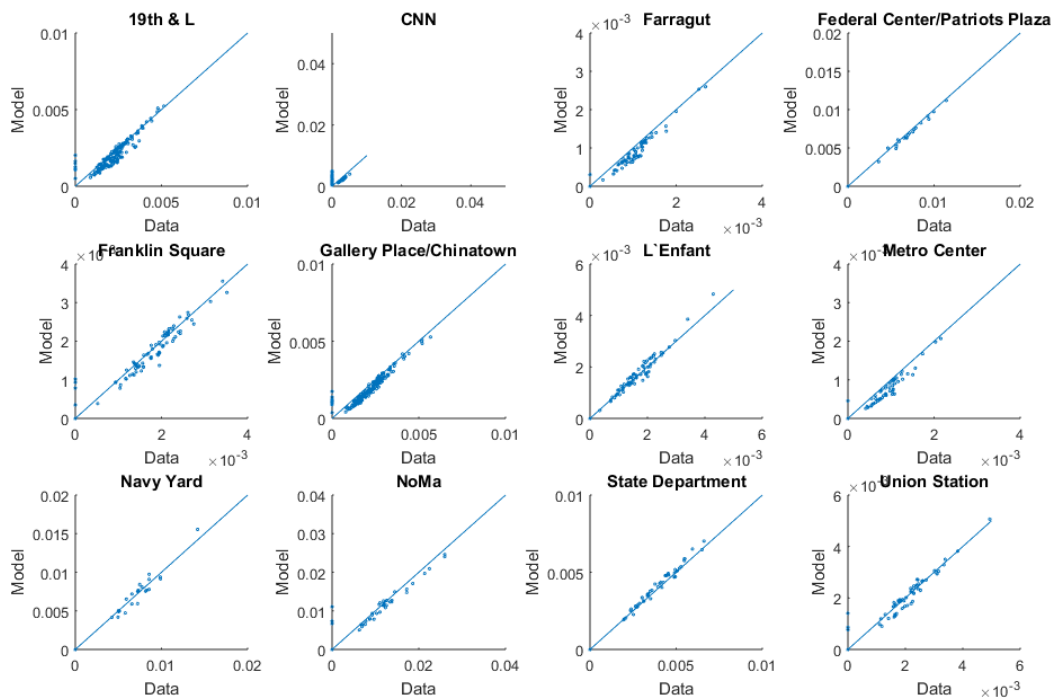
(b) Model Fit of Market Shares.

Figure 1.9: Model Fit at Competition Level (Second Stage of Model).

49

Figure 1.10 shows the panel of location choice fit checks for each location, with the conditional logit estimated choice probabilities from the Twitter data on the horizontal axis and the model predicted choice probabilities on the vertical axis. The model seems to do fairly well in some locations and less well in others. Recall, the issue with the CNN entry costs. This is reflected in the CNN panel of Figure 1.10. Firstly, the choice probabilities are small in general (all less than 0.03 percent) and also the conditional logit model predicts a near zero choice probability for a lot of trucks. The other location with predictions that don't seem to be doing very well is Gallery Place. Here for a lot of the trucks, the Twitter data implies a much higher choice probability than the model predicts. Further examination shows that the trucks that the model is not performing well for are disproportionately *Asian* and *Latin American* trucks. There must be some unmodeled reason that these genres of trucks enter non-lottery but very active locations with a higher probability than the other genres.
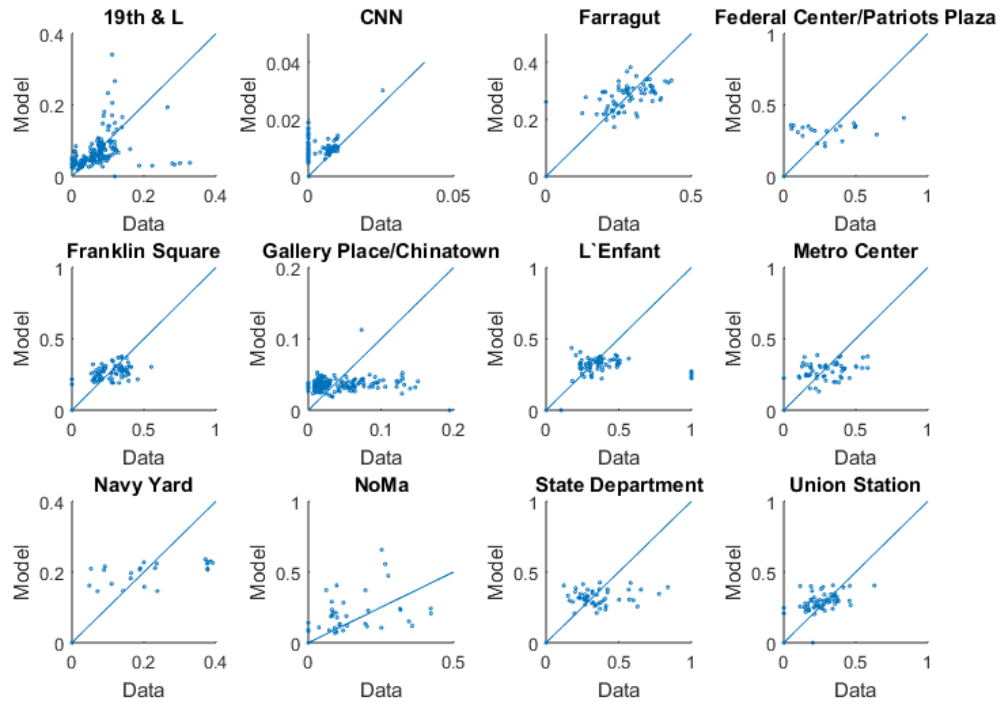
Figure 1.10: Model Fit of Location Choice Probabilities

1.7 Conclusion

In this chapter I have collected data on and estimated a model of the food truck industry in Washington, DC. Due to the novelty of this industry, regulators in practice fall back on ad hoc policy regimes without understanding the consequences of potential regulations and/or regulations already implemented. Theoretically, it is also a challenging landscape to model and analyze due to the large number of firms, locations, and externalities arising out of who the locations are the allocated to. Building on a large and well-developed literature dealing with endogenous product positioning and firm entry, chapter 1 offers a structural model as an example of how, despite the challenges, this industry can be analyzed and scrutinized in a

rigorous way.

I find that consumers preferences are only moderately correlated within cuisine genres ($\hat{\sigma} = 0.241$) and that trucks' own price elasticity's are on average about -1.61. Firms price-cost margin is on average 0.66, which closely resembles industry insiders rule of thumb for a truck that can operate into the foreseeable future. I find that the choice probabilities for lottery locations are higher than that of the non-lottery locations given both are in a truck's choices set. In my model, this implies that given the lottery, entry costs for the lottery locations are lower that of the non-lottery locations. However, the heterogeneity of each lottery location is still reflected in my estimates.

My model seems to perform well in predicting post-entry competition but is worse at predicting the location choice and entry stage of the model. This is due to the myriad of factors that the model fails to capture completely such as truck specific location choice decision rules, lottery allocations forgone and swapped by trucks, and any other general non-adherence to the allocation mechanism.

In chapter two I will use these estimates and the to consider two counterfactual policy regimes. Firstly, a scenario in which some locations are removed from the lottery and open to entry to all of the trucks and secondly, a counterfactual policy where the parking capacity of the non-lottery locations are increased. This counterfactual policy will be compared with the model prediction to assess the welfare effects of going from the current regime to another.

Chapter 2 Counterfactual Policy Analysis: Examining the Welfare Impacts of Washington, DC's Mobile Roadside Vending License Program

2.1 Introduction

The food truck industry in Washington, DC relied on a propriety lottery run by the DC Department of Consumer and Regulatory Affairs (DCRA), namely the Mobile Roadside Vending License Program. The program allocates registered trucks to a set list of locations (see Figure 1.2). I assess the welfare impacts of two alternative regulatory policies in the Washington, DC food truck industry. Firstly, a counterfactual policy where some locations are removed from the lottery, such that trucks are allowed to enter more locations without being subject to the lottery, up to the location specific capacity constraint being satisfied in expectation. Secondly, I will consider a counterfactual policy where the number of parking spaces are increased by 2 spots in each non-lottery location (*CNN*, *L'Enfant*, *Gallery Place* in my model). In Chapter 1 I have introduced and estimated the model that I will utilize in this counterfactual analysis. The importance of these counterfactual policy questions are not limited to academic curiosity, it is also practical for policy makers in many cities all over America. This dissertation chapter will assess real policies that can be implemented by a regulatory body. Such motivation is quite novel in the literature.

The initial motivation for the implementation of the current lottery system, was

both to protect brick-and-mortar stores and to stop trucks from engaging showing up earlier and earlier to secure a space at favorable locations. In 2018 it has now been 5 years since the inception of this policy and some negative side-effects seemed to have emerged. Not only are trucks continuing to show up early to secure locations in the non-lottery spots anyway [MacFarlane et al., 2018] while the DCRA is slow to respond by adding such locations to the official lottery. There seems to be a slowing down of the industry with truck owners reporting declining revenues. Anecdotal evidence suggests this is due to a lack of innovation and market shares being taken by food trucks that are similar and low quality but are able to stay in business anyway by simply obtaining the right to operate in the lottery designated locations, which are the most popular locations in DC [Hayes, 2018].

This anecdotal consensus combines both long-run and short-run arguments. In the long-run, allowing the market to be overrun by similar and low quality trucks is bad for the entire industry as the appeal of food trucks in general may decrease. In the short-run this suggests a benefit to some of the food trucks, as it is allowing the non-differentiated low quality trucks to survive. In my counterfactual analysis I focus on relatively marginal changes to policies. This is because my model is geared to predict short-run states of the market, as key important long-run aspects of the industry are not modelled. For example, I don't model the stage of a food truck entrepreneurs decision to enter the industry in general and his choice of cuisine genre and quality. I take these long-run aspects as exogenous and solely focus on the pricing and location choice aspects of the industry. That said, with my model, by looking at how the lottery is impacting consumer surplus and the distribution of trucks operating in the modelled locations, we can get an idea of the direction of the long-run implications.

The first counterfactual policy will open up some of the currently lottery locations, which are in general the more popular locations to entry, by more trucks which in the model should increase the entry costs to these locations and decrease the entry costs to other locations. My second counterfactual policy will slacken up the capacity constraint for each non-lottery location while keeping the set of lottery locations unchanged. This will reduce entry costs for the trucks but make competition fiercer at these locations (more firms competing for the same market). As the parking spot allocation mechanisms and capacity constraints change, there are various trade-offs that must be considered. In the current regime, the various costs that a truck may experience to secure a spot for the most popular locations are artificially decreased by removing these locations from the choice sets of many trucks and reducing the number of potential entrants. From the truck's perspective this is a good thing as long as the entry costs to the location without the lottery is higher than the costs to register for the lottery. But the lottery may have an adverse impact for the truck's too. For example, if you are a high quality and popular truck that has high market shares, your potential profits may be decreased due to the fact that the lottery is not allowing you to enter only the most populous locations every day. Effectively the lottery may be propping up lower quality trucks in two ways; by decreasing the entry costs to a location which would be unaffordable for a low quality truck without the lottery, and also allowing them to face a less competitive market place once they have entered the popular location by distributing these high profit parking spots to arbitrary trucks that are not the most competitive. In this scenario the lottery would be benefiting low quality trucks and hurting high quality trucks. In other words, the lottery's and capacity constraint's implications on industry profits and utilities are ambiguous.

In terms of prices, the lottery should be reducing the downwards pressure on prices by restricting the intensity of competition. This again may allow lower quality trucks that otherwise wouldn't be able to compete to survive resulting in higher prices and lower average quality for the consumers. There are also implications of the parking space allocations on the distribution of truck types. If the distribution of the cuisine genre is not even across genres (like in reality), a lottery that randomly allocates a set of trucks to a location may result in a distribution that is not preferred by the trucks and/or by the consumers (for example, if demand resembles nested logit utilities). In this case getting rid of the lottery will help even out the variety of trucks across the different locations.

2.1.1 Method of Comparison for Counterfactual Scenarios

The counterfactual scenarios will be compared to the status quo across several dimensions. Below I outline them before heading into the comparison itself.

Supply Side: Trucks

I will assess how the counterfactual scenario has impacted the supply side of the model by looking at the entry costs, prices, and expected profits. The entry costs are calculated by matching the location average choice probability implied by the locations capacity constraint and the analogous model moments. For example if there are 200 trucks active, then the average choice probability for *Union Station* would be $Pr_{Union_{Station}} = 14/200$ because 14 is the capacity constraint at *Union Station*. In other words, the entry costs will adjust the model's location choice probabilities such that on average, the capacity constraint will be satisfied for

every location.[1] Looking at the change in prices will help me gauge the impact of the change in competitive structure due to the counterfactual scenario and the changes in expected profits will assess the net effect of the changes on the trucks.

Demand Side: Consumers

To calculate the net welfare effect of the policy change we must also calculate the net effect on consumer surplus. Given the nested logit demand structure, following [Train, 2009] I can calculate the expected utility of a whole location in the standard 'inclusive value' formulation, for each $l \in 1, ..., L$ and denoting the outside option as $g = 0$:

$$I_l = \ln \left(\sum_{g=0}^{G} \left(\sum_{k \in \Gamma_{gl}} \exp \left(\frac{v_k + \alpha \ln(\bar{Y}_l - p_k)}{1 - \sigma} \right) \right)^{1-\sigma} \right) \tag{2.1}$$

Where an approximation of the inner sum is obtained in solving the model (i.e. $\hat{I}_{gl}$) I can then obtain the dollar representation of the expected utility by multiplying I_l by the inverse of the marginal utility of income. In my model, the marginal utility of income is:

$$MU_{lY} = \frac{\alpha}{(Y_l - p)} \tag{2.2}$$

In my calculation I simply use the average price of all the trucks that are operating in the modelled locations in the equilibrium.

---

[1] The standard deviation of the number of trucks at each location during the simulation is reasonable. The location with the most spread is *L'Enfant* with a standard deviation of 4.19. This implies that a capacity constraint violation of more than 4 trucks is quite unlikely.

2.2 The Impact of the Counterfactual Policies

2.2.1 Counterfactual I: Reducing the Reach of the Lottery

Supply Side: Trucks

The counterfactual policy I consider first is for the DCRA to reduce the reach of the lottery by removing some key locations as a lottery designated location. During estimation, the truck's choice sets were restricted and the model took what was determined by the publicly available DCRA lottery outcomes as given. In contrast, in this counterfactual scenario some lottery locations are added to every trucks' choice set. Given these expanded choice sets, the trucks make entry decisions and choose prices such that their profits are maximized as outlined in Chapter 1.

I solve the model for two scenarios for this counterfactual analysis. Firstly, releasing *L'Enfant* from the lottery and secondly, releasing *L'Enfant* and *Franklin Square*. These locations have been chosen because they large and prominent locations in the DC food truck industry.

The changes to the entry costs can be seen in Table 2.1. We can observe the entry cost increases in the hypothetically excluded locations. Column (2) which show the entry costs when only *L'Enfant* is excluded shows that entry costs will increase from \$123 to \$282.62. On the other hand, all the other locations' entry costs decrease. This is expected, as *L'Enfant* is now accessible to all trucks, there is more pressure on *L'Enfant*'s capacity constraint whereas there is now less pressure on the capacity constraint of the other locations. From column (3), which shows the entry costs if *Franklin Square* is also excluded from the lottery we can see a similar change. The entry costs for *Franklin Square* increases while the entry costs

for the other locations fall. To assess the impact of the counterfactual policies on trucks, we will need to look at the equilibrium prices and profits.

		Entry Costs ($)		
			L'Enfant	L'Enfant + Franklin
	Lottery	Current	Excluded	Excluded
		(1)	(2)	(3)
19th & L	No	499.95	493.76	492.76
CNN (First St NE)	No	657.15	652.17	650.39
Farragut Square	Yes	101.24	84.10	84.56
Federal Center / Patriots Plaza	Yes	82.38	60.21	57.41
Franklin Square	**Yes**	**141.93**	**129.97**	**287.74**
Gallery Place / Chinatown	No	400.54	396.40	394.50
L'Enfant	**Yes**	**123.00**	**282.62**	**281.08**
Metro Center	Yes	68.50	64.29	54.44
NOMA	Yes	155.67	146.52	144.96
Navy Yard	Yes	440.17	425.31	416.85
State Department	Yes	167.56	159.83	148.22
Union Station	Yes	128.37	128.28	115.42

Table 2.1: Entry Cost Changes Under the Counterfactual Policy

Figure 2.1 shows the prices in the scenarios I am considering. Prices do not seem to have changed too much in the counterfactual scenario.

(a) L'Enfant Excluded   (b) L'Enfant + Franklin Sq. Excluded
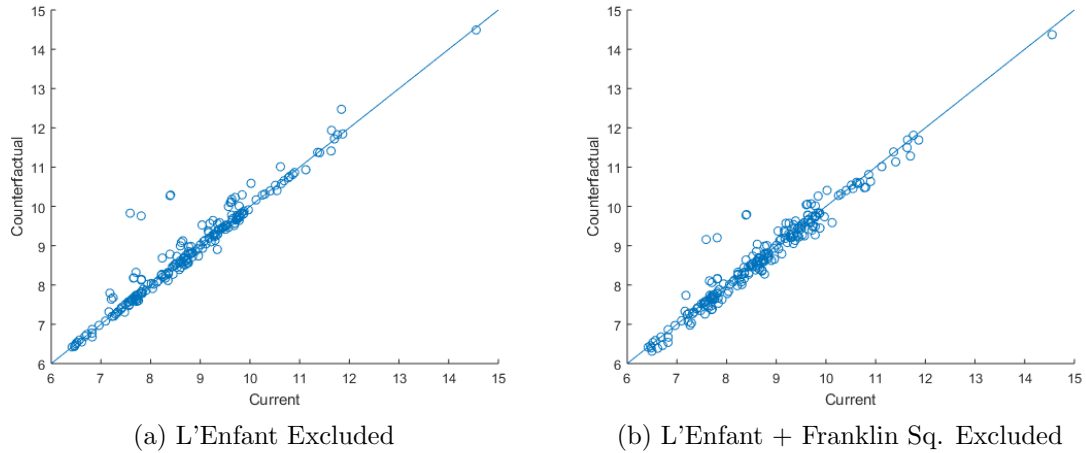
Figure 2.1: Current Regime Model Predicted Prices ($) VS. Counterfactual Regime Prices ($)

While prices do not change much, profits for the trucks decrease as the reach of the lottery is reduced (Figure 2.1). which shows a scatter plot of the simulated trucks' profits under the current and counterfactual scenarios. Current regime profits are higher than the counterfactual profits and we can see that as more locations are taken out of the lottery, profits fall lower. How much a truck's profit is impacted in the counterfactual scenario is dependent on many factors such as your quality, genre, and marginal costs but on average it seems like it is a net negative impact on profits. Taking a closer look at how the counterfactual regime impacts on profits, I find that having a lot of trucks in your own genre has large impacts on your profits. The trucks that are impacted the most by the removal of these locations from the lottery were *American*, the most common genre.

(a) L'Enfant Excluded

(b) L'Enfant + Franklin Sq. Excluded
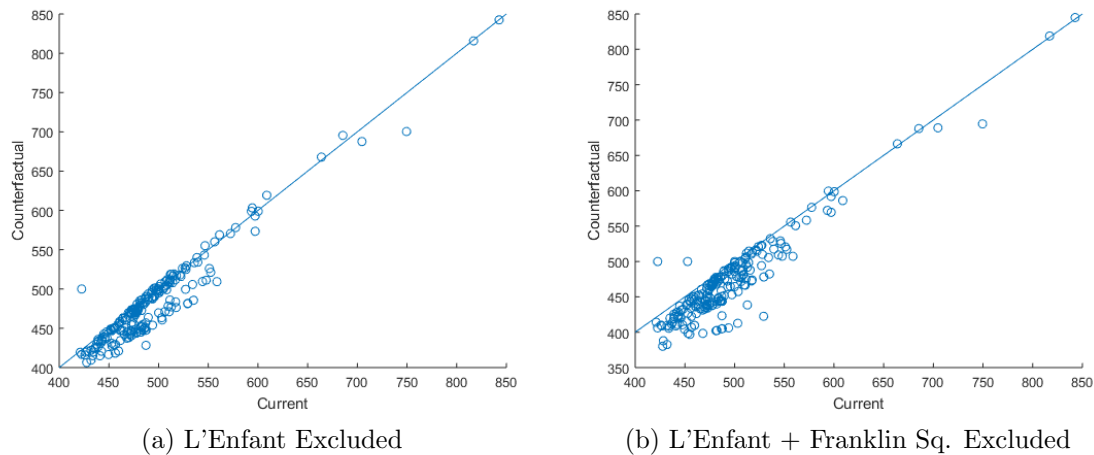
Figure 2.2: Current Regime Model Predicted Profits ($) VS. Counterfactual Regime Profits ($)

There are some trucks that stop operating in the modelled locations and this impacts the distribution of genres in these locations (i.e. no feasible price to the profit maximization problem). In particular as it is the more common cuisine genres that will be affected the most by the lack of a lottery, the model would suggest that the distribution of genres become more even distributed. Another force that will impact whether a truck continues operating or not is the truck's quality i.e. given a distribution of genre types, lower quality trucks will be the quickest to stop operating as entry costs and competition increases. If there is some strong correlations between a particular genre and quality, the lottery may not necessarily make the distribution of trucks more even. This is what I observe in my counterfactual analysis.

Removing one location from the lottery results in one truck not operating in the modelled locations. In particular this truck is a low quality *Indian* truck and in the scenario with two locations removed from the lottery we see an additional *American* truck cease to operate in the modelled locations. These observations suggest

that *Indian* trucks must on average be low quality trucks and this is a stronger force than the more intense competition that *American* trucks' experience from other trucks of their genre. In other words, *Indian* truck qualities should be much lower than the *American* trucks. We can confirm this is indeed the case, by looking at the estimated quality measures (i.e. the empirical parameterization of v_j estimated from the demand estimates). Figure 2.2 shows that *Indian* trucks have low quality, compared to the *American* trucks. Given that I only find the nesting parameter in the nested logit model to be 0.241 it is reasonable that the impact of there being a lot of trucks in your own genre isn't a strong driver of who exits or stays in the market but despite this I observe an *American* truck which on average has relatively high quality choosing the outside option location.
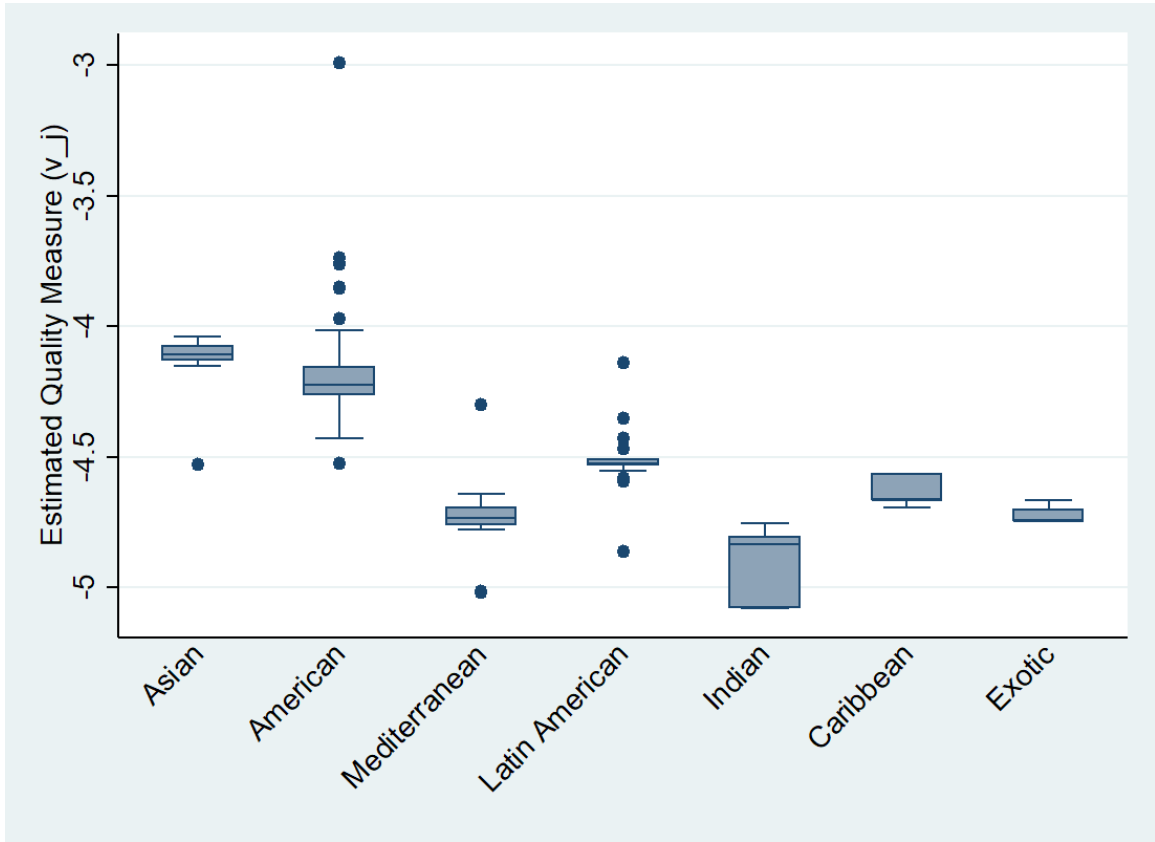
Figure 2.3: Box Plots of Estimated Quality (v_k) by Cuisine Genre

With all these considerations, the decrease in total industry profits from the counterfactual policy amount to $2,698.06 and $5,289.85 respectively for the two scenarios.

Demand Side: Consumers

Table 2.2 shows the computed average consumer surplus for each location. The interpretation of column (1), (2) or (3) is the dollar value of the expected utility of having lunch at each location. The counterfactual scenarios change consumer surplus due to various reasons and the effect shown in the table is the net impact of a combination of lower prices, higher value trucks, and consumers demand being

factored in with less constraint when firms decide which location to enter. The implied increase in consumer surplus after getting is $4,992.24 and $5,435.83 each day respectively for the two scenarios. To separate the effect of prices and other factors that influence consumer surplus, I calculate the change in consumer surplus under the counterfactual equilibrium but holding prices at the status quo equilibrium. This change should capture the consumer surplus gains from non-price impacts such as variety and quality. I find the in both scenarios the consumer surplus increase from non-price effects are approximately $3,938 out of the total aforementioned change.

Considering the changes in consumer surplus with the change in truck profits, in the first scenario the net impact on welfare is an increase of $2,294.18 and in the second scenario an increase of $145.98.

	Current	L'Enfant Excluded	L'Enfant + Franklin Excluded	Market Size	Implied Change in CS ($)	
	(1)	(2)	(3)	(4)	((2)-(1)) × (4)	((3)-(1)) × (4)
19th & L	0.289	0.288	0.291	24,583	-20.12	59.38
CNN (First St NE)	0.086	0.051	0.051	16,365	-576.07	-569.17
Farragut Square	0.309	0.363	0.366	23,779	1,280.82	1,358.19
Federal Center / Patriots Plaza	0.087	0.133	0.136	3,659	168.72	178.07
Franklin Square	**0.355**	**0.410**	**0.422**	**14,317**	**799.73**	**968.36**
Gallery Place / Chinatown	0.173	0.128	0.129	10,783	-491.58	-471.18
L'Enfant	**0.267**	**0.367**	**0.372**	**16,901**	**1,684.53**	**1,777.21**
Metro Center	0.279	0.338	0.338	23,020	1,362.39	1,359.18
NOMA	0.265	0.190	0.192	6,074	-453.40	-443.10
Navy Yard	0.193	0.192	0.192	2,898	-2.50	-4.28
State Department	0.150	0.214	0.216	9,358	589.88	609.06
Union Station	0.200	0.241	0.244	14,124	579.84	614.11
Total					4,922.24	5,435.83

Table 2.2: Consumer Surplus (CS) per Consumer by location and Implied Total Change in CS

2.2.2 Counterfactual II: Increasing Parking Capacity at Market Locations

In this counterfactual scenario I increase the parking space constraints at non-lottery modelled locations. This will have impacts on entry costs to both lottery and non-lottery locations. The locations directly impacted are *19th & L*, *CNN*, and *Gallery Place*, where I increase the number of parking spaces here by 2 spaces. We would expect to see entry costs to these locations decrease directly as a function of the capacity constraint not binding as tightly. Also, as more trucks can enter these locations, the demand for the other locations will decrease and we should see a indirect decrease in entry costs to the locations where the capacity remains the same. This is a benefit to the trucks, as a crucial input (parking space) to doing business is less scarce and consequently costs to entry have reduced. However, there is a trade-off. With more parking spaces at these locations there are more trucks and hence competition within the location is fiercer. The simulations of this counterfactual scenario will shed light on the magnitude of such effects.

Supply Side: Trucks

Table 2.3 shows the changes in entry costs under the counterfactual scenario. We can see that the entry cost decreases due to the direct effect of a location obtaining more parking spaces is larger than the indirect effect that the other locations experience. *CNN* especially experiences a large entry cost change as it is a very constrained location to begin with (only has 4 spaces where trucks can park), while *19th & L* and *Gallery Place* see changes in the magnitude of about \$40-\$50. The lottery locations only see entry cost changes of approximately \$15-\$17.

	Entry Costs ($)		
	Current	Counterfactual	Difference
	(1)	(2)	(1)-(2)
19th & L	**499.95**	**457.20**	**42.75**
CNN (First St NE)	**657.15**	**537.44**	**119.71**
Farragut Square	101.24	84.69	16.55
Federal Center / Patriots Plaza	82.38	65.40	16.98
Franklin Square	141.93	125.37	16.55
Gallery Place / Chinatown	**400.54**	**350.93**	**49.61**
L'Enfant	123.00	107.28	15.72
Metro Center	68.50	51.85	16.66
NOMA	155.67	138.00	17.67
Navy Yard	440.17	422.93	17.24
State Department	167.56	151.81	15.75
Union Station	128.37	111.99	16.38

Table 2.3: Change in Entry Costs Under Counterfactual Parking Capacities.

Figure 2.4 shows the change in prices as the additional parking spaces become available. Prices fall considerably for a moderate number of trucks while for most trucks prices aren't changing. For most trucks, their choice probabilities of entering these non-lottery locations are quite low hence it makes sense that the impact of prices aren't ubiquitous across all trucks.
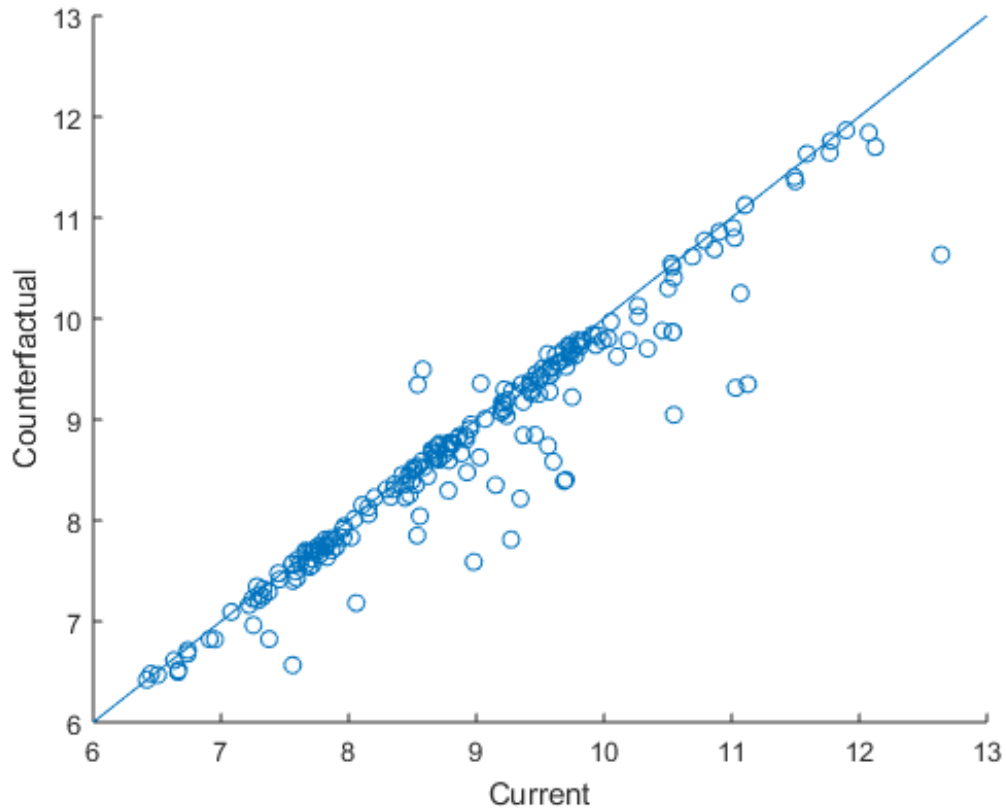
Figure 2.4: Current Capacity Model Predicted Prices ($) VS Counterfactual Capacity Prices ($)

Investigating how expected profits have changed in Figure 2.5, profits have on average increased under the counterfactual scenario. The figure suggests that the impact of increased competition due to the increased capacity does not erode away the benefits of the reduced entry costs (the average profit increase is $8.08) and results in a net profit increase to the trucks. The total increase in profits in the industry is $1,810.22.
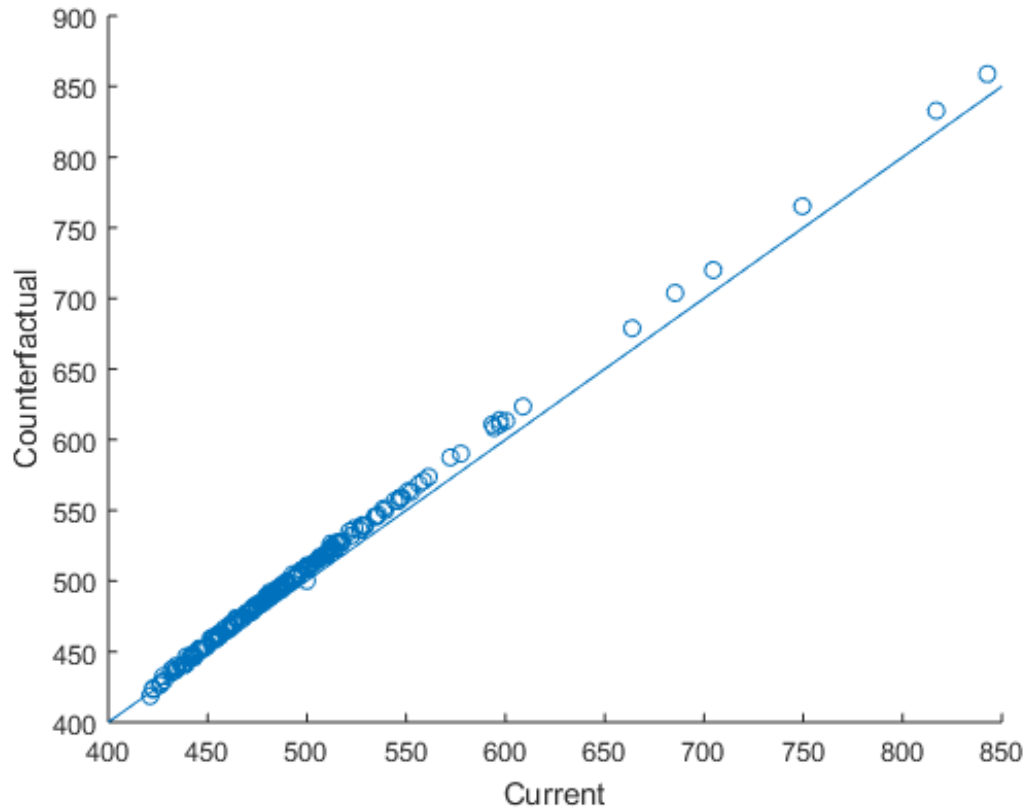
Figure 2.5: Current Capacity Model Predicted Profits ($) VS Counterfactual Capacity Profits ($)

Demand Side: Consumers

I calculate consumer surplus in the same way as above and these calculations are shown in Table 2.4. I also plot the equilibrium prices under the current and counterfactual capacity. With the counterfactual parking capacities, prices fall and consumer surplus increases.

	Current	Counterfactual	Market Size	Implied Change in CS ($)
	(1)	(2)	(3)	((2)-(1)) ×(3)
19th & L	**0.29**	**0.32**	**24,583**	**755.26**
CNN (First St NE)	**0.09**	**0.09**	**16,365**	**11.39**
Farragut Square	0.31	0.36	23,779	1,276.56
Federal Center / Patriots Plaza	0.09	0.13	3,659	171.76
Franklin Square	0.35	0.41	14,317	852.88
Gallery Place / Chinatown	**0.17**	**0.15**	**10,783**	**-212.45**
L'Enfant	0.27	0.35	16,901	1,368.49
Metro Center	0.28	0.34	23,020	1,432.57
Navy Yard	0.26	0.20	6,074	-396.55
NOMA	0.19	0.19	2,898	-1.94
State Department	0.15	0.21	9,358	591.58
Union Station	0.20	0.24	14,124	600.47
Total				6,450.04

Table 2.4: Consumer Surplus (CS) per Consumer by location and Implied Total Change in CS

Again, keeping the prices at the status quo equilibrium to calculate utilities to gauge the impact of the non-price effects of the counterfactual scenario I find that $6,450.04 of the $5,112.67 consumer surplus increase is due to such non-price changes. In contrast to the prior counterfactual scenario, this counterfactual gives rise to an unambiguous welfare increase of $8,260.26.

2.2.3 Conclusion

I find that the net change in welfare going from the current policy to counterfactual policies I consider are positive. I consider counterfactual scenarios where fewer locations are a part of the lottery and where the capacity of food truck lunch locations parking spaces is increased. Note that this doesn't account for a variety of other factors that may be considered in a broader definition of welfare. For example, how the DCRA is spending the revenue generated from the lottery. It may be that the enforcement of the parking capacities and increasing the safety of passers-by in these parking locations generates considerable welfare not quantified in my analysis.

The short-run implications of the lottery are clear from the results of my first counterfactual analysis. The lottery is facilitating the survival of trucks that would otherwise not enter any of the modelled locations (recall, that I model most of the most popular and largest locations). My results agree with the anecdotal consensus of veteran truck owners in DC as interviewed in [Hayes, 2018]. The article paints a picture of deeply worried food truck vendors that have been operating in DC for much longer than the lottery has been around, for example, Kirk Francis who co-owns the *Captain Cookie & the Milk Man* food trucks has been quoted to claim:

"If you went to *Franklin Square* five years ago during lunch, 12 out

71

of 15 trucks would be "great" and "chef-driven" such as *Cap Mac,*
TaKorean and *Dangerously Delicious Pies.* Only three would be ...
"budget trucks." Now it's the reverse".

This is exactly what I find the effect of the lottery to be. The lottery is depressing the quality and variety of the trucks in the market to the extent that quality is correlated with genre. The long-run implications of these results on consumer surplus and the existing trucks are beyond the scope of this chapter, but as the expected utility from the food truck segment decreases due to the lottery the market may not be able to support the large number of heterogeneous trucks in the market going into the future.

My second counterfactual analysis considering the expansion of parking facilities for food trucks at non-lottery locations suggests that this is a Pareto-improvement. Both profits and consumer surplus increases. The total benefit that arises from lower entry costs and greater competition generates a surplus to both sides of the market. The surplus increase sum to \$8,260.26. However, to capture these benefits, Washington, DC will need to add a total of 6 parking spaces across 3 already very congested locations.

My findings suggest that reducing the reach of the lottery to include less locations and adding additional parking spaces, will be beneficial for the food truck industry of Washington, DC.

Chapter 3   The Impacts of Large Disruptions on Long-Run Public Transit Ridership: An Analysis of Washington, DC's Subway Transit System.

3.1   Introduction

Public transit systems are very important in congested cities for many reasons. The positive externalities of a well-functioning public transit system are primarily based on improved efficiencies with respect to; congestion, environmental harm, and labor mobility. Hence, public transit is of direct interest to many parties, including but not limited to; governments, engineers, environmental scientists, and economists, consequently there is a large body of interdisciplinary work attempting to model and understand public transit. Modeling demand for transport is comprehensively outlined in Domencich and McFadden [1975], and ever since, countless authors in various locale and continents have applied discrete choice models to understand the behavioral and economic aspects of transport mode choice.[1] Similarly, there are abundance of research on the costs and benefits of a successful public transit system.[2] This chapter does *not* attempt to model the decision making process of a consumer of transportation and/or attempt to uncover the various costs and benefits of a public transit system.

This chapter explores more specific questions that must be considered for agen-

---

[1]Some examples include; Beirão and Cabral [2007], Paulley et al. [2006], dell'Olio et al. [2011], Paulley et al. [2004].

[2]Some examples include;Litman [2017], Rissel et al. [2012], Geurs and van Wee [2004].

cies that manage large metropolitan subway systems using reduced form methods. Transportation agencies need to decide whether to invest in quality improvements, such as making repairs to improve reliability and reduce delays. There are two types of costs that the agency must consider. Firstly, the direct cost of the repairs and maintenance and secondly, any reductions in ridership that occur before, during, and after the repairs. This chapter focuses on the latter type of costs an agency must consider. I will be exploring impacts on ridership of a large-scale system repair program geared to increase quality. Lastly, I will also quantify the magnitude of these impacts in terms of price changes using an observed fare increase. The background for this analysis is the DC Metrorail system, operated by the Washington Metropolitan Area Transit Authority (WMATA) and their year-long system repair program SafeTrack which lasted between June, 2016 to June, 2017. SafeTrack systematically closed down and "single-tracked"[3] segments of the subway system. There are policy implications regarding fare increases and large-scale maintenance that arise out of the analysis that may be of value when designing future large repair programs and fare hikes.

My analysis focuses on the "AM Peak" riders to focus on the morning commuter market. I find that the price hike implemented at the end of June, 2017 of $0.10 on every 'tier' of AM Peak fares decreased average monthly ridership by 2.05%. I also find that the disruptions caused by SafeTrack have a dynamic impact on ridership. There is initially an anticipatory effect as scheduled repair dates loom. This is expected as WMATA was encouraging people to look for alternative modes of commuting well before the actual beginning of the repairs. One month prior to the repairs on a segment, the original-destination station pairs that are

---

[3]Where only one side of the rail tracks are used between stations for trains going in either direction. This induces major delays and unpredictability in train schedules.

directly affected, see an average decrease of 2% in ridership. During the month of the repairs the station pairs that are directly affected see an average of 9.11% decrease in ridership and I find that there is a persistent decrease to ridership even after the repairs have been complete. For about 10 months after the repairs have been finished, ridership doesn't fully recover to pre-repair levels, persistently being about 1.68% lower than pre-repair periods.

This chapter will be organized as follows. In section 3.2 I will discuss the institutional background. Section 3.3 will describe the data and illustrate descriptive patterns in the data. Sections 3.4 will discuss my estimation methods and estimates, and Section 3.5 will conclude.

3.2 Institutional Background

Metrorail, a subway system whose network spreads across DC, MD, and VA has seen a large decline in ridership in the last 4-5 years, and the cause[4] of this exodus of metro riders seems to be the persistent degradation of service quality (including fatal accidents [Jansen, 2016]). The National Transportation Safety Board (NTSB) has voiced concern over the state of Metrorail and SafeTrack was WMATA's response to the situation. WMATA describes SafeTrack in the following way on their website:[5]

What is SafeTrack?

SafeTrack is an **accelerated track work plan** to **address safety**

---

[4]Simultaneously, with the system's degrading quality over time, Washington, DC embraced the advent of the car sharing industry [Moylan and Graves, 2015] which worsened the situation. However, the impact of car sharing on public transit ridership is not the focus of the chapter.

[5]https://www.wmata.com/service/SafeTrack.cfm

recommendations and rehabilitate the Metrorail system to improve safety and reliability. Through SafeTrack, Metro will complete approximately three years' worth of work into approximately one year. The plan significantly expands maintenance time on weeknights, weekends and midday hours and includes 16 "Safety Surges" - **long duration track outages for major projects in key parts of the system**.

Why is SafeTrack necessary?

Metrorail is currently open 135 out of 168 hours per week, leaving insufficient time for maintenance and other necessary track work. By closing the system at midnight on weekends and **expanding weekday maintenance opportunities**, SafeTrack addresses FTA and NTSB safety recommendations and deferred maintenance backlogs while restoring track infrastructure to good health. In addition the 16 "Safety Surges" will utilize **long-duration track outages through around-the-clock single tracking or line-segment shutdowns that will impact rush hour commutes**.

How will SafeTrack impact my commute?

Due to reduced capacity and expected longer travel times, **Metrorail riders are encouraged to consider using alternate travel options while safety surge work is scheduled on their line**. Trains and platforms may be extremely crowded during peak periods and customers may experience extended delays. Review the links below for more information about each surge project and potential impacts on

your commute.

I have emphasized the parts of WMATA's SafeTrack description that is directly relevant to this chapter. SafeTrack was undoubtedly an important and ambitious undertaking. I will be exploring how ridership has been evolving before, during, and after SafeTrack to better understand the impact and results of this program. WMATA explicitly states to metro users that commute times will be longer during SafeTrack impacted stations and that the consideration of alternative travel options are encouraged. This potentially gives rise to dynamic impacts, which I explore.[6]

Table 3.1 and Figure 3.1 shows the schedule of repairs and the network segments that fell under SafeTrack. We can see that not all of the repairs are of the same magnitude, some are much larger than others. There are two types of repairs. Repairs can either lead to trains on a segment shutting down or to single tracking. During the repairs that require a full segment shut down, a free shuttle service is provided between the closed stations. Also, on June 25th, 2017 a fare increase of $0.10 across all tier of peak time fares was implemented.

---

[6]My analysis ignores the indirect impacts of a disruption. The subway system is a large network where there is a persistent disruption through the entire system even if only a small segment of the system was disrupted. These impacts are ignored.

|	Date	Duration (Days)	Line	Type of Impact	Segment Affected
1	Jun 4 - 16, 2016	13	Orange	Single Track	East Falls Church to Ballston
2	Jun 18 - July 3, 2016	16	Orange	Shutdown	Eastern Market to Minnesota Ave & Benning Road
3	Jul 5 - 11, 2016	7	Yellow	Shutdown	National Airport to Braddock Road
4	Jul 12 - 18, 2016	7	Yellow	Shutdown	Pentagon City to National Airport
5	Jul 20 - 31, 2016	12	Orange	Single Track	East Falls Church to Ballston
6	Aug 1 - 7, 2016	7	Red	Single Track	Takoma to Silver Spring
7	Aug 9 - 21, 2016	13	Red	Single Track	Shady Grove to Twinbrook
8	Aug 27 - Sep 11, 2016	16	Yellow	Single Track	Franconia-Springfield to Van Dorn Street
9	Sep 15 - Oct 26, 2016	42	Orange	Single Track	Vienna to West Falls Church
10	Oct 29 - Nov 22, 2016	25	Red	Shutdown	Fort Totten to NoMa
11	Nov 28 - Dec 20, 2016	23	Orange	Single Track	East Falls Church to West Falls Church
12	Feb 11 - 28, 2017	18	Blue	Shutdown	Rosslyn to Pentagon
13	Mar 4 - Apr 12, 2017	40	Blue	Single Track	Braddock Rd to Huntington/Van Dorn St
14	Apr 15-May 14, 2017	30	Green	Shutdown	Greenbelt to College Park/Prince George's Plaza
15	May 16 - Jun 15	31, 2017	Orange	Shutdown	New Carrollton to Stadium-Armory
16	Jun 17 - 25	9, 2017	Red	Shutdown	Shady Grove to Twinbrook

Table 3.1: SafeTrack Repairs Schedule.
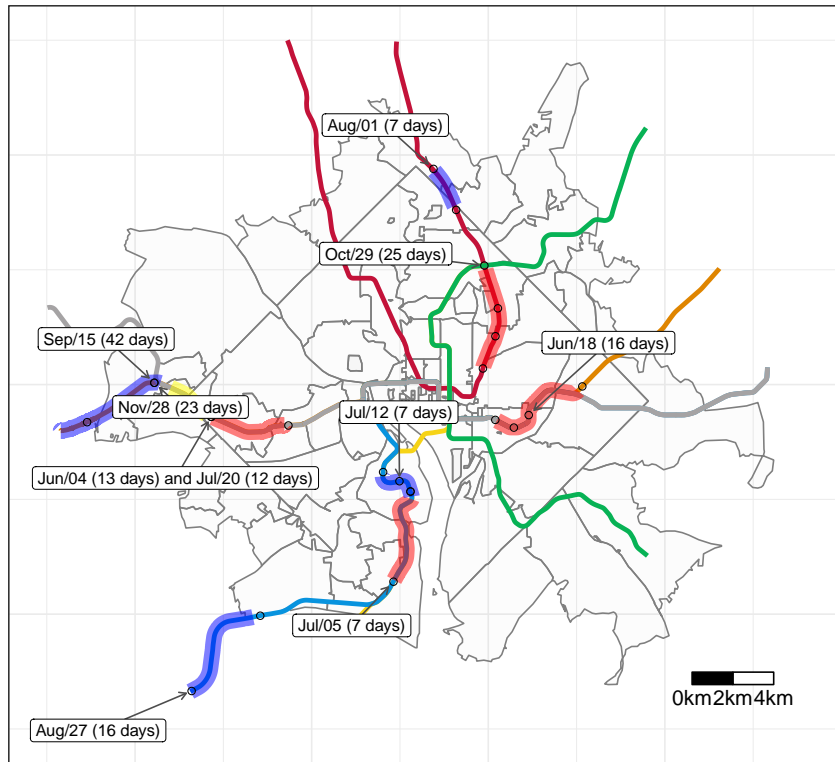
(a) 2016 SafeTrack Repairs. Repair Segments Highlighted Red and Blue for Clarity.
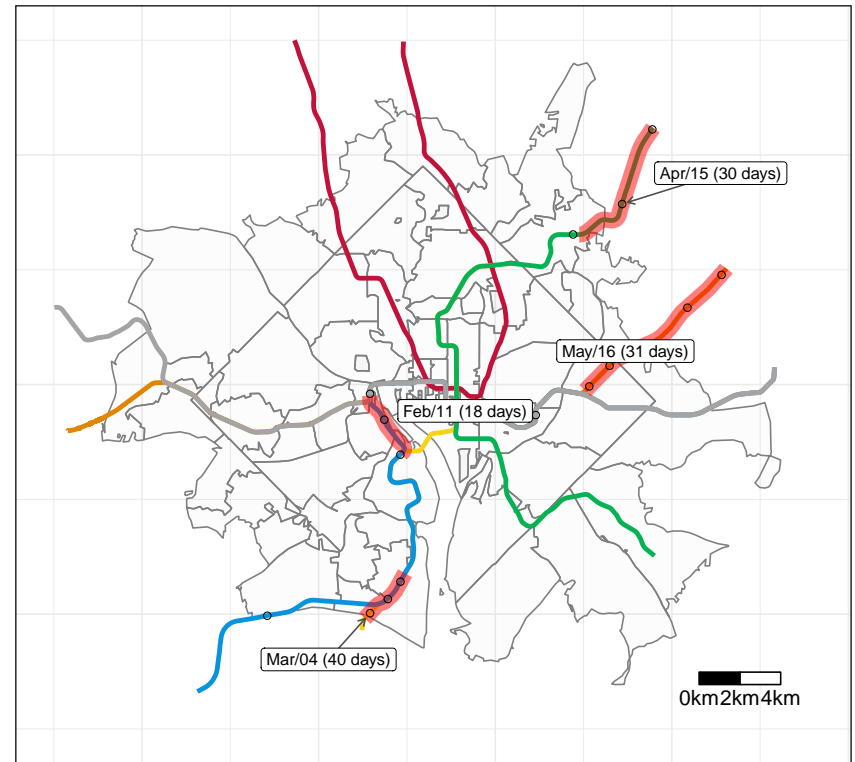
(b) 2017 SafeTrack Repairs.

Figure 3.1: The Metrorail system and SafeTrack Repair Segments

3.3 Data

3.3.1 Data Sources

The data I use for the analysis is from two sources. One is a data set obtained directly from WMATA of monthly average ridership between an origin-destination pair, by time of day divided into 5 time-intervals and by service type consisting of Weekday, Weekends and various others (for example public holidays). I only use the weekday data to focus on commuters. There are a couple of reasons for this. Firstly, AM and PM peak ridership constitutes on average 56.2% of the riders while only taking 44.7% of operating time of the system. Secondly, to capture the persistent impacts of a repairs program such as SafeTrack, I want to focus on the segment of the market that is actively and repeatedly making a transport mode choice. The daily time intervals in the data are *AM Peak*, *Midday*, *PM Peak*, *Evening*, and *Late Night Peak*. This monthly data goes from June, 2013 to October, 2017. Out of these I only consider the *AM Peak* ridership, because the *PM Peak* ridership seems to be strongly correlated with the *AM Peak* ridership. To verify this, I run a regression of log monthly average ridership on each station pair's corresponding AM-PM commute origin-destination match (for example, regress AM ridership of *Crystal City-Dupont* on PM ridership of *Dupont-Crystal City*). The coefficient on this regression is 0.97 and is not statistically significantly different from 1 where standard errors are calculated using pairwise clustering on origin-destination. Given this, I carry out the rest of the analysis using only the *AM Peak* data. In the analysis below, the treatments I focus on are repairs 1-13 in Table 3.1. I avoid the others because the data set ends not long after these events which means that I am unable to track the long-run impacts of these treatments.

The second data set is from a publicly available data set, that contains data on the length of every delay that the Metrorail system experienced from September, 2012 to September, 2016 categorized by cause of the delay. Unfortunately, this data set only encompasses the very beginning of SafeTrack so I am unable to observe actual quality changes after SafeTrack is complete. However, this data can shed light on the relationship between delay (a primary quality measure for consumers) and ridership. I use this data to control for the effect of delays on ridership when I estimate how the price increase implemented at the end of SafeTrack impacted ridership.

3.3.2 Descriptive Statistics

From the monthly ridership data, it can clearly be observed that ridership has been falling for the last 4-5 years. Figure 3.2 shows the sum over all station pairs' monthly average daily ridership and the average daily system wide delays over time, the latter number is calculated by adding all the reported service delay minutes over a given month and dividing by $60 \times 30 = 1800$ to obtain average daily delays in hours. The sum over all station pairs' ridership approximately represents the average system wide daily ridership for a given month. The raw data is very jagged, because there are strong seasonal effects in the data (people ride the metro more during the warmer months as opposed to winter, also during the holiday season *AM Peak* ridership is limited) so I plot the 6 month moving average to bring out the more important features of the data. Some features of note are the gradual decline in total ridership from late 2013 until the beginning of SafeTrack that and the clear dip in ridership during SafeTrack starting June, 2016 and somewhat recovering by March, 2017, by when most of the repairs have been executed.

81

Figure 3.2: 6 Month Moving Average of the Sum of Average Ridership At Every Origin-Destination Pair. and Average system wise delays.

The increase in total daily delays the whole system accrues can be seen to sharply increase around March, 2014. This coincides with when the declining ridership trend begins. Wait time and timeliness seems to be a strong determinant in consumers commute mode choice. The causes listed for the increase in delays are primarily due to increased train and infrastructure related issues. In light of these descriptive figures, it seems that a large-scale repair program seems to have been the right solution. It makes sense to drive delays down, to decrease the rate of ridership decline or even increase ridership to 2013-2014 levels. However, it is difficult to eyeball these effects. The next section will assess the impacts of SafeTrack and the following fare increase in a more systematic way.

3.4 Estimation and Analysis

3.4.1 The Impact of SafeTrack on Ridership

The Magnitude of Impact and SafeTrack Characteristics

We have seen on Table 3.1 that not all SafeTrack repairs are born equal. There is heterogeneity in the treatment. For example, the treatment for repairs that last 7 days on continuous single tracking are likely to have different impacts on ridership to repairs that last 25 days with complete segment shut downs. I explore the relationship between these heterogeneities and the magnitude of impact on ridership.

Firstly, I estimate a regression equation with SafeTrack repairs 1-13 in Table 3.1 each coded as separate treatments.[7] The corresponding coefficients on these treatment dummies can be interpreted as the impact of each repair treatment on the corresponding treated stations pairs and will uncover the differences in magnitude of the effect of each treatment. Secondly, I will consider each treatment's affect in light of the treatment's characteristics to discern any relationships between the magnitude of impact at treatment characteristics.

The regression equation estimated is:

$$\ln(Ridership_{ijt}) = \alpha + \sum_{1}^{I\times(I-1)} O.D.\ FE_{ij} + \sum_{1}^{T} Time\ FE_{t}$$

$$+ \sum_{k=1}^{12} \beta_k SafeTrack\ Repair\ k_{ijt} + \epsilon_{ijt} \qquad (3.1)$$

---

[7]Repairs 3 and 4 are combined into one treatment because they are chronologically and geographically contiguous. This gives me 12 treatments to consider.

Subscript ij denotes a original-destination (O.D.) pair respectively for $i \neq j$ and $i, j \leq I$. $SafeTrack\ Repair\ k_{ijt}$ is the treatment variable, a dummy variable that is equal to 1 if the station pair ij in period t is "affected" by scheduled SafeTrack repair k where k indexes the treatment number (i.e. SafeTrack repair). I define "affected" observations as station pairs where a segment scheduled for SafeTrack repair lies in between the origin and destination that defines the observation and if the origin station is *not* a inner-Washington station. This is because I am focusing on the AM commuters that are commuting into Washington. How the treatment variable is formulated is illustrated in Figure 3.3. I control for the time trend not at the pair specific level (there are too many pairs) but by sample groups categorized by level of ridership. Pairs of stations are categorized as; HIGH if there are more than or equal to 400 riders, MEDIUM if there are between 3 and 400 and LOW if there are 2 or less riders. These cutoffs were determined by eyeballing the data and it is reasonable to think there are different time trends for stations with less than 3 riders vs. a station with more than 400 riders. Each group LOW, MEDIUM, and HIGH make up approximately 55%, 44%, and 1% of the stations pairs respectively. The identification strategy to estimate equation (3.1) is a difference-in-differences which requires a common pre-treatment trend assumption.
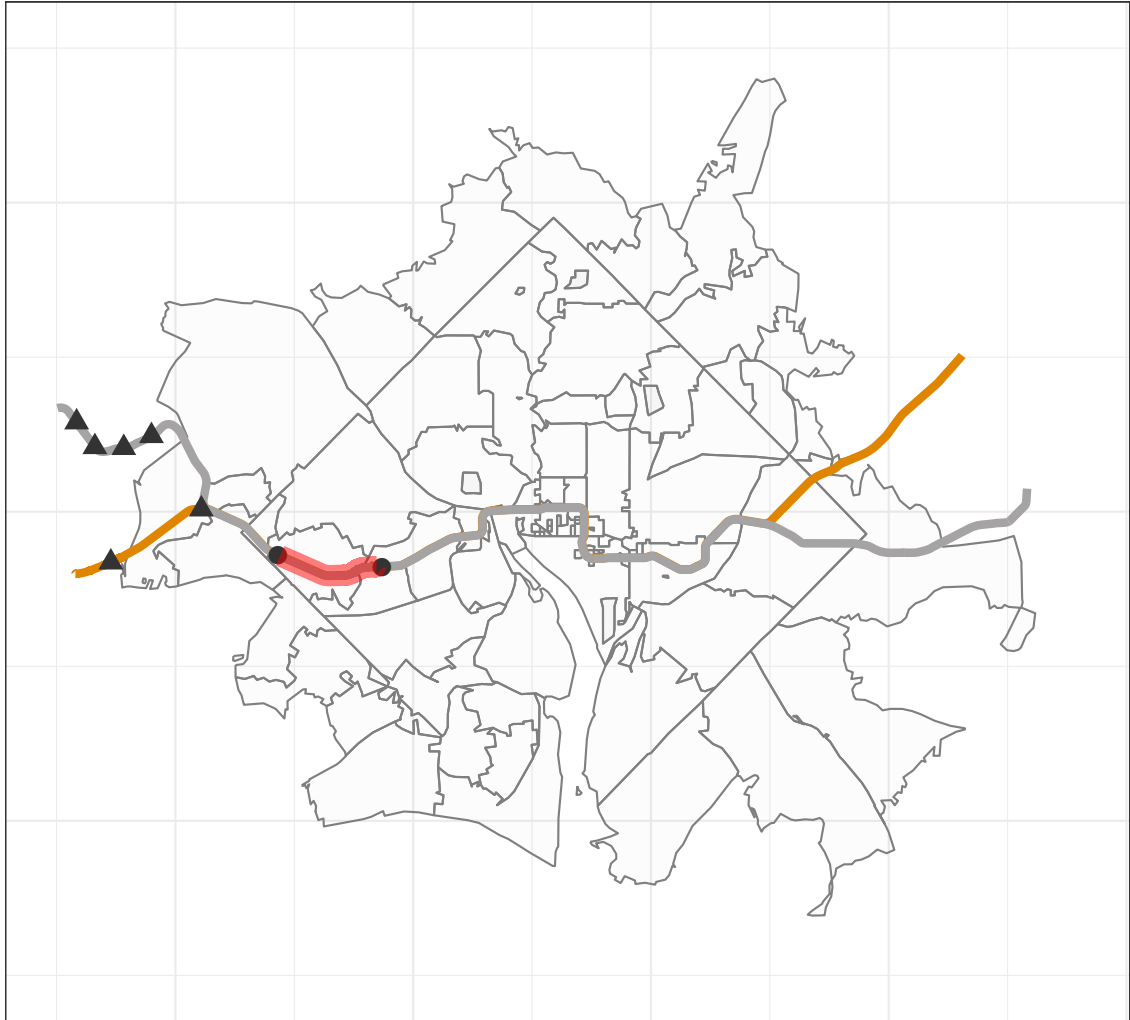
Figure 3.3: Graphical Interpretation of Treatment 1. Station Pairs where the Origin Station is Not on the Inner-Washington" Side of the DC, MD, VA Area and Hence "Affected" (Triangles) are Defined as Treated, Along with the Stations Directly Affected by the Repairs (Dots).

	OLS
Treated	0.00127
	(0.00133)
Constant	-0.0296***
	(0.000991)
Observations	154,000
R-squared	0.000

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.2: OLS Regression of Pre-treatment Growth Rates for Treated and Control Stations Pairs

Another important note is how we should calculate the standard errors. Assuming each station pair is independent is not a very good assumption because there may be arbitrary correlation between pairs if the same origin or destination is part of another observation. Given this, I cluster observations across two dimensions, original and destination following Cameron et al. [2006], to account for the fact that each observation falls into both of these dimensions. In Table 3.2 I show the results of a regression on the growth rates[8] on a treatment dummy variable to investigate whether the common pre-existing trends assumptions are met for identification of the above equation. I find that there is no significant difference in the pre-treatment trends. Table 3.3 shows the estimation results for the above equation. Controls and fixed effects coefficients are not reported.

---

[8]Calculated as the first difference of log ridership.

Treatment k=	(1) Quadratic Time Trend	(2) Linear Time Trend	(3) No Time Trend
1	-0.101***	-0.101***	-0.115***
	(0.0260)	(0.0260)	(0.0313)
2	-0.0553	-0.0553	-0.0615
	(0.0403)	(0.0402)	(0.0488)
3	-0.0572	-0.0572	-0.0616
	(0.0362)	(0.0361)	(0.0445)
5	-0.0637**	-0.0636**	-0.0702**
	(0.0301)	(0.0301)	(0.0350)
6	-0.0138	-0.0139	-0.0175
	(0.0134)	(0.0134)	(0.0164)
7	-0.0293	-0.0294	-0.0322
	(0.0642)	(0.0643)	(0.0739)
8	-0.0224	-0.0224	-0.0223
	(0.0254)	(0.0255)	(0.0264)
9	-0.429***	-0.429***	-0.518***
	(0.131)	(0.131)	(0.156)
10	-0.0411	-0.0412	-0.0559
	(0.0348)	(0.0348)	(0.0421)
11	-0.0891**	-0.0892**	-0.102**
	(0.0392)	(0.0391)	(0.0462)
12	-0.0623***	-0.0621***	-0.0640**
	(0.0209)	(0.0210)	(0.0253)
13	-0.195***	-0.196***	-0.238***
	(0.0416)	(0.0415)	(0.0510)
Observations	361,199	361,199	361,199
R-squared	0.235	0.235	0.094
Number of OD_id	8,183	8,183	8,183

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

*Note:*Repair $k = 3$ combines repair 3 and 4 in Table 3.1.

Table 3.3: Heterogeneity in the Impact of SafeTrack

The results are robust to different specifications, and we can clearly observe the heterogeneity in each repair. There are some obvious relationships that can be drawn. Firstly, as expected the treatments with the largest negative impacts

are the ones with the longest duration. For example, treatment 9 and 13 are the repairs that lasted 42 and 40 days respectively. Also, treatments with line segment shut downs seemed to have non-significant effects, with the exception of treatment 12. This may seem counter-intuitive, but it could simply reflect the fact that the shuttle service between closed segments are quite effective in moving the passengers between the segment that is shut down. Table 3.4 summarizes some of these comparisons.

Treatment k=	Index to Table 3.1	% Change	Statistically Significant?	Shutdown?	Duration	State
1	1	-10.10%	Yes	No	13	VA
2	2	-5.53%	No	Yes	16	MD
3	3+4	-5.72%	No	Yes	7	VA
4	5	-6.36%	Yes	No	12	VA
5	6	-1.39%	No	No	7	MD
6	7	-2.94%	No	No	13	MD
7	8	-2.24%	No	No	16	VA
8	9	-42.87%	Yes	No	42	VA
9	10	-4.12%	No	Yes	25	MD
10	11	-8.92%	Yes	No	23	VA
11	12	-6.21%	Yes	Yes	18	VA
12	13	-19.56%	Yes	No	40	VA

Results from Specification (2) in Table 3.2 are used to construct this table.

Table 3.4: Treatment Characteristics and Estimation Results.

However, the most striking and robust relationship between treatment characteristic and magnitude of effect is geographic in nature. It seems that all of the statistically significant decreases are treatments that are on the VA side of the Metrorail system. Only one of the VA treatments (treatment 3) is insignificant. Treatment 3 is found to have an insignificant effect, but in all 3 specifications the $t-stat$ is very close from being large enough for significance despite the fact that this treatment only lasts 14 days. My findings suggest that commuters commuting

in from VA substituted the most heavily away from riding the subway.

The Dynamic Impact of SafeTrack on Ridership

There is reason to believe that SafeTrack has an impact on ridership that follows a dynamic path that starts and continues, before and after the actual repairs. Firstly, there is potential for an anticipatory effect. SafeTrack repairs did not start suddenly with no warning. WMATA announced SafeTrack one month prior to the first set of repairs beginning and encouraged commuters to look for alternative modes of transport. Secondly, during the repairs we would see a big decrease in ridership, both due to the fact that there is literally less trains moving less passengers between origin and destination, and also because less people are choosing this mode of transport for their daily commute as there are more delays.[9] Thirdly, we could expect to see a persistent impact on ridership for the stations that went through the repairs. For example, if there are significant switching and re-optimization costs to researching to switch to another form of transport for the daily commute, then we would expect a large negative shock to utility from programs like Safe-Track would push these people to actually switch to another mode of commuting and not return to the metro even if the delays have shortened and quality has increased. If this is indeed the case, simply assessing the impact of SafeTrack with a treatment dummy variable as in equation (3.1) will fail to capture these dynamics. To capture these effects, I estimate the following regression equation:

---

[9]WMATA's data driven blog documents big surges in Metrobus utilization during repairs [Catherine, 2016]

$$\ln(Ridership_{ijt}) = \alpha + \sum_{1}^{I \times (I-1)} O.D.\ FE_{ij} + \sum_{1}^{T} Time\ FE_t + \beta_0 Repair\ Month_{ijt}$$

$$+ \sum_{l \geq L}^{-1} \beta_l Repairs\ to\ begin\ l\ periods\ later_{ijt}$$

$$+ \sum_{k \geq 1}^{K} \beta_k Repairs\ completed\ k\ periods\ ago_{ijt} + \epsilon_{ijt} \qquad (3.2)$$

The subscripts for station pairs and time are as in equation (3.1). With equation (3.2), I am tracing out the full path of how a treatment of getting a repair done impacts the station pair's ridership. My choice of L and K are -2 and 10 respectively. By looking at these before, during, and after repair dummies we can get a better understanding of whether the aforementioned dynamics are present in the data. The time fixed effects will capture unobserved month-year specific variables that impacted the entire metro system, such as seasonal effects and the impact of delays. Later, when I explore the impact of delays and the post-SafeTrack fare increase, I use these month-year fixed effect coefficients as the left-hand side variable.

Table 3.2 shows the set of estimations (fixed effects are not reported) and Figure 3.2 visually depicts the estimates from the linear time trend estimates. The variable on the left-hand side is a log transformation and the coefficients are on dummy variables so can be interpreted as a percentage change given the dummy is turned on, for example, $\beta_{-1} = -0.0251$ can be interpreted as, there being on average 2.5% less riders one month before a station pair is scheduled for SafeTrack repairs.

The estimates are robust to different ways of controlling for the time trend and we can see the dynamic path of ridership as a station pair experiences a SafeTrack repair. Ridership fell about 2.5% a month before the actual repairs (i.e. anticipatory effect) and it takes about 2 months for ridership to recover from the shock. I observe a consistently negative point estimate of the post-repair coefficients although some of the coefficient aren't significantly different from 0. However, an *F-test* for joint significance for all the coefficients for $k > 2$ suggest that jointly, the coefficients are all indeed significant ($F - Statistic = 31.97$). There is indeed a long-run impact from the disruptions caused by SafeTrack. About 1.68% percent of former Metrorail commuters seem to have switched to another mode of transport for good.
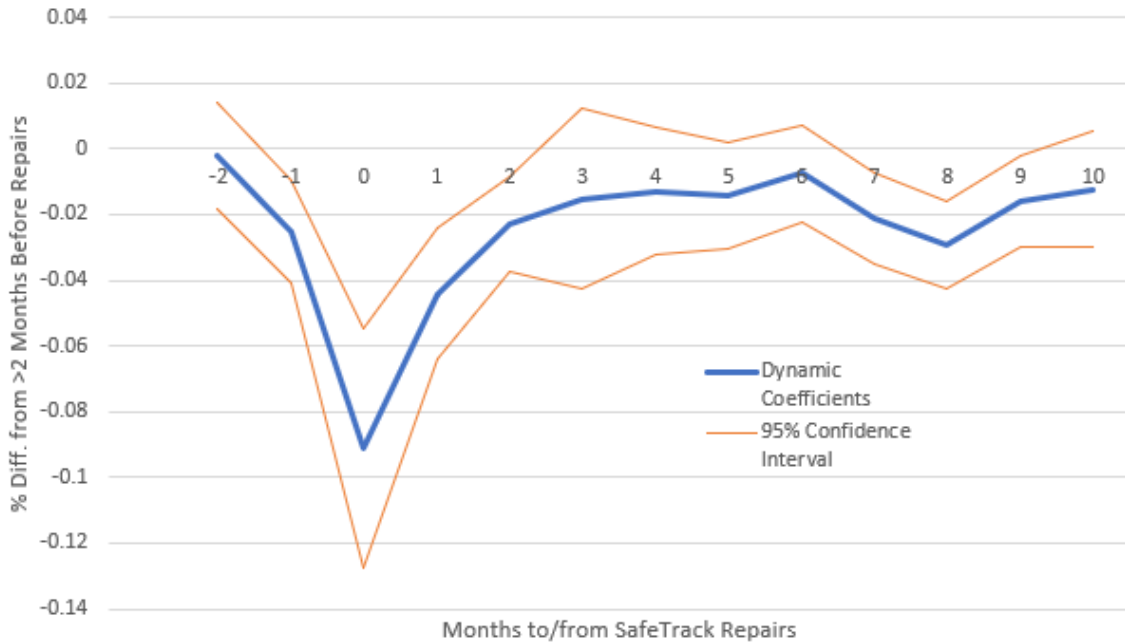

Figure 3.4: Graphical Interpretation of Regression Coefficients and Confidence Intervals.

Months to/from Repair	(1) Quadratic time trend	(2) Linear time trend	(3) No time trend
-2	-0.00221	-0.00223	-0.00342
	(0.00825)	(0.00826)	(0.00992)
-1	-0.0251***	-0.0251***	-0.0305***
	(0.00793)	(0.00793)	(0.00936)
0	-0.0911***	-0.0911***	-0.108***
	(0.0185)	(0.0185)	(0.0221)
1	-0.0441***	-0.0441***	-0.0545***
	(0.0102)	(0.0102)	(0.0121)
2	-0.0230***	-0.0230***	-0.0316***
	(0.00742)	(0.00741)	(0.00793)
3	-0.0151	-0.0151	-0.0183
	(0.0139)	(0.0139)	(0.0160)
4	-0.0130	-0.0130	-0.0169
	(0.00992)	(0.00989)	(0.0113)
5	-0.0141*	-0.0141*	-0.0161*
	(0.00832)	(0.00829)	(0.00933)
6	-0.00740	-0.00741	-0.0110
	(0.00754)	(0.00749)	(0.00845)
7	-0.0213***	-0.0213***	-0.0236***
	(0.00708)	(0.00703)	(0.00801)
8	-0.0294***	-0.0294***	-0.0364***
	(0.00682)	(0.00678)	(0.00778)
9	-0.0158**	-0.0157**	-0.0189**
	(0.00714)	(0.00710)	(0.00762)
10	-0.0124	-0.0123	-0.0148
	(0.00906)	(0.00902)	(0.00994)
Observations	361,199	361,199	361,199
R-squared	0.234	0.234	0.093
Number of OD_id	8,183	8,183	8,183

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.5: The Dynamic Impact of SafeTrack on Ridership.

3.4.2 The Impact of Service Quality and Fares on Ridership

In this section I look at how service quality and fares impact ridership. Due to the absence of delay data during and after SafeTrack, it is not possible to assess exactly how much of an impact SafeTrack has had on actual level of delays. I replace the missing data, with zeros and then account for the missing time periods with a dummy variable. This should allow me to capture the impact of delays on ridership for the parts where data is not missing, and then when it is missing, account for it in levels.

The left-hand side variable I use are the estimated the month-year dummy coefficients from the previous regression estimation (in particular, specification (1)). These coefficients capture the unobserved characteristics (characteristics not included in the previous regression) of the monthly average ridership in a particular month-year. Recall, the regression in the previous section is in log scale, which means that the interpretation of the coefficient estimates in this section are in percentage changes per unit change in the right-hand side variable. Firstly, I estimate the following equation:

$$Time\ FE_t = Month\ FE_t + \alpha + \beta_1 Delay_t + \beta_2 Post\ fare\ increase_t \quad (3.3)$$
$$+ \beta_3 SafeTrack + \beta_4 Missing + +\beta_5 time + \epsilon_t$$

The month fixed effects should capture any seasonal fluctuations that are a function on month of year. The effect of delay on ridership is captured by the β_1 (delays are a continuous measure). β_2 captures the impact of the fare increase on ridership. This should capture solely the fare increase because firstly, the time

fixed effect has a lot of the variation coming from other relevant dependent variables stripped out from it already. Secondly, the delays and month fixed effects are controlling for any variation in the time fixed effects that don't vary within month-year but do vary across month-year. These variations would have all been pushed into the time fixed effects we are using in the estimation of equation (3.1). β_3 captures the impact of SafeTrack at the system level. Note that the treatments in equation (3.1) are determined at the origin-destination pair level so indirect, system wide impacts of SafeTrack are not fully extracted from the time fixed effects. β_4 captures the levels difference of ridership while delays are missing. Lastly, β_5 controls for the time trend.

	OLS
Fare Increase	-0.0205
	(0.0176)
Delays	-0.001916
	(0.0091)
Safetrack	-0.0298847**
	(0.0110)
Missing Delays Dummy	0.0118
	(0.0290)
Constant	1.8486***
	(0.2569)
Time (months)	-0.0030***
	(0.0004)
Observations	46
R-squared	0.947

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.6: The Impact of Delays and Fare Increases on Ridership.

Table 3.6 shows the estimated coefficients for equation (3.3), because of the

correlation between delays and the time trend variable, we see that delays are not found to be statistically significant in the estimated equation. The estimates suggest that per year, ridership has been falling on average 3.6%. I have also examined the impact of lagged delays on ridership, however, lagged delays do not affect ridership at the time interval the data comes in (i.e. one month intervals). This suggests that most commuter transport mode substitutions happen within a month. The fare's impact on ridership is insignificant. The point estimate suggests that post-fare increase, ridership is on average 2.05% lower than before the fare increase. The absolute value of the change in fares is $0.1 for any destination originating from a peak time entry.[10] The actual average percentage change in fare is more convoluted to calculate because the fares are calculated in a non-linear way, as a function of distance travelled on the track and the coordinate distance between two stations. However, the price change in percentages ranges from being a 4.56% increase ($2.15 to $2.25) to a 1.89% increase ($5.90 to $6.00) and in 2013 the average fare collected by WMATA was $2.90 [Duggan, 2013]. This would imply an average effective fare increase of 3.45% which suggests ridership elasticity with respect to fare is -0.59. This is in line with the literature's prior findings. Paulley et al. [2004] and Litman [2017] both offer a comprehensive survey on transit elasticities. The authors in both surveys suggest that metro ridership's short-run elasticities are in the range of -0.3 to -0.6, while long-run elasticities are close to 1.

To compare these estimates with the persistent impacts of SafeTrack, taking the back of the envelope elasticity calculation above, the estimates suggest that

---

[10]Price changes on weekly and monthly passes are different. I couldn't get data on the market share for these pass holders and single trip pay-as-you-go card holders, but it is reported at **https://planitmetro.com/?s=smartbene** that 87% of Federal Metrorail customers are paying for the usage using a pay-as-you-go standard fare payment method. This may suggest that the aforementioned fare increase is relevant to the vast majority of Metrorail users.

the impact that SafeTrack had on ridership in the long-run is similar to an average fare increase of 2.85%.[11] The average daily ridership across all treated origin-destination pairs for the time periods before SafeTrack (June, 2013 to May, 2017) is 135,767.4. My dynamic coefficients then imply there is about 2,279 ($135,767.4 \times 0.0168 = 2,279.36$) less *AM Peak* commuters riding *Metrorail* to get into DC in the morning. Assuming that an average commuter pays $3.00 for his/her commute, this a fare revenue loss of about $204,840$ per month from the AM commute market. Assuming that the commuters who substituted away from the subway may also not be riding in the *PM Peak* time period, the total daily impact is possibly closer to double the amount above.

3.5   Conclusion

I find that the commuter origin-destination pairs that suffered the largest losses during SafeTrack are ones that are originating from Virginia, quantifying why this might be the case goes beyond the scope of the current chapter. There is an indirect cost to investing in a large maintenance program such as SafeTrack in the form of persistent rider losses. If consumers face switching costs, the disruptions caused by maintenance may provide a sufficiently large negative utility shock such that riders consider other options for transportation. This chapter explored this indirect cost to assess the persistent impact that SafeTrack had on Metrorail ridership in the DC, Maryland, Virginia (DMV) area. Considering how delays were increasing and ridership was decreasing consistently since 2013-2014 a big push to increase reliability and timeliness seems to have been a good idea. However, I find evidence of persistent ridership losses. Ridership for the subway system in the DMV area does not fully recover to pre-SafeTrack levels even up to 10 months after a segment

---

[11] $\frac{\sum_{k=2}^{10} \hat{\beta}_k/9}{0.59} = 2.85$. Using estimates from specification (2) in Table 3.5.

has been worked on. My estimates suggest that 2 to 10 months after the repairs ridership is on average 1.68% lower compared to 2 months before the repairs. This level of persistent ridership loss translates approximately to losses of about $410,000 per month. I also find that ridership is inelastic with respect to fares. This is in line with the literature's findings.

Bibliography

E. Anenberg and E. Kung. Information technology and product variety in the city: The case of food trucks. *Journal of Urban Economics*, 90(C):60–78, 2015. URL http://EconPapers.repec.org/RePEc:eee:juecon:v:90:y:2015:i:c:p:60-78.

G. Beirão and J. S. Cabral. Understanding attitudes towards public transport and private car: A qualitative study. *Transport Policy*, 14(6):478 – 489, 2007. ISSN 0967-070X. doi: https://doi.org/10.1016/j.tranpol.2007.04.009. URL http://www.sciencedirect.com/science/article/pii/S0967070X07000522.

C. L. Benkard, P. Jeziorski, B. V. Roy, and G. Y. Weintraub. Nonstationary oblivious equilibrium. Economics Working Paper Archive 568, The Johns Hopkins University,Department of Economics, Aug. 2008. URL https://ideas.repec.org/p/jhu/papers/568.html.

S. Berry and A. Pakes. Some applications and limitations of recent advances in empirical industrial organization: Merger analysis. *The American Economic Review*, 83(2):247–252, 1993. ISSN 00028282. URL http://www.jstor.org/stable/2117672.

S. T. Berry. Estimation of a Model of Entry in the Airline Industry. *Econometrica*, 60(4):889–917, July 1992. URL https://ideas.repec.org/a/ecm/emetrp/v60y1992i4p889-917.html.

S. T. Berry. Estimating Discrete-Choice Models of Product Differentiation. *RAND Journal of Economics*, 25(2):242–262, Summer 1994. URL https://ideas.repec.org/a/rje/randje/v25y1994isummerp242-262.html.

T. F. Bresnahan and P. C. Reiss. Entry in monopoly markets. *The Review of Economic Studies*, 57(4):531–553, 1990. ISSN 00346527, 1467937X. URL http://www.jstor.org/stable/2298085.

A. C. Cameron, J. B. Gelbach, and D. L. Miller. Robust inference with multiway clustering. Working Paper 327, National Bureau of Economic Research, September 2006. URL http://www.nber.org/papers/t0327.

Catherine. Tens of thousands of customers relied on metrobus during safetrack surges 3 and 4. Blog Post 48, PlanItMetro, 2016. URL https://planitmetro.com/2016/09/12/tens-of-thousands-of-customers-relied-on-metrobus-during-safetrack-surges-3-and-4/.

DCRA. Vending handbook. 2013. URL https://dcra.dc.gov/sites/default/files/dc/sites/dcra/publication/attachments/VendingHandbook.pdf.

L. dell'Olio, A. Ibeas, and P. Cecin. The quality of service desired by public transport users. *Transport Policy*, 18(1):217 – 227, 2011. ISSN 0967-070X. doi: https://doi.org/10.1016/j.tranpol.2010.08.005. URL http://www.sciencedirect.com/science/article/pii/S0967070X10001009.

T. Domencich and D. McFadden. *Urban travel demand: a behavioral analysis.* Contributions to economic analysis. NORTH-HOLLAND, 1975. URL https://books.google.com/books?id=Nx4pAQAAMAAJ.

P. Duggan. Proposed hikes in metro rail fares would make a pricey subway system even pricier. 2013. URL https://www.washingtonpost.com/local/trafficandcommuting/proposed-hikes-in-metro-rail-fares-would-make-a-pricey-subway-system-even-pricier/2013/12/07/be6e35ba-5e86-11e3-bc56-c6ca94801fac_story.html?utm_term=.b3178762ea38.

K. T. Geurs and B. van Wee. Accessibility evaluation of land-use and transport strategies: review and research directions. *Journal of Transport Geography*, 12(2):127 – 140, 2004. ISSN 0966-6923. doi: https://doi.org/10.1016/j.jtrangeo.2003.10.005. URL http://www.sciencedirect.com/science/article/pii/S0966692303000607.

L. Hayes. No longer trendy, food trucks facing declining revenue find ways to survive. 2018. URL https://www.washingtoncitypaper.com/food/article/20986600/no-longer-trendy-food-trucks-facing-declining-revenue-find-ways-to-survive.

Intuit. Food trucks motor into the mainstream. Report, Intuit., 2012. URL http://gourmetstreets.com/wp-content/uploads/2013/12/Free-Food-Truck-Industry-Report-3.pdf.

B. Jansen. Investigators: D.c. metro ignored safety provisions for years. 2016. URL https://www.usatoday.com/story/news/2016/05/03/washington-metropolitan-area-transit-authority-subway-death/83864546/.

T. A. Litman. *Evaluating Public Transit Benefits and Costs: Best Practices Guidebook*. Victoria Transport Policy Institute, 2017. URL http://www.vtpi.org/tranben.pdf?b81542c0?db0c3fd8.

S. MacFarlane, R. Yarborough, S. Jones, and C. Decker. Group of food trucks squatting on desirable parking spaces in southwest d.c. 2018. URL https://www.nbcwashington.com/investigations/Group-of-Food-Trucks-Squatting-on-Desirable-Parking-Spaces-in-Southwest-DC-475185583.html.

M. J. Mazzeo. Product choice and oligopoly market structure. *RAND Journal of*

Economics, 33(2):221–242, 2002. URL http://EconPapers.repec.org/RePEc:rje:randje:v:33:y:2002:i:summer:p:221-242.

A. Moylan and Z. Graves. Ridescore 2015: Hired driver rules in u.s. cities. Report 48, R Street Institute, 2015. URL http://www.rstreet.org/wp-content/uploads/2015/12/RSTREET48.pdf.

A. Nevo. Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342, 2001. ISSN 1468-0262. doi: 10.1111/1468-0262.00194. URL http://dx.doi.org/10.1111/1468-0262.00194.

NPD Group's. Recent menu price hikes, more telecommuters, and shopping online contribute to foodservice lunch visit declines, September 2016. URL https://www.npd.com/wps/portal/npd/us/news/press-releases/2016/recent-menu - price - hikes - more - telecommuters - and - shopping - online - contribute - to-foodservice-lunch-visit-declines/. [Online; posted 13-September-2016].

N. Paulley, R. Balcombe, R. Mackett, H. Titheridge, J. Preston, M. Wardman, and J. Shires. The demand for public transport: a practical guide. Report 593, Transportation Research Laboratory: London, UK, 2004. URL http://discovery.ucl.ac.uk/1349/.

N. Paulley, R. Balcombe, R. Mackett, H. Titheridge, J. Preston, M. Wardman, J. Shires, and P. White. The demand for public transport: The effects of fares, quality of service, income and car ownership. *Transport Policy*, 13(4):295 – 306, 2006. ISSN 0967-070X. doi: https://doi.org/10.1016/j.tranpol.2005.12.004. URL http://www.sciencedirect.com/science/article/pii/S0967070X05001587. Innovation and Integration in Urban Transport Policy.

S. Qi. The impact of advertising regulation on industry: the cigarette advertising ban of 1971. *The RAND Journal of Economics*, 44(2):215–248, 2013. ISSN 1756-2171. doi: 10.1111/1756-2171.12018. URL http://dx.doi.org/10.1111/1756-2171.12018.

C. Rissel, N. Curac, M. Greenaway, and A. Bauman. Physical activity associated with public transport use—a review and modelling of potential benefits. *International Journal of Environmental Research and Public Health*, 9 (7):2454–2478, 2012. ISSN 1660-4601. doi: 10.3390/ijerph9072454. URL http://www.mdpi.com/1660-4601/9/7/2454.

M. Saeedi. Reputation and adverse selection: Theory and evidence from ebay. Discussion paper, The Ohio State University, 2014.

R. Schmalensee. Sunk costs and market structure: A review article. *The Journal of Industrial Economics*, 40(2):125–134, 1992. ISSN 00221821, 14676451. URL http://www.jstor.org/stable/2950504.

K. Seim. An empirical model of firm entry with endogenous product-type choices. *The RAND Journal of Economics*, 37(3):619–640, 2006. ISSN 07416261. URL http://www.jstor.org/stable/25046263.

D. Staiger and J. H. Stock. Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586, 1997. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/2171753.

J. Suzuki. Land Use Regulation as a Barrier to Entry: Evidence from the Texas Lodging Industry. Working Papers tecipa-412, University of Toronto, Department of Economics, Oct. 2010. URL https://ideas.repec.org/p/tor/tecipa/tecipa-412.html.

A. Sweeting. A model of non-stationary dynamic price competition with an application to platform design. Working Paper 15-03, NET Institute, 2015.

K. E. Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

T. Wollmann, D. Pollmann, M. Shepard, M. Sinkinson, H. Tabakovic, and E. Tamer. Trucks without bailouts: Equilibrium product characteristics for commercial vehicles job market paper. 2014.

Y. D. Xu. A structural empirical model of r&d, firm heterogeneity , and industry evolution. 2008 Meeting Papers 744, Society for Economic Dynamics, 2008. URL https://ideas.repec.org/p/red/sed008/744.html.