DATA-DRIVEN MODELS OF CUSTOMER BEHAVIOR TO IMPROVE OPERATIONAL EFFICIENCY IN SERVICE SYSTEMS

Pradeep Pendem

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Operations in the Kenan-Flagler Business School.

Chapel Hill
2018

Approved by:

Vinayak Deshpande

Bradley Staats

Saravanan Kesavan

Adam Mersereau

Vidyadhar Kulkarni

To my parents and sister

ACKNOWLEDGMENTS

First and foremost, I am thankful for my advisor Prof. Vinayak Deshpande for being my dissertation committee chair and constantly guiding me throughout my work during my program at Kenan-Flagler Business School, University of North Carolina at Chapel Hill. He has been a source of immense strength to me in my academic work, and also guided me to work on interesting data-driven problems which have been receiving significant attention lately in the Operations Management community. Despite his busy schedule, he always allotted time to meet whenever I needed it.

My sincere thanks and gratitude go to Prof. Bradley Staats with whom I collaborated on a paper during the last 4 years of my PhD program and agreed to be a committee member. He also has been a source of immense moral strength to me, and boosted my spirits during tough times over the past four years. I gratefully acknowledge his contribution and support. I also thank my other collaborators Paul Green (University of Texas at Austin) and Prof. Francesca Gino (Harvard Business School). I am also grateful to Prof. Saravanan Kesavan with whom I worked as a Research Assistant on retail data analysis, Prof. Adam Mersereau for evaluating my research and providing feedback on my dissertation topic and Prof. Vidyadhar Kulkarni (UNC STOR) who guided me on the technical parts of my research. Thanks to each one individually for agreeing to serve on my dissertation committee. Finally, I thank Operations area assistants Erin Leach and Holly Guthrie for their help on administrative issues during my stay here.

This dissertation would not have been possible without the cooperation and help of many individuals. I was a part of wonderful cohort of researchers, friends and their immediate family members who have been a huge moral support to me in the last 5 years. To name a few, Nagarajan Sethuraman (his wife and son - Gowri and Aditya Nagarajan), Deepak Jena (his wife Reema Seksaria), Soumya Gayathri (her husband - Kamakhya Mishra), Oguz Cetin, Ying Zhang (Maggie), Herbie Baiyang Huang, Ting Yao, Hyun Seok Lee, Anand Bhatia, Neha Jha, Dayton Steele and many more. I wish all of them happy and successful careers.

Special thanks go to two people, Prof. Aditya Jain (Zicklin School of Business, Baruch College, CUNY) and Prof. Sarang Deo (Indian School of Business, Hyderabad), who motivated me and helped me to prepare well and build a strong foundation before I embarked on my PhD journey.

Finally and most importantly, I am grateful for and indebted to my parents Padmavathi Pendem and Sankara Rao Pendem and my sister Anu Radha Namani for their unconditional love and all the sacrifices they have made to make me realize my goals. They have always been with me to share my joys and sorrows, and to shower their blessings on me. Finally, my thankful prayers goes to the beloved and revered Lord Sri Venkateswara Swamy for giving me the opportunity, courage, and endurance to successfully traverse the doctoral program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: DISSERTATION OVERVIEW

In 2015, on a global level, the service industry represented 66% of GDP and employed over 51% of the total working population making it one of the largest industry[1]. Some of the sectors in this industry include finance, transportation, call centers, retail (on-line, brick and mortar), and health care. Customers and service providers are key players in this industry. A successful transaction between these two results in a valued service for the customer and revenue to the provider. The primary objective of the providers therefore is to understand the customers needs, meet their requirements, provide quality service and achieve customer satisfaction.

Service systems are inherently complex environments with multiple operational challenges. One of the challenges providers face is predicting true customer demand, because demand is highly variant in time, space, and quantity; most importantly, each customers needs are heterogeneous. Predicting true demand is important since the ability to do so helps providers plan to have the correct amount of resources and to offer better service to their customers. Many service systems exhibit a demand irregularity pattern called non-stationarity (i.e., customer arrivals vary by time of the day). For example, in most call centers, customer demand is low at start of the day, reaches a peak level and drops gradually towards end of the day. Similar non-stationarity patterns are also observed in other service systems, such as customer traffic to a brick-mortar (or on-line) retail store and patient arrivals to the hospital outpatient department. Along the space dimension, customers choose, travel (or are transferred) to the location of the provider to receive service. Typically call

[1]https://data.worldbank.org/

centers provide services with agent response in multiple languages. For example, customers who request to speak with an agent in Spanish are transferred to call centers located in Spanish-speaking countries although the call may originate from a country where Spanish is not the primary language. Similarly, in retail, customers are likely to seek substitutes by using competing stores if the products they wish to purchase are not available in the focal store. Hence, customer demand on a provider also has a spatial component. Lastly, along the quantity dimension, each customer has different requirements which requires customized service delivery. Consider two different customer calls to a financial call center to clarify a money balance calculation. One customer might need a more detailed explanation than the other due to their varied experience in and knowledge of accounting. Although the calls are of the same type, one customer demands more time from the provider than the other does. Similarly, in retail context, each customer may purchase different items or the same items as another customer but in different quantities. Hence, predicting true demand by incorporating time, space, quantity and each customers needs is a challenging problem.

Managers typically use demand models which are aggregated in either one dimension or multiple dimensions. These models consider total volume of demand without incorporating time, spatial dynamics and customer preferences. For example, consider a call center with daily customer demand of 50 calls per hour and 100 calls per hour during non-peak and peak periods, respectively. Scheduling agents for an aggregated (average) demand of 75 calls per hour and not accounting for non-stationarity can lead to under-utilization of agents during the non-peak period and longer periods of customer waiting during the peak period. Similarly, in the retail context, allocating incorrect amounts of product inventory across spatially distributed stores can lead to high inventory cost and lost sales. Operational planning using aggregated demand models often leads to inefficiencies such as under- or over-utilization of resources, high waiting costs, and inadequate levels of service. As a result, customers are likely to be unhappy with the service and experience lower satisfaction. With customers expectations changing over time and increasing competition between providers, dissatisfied customers are likely to switch to competitive providers. This results in loss of future demand on and revenue for the focal provider. To provide better service, it is important that managers execute operational planning using micro-demand models which incorporate customer preferences, time and spatial dynamics. To achieve this, firms need to embrace a data-driven culture of collecting and analyzing the granular data of each customer transaction.

Over the last decade, owing to the lower costs of data storage, firms now collect more granular information for each service transaction (e.g., customer characteristics, time of transaction and customer location). **Big data**, the term widely used in industry and academia, is reshaping the service industry. More data has been created in the past 2 years than in the entire history of the human race (Cohen [2017]). For example, Walmart handles more than 1 million customer transactions every hour, the accumulated data storage for which is estimated to be more than 2.5 petabytes (Cukier [2010]). Each transaction includes information on the customer, date of visit, items purchased, and the quantity of each item. Using this information, Walmart can now analyze the history of each customers transaction, predict their next arrival and promote products to increase sales. In October 2016, Uber had 40 million monthly riders worldwide[2]. Uber completed its first billion rides between 2009 and 2015, and required only an additional 6 months to reach its two-billionth ride[3]. Each transaction includes information on the customer, a date/time stamp, location of pickup, and the location of drop-off. Data at such a granular level can help Uber learn about each customers requirement in time and space thereby providing more customized services. Uber could analyze each customers sensitivity to waiting and dispatch suitable drivers to increase overall matches as well as customer and driver satisfaction. A common contribution that big data provides regardless of the context is information on each customers needs and evolution

[2]http://fortune.com/2016/10/20/uber-app-riders

[3]https://techcrunch.com/2016/07/18/uber-has-completed-2-billion-rides

of preferences over time and space, which was not available a decade ago. Managers can now analyze these large datasets and build micro demand models incorporating customer behavior, time, and spatial dynamics. Operational planning using enhanced demand models leads to efficient resource allocation whereby providers can offer better service to their customers.

In this thesis, I utilize large data sets on customer-provider transactions to study two important issues. First, I build micro-demand models to predict true demand incorporating customer behavior, time and spatial dynamics. I utilize the predicted demand to optimally allocate the resources for improved operational performance. The study contexts I focus on are Bike Sharing systems and Street-Hail Taxi services.

Bike Sharing systems and Street-Hail Taxi services are unique contexts compared to other traditional systems like call centers or retail. Bike Sharing systems are self-initiated, short-term bike rental services. Consumers arrive at a station to pick a bike, take a trip and park the bike at their selected destination station. The demand at each station is both bike pickups and drop-offs. In a traditional service system, high demand leads to only reduction of available resources or inventory. For example, in a retail store, high demand for a product leads to only depletion in its inventory. Increase in inventory is possible if there are product returns, however typically the frequency of such is extremely low. Bike inventory at a station can either increase or decrease with sizable proportion of pickups and drop-offs, hence the number of bikes at a station depends on both pickups and drop-offs. On the other hand, Street-Hail Taxi services are urban transit systems which provide taxi trip services to customers across different locations. A trip is realized when a customer hand hails for an available taxi, a driver picks them up, and drops them off at their chosen location. The key difference between this system and the Bike Sharing system is that taxi drivers need not stay at a location for pickup and they might chose to move to a different location. However in Bike Sharing systems, bikes continue to remain at the station until a pickup is realized. Hence the number of available taxi drivers at a location depends on pickups, drop-offs and their patience level. This key irregularity in both Bike Sharing systems and Street-Hail taxi services is equivalent to variable servers in Queuing theory. Typically, Queuing models are analyzed with fixed servers. Analyzing systems with variable servers makes it an operationally challenging problem.

Second, I build micro models to understand the factors driving customer provided satisfaction measure and their impact on purchase probability in E-commerce platforms. On-line platforms are large scale service systems where sellers sell products and customers purchase the products. Sellers rely on the ratings to signal product quality, while customers use this information for making the purchase decision. One common feature of these three systems is that they depict high degree of temporal and spatial distribution of demand. I analyze large and granular transactional data, build micro demand models incorporating customer behavior and incorporate the models into planning to improve operational efficiency. I briefly describe each chapter of my thesis in the following paragraphs.

1.1 DEMAND ESTIMATION AND ALLOCATION OF BIKES TO MAXIMIZE RIDERSHIP IN BIKE-SHARING SYSTEMS

In the first chapter, I study demand estimation and resource allocation in Bike Sharing systems. Bike Sharing systems are self-initiated, short-term bike rental service available to users on a shared basis. The performance of these systems are deeply impacted by users spatial and temporal preferences as well as the allocation of resources over the network. Unavailability of bikes (stock-out) for pick-up and empty docks (full-capacity) to drop-off are a common phenomenon due to spatial and temporal imbalance in supply (of bikes) and demand (from users). As a result, users either substitute in the vicinity of stations or are lost from the network to a different mode of transport. I utilize large datasets from the Bay Area Bike Sharing System on censored trips ($\approx$ 300k records), minute-level inventory information at stations ($\approx$ 16 million records), and walking distances between stations to analyze the performance of this system. I first develop a model to predict demand at stations that incorporates spatial and temporal fluctuations for pickups and drop-offs, as well as consumer substitution between stations under stockout scenarios. I find that users utility of station substitution decreases with an increase in walking distance between stations. This demand model is then used to analyze the optimal allocation of bicycles in the network stations to improve ridership. A dynamic program is developed to determine the optimal allocation of bicycles across stations at the start of the day while incorporating non-stationary demand and station substitution. Derivation of performance analysis relies on transient analysis during an operating day, instead of steady state performance measures. I then examine the optimal timing and quantity of re-allocations of bikes in the network to maximize ridership. I find that an optimal allocation policy can improve ridership and service level by 7.60% and 1.69% (respectively) compared to the current policy.

1.2 LOGISTICS PERFORMANCE, RATINGS, AND ITS IMPACT ON SALES IN E-COMMERCE PLATFORMS

In the second chapter, I examine the impact of logistics performance metrics such as delivery delays, customer's promised speed of delivery, order split, etc. on logistics service ratings of sellers on an e-commerce platform. Further, I analyze and quantify the impact of logistics service ratings and performance on customer purchasing behavior and sales. Prior work on online ratings in e-commerce platforms have largely analyzed customer response to product functional performance and biases that exist with-in. This study contributes to this stream by examining customer experience from a service quality perspective by analyzing logistics performance, logistics ratings and its impact on customer purchase behavior. The insights from this study are relevant to independent sellers as well as e-commerce platform managers who aim to improve long-term online traffic and sales. Using a large data set of customer orders ($\approx$ 21 million records) from the Tmall platform and the Cainiao network, I utilize ordered regression to understand the variation in logistics ratings and its drivers. I then use a customer utility model to quantify the impact of logistics ratings on customer purchasing behavior. I find, Logistics ratings are negatively impacted by delivery delays, but positively impacted by faster promised speed of delivery and total order amount paid. The impact of delivery delays on logistics ratings are moderated by the total order amount paid but not by faster promised speed of delivery. For example, a customer who paid a higher order amount is likely to give a more negative rating to a delayed order compared to an on-time order than a customer who paid a low order amount. The results also show that splitting an order into multiple shipments, so that a part of the order is delivered on-time even if the overall order is delayed, does not improve logistics ratings. I also find that logistics ratings impact customer purchasing behavior positively. Lastly, I show that a reduction in delivery delay by one day can improve the average weekly sales by as much as 2.5%. This study emphasizes that logistics performance and ratings which measure service quality are important drivers of customer purchase behavior on e-commerce platforms. Hence, e-commerce platforms and sellers should pay attention to logistics performance, in addition to product performance, in driving traffic and sales online.

1.3 DEMAND ESTIMATION AND OPTIMAL SHIFT TIMING IN STREET-HAIL TAXI SERVICES

In the third chapter, I study passenger demand estimation and the taxi driver optimal shift timing problem in Street-Hail Taxi services. Improving taxi services efficiency is an important problem as it affects drivers income, passenger service and importantly transportation revenue. Taxi regulations limit the number of hours a driver can operate. Driver service intensity is endogenous as they choose the location, start time and end time of their shift. An absence of centralized control under regulation and endogenous service intensity can lead to a sizable number of drivers starting or ending their shift during the same time period. Increased delay in shift change time when drivers share a common resource (taxi) can lead to low availability affecting service level and revenue. The impact can be significant if the shift changes overlap with peak demand period. I utilize large-scale datasets of the

GPS information of pick-ups and drop-offs ($\approx$ 14 million records) from New York Yellow Taxi services during June 2013 for this study. I first develop a stochastic matching model (double-ended queue) to predict passenger demand in location and time. The stochastic model allows for non-stationarity, randomness in arrivals and reneging behavior of both drivers and passengers. Using sample path information along with Maximum Likelihood Estimation, I estimate potential passenger demand as well as drivers and passengers relocate rates. The predicted demand is then used to analyze the optimal timing of drivers changing their shift to maximize revenue under current status quo of delay in shift changeover. Also, I quantify the amount by which revenue could be improved if the shift changeover is instantaneous (*ongoing*).

CHAPTER 2: DEMAND ESTIMATION AND ALLOCATION OF BIKES TO MAXIMIZE RIDERSHIP IN BIKE-SHARING SYSTEMS

2.1 Abstract

Bike Sharing systems are self-initiated, short-term bike rental service available to users on a shared basis. The performance of these systems are deeply impacted by users spatial and temporal preferences as well as the allocation of resources over the network. Unavailability of bikes (stock-out) for pick-up and empty docks (full-capacity) to drop-off are a common phenomenon due to spatial and temporal imbalance in supply (of bikes) and demand (from users). As a result, users either substitute in the vicinity of stations or are lost from the network to a different mode of transport. I utilize large datasets from the Bay Area Bike Sharing System on censored trips ($\approx$ 300k records), minute-level inventory information at stations ($\approx$ 16 million records), and walking distances between stations to analyze the performance of this system. I first develop a model to predict demand at stations that incorporates spatial and temporal fluctuations for pickups and drop-offs, as well as consumer substitution between stations under stockout scenarios. I find that users utility of station substitution decreases with an increase in walking distance between stations. This demand model is then used to analyze the optimal allocation of bicycles in the network stations to improve ridership. A dynamic program is developed to determine the optimal allocation of bicycles across stations at the start of the day while incorporating non-stationary demand and station substitution. Derivation of performance analysis relies on transient analysis during an operating day, instead of steady state performance measures. I then examine the optimal timing and quantity of re-allocations of bikes in the network to maximize ridership. I find

that an optimal allocation policy can improve ridership and service level by 7.60% and 1.69% (respectively) compared to the current policy.

2.2 Introduction

The transportation sector in the United States was the second largest contributor (26%) of greenhouse gas emissions in 2014 next to electricity generation (30%) (Brown et al. [2016]). The two largest sources within transportation - light duty vehicles (e.g. passenger cars, motor bikes) and medium and heavy duty trucks - contributed to 61% and 23% of these greenhouse gas emissions respectively. During the period 1990-2014, emissions from the transportation sector increased more in absolute terms (17%) than any other sector (electricity generation, industry, agriculture, commercial and residential). The primary greenhouse gas emitted by contributors in the transportation sector is carbon dioxide (96%) which is one of the prominent drivers of increasing global temperatures and climate change. A recent report[1] published by the United States Environmental Protection Agency (U.S. EPA) proposed various opportunities for the reduction of greenhouse gas emissions within the transportation sector. These opportunities include switching to low-carbon fuels, usage of electric vehicles, developing technologies to improve fuel efficiency, improving existing operating practices (e.g., reducing average taxi time for aircraft, engine idling, etc.) and reducing the demand for travel in light duty vehicles. To reduce travel demand, government and urban planners have implemented zoning for mixed use areas, so that residences, schools, and businesses are close together and have initiated pedestrian and car/bike share programs.

The traditional model of public transit has been to pick people up at designated locations on a set schedule and according to a pattern. With cities becoming denser than ever, additional means of transit have become a necessity to keep people moving and cut down travel time. In recent years, several firms like Uber and Lyft have entered the transportation

[1]https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions#transportation

industry and challenged the traditional model. These firms started offering carpool services (Uber Pool, Lyft Line) which provide easy and affordable on-demand support so that more rides can be achieved with fewer cars. Complementing this wave of new private firms, governments all over the world through private-public partnerships are implementing a non-traditional means of mobility - Bicycle Sharing Programs.

As previously mentioned, Bicycle Sharing Programs are self-initiated, short-term bicycle (hereafter "bike") rental services made available to consumers on a shared basis. These systems are located mostly in urban areas with a high population density and comprise networks of stations where bikes are docked. Consumers arrive at a station to check out a bike and drop it at their selected destination when they complete their trip. These state of the art systems (which can be found worldwide) are equipped with electronic sensors in both bikes and stations. These sensors make it easy and inexpensive to collect data in real-time. Most programs typically provide the first 30 minutes of the ride at no cost to encourage participation by consumers.

Bike Sharing Program was first introduced for use without cost to consumers in 1965 in Amsterdam. The program failed due to theft and damage. However, over the 53 years since then, with developments in security these programs have spread across Europe to Copenhagen, the United Kingdom, as well as France and Spain. China launched its first program in 2008 with 2,800 bikes in the city of Hangzhou. Today, it is the worlds largest program with more than 78,000 bikes. In the United States, most programs are concentrated in major cities and university towns. In 2010, Capital Bike share was launched in Washington D.C, Nice Ride in Minneapolis, and B-cycle in Denver. In 2013, Citi Bike was launched in New York with 6,000 bikes fully funded by corporate sponsorship. By end of 2016, the largest program in the United States was Citi Bike followed by Capital Bike. Today, there are nearly 900 programs operating worldwide[2] and this number is expected to increase in

[2]http://www.citylab.com/city-makers-connections/bike-share/)

the near future given the high degree of acceptance from governments and general public[3].

Bike Sharing program makes system access affordable for short trips which otherwise would have required motorized public or private vehicle transport. These programs address the “first/last mile” problem in transportation/supply chain management by increasing the accessibility to public transit networks (Liu et al. [2012]). The typical sources of revenue for these programs are funding from private organizations, subscription and fee from usage. These programs incur significant operational cost in terms of managing resources across their networks. Increasing ridership at a modest (if any) profit is the primary goal of these programs as they have significant social benefits. Most of the social benefits include increased transit use (Ma et al. [2015]), lower greenhouse gas emissions, improved public health (Woodcock et al. [2014]) and a safe mode of transportation (Martin et al. [2016]). As a result, most programs even provide the first 30-45 minutes of a ride at no cost to encourage participation by consumers.

Although a Bike Sharing program provides multiple benefits and promotes an alternative, attractive mode of transportation, these systems are subject to major operational challenges. One major challenge is the spatial and temporal imbalance of supply (of bikes) and demand (from consumers). These imbalances result from consumers' spatial and temporal preferences and allocation of resources over the network. Consumer behavior and choice of station use can lead to asymmetric demand flow in the network. For example, more consumers use the program while traveling from their residence to work in the morning. This leaves stations in the vicinity of residential locations empty (stock-out) and those near work locations filled with bikes (full-capacity), which creates a spatial imbalance. In the evening, not as many as morning consumers participate in riding bikes in the opposite direction. These inefficiencies cause frequent unavailability of bikes for an incoming consumer to initiate a trip, as well as unavailability of empty docking spaces at destination stations to complete the trip (Brussel

[3]The Bike Share ridership in the US has grown by approximately 10 fold from 2.3 million in 2011 to 28 million in 2016 - NACTO 2016

Nieuws (2010), Shaheen and Guzman [2011], Tusia-Cohen (2012)). As a result, demand from this consumer is either transferred to nearby stations, or lost to the bike network by being realized in a motorized public transport. The data from Capital Bike Share in Washington D.C shows that an average station is unable to handle a transaction (either pickup or drop-off) several times per day [4].

The unavailability of bikes for pickup is a significant cost to managers as the opportunity for a trip realization is lost. Operators could increase the likelihood of availability by allocating more bikes in the network. However, this can result in lower utilization and lower levels of service for consumers who seek to drop-off at destination stations but who cannot find an empty dock. Frequent unavailability of bikes or empty docks to regular consumers can lead to loss of interest and eventually deters them from participating in these programs, leading to lower ridership in the long term. Hence managing bike inventory and maintaining required service levels in the network of stations on a short-term basis is an important operational problem in Bike Sharing programs.

In this paper, I study the following four research questions: (i) How to estimate intrinsic consumer demand at each station in the network under an assumption of non-stationarity while incorporating consumer substitution across stations during stock-outs? An implied question that I answer is: Do consumers substitute across the stations when they experience stock-outs? If so, what is the impact of station pair distance on the likelihood of substitution? (ii) How should operators allocate inventory of bikes at the beginning of the day to maximize ridership? (iii) If the operators wish to re-allocate inventory multiple times in a day, when and in what quantities should the reallocation be performed to improve ridership?, (iv) What is the impact of ignoring station substitution on inventory allocation and ridership?

The rest of my paper is organized as follows. Section 2.3 provides details on my review of the literature and contribution of my work to the field of Operations Management. Section

[4] https://operationsroom.wordpress.com/2011/09/21/the-logistics-of-bike-sharing/

2.4 discusses the selection of a Bike Sharing program in United States, its operational data and necessary summary statistics. In Section 2.5, I develop a demand model integrating both stochastic models and large scale data to empirically estimate true or intrinsic demand. In Section 2.6, I develop closed form expressions for ridership and service level for single and multiple periods. Section 2.7, 2.8, provides an inventory allocation optimization model and results. Section 2.9 provides rigorous analysis of the value of incorporating substitution in demand estimation, optimal inventory allocation and ridership. Section 2.10 concludes the paper with key managerial implications and suggestions for future work.

2.3 Literature

My work builds on the prior literature in bike sharing, stock-out based substitution in demand systems, and stochastic modeling of service systems. I discuss the prior research in each of these areas and describe the contribution of my paper. The work in bike sharing can be classified into the following three areas: (i) Strategic design problems - number of stations, locations and capacity decisions; (ii) Demand analysis; and, (iii) Inventory management and repositioning of bikes.

First, on design problems, Lin and Yang [2011] studied the decisions on number of stations, locations in a bike sharing network using service level through demand between stations, travel and setup costs of lanes. García-Palomares et al. [2012] examined the station capacity decision problem in addition to number and location using GIS methodology. Dell'Olio et al. [2011] proposed a model using GIS information for setting pickup and drop-off locations as well as the maximum tariff to charge for commuters. My work contributes towards studying short-term operational decisions - allocation of bikes in the network after the station count, location and capacity numbers are decided.

Second, in demand analysis, Kaltenbrunner et al. [2010] used real-time data to build a predictive model of bike availability at any given station for defined periods of minutes in advance. Lathia et al. [2012] studied the impact of allowing casual users to access the London shared bike scheme. Faghih-Imani et al. [2015] explained the variation in bike pickup and drop-off demand using socio-demographic and land-use characteristics of multiple Bike Sharing programs in Spain. Singhvi et al. [2015] studied the problem of predicting station pair demand during rush hour for Citi Bike using variables for temporal, demographic, weather and taxi usage as covariates.

Kabra et al. [2016] and Zheng et al. [2018] modeled demand in Paris and London Bike Sharing systems (respectively) as aggregations of consumer preferences for station accessibility from their origin location, bike/dock availability and distance between stations, but incorporated only the consumer's first choice of station (i.e., all assessments were from their origin location to a station). My work extends this to incorporate the consumer's choice of first and second station (which may be a necessary substitution due to stock-out at the first). I also consider demand at each station as the sum of its true demand process (unobserved) and substitution spill-over from vicinity of stations during stock-outs, which neither Kabra et al. [2016] and Zheng et al. [2018] did. My model is thus rich in incorporating the interdependence of demand and spill-over across stations in the network. Henderson et al. [2016] addressed the problem of censoring in trip data but estimated true demand solely based on the re-balancing operations. During a selected time duration (e.g., rush hour), they tracked the end period bike inventory at every station; if the value was 0, they assumed that censoring occurred, thus they computed true demand as an average of flow rates for only the days when end period bike inventory was positive. This approach results in biased estimates of true demand due to sample selection bias. In addition, authors attribute the difference between the true rates and censored rates to lost demand. In my work, I did not exclude any data but instead I used every detail in the *minute-level inventory* data to define exact instances of stock-out and full-capacity, thus I can define the difference between true and censored rates as a combination of lost demand and station substitution. To my knowledge, mine is the first study to model demand substitution under stock-outs and interdependence of demand across stations in the network of Bike Sharing programs.

Third, in inventory management, Raviv and Kolka [2013] proposed an inventory model for a single station with bike allocation decisions made at the start of the day and derived an expression for the expected number of shortages under a Poisson model of the arrival and departure process. Shu et al. [2013] developed a network flow model under an assumption of constrained trips based on the initial allocation of bikes at each station. Utilizing data from Singapore MRT, the authors conducted a feasibility study and concluded 16% of demand for commute short trips can potentially be substituted by the Bike Sharing program. Henderson et al. [2016] extended the work of Raviv and Kolka [2013] with joint decision of bikes and docks in Citi Bike Sharing for single period (rush hour) under steady state. I extend the work of both Raviv and Kolka [2013] and Henderson et al. [2016] to multiple periods each of fixed duration, non-stationary demand and station substitution; I also include the following components as contributions to the literature : (i) I determine the optimal allocation of bikes across stations to maximize ridership using a dynamic program; and, (ii) I use transient analysis instead of steady state to model inventory transition between two subsequent periods. My modeling approach thus can be generalized to analyze service systems for any selected period length.

Fourth, with respect to the bike re-balancing studies in the vehicle rental systems, scholars have examined two types of questions : the quantity of bike inventory to be repositioned among stations and the optimal vehicle routes (Nair and Miller-Hooks [2011], Raviv et al. [2013], Schuijbroek et al. [2013], Nair et al. [2013]). The prior work on re-balancing operations can be replicated using my novel sources of data to improve estimates of performance measures.

Until now, studies of empirical demand estimation in bike sharing have considered only censored trip data[5] Further, analytical research in service systems has traditionally focused on improving the performance of operational measures (e.g. waiting time, cost, profit) for a single period under steady state. However, most service systems are inherently non-stationary and true behavior can be transient between successive time periods. I estimate true demand utilizing both stochastic models and large-scale datasets on minute-level inventory *and* censored trip data. To my knowledge, this is the first study to contribute in the following areas: (i) utilizing granular operational data to address censoring problems and substitution in demand estimation; (ii) incorporating non-stationarity in demand estimation and capture dynamics realized over successive periods through transient analysis; and, (iii) providing a methodological framework integrating stochastic models and large scale datasets to address operational issues in Bike Sharing programs contributing towards Big Data Analytics studies in Operations Management.

My work also contributes to the literature on demand estimation and inventory planning in operational systems under stock-out based substitution. Research on product stock-out and its implications for operational/financial outcomes (e.g., sales) has been primarily focused in the retail context. My research work is classified into three different areas such as accurate demand estimation for substitutable products, assortment/inventory planning, and the impact on financial outcomes (e.g., lost sales, profitability, margins). My research extends existing work towards modeling and estimating demand under substitution to systems operating as a *network*.

Anupindi et al. [1998] was one of the early papers on demand planning in substitutable products under stock-out in retail vending. They applied an expectation-maximization (EM) algorithm to a periodic inventory data (daily level) to obtain demand estimates considering stock-out time as missing data and showed that avoiding stock-out information can lead to

[5] Trip information in the demand data is recorded only when an incoming customer finds a bike at the station and initiates the trip. No information on demand from a potential consumer arriving during a station stock-out is captured. Hence the recorded trip information in Bike Sharing programs is censored.

biased demand estimates even for items with low stock-out frequency. Conlon and Mortimer [2013] built on the work by Anupindi et al. [1998] by collecting periodic information on product availability and sales at a higher frequency (four hours). They reached similar conclusions with respect to the bias of demand estimates. My work extends this to studying systems with continuous (minute-level) inventory data.

Musalem et al. [2010] built a structural model which allowed substitution between products based on consumer choices under a condition of stock-out. For shampoo products, they found average lost sales varied between 5.9% and 20.2% of full availability sales and contribution margins decreased by between 30% and 65% for most products with 20% or below availability. Vulcano et al. [2012] proposed a method to estimate substituted and lost demand when data on sales and product availability are observed for related products; similar work is found in van Ryzin and Vulcano [2014] and references within. A long line of research has studied the impact of the demand substitution effect on optimal inventory decisions (Smith and Agrawal [2000], Mahajan and Van Ryzin [2001], Kök and Fisher [2007], Honhon et al. [2010]) in retailing. The key conclusion of these studies is that if substitution is significant, avoiding its effect on demand estimation and inventory planning leads to incorrect inventory policies and hence lower profits. For a more extensive review, I refer the reader to Kök et al. [2015] in demand estimation, substitution effects in assortment and inventory planning. My work corroborates their conclusion and I additionally quantify its effect in Bike Sharing Programs which have the additional features of demand non-stationarity and network-based substitution traditionally not modeled in this literature.

2.4 Data & Summary Statistics

I utilize publicly available data[6] from the Bay Area Bicycle Sharing program to address my research questions. I provide a brief description of its operations and my reasons for

[6]http://www.bayareabikeshare.com/open-data

studying this system.

Bay Area Bicycle Sharing is a public program which began its operations in August 2013 with 700 bikes and 70 stations. Half of the total stations (35) are located in the city of San Francisco (hereafter "SFO"), and the remainder in Redwood City, Palo Alto, Mountain View and San Jose. The organization published granular data of its operations for the period 1^{st} September 2014 - 31^{st} August 2015. First, the published information includes the *trip* data which describes trips carried out by consumers between stations; this data comprises uniquely identified columns such as : *trip id*, *start and end date time of trip*, *start and end station id* (trip origin and destination), *consumer type* and uniquely identified *bike id*. Consumer type is classified based on length of the period they subscribe for service. Consumers subscribe for two types of membership: annual ($88) and casual (24 hour - $9/3 day - $22). Casual consumers request a new ride code for each trip by swiping their credit or debit card at the kiosk. The total number of trips realized during the period were 354,152.

Second, the published information also includes *minute-level inventory* data for each station. This data comprises uniquely identified columns such as: *station id*, *number of bikes*, *number of empty docks*, and *date time stamp*. The total number of records included in this data are 36,647,622. The third and the final dataset concerns *station specifications*, which comprises uniquely identified columns such as: *station id*, *latitude/longitude coordinates*, *city* (location) and *dock capacity*. The number of records in this data is 70, which corresponds to the total number of stations in the network. I utilize this data in conjunction with Google maps to obtain the walking and biking distance between every station pair.

Analyzing the *trip* data, I found over 90% of trips are realized within SFO and the remainder across the other cities. Hence I focused my analysis on the data after limiting it to SFO city. Similarly, I limited the *inventory* data to stations in SFO, which resulted in 18,321,582 records. For a selected station and day of operation, I expected to see 1,440 records (1440 minutes in a day) in the inventory data, however, for 35 stations and 13 days, I found fewer than 1000 records. I believe this might have happened due to technical failure which led to a cessation in information recording by electronic sensors. In addition, I found an abnormally low number of trips for 32 days which were holidays. I excluded the trip and inventory data for the 13 days with insufficient records and 32 days with low volume as these days did not represent a typical operational day. I made efforts to reduce any inventory inaccuracies that could be present in the data. The *number of bikes* column in the inventory data represent available bikes for use, however some of them may be in an unusable condition (tire puncture, broken chain, etc.). For example, if a station has 5 bikes and 2 of them are not in a usable condition, the data inaccurately captures 5 even though actual availability is 3. Henderson et al. [2016] and Kabra et al. [2016] considered stations to be empty if the number of bikes < 5. I performed an inventory inaccuracy correction by reducing the count of bikes if the idle time for a given bike-id at a given station is ≥ 48 hours (99^{th} percentile). After the data was cleaned, the number of records in *trip* and *inventory* data is 294,631 and 16,128,800 (320 days * 35 stations * 1440 minutes), respectively. I decided to focus on Bay Area Bike Sharing program for my work because it is the only program which made *inventory* data publicly available.

2.4.1 Trip Demand

I first summarize the cleaned *trip* data. I found 22 stations ($\approx$ 63% of total stations) contributed to 80% of the total trip demand. The three largest stations in descending order of their demand are SFO caltrain - Townsend at 4th (station code - 70), SFO caltrain 2-330 Townsend (station code - 69), and Harry Bridges Plaza-Ferry Building (station code - 50). These three stations combined represented 20% of the total trips.

Figure 2.1 displays inter-day variation in the number of trips in the station network. High demand is observed on weekdays and low demand on weekends. I refer to weekdays as “busy” and weekends as “non-busy” below. During busy days, I find Tuesday had more trips on average followed by Thursday, Wednesday, Monday and Friday.

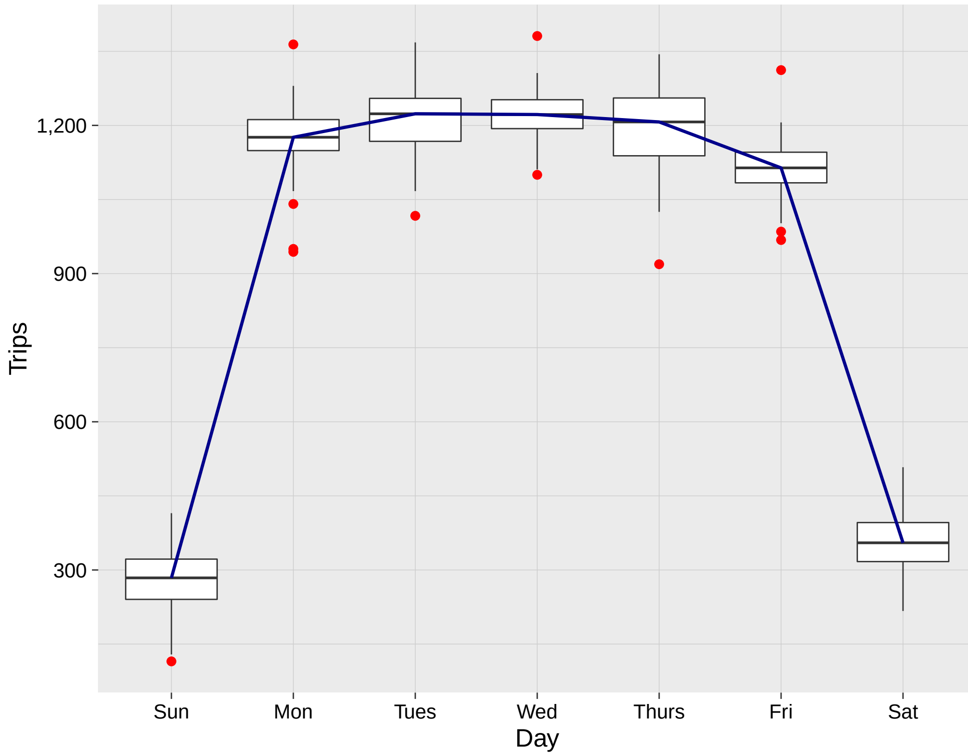
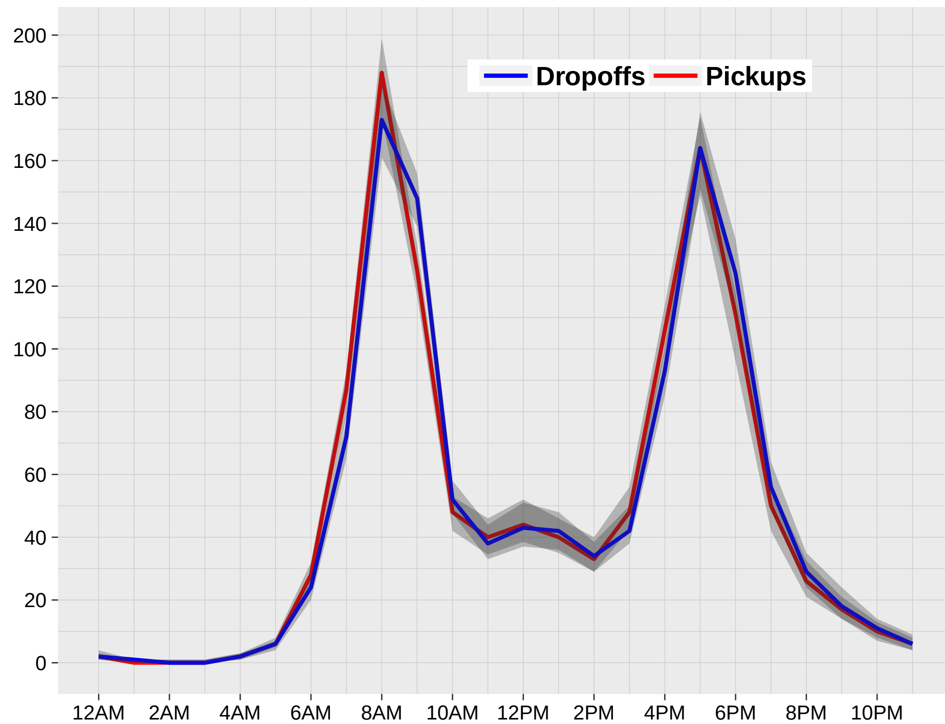


Figure 2.1: Inter-day trips in the station network

As Figure 2.1 indicates distinct patterns for busy and non-busy days, I analyzed the intra-day variation for both these day types separately. Figure 2.2 shows median trips in the station network by each hour for the entire day. During busy days, I find two peak periods, 8am-10am and 4pm-6pm. Close analysis of the first plot indicates that pickups are higher during 8am-9am compared to 4pm-6pm. This indicates the program usage by consumers is high on average during the morning compared to the evening. The demand pattern for non-busy days closely resembles a bell-shaped curve. The bands around median represent 25^{th} and 75^{th} percentile. Interestingly, I find bike share usage during 12am-6am on weekends is higher than weekdays.

I next analyzed variation in trip time between any station pair demand. Figure 2.3 shows the distribution of the trip time in minutes. The distribution is right skewed with 99% of trips less than 30 minutes. The program follows a pricing policy of unlimited trips up to 30



minutes at no additional cost. Trips of more than 30 minutes incur overtime fees. Overtime fee includes $4 for additional 30-60 minutes conditional on 30 minutes of trip completion time, $7 for every additional 30 minutes conditional on 1.5 hours of trip completion time.

2.4.2 Stock-Out and Full-Capacity

I summarize here the *inventory* data and explain two important states of stations derived from the value of the number of bikes available at any time moment. I first created a variable which represents a 2-hour time slot for each row. For example, time stamps between 12am-1:59am (extremes included) are coded as "12am", 2am-3:59am as "2am", 4am-5:59am as "4am" and similarly for the remaining part of the day. I then created two more variables: "*stock-out*" and "*full-capacity*". If number of available bikes = 0, the *stock-out* variable takes a value of 1, else it is set to 0. Similarly, if empty docks = 0, *full-capacity* variable takes a value of 1, else it is set to 0. I chose 2-hour time slots to ensure a sufficient number of trips

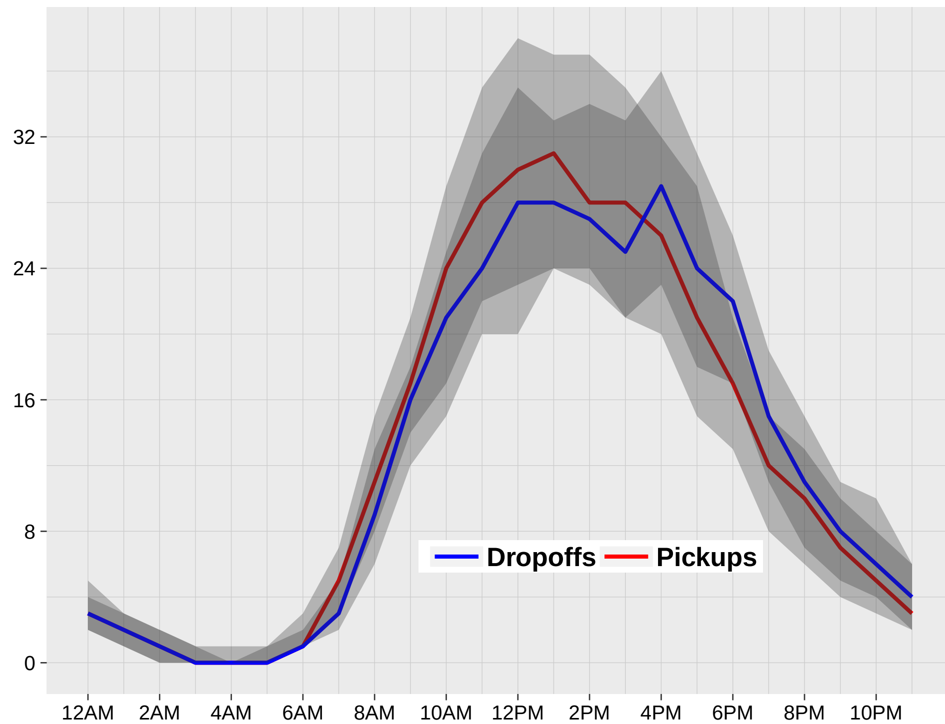


Figure 2.2: Intra-day trips in the station network (busy & non-busy days)

were realized during the slot in the entire dataset.

In Table 2.1, I summarize the variation in the proportion of stock-out stations and the stock-out time by each 2-hour slot for busy days. During peak period, I find 20-23% median number of stations were stocked out and the median stock-out time varied between 17-25 minutes. The proportion of stock-out stations is high during the 8am and 4pm slots compared to the remaining slots, however the average times are shorter.

Similarly, in Table 2.2, I summarize the variation in proportion of full-capacity stations and its time by each 2-hour slot for the busy days. During the entire day (6am-10pm), I find a little more than 3-9% median number of stations in full-capacity state and the average time varies between 11-26 minutes.

In the analytical research literature in service operations, scholars typically assume arrival/departure process following a Poisson distribution. For example, Gans et al. [2003]

Table 2.1: Summary of stocked out stations, stock-out time on busy days

	Proportion of stations (%)			Stock-out time (mins)		
Slot	P25	P50	P75	P25	P50	P75
6am	6	11	17	11.04	24.96	45.48
8am	20	23	26	9.00	18.96	33.00
10am	6	9	11	9.96	20.04	36.96
12am	3	6	6	10.77	20.04	35.04
2pm	3	3	6	9.96	23.04	47.04
4pm	14	20	23	8.04	17.04	29.04
6pm	14	20	26	11.28	24.00	42.00
8pm	3	3	6	8.04	18.00	34.50

Table 2.2: Summary of full-capacity stations, full-capacity time on busy days

	Proportion of stations (%)			Full-capacity time (mins)		
Slot	P25	P50	P75	P25	P50	P75
6am	4	6	9	5.04	12.00	27.72
8am	6	9	9	6.00	12.96	24.96
10am	3	6	9	9.00	18.96	36.24
12pm	3	3	6	6.00	17.04	32.04
2pm	3	3	6	10.77	26.52	48.96
4pm	3	6	9	5.04	11.04	21.96
6pm	6	9	9	6.96	15.96	30.48
8pm	3	5	6	7.77	15.00	26.04

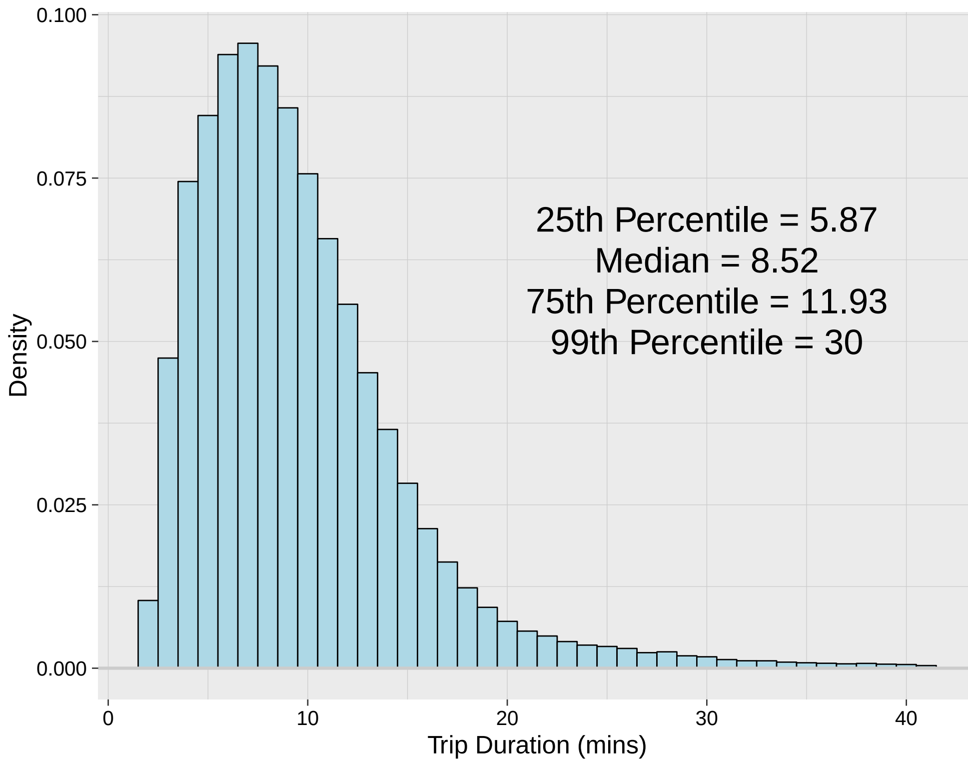


Figure 2.3: Trip time between pair of stations

considered customer arrivals at a contact center as a Poisson process. This specific distributional assumption helps to build tractable analytical models. In my work, using real data, I can statistically test and verify the distributional assumption of pickup and dropoff events at each station and a 2-hour slot. I use the familiar Kolmogorov-Smirnov one-sample test applied in the contact center setting by Brown et al. [2005] to verify if the inter-arrival time distribution follows the hypothesized exponential distribution. I implement the test on pickup and dropoff events individually for observations in each group formed by combination of station, day (Monday - Friday), 2-hour slot (6am, 8am, 10am, 12pm, 2pm, 4pm, 6pm and 8pm). For 35 stations, 5 days and 8 time slots, I have a total of 1400 groups. The hypothesis test specification for a group is provided below.

$$\mathrm{H}_0 : F(x) = F^*(x) \quad \forall\, x \geq 0$$

$$\mathrm{H}_1 : F(x) \neq F^*(x) \quad \text{for at least one value of } x$$

where $F(x)$ is unknown distribution of the observations in the group and $F^*(x)$ is hypothesized distribution (exponential).

An example test is presented for pickup events at station Townsend 7th on Thursday during 8am slot and dropoffs at An example test is presented for pickup events at station Townsend 7th on Thursday during the 8am slot and dropoffs at Townsend at 4th on Monday during the 10am slot in Table 2.3. For both pickups and dropoffs, the p-value > 0.05. Hence, I fail to reject the null hypothesis that my sample distribution and the hypothesized distribution are the same. at 4th on Monday during 10am slot in Table 2.3. For both pickups and dropoffs, the p-value > 0.05. Hence, we fail to reject the null that sample and hypothesized distribution are same.

Table 2.3: Goodness-of-fit test for Poisson distribution

	K-S statistic	p-value
Pickup events	0.0599	0.2273
Dropoff events	0.0956	0.1261

Among the 1400 groups, I find 80.0% pickup and 82.2% dropoff groups following an exponential distribution. To ensure robustness, I performed Chi-Square goodness of fit test for a Poisson distribution and obtained similar results. The robustness results are provided in Appendix 6.1. Hence, in the remaining part of my work, I assume pickups and dropoffs at each station during each 2-hour slot follow the Poisson distribution.

2.4.3 Station Specifications

I summarize here the *station* data. The data comprises information on station id, latitude/longitude coordinates, city to which station belongs and capacity. I implemented an algorithm in R which takes latitude/longitude coordinates and interacts with Google Maps to obtain the walking and biking distance between each station pair. In Table 2.4, I summarize station capacity, walking and biking distance between stations in miles. Walking and biking distance (or time) need not necessarily be the same for bi-directional station pair. Hence I differentiate between the two directions. I find median walking or biking distance to be little over a mile and mean/median station capacity of 19 in the network.

Table 2.4: Stations summary

	Min	P25	Median	Mean	P75	Max
Capacity	15	15	19	19	21	27
Walk (miles)	0.012	0.644	1.016	1.088	1.492	2.979
Bike (miles)	0.012	0.722	1.161	1.241	1.688	3.580

In Section 2.5, I develop, explain demand model and empirically estimate the necessary parameters of interest.

2.5 Demand Estimation

My objective is to develop a demand model under substitution for each station at two levels (bike pickup and dropoff). I propose that estimating demand either only from *trip* or *inventory* data can lead to biased estimates. My reasons for the above assertion are as follows.

Utilizing only trip data is likely to underestimate the true or intrinsic demand because all the trip pickups recorded in the data occur only when there is a minimum of one bike available. For example, if 30 pickups are recorded in the trip data during 8am slot at a station, it is reasonable to believe the true pickup demand is 30. However, assume inventory data now indicates that there was no bike available for 0.5 hours. This implies the station was in stock-out state for 0.5 hours. I now arrive at the conclusion that 30 pickups were realized during the 1.5 hours. A consumer arriving for pickup at the station during 0.5 hours of stock-out time would not find a bike. Demand from this potential consumer is either lost to an outside mode of transport or substituted in the vicinity of stations. Assuming pickup demand arrives uniformly and is lost upon realization of stock-out, the true pickup demand would be 40 pickups ($= \frac{2*30}{1.5}$). Hence, failing to incorporate the inventory information of 0.5 hours stock-out time leads to an estimate of pickup demand as 30 (censored) versus 40 (true value). Hence estimating demand only from trip data can lead to a lower estimate than the true value.

Similarly, utilizing only inventory data is likely to overestimate the true demand. An analyst can estimate demand from inventory data in the following way. A trip pickup (dropoff) is realized if the bike inventory decreases (increases) by 1 between two successive time stamps. However, inventory transitions alone do not completely indicate demand. Operators regularly re-balance bikes from stations with high dropoff demand to stations with high pickup demand. Inventory transitions due to re-balancing are captured in the data, but are not distinctly identified. Thus, the demand computation based on inventory transitions includes both true demand and re-balances. Failing to incorporate trip information and delineation of re-balance transition can result in a higher estimate of true demand. Hence, it is important and necessary to use both the *trip* and *inventory* data together to estimate true or intrinsic demand at each station.

I define intrinsic demand (pickups or dropoffs) as the maximum rates of bike departure or arrivals that could have occurred under no inventory or capacity constraints. I propose a model for both pickup and dropoff under stock-out and full-capacity individually. I then use the demand model in conjunction with the trip and inventory data to empirically estimate true demand.

2.5.1 Pickup Demand

I here develop a station bike pickup demand model under stock-out. I consider the following five plausible assumptions. (i) First, consumers arrive randomly to a station to pick a bike, commute and complete the trip by dropping off at the same or different station. (ii) Second, pickups from each station during any time slot follow Poisson process[7]. (iii) Third, consumers upon arrival to a station and experiencing stock-out probabilistically substitute to any station located within 10 minutes[8] of walking time during the same time slot. (iv) Fourth, consumers are informed on the inventory status of remaining stations and they find a bike upon substitution to a different station[9].

I now define necessary parameters which form basis of my model. Let "S" represent set of stations in the network, "t" for 2-hour time slot (e.g. 6am).

- μ_{it} - True or intrinsic pickup rate from station "i" during "t". We do not directly observe this process in the data. Hence, we empirically estimate its value
- $\overline{\mu_{it}}$ - Realized pickup rate from "i" during "t". We observe this process in the *trip* data
- τ_{it} - Proportion of stock-out time at station "i" during "t". For example, during 6am

---

[7] I statistically test and verify Poisson distribution assumption in Section 2.4.2

[8] Kabra et al. [2016] considered 300 meters from consumer origin location as their walking neighborhood for the measure of accessibility. I chose 10 minutes as a measure of distance based on the evidence from O'Sullivan and Morrall [1996], Zhao et al. [2003] which represents first step (consumer residence or work location to a bike station) transit accessibility distance. In my context, walking distance between station pairs is the second step (bike station to a bike station) of transit access as a result of bike unavailability in the first step. Consumers choice of neighborhood is likely to be different in both cases. In the latter case, how far the consumer prefers to walk to obtain a substitute depends on the trade-off between walking and trip time. It is fair to assume consumers do not have an incentive to substitute by walking if the walking time is greater than trip time even though the distance is < 300 meters. I found that median trip time to be 8.5 minutes and 99% of trips are completed in 30 minutes. I chose 10 minutes as it includes larger choice set of stations for substitution. I also tested this by using values of 5 and 8.5 minutes and found similar conclusions compared with 10 minutes. Results are provided in Appendix 6.2. Interestingly, I found that at 5 minutes of walking time the distance covered had a maximum of up to 411.99 meters which bounds the choice of distance considered by Kabra et al. [2016].

[9] Bike Sharing programs typically provide apps which consumers can install to be informed of the real-time status of bike availability at any station.

slot, if number of bikes at "i" is 0 for 15 minutes, then $\tau_{it} = 15/120 = 0.125$. We observe this process in the *inventory* data

- p_{jit} - Probability that a consumer arriving to station "j" experiences stock-out and substitutes to station "i" during "t". This parameter takes value in $(0, 1]$ when $\tau_{jt} > 0$. If $\tau_{jt} = 0$, we force $p_{jit} = 0$ as consumer has no incentive to substitute to "i" due to availability of bikes at "j".

The realized pickup rate at station "i" during "t" is specified by the following equation

$$\overline{\mu_{it}} = \left(\mu_{it} + \sum_{j \in Q_i} p_{jit} \cdot \mu_{jt} \cdot \tau_{jt}\right) \cdot (1 - \tau_{it}) \tag{2.1}$$

$Q_i = \{j \in S \mid \text{walk\_time}_{ji} \leq 10\,\text{mins}\}$. Q_i is the set of all stations in the network from which consumer substitution is feasible to "i". The factor $1 - \tau_{it}$ is incorporated in the expression to correct that pickups recorded in the data only during the non stock-out time.

The unknown parameters in the expression are $\{\mu_{it}, p_{jit} \mid i \in S, j \in Q_i\}$. The identification requirement for $\overline{\mu_{it}}$ is the station "i" should not be stocked-out for a large fraction of time ($\tau_{it} \neq 1$), for p_{jit} is the station "j" should be stocked-out for a large fraction of time ($\tau_{it} >> 0$). For the entire 35 station network and a slot "t", if each station "i" has average count of set Q_i as 5, this results in estimating 175 substitution probability parameters (p_{jit}). my inventory data does not provide sufficient variation is stock-out time to estimate the entire 175 substitution probability parameters. To reduce the parameter dimension and computation complexity, I parameterize substitution probability using the Multinomial Logit Choice Model.

Consider a consumer "c" arriving to a station "j" experiences stock-out and is confronted with a choice of stations, R_j where $R_j = \{o \in S \mid \text{walk\_time}_{jo} \leq 10\,\text{mins}\}$. R_j is the set of feasible stations in the network to which consumers substitute from "j". The consumer has preference or utility of an alternative $i \in R_j$ given by

$$U_{cjit} = \alpha_t + \beta_t \cdot \text{walk\_dist}_{ji} + \epsilon_{ct} \tag{2.2}$$

where α_t represents mean consumers utility to substitute, $\text{walk\_dist}_{ji}, \beta_t$ represents walking distance from station "j" to "i", consumers sensitivity to walking distance, ϵ_{ct} is a random component capturing $c^{'}s$ utility varying from alternative to alternative in R_j and difference in consumers innate preferences to substitute. Under the assumption of Gumbel distribution for random component, the consumer $c^{'}s$ probability of choosing a station $i \in R_j$ is given by

$$p_{jit} = \frac{e^{\alpha_t + \beta_t \cdot \text{walk\_dist}_{ji}}}{1 + \sum_{o \in R_j} e^{\alpha_t + \beta_t \cdot \text{walk\_dist}_{jo}}} \tag{2.3}$$

Substituting p_{jit} of expression (2.3) in (2.1) results in

$$\overline{\mu_{it}} = \left(\mu_{it} + \sum_{j \in Q_i} \frac{e^{\alpha_t + \beta_t \cdot \text{walk\_dist}_{ji}}}{1 + \sum_{o \in R_j} e^{\alpha_t + \beta_t \cdot \text{walk\_dist}_{jo}}} \cdot \mu_{jt} \cdot \tau_{jt} \right) \cdot (1 - \tau_{it}) \tag{2.4}$$

The pickup demand model (2.4) is highly non-linear with different cardinality of sets Q_i, R_j for each i, j. The model incorporates interdependence of demand across stations and spill-over as a function of walking distance.

2.5.2 Estimation - Pickup

I estimate the parameter set - intrinsic pickup rates and choice model parameters $\{\mu_{it}, \alpha_t, \beta_t \mid i \in S\}$ using the stochastic model together with the trip and inventory data. In the remainder of this chapter, I focus my analysis only on busy days as the program is used more during weekdays than on weekends. During the data analysis, I found there is insufficient variation on the proportion of stock-out time τ_j, $\forall j \in S$ for each day (Monday, Tuesday, Wednesday, Thursday, Friday) when analyzed individually (Monday different from any other day). Hence, I decided to estimate parameters for a uniform busy day rather than differentiating for each day.

I performed a Maximum Likelihood Estimation using the model and data. Let "D" represent the set of all "busy" days, g_{ilt}, τ_{ilt} be realized pickups and the proportion of stock-out time from station "i" on busy day "l" during "t". The optimization problem is given by:

$$\max_{\mu_{it},\alpha_t,\beta_t} \prod_{l=1}^{D} \prod_{i \in S} \frac{e^{-\left(\overline{\mu_{it}} \cdot T\right)} \left(\overline{\mu_{it}} \cdot T\right)^{g_{ilt}}}{g_{ilt}!}$$

subject to

$$\mu_{it} > 0 \quad \forall\, i \in S$$

where $\overline{\mu_{it}}$ is given by equation (2.4)

$$\overline{\mu_{it}} = \left(\mu_{it} + \sum_{j \in Q_i} \frac{e^{\alpha_t + \beta_t \cdot \text{walk\_dist}_{ji}}}{1 + \sum_{o \in R_j} e^{\alpha_t + \beta_t \cdot \text{walk\_dist}_{jo}}} \cdot \mu_{jt} \cdot \tau_{jt}\right) \cdot (1 - \tau_{it})$$

For a 35 station network and specified slot "t", the number of parameters to be estimated are 37 : 35 pickup rates ($\mu_{it} \mid i \in S$) + 2 Multinomial choice parameters (α_t, β_t). The estimation process is run for each time slot, $t \in$ {6am, 8am, 10am, 12pm, 2pm, 4pm, 6pm, 8pm} (8 times).

To realize the value of incorporating inventory data into our demand estimation process, we ran the optimization for any slot "t" under three different cases. Table 2.5 lists the three cases.

Case 1 - Absence of inventory data. In this setting, I use only censored trip data. The proportion of stock-out time and the substitution probability parameters are excluded from

Table 2.5: Exhaustive cases of pickup demand estimation

Case	Inventory information	Substitution	$\tau_{it}, \forall i \in S$	p_{jit}
1	No	No	0	0
2	Yes	No	> 0	0
3	Yes	Yes	> 0	> 0

the demand model. The model is then simplified as the intrinsic demand rate is equal to the realized demand rate. The likelihood function is a product of the probability mass function of Poisson distribution for each station and day. The function can then be further simplified as the sum of log likelihood functions for each station independently. The estimated pickup rate at a station "i" and slot "t" is given by:

$$\mu_{it} = \frac{\sum_{l=1}^{D} g_{ilt}}{D \cdot T} \tag{2.5}$$

Case 2 - Presence of inventory information with no substitution (i.e., lost sales/demand in the traditional retail context). In this setting, I use both the censored trip data and the inventory data. When a consumer arrives at a station for a bike pickup and faces stock-out, (s)he chooses to opt for an alternative mode of commute (realistically, motorized transport). The estimated pickup rate at a station "i" and slot "t" is given by:

$$\mu_{it} = \frac{\sum_{l=1}^{D} g_{ilt}}{(D - \sum_{l=1}^{D} \tau_{ilt})T} \tag{2.6}$$

Comparing Case 1 and Case 2, I find failing to incorporate inventory information and the estimated pickup rates will differ by $\frac{\sum_{l=1}^{D} \tau_{ilt}}{D}$.

Case 3 - Presence of inventory information with substitution. This case refers to the lost sales/demand and station substitution. In this setting, I again use both the censored trip data and the inventory data. The likelihood function cannot be further simplified to have a closed form. Hence I solve the optimization problem using an iterative approach. I implemented the procedure in R.

I summarize the demand estimates for the two largest stations in the network, SFO caltrain 2-330 Townsend (code 69) and SFO caltrain-Townsend at 4^{th} (code 70), which are close substitutes to each other[10]. Figure 2.4 and 2.5 display the pickup rates of station 69 and 70 for the above three cases. The blue line refers to Case 1 - Censored demand, orange line to Case 2 - Lost demand and the black line to Case 3 - Substitution.

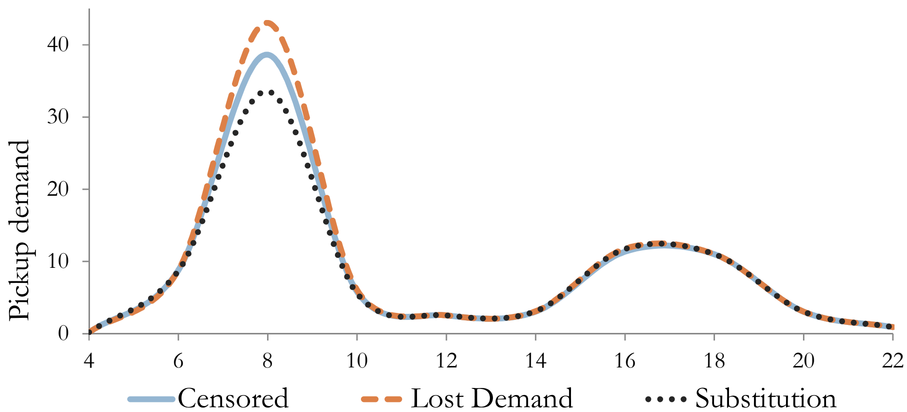


Figure 2.4: Busy day pickup rate at station 69

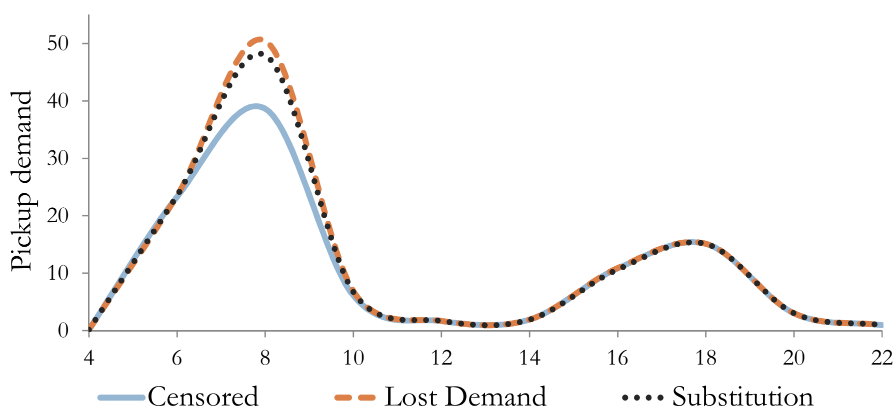


Figure 2.5: Busy day pickup rate at station 70

The common conclusions derived from both the plots are as follows: (i) a visual difference

[10]Walking distance between stations 69 and 70 is 18.99 meters

in the estimates in the three cases exists only during the 8am slot, which is the morning peak period; (ii) the demand estimates under Case 2 are higher than those under Case 1 (for station 69, I find true pickup rate under substitution is lower is compared to lost demand, but for station 70, I find true pickup rate under substitution is higher compared to lost demand, which implies that more consumers substitute from 69 to 70 than 70 to 69)[11]. I did not find significant difference in the estimates for the remaining slots in the day. If a manager decides to reset the inventory allocation to maximize pickups before the 8am slot and utilizes the demand estimates from censored demand (Case 1), (s)he would allocate more (less) inventory to station 69 (70) compared to their true optimal values[12]. Hence, my demand estimation process signifies the importance of incorporating inventory information with the trip data. Failing to incorporate this process in optimizing the system can likely lead to sub-optimal allocation. Also, I performed a Likelihood Ratio test between Case 2 and Case 3 to assess the value of substitution in the estimation process. I found that the Log Likelihood was statistically higher under Case 3 than under Case 2 during the 6am, 8am, 10am, 4pm and 6pm slots.

In addition to the demand rates, the optimization procedure provides estimates of Multinomial Logit Choice model parameters $\{\alpha_t, \beta_t\}$. Table 2.6 summarizes the parameters for each slot for the entire day.

Table 2.6: Multinomial logit model estimates (10 mins of walking time)

Variable	6am	8am	10am	12pm	2pm	4pm	6pm	8pm
Walking distance	-0.385***	-6.738***	-5.627***	-0.003	-0.026	-3.571***	-7.07***	-0.011
	(0.097)	(0.046)	(0.027)	(0.019)	(0.028)	(0.501)	(0.007)	(0.048)
Constant	-0.367***	2.672***	1.3297***	0.0067	0.0018	0.2971**	1.207***	-0.0099
	(0.046)	(0.249)	(0.022)	(0.02)	(0.061)	(0.144)	(0.002)	(0.045)
-LogLL	-18828.14	-109420.70	-720.54	749.83	701.52	-69796.74	-24238.48	5911.24

*p<0.1; **p<0.05; ***p<0.01

[11] Median stock-out time of station 70 is 29 minutes, almost double that of station 69 (15.96 minutes)

[12] I request the reader to follow Section 2.9 for more details on the impact of demand substitution on optimal inventory allocation decisions and ridership.

I find that both the constant and coefficient of walking distance were statistically significant during the peak periods (6am, 8am, 10am, 4pm and 6pm) and non-significant during non-peak periods. My findings provide statistical evidence for the existence of consumer substitution under stock-out. The direction of the distance coefficient is negative, which implies that a greater walking distance is perceived by the consumer as a barrier to substitute (in other words, if they have to walk farther to substitute they are less likely to do so). Comparing the coefficients in all the time slots, I find this lack of utility for walking is higher during peak periods than it is for the non -peak periods.

Utilizing the estimated choice model parameters $\{\hat{\alpha}_t, \hat{\beta}_t\}$, I compute the probability of consumer substituting station "j" to "i" during "t" given by

$$p_{jit} = \frac{e^{\alpha_t + \beta_t \cdot \text{walk\_dist}_{ji}}}{1 + \sum_{o \in R_j} e^{\alpha_t + \beta_t \cdot \text{walk\_dist}_{jo}}}$$

Table 2.7 summarizes the probability of substitution across stations in the network for each time slot. During peak period, I find consumers are more likely to substitute in morning compared to the evening.

Table 2.7: Summary of substitution probability across stations in the network

Variable	6am	8am	10am	12pm	2pm	4pm	6pm
Mean	0.735	0.917	0.797	0.818	0.816	0.698	0.701
Stdev	0.113	0.089	0.119	0.092	0.093	0.120	0.141

2.5.3 Dropoff Demand

My next objective is to develop and estimate a dropoff demand model under full-capacity. Similar to the argument made for pickups, I propose that computing station pair-wise demand only from *trip* data does not yield the true demand. My reasons for the above assertion result from a combination of capacity and trip completion. For example, let me consider the following statements: (i) A three station network indexed as {i, j, k} and "k" is nearest to "j". (ii) A total of 50 trips originated from "i" and "i" is never stocked out. The 50 trips are distributed as 30 from $i \rightarrow j$ and 20 from $i \rightarrow k$ in the trip data. (iii) Station "j" is filled with bikes to its capacity (i.e., it is in full-capacity state) for 0.5 hours . (iv) Station "k" is available with empty docks through the entire slot period, implying any consumer can find an empty slot to complete the trip at "k". A consumer originating a trip $i \rightarrow j$ during the "j" full-capacity time would not find an empty dock and hence cannot complete the trip. Consumer then bikes from $j \rightarrow k$ and completes the trip. This indicates a part of the $i \rightarrow j$ demand is substituted to $i \rightarrow k$. Under the assumption of linear rate, the true demand for from $i \rightarrow j$ and $i \rightarrow k$ are 40 and 10, respectively. Hence, failing to incorporate the inventory information of 0.5 hours full-capacity time leads to an estimate of $i \rightarrow j$ demand as 30 instead of 40 (the true value). This example shows that failing to account for inventory information at full-capacity time can lead to biased estimates of each arc demand in the network.

Table 2.2 summarizes variation in proportion of stations in full-capacity state and its time by each 2-hour slot for busy days. During the entire day (6am-10pm), I find a little over 3-9% median number of stations in full-capacity state which amounts to 1 to 3 stations out of 35. This left me with extremely little variation in full-capacity time from the inventory data to use to build and estimate the dropoff demand model. Hence, I chose to use the linear distribution of flow rates in the trip data and apply them to pickup demand to estimate dropoff demand parameters. q_{jit} is proportion of trips realized from source station "j" to destination "i" during "t". I differentiate q_{jit} for every time slot "t" due to temporal variation in demand, μ_{jt} are the pickup rates estimated under presence of inventory information and station substitution (Case 3) and λ_{it} are the estimated dropoff rates. For the optimization model in the latter part of my dissertation, I utilize the demand estimates from Case 3[13].

Up to this point, I have been addressing my first research question concerning how to estimate true demand while taking into account spatial and temporal imbalance and substitution across stations during stock-outs. In the next section I develop expressions for ridership and service level for single and multiple periods.

2.6 Ridership and Service level

I here develop a closed form expression for expected pickups, dropoffs and their service level for a station and time slot. Consider a station with capacity "C" which is subjected to pickup and dropoff at Poisson rates μ and λ (respectively). Under these assumptions, bike inventory status, $X(u)$, is equivalent to the number of consumers in a $M/M/1/C$ queuing system (Figure 2.6), where inter-arrival and service times are exponentially distributed with parameters λ, μ and "C" is the capacity of the queuing system (waiting + service). The state space is given by $SS = \{0, 1, 2, 3, \ldots, C\}$.

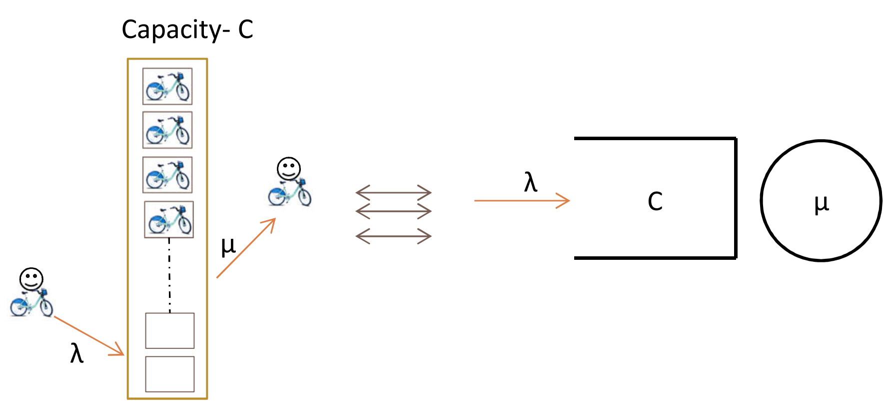


Figure 2.6: Equivalence of Bike station and M/M/1/C Queuing system

For a slot of length "T", I develop an expression for the expected occupancy time in inventory "n" conditional on the initial inventory "m". Let $Z_n(u)$ be an indicator variable

[13] I request the reader to follow Appendix on the measures for out of sample model fit

whose value is 1 if inventory at time u is "n", 0 otherwise. Mathematically,

$$Z_n(u) = \begin{cases} 1 & \text{if } X(u) = n \\ 0 & \text{otherwise} \end{cases}$$

Occupancy time, $V_n(T)$ in inventory "n" during $[0, T]$ is given by

$$V_n(T) = \int_0^T Z_n(u)\, du$$

Expected occupancy time, $M_{mn}(T)$ in inventory "n" conditional on initial inventory "m" is given by

$$\begin{aligned} M_{mn}(T) &= E\left(V_n(T)\middle| X(0) = m\right) \\ &= E\left(\int_0^T Z_n(u)\, du\middle| X(0) = m\right) \\ &= \int_0^T E\left(Z_n(u)\middle| X(0) = m\right) du \\ &= \int_0^T P\left(X(u) = n\middle| X(0) = m\right) du \\ &= \int_0^T p_{mn}(u)\, du \end{aligned}$$

where $p_{mn}(u)$ represents probability of inventory "n" after $u \in (0, T]$ conditional on initial inventory "m" and $p_{mn}(0) = \mathbf{1}_{n=m}$. Customer pickups occur only during the time inventory is non-negative, the expected occupancy time in inventory $n > 0$ during $(0, T]$ given by

$$\begin{aligned} \int_0^T \sum_{n=1}^{C} p_{mn}(u)\, du &= \int_0^T (1 - p_{m0}(u))\, du \\ &= T - \int_0^T p_{m0}(u)\, du \end{aligned}$$

Hence, the expression for expected pickups is given by $\mu \cdot \left(T - \int_0^T p_{m0}(u)\, du\right)$ which is interpreted as difference between maximum pickups that is likely to occur and lost demand. The number of pickups that are likely to happen during the stock-out time is $\mu \cdot \int_0^T p_{m0}(u)\, du$. Using the similar approach, we can derive expected dropoffs given by $\lambda \cdot \left(T - \int_0^T p_{mC}(u)\, du\right)$ where $\int_0^T p_{mC}(u)\, du$ is expected occupancy time in state "C" (full-capacity) during $(0, T]$ conditional on initial inventory "m". We formally represent these two expressions as follows

$$\begin{aligned} f(m) &= \mu \cdot \left(T - \int_0^T p_{m0}(u)\, du\right) \\ g(m) &= \lambda \cdot \left(T - \int_0^T p_{mC}(u)\, du\right) \end{aligned} \tag{2.7}$$

2.6.1 Service Level

The concept of service level in bike sharing programs is important both from the planning and practice perspective. I define Service level as the probability that a customer finds an opportunity to complete a transaction (pickup or drop-off). From the planning view, the decision of bike inventory allocation impacts the service level for pickup and dropoff. For a specified demand rate (μ, λ), if the allocation is too low, the station offers more empty docks for dropping a bike, however, the probability of station transitioning to stock-out state is high. This results in a greater chance of bike unavailability for a pickup which causes problems for the pickup service level. Similarly, if the allocation is too high, the station offers more bikes for pickup, however, the probability of station transitioning to full-capacity state is higher, which results in a greater chance of empty dock unavailability for a dropoff thereby lowering the dropoff service level.

From the practice view, the revenue realized by most of the Bike Sharing programs is the sum of subscription and overtime fees (realized only when trip time > 30 minutes). Currently 99% of the trips fall within 30 minutes (Figure 2.3) which implies much of the cash inflow is through subscription. Hence for the program to maintain an uninterrupted inflow of cash, it is important for managers to maintain adequate pickup and dropoff service level for consumers. Failing to maintain a sufficient service level results in more frequent instances of stock-out and full-capacity. A potential consumer arriving at the station is likely to not complete a transaction and this leads to lower satisfaction. In addition, if the consumer faces these instances more frequently, it could deter them from participating in the program thus leading to less ridership in the long term. Hence, Service level is an important attribute for better management of bike sharing programs. The expressions for the service level are given by:

$$\text{Pickup service level} = \frac{\text{E[Realized pickups]}}{\text{E[Maximum pickups]}} = \frac{\mu \cdot \left(T - \int_0^T p_{m0}(u)\, du\right)}{\mu \cdot T} = 1 - \frac{\int_0^T p_{m0}(u)\, du}{T}$$

$$\text{Dropoff service level} = \frac{\text{E[Realized dropoffs]}}{\text{E[Maximum dropoffs]}} = \frac{\lambda \cdot \left(T - \int_0^T p_{mC}(u)\, du\right)}{\lambda \cdot T} = 1 - \frac{\int_0^T p_{mC}(u)\, du}{T}$$

Formally, a generalized expression is given by

$$sl(m, n) = 1 - \frac{\int_0^T p_{mn}(u)\, du}{T} \tag{2.8}$$

where $sl(m, 0)$, $sl(m, C)$ represent pickup and dropoff service level respectively.

2.6.2 Ridership and Service Level trade-off

We state the trade-off managers face between ridership and service level. For a specified demand (μ, λ), expected pickups, dropoff service level are non-decreasing, non-increasing functions of starting inventory. Similarly, expected dropoffs, pickup service level are non-increasing, non-decreasing functions of starting inventory.

Lemma 1. *For given μ and λ at a bike station with capacity "C",*

1. $f(m+1) \geq f(m)$ *and* $sl(m+1, C) \leq sl(m, C)$

2. $g(m+1) \leq g(m)$ and $sl(m+1,0) \geq sl(m,0)$

2.6.3 Ridership under Non-stationarity

Operational measures in service systems, for e.g. inbound call volume to contact center (Aksin et al. [2013]), traffic to retail store (Perdikaki et al. [2012]) exhibit high degree of non-stationarity. Similarly in our setting, we find both pickups and dropoffs are highly non-stationary (E.g. Station 69 and 70 in Figure 2.4, 2.5). Hence, we develop expressions for expected pickups, dropoffs and their respective service levels for a single station and multiple periods as function of starting period inventory.

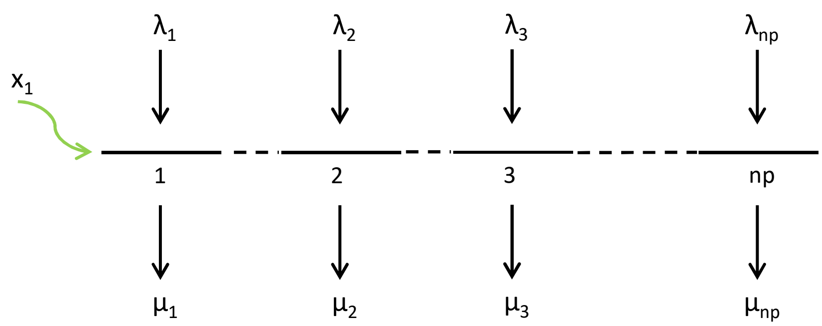


Figure 2.7: Demand rates under non-stationary process

A schematic representation of operational day with "*np*" slots/periods is displayed in Figure 2.7. Each period is distinguished with different pickup and dropoff rates. Let the starting bike inventory be x_1. We define a sequence of functions $f_t(x_t)$ for $t = 1, 2, 3, \dots, np$ which represents expected pickups over current period "t" and subsequent periods conditional on inventory state x_t at start of period "t". The value function for a given inventory at any previous time slot can be calculated through backward induction given by

$$f_t(x_t) = \mu_t \cdot \left(T - \int_0^T p^t_{x_t 0}(u)\, du \right) + \sum_{x_{t+1}=0}^{C} \left([P^{(t)}]_{x_t\, x_{t+1}} \cdot f_{t+1}(x_{t+1}) \right) \tag{2.9}$$

where $p^t_{mn}(u)$ is probability of inventory "n" after time "u" conditional on initial inventory "m" and $P^{(t)}$ is transition probability matrix during period "t". The expected pickups over "np" periods is given by $f_1(x_1)$.

Similarly, we can define a sequence of functions $g_t(x_t)$ for $t = 1, 2, 3, \ldots, np$ which represents expected dropoffs over current period "t" and subsequent periods conditional on inventory state x_t at start of period "t". The expected dropoffs over "np" periods is given by $g_1(x_1)$ where

$$g_t(x_t) = \lambda_t \cdot \left(T - \int_0^T p^t_{x_t 0}(u)\, du\right) + \sum_{x_{t+1}=0}^{C} \left([P^{(t)}]_{x_t\, x_{t+1}} \cdot g_{t+1}(x_{t+1})\right)$$

The expected pickup and dropoff service levels are given by $\frac{f_1(x_1)}{\sum_{t=1}^{np} \mu_t \cdot T}$ and $\frac{g_1(x_1)}{\sum_{t=1}^{np} \lambda_t \cdot T}$. The expressions $f_t(x_t)$ and $g_t(x_t)$, involve two major components, $p_{mn}(T)$ and $\int_0^T p_{mn}(u)\, du$. The Lemma below provides a closed form expression for transition probability, $p_{mn}(T)$ and subsequently $\int_0^T p_{mn}(u) du$ can be derived.

Lemma 2. *A closed form expression for $p_{mn}(u)$ exists (Morse [2004]) and given by*

$$p_{mn}(u) = \pi_n + \frac{2 \cdot \rho^{\frac{n-m}{2}}}{C+1} \cdot \sum_{s=1}^{C} \frac{\mu}{k_s} \cdot K_s \cdot e^{-k_s u}$$

where

$$K_s = \left(\sin \frac{sm\pi}{C+1} - \sqrt{\rho} \cdot \sin \frac{s(m+1)\pi}{C+1}\right) \cdot \left(\sin \frac{sn\pi}{C+1} - \sqrt{\rho} \cdot \sin \frac{s(n+1)\pi}{C+1}\right)$$

$$k_s = \lambda + \mu - 2 \cdot \sqrt{\lambda \cdot \mu} \cdot \cos\left(\frac{s\pi}{C+1}\right)$$

$$\rho = \frac{\lambda}{\mu}$$

$$\pi_n = \begin{cases} \frac{1}{C+1} & \textit{if } \rho = 1 \\ \frac{1-\rho}{1-\rho^{C+1}} \cdot \rho^n & \textit{otherwise} \end{cases}$$

Definite integral $\int_0^T p_{mn}(u)\, du$ is given by

$$= \pi_n \cdot T + \frac{2 \cdot \rho^{\frac{n-m}{2}}}{C+1} \cdot \sum_{s=1}^{C} \frac{\mu}{k_s^2} \cdot K_s \cdot (1 - e^{-k_s T})$$

We provide proof for the expression $p_{mn}(T)$ in Appendix 6.4. The transition probability has two parts - steady state and transient adjustment. The transient part ensures the dependence between starting inventory of first period and future periods. Hence, we note that it is important to incorporate transient analysis in building the operational metrics.

2.7 Empirical Application

In the traditional retail context, satisfying incoming demand leads to only reduction of inventory. Hence it is necessary to have an inventory ordering policy at the start of every period. Inventory transitions in a bike sharing network are different from those in a retail setting. The randomness in pickups and dropoffs, both in sizable proportion, results in depletion and superfluity of inventory. Hence, it may not be necessary to reset the inventory at the start of every period.

I build an optimization model to maximize pickups at each station subject to a constraint on dropoff service level and the number of bikes in the network. I choose to maximize pickups because unavailability of bikes for pickup is a significant cost to managers as the opportunity for a trip realization is lost. Figure 2.8 shows the temporal variation in pickups and dropoffs at station 69 over the operational day (6am-10pm). I find significant difference in pickups and dropoffs during the slots between 6am-10am and 2pm-8pm. During 10am-2pm, the difference is fairly minimal across all stations in the network. As the imbalance starts at

6am and 2pm slots, I choose to allocate inventory at these two time points on each day. For analysis, I divide the operational day into two parts, 6am-2pm and 2pm-10pm. The former is referred to as “morning” and the later as “evening.” I define the necessary variables below which form the basis for my optimization model.

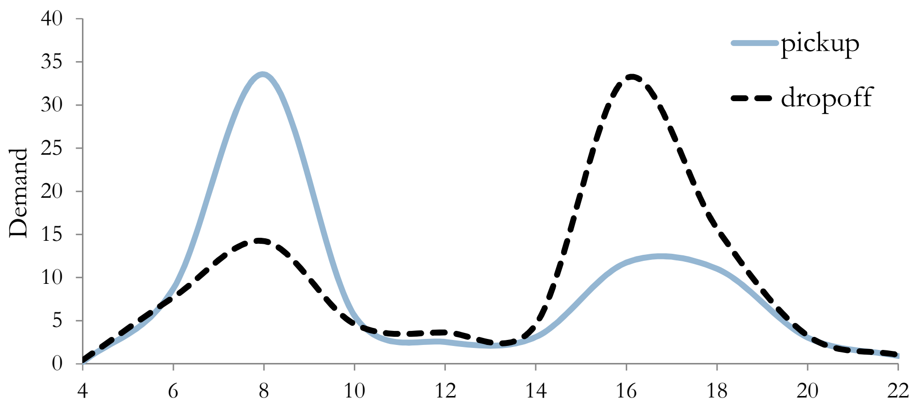


Figure 2.8: Busy day demand at SFO caltrain 2-330 Townsend (69)

- x_{i1} - inventory decision to station “i” at start of “morning” (6am)
- $f(\text{morning}, x_{i1})$ - expected pickups realized during 6am-2pm with starting inventory x_{i1}
- $ds(\text{morning}, x_{i1})$ - expected dropoff service level realized during 6am-2pm with starting inventory x_{i1}
- x_{i2} - inventory decision to station “i” at start of“evening” (2pm)
- $f(\text{evening}, x_{i2})$ - expected pickups realized during 2pm-10pm with starting inventory of x_{i2}
- $ds(\text{evening}, x_{i2})$ - expected dropoff service level realized during 2pm-10pm with starting inventory of x_{i2}

Our optimization formulation is given as follows

$$\max_{x_{i1},x_{i2}} \sum_{i\in S} f(\text{morning}, x_{i1}) + \sum_{i\in S} f(\text{evening}, x_{i2})$$

subject to

$$ds(\text{morning}, x_{i1}) \geq ds(\widehat{\text{morning}}, i) \quad \forall\, i \in S \tag{2.10}$$

$$ds(\text{evening}, x_{i2}) \geq ds(\widehat{\text{evening}}, i) \quad \forall\, i \in S \tag{2.11}$$

$$\sum_{i\in S} x_{i1} = \sum_{i\in S} x_{i2} = \text{nb} \tag{2.12}$$

$$0 \leq x_{i1}, x_{i2} \leq C_i \tag{2.13}$$

The functional forms for $f(\text{morning}, x_{i1}), f(\text{evening}, x_{i2}), ds(\text{morning}, x_{i1})\text{and}ds(\text{evening}, x_{i1})$ are provided in the section 2.6.3. Constraints 10, 11 ensure a required minimum dropoff service level is maintained across all the stations during morning and evening. The right hand side of the constraints are empirically estimated from data. Managers can also set the dropoff service level which makes our problem specification more flexible. Constraint 12 ensures all the resources ("*nb*" bikes) are allocated in the network. Excluding this constraint, the problem simplifies to individual station optimization. Constraint 13 ensures allocation at each station is non-negative and bounded by its capacity.

The objective function and two dropoff service level constraints in the problem are highly non-linear and posses a dynamic structure leaving a challenging task to solve the problem. We transform the optimization problem into a Binary program. For each "morning/evening", we enumerate pickups, dropoff service level for exhaustive choice of starting inventory across all the stations. Let "*ncap*" be the sum of capacity of stations in the network. The Binary program is given below. The bold letters represent either vectors or matrices.

$$\max_{\mathbf{y},\mathbf{z}} \left(\mathbf{f}_{\text{morning}}^{\mathsf{T}} \cdot \mathbf{y} + \mathbf{f}_{\text{evening}}^{\mathsf{T}} \cdot \mathbf{z} \right)$$

$$\text{subject to}$$

$$\mathbf{A} \cdot \mathbf{y} \geq \mathbf{b}_{\text{morning}}$$

$$\mathbf{A} \cdot \mathbf{z} \geq \mathbf{b}_{\text{evening}}$$

$$\mathbf{C} \cdot \mathbf{y} = \text{nb}$$

$$\mathbf{C} \cdot \mathbf{z} = \text{nb}$$

$$\mathbf{f}_{\text{morning}} = \left[\text{Morning Pickups}_{(ij)} \right]_{(ncap+N)*1}$$

$\mathbf{f}_{\text{morning}}$ is column vector of order $(ncap+N)$. Each element of the vector, Morning Pickups$_{(ij)}$ represents expected pickups at station "i" conditional on starting inventory "j" during morning. Similarly

$$\mathbf{f}_{\text{evening}} = \left[\text{Evening Pickups}_{(ij)} \right]_{(ncap+N)*1}$$

Matrices **A** and **C** are given by

$$\mathbf{A} = \left[\mathrm{a}_{k,(ij)} \right]_{N*(ncap+N)} \qquad \mathrm{a}_{k,(ij)} = \begin{cases} 1 & \text{if } k = i \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{C} = \left[\mathrm{c}_{k,(ij)} \right]_{N*(ncap+N)} \qquad \mathrm{c}_{k,(ij)} = \begin{cases} j & \text{if } k = i \\ 0 & \text{otherwise} \end{cases}$$

where row of each element in the matrix is station and column is combination of station and its inventory level.

$$\mathbf{b}_{\text{morning}} = \Big[\text{Morning Dropoff Service Level}_{(ij)}\Big]_{(ncap+N)*1}$$

$\mathbf{b}_{\text{morning}}$ is column vector of order $(ncap + N)$.

Each element of the vector, Morning Dropoff Service Level$_{(ij)}$ represents expected dropoff service level at station "i" conditional on starting inventory "j" during morning. Similarly

$$\mathbf{b}_{\text{evening}} = \Big[\text{Evening Dropoff Service Level}_{(ij)}\Big]_{(ncap+N)*1}$$

The decision variables are given by

$$\mathbf{y} = \big[\mathrm{y}_{(ij)}\big]_{(ncap+N)*1} \qquad \mathrm{y}_{(ij)} \in \{0, 1\}$$

$$\mathbf{z} = \big[\mathrm{z}_{(ij)}\big]_{(ncap+N)*1} \qquad \mathrm{z}_{(ij)} \in \{0, 1\}$$

where $i, k \in \{1, 2, 3, \ldots N\}$ and $j = \{0, 1, 2, 3, \ldots C_i\}$. I code the exhaustive enumeration of pickups/dropoffs service level for each choice of inventory level and solve the Binary program in R and Matlab. The optimal allocation and hence ridership are computed subsequently.

2.8 Results:

I initially present the results of optimization problem resolution where the inventory is reset at 6am in morning and 2pm in evening period. Table 2.8 presents optimization results for each reset time and operational day. The optimal inventory allocation results in ridership of 554.656 and a service level of, 0.959 during morning. Similarly, results follow for evening and the entire operational day.

Table 2.8: Optimal Ridership, Service Level

Period	Reset time	Ridership (Pickups)	Service Level (Dropoff)
Morning	6am	554.656	0.959
Evening	2pm	533.480	0.970
Operational day	6am, 2pm	1088.136	0.965

2.8.1 Current vs Optimal policy

I next compare the managers current policy to an optimal policy suggested by my optimization model. To find current policy, I analyzed bike re-balancing operations from the *inventory* data. I found managers were re-balancing bikes during the 8am, 10am, 4pm and 6pm slots. To compute the ridership and service level under current policy, I extract the starting distribution of bikes from the *inventory* data for each station during the slots. I then apply the distribution to my stochastic model. To ensure robustness in the values of starting distribution, I average the inventory values during the first 30 minutes for each of the 8am, 10am, 4pm and 6pm slots instead of using the values at the exact start. Table 2.9 provides changes in policy on ridership and service level for morning. Ridership and service level can be improved by 7.74% and 1.91% (respectively) at the same reset times. If managers change their policy from an 8am and 10am allocation to only at 6am, ridership can be improved by 2.32% but the service level drops by 1.21%. Although managers can avoid cost of an additional reset, such an improvementis achieved at the cost of drop in service level. Similarly, results follow for evening and the entire operational day in Table 2.10 and 2.11.

This concludes the analysis intended to address my second research question concerning how should managers allocate inventory of bikes to maximize ridership.

I next address my third research question, if the managers wish to re-allocate inventory multiple times in a day, when should the re-allocation be done to improve ridership and service level? I analyze this question for morning and evening periods individually. Tables

Table 2.9: Current vs Optimal policy - Morning

Policy	Reset time	Ridership (Pickups)	Service Level (Dropoff)	Change (Current → Optimal)	
Current	8am, 10am	542.084	0.971		
Optimal	8am, 10am	584.056	0.990	7.74%	1.91%
	8am	571.833	0.983	5.49%	1.18%
	6am	554.656	0.959	7.622	2.32%

Table 2.10: Current vs Optimal policy - Evening

Policy	Reset time	Ridership (Pickups)	Service Level (Dropoff)	Change (Current → Optimal)	
Current	4pm, 6pm	516.188	0.973		
Optimal	4pm, 6pm	554.641	0.987	7.45%	1.46%
	4pm	540.216	0.982	4.65%	0.93%
	2pm	533.480	0.970	3.35%	−0.29%

Table 2.11: Current vs Optimal policy - Operational day

Policy	Reset time	Ridership (Pickups)	Service Level (Dropoff)	Change (Current → Optimal)	
Current	8am, 10am, 4pm, 6pm	1,058.272	0.972		
Optimal	8am, 10am, 4pm, 6pm	1,138.697	0.988	7.60%	1.69%
	8am, 4pm	1,112.049	0.982	5.08%	1.06%
	8am, 2pm	1,105.313	0.976	4.45%	0.45%
	6am, 4pm	1,094.872	0.971	3.46%	−0.14%
	6am, 2pm	1,088.136	0.965	2.82%	−0.75%

2.12, 2.13 provide exhaustive combinations of single, double and triple re-allocation policies and ranking of ridership. If managers wish to reset once during morning, the best time slot is 8am. For double reset and triple reset, the best time slot combinations are (8am, 10am) and (6am, 8am, 10am), respectively. Similarly, if managers wish to reset only once in the evening, the best time slot is 4pm. For double reset and triple reset, the best time slot combinations are (4pm, 6pm) and (2pm, 4pm, 8pm), respectively. The necessary results are provided in Tables 2.14 and 2.15.

Table 2.12: Single & Double Re-allocation - Morning

Single				Double			
Reset time	Ridership (Pickups)	Service Level (Dropoff)	Rank	Reset time	Ridership (Pickups)	Service Level (Dropoff)	Rank
6am	530.285	0.923	4	6am, 8am	574.533	0.988	3
8am	571.833	0.983	1	6am, 10am	567.820	0.982	4
10am	540.680	0.954	2	6am, 12pm	558.344	0.981	5
12am	535.480	0.958	3	8am, 10am	584.056	0.990	1
				8am, 12pm	575.852	0.989	2
				10am, 12pm	549.839	0.979	6

Table 2.13: Triple & Quadruple Re-allocation - Morning

Triple				Quadruple		
Reset time	Ridership (Pickups)	Service Level (Dropoff)	Rank	Reset time	Ridership (Pickups)	Service Level (Dropoff)
6am, 8am, 10am	586.756	0.993	1	6am, 8am, 10am, 12pm	530.889	0.996
6am, 8am, 12pm	578.552	0.993	3			
6am, 10am, 12pm	568.007	0.989	4			
8am, 10am, 12pm	584.243	0.993	2			

Table 2.14: Single & Double Re-allocation - Evening

Single				Double			
Reset time	Ridership (Pickups)	Service Level (Dropoff)	Rank	Reset time	Ridership (Pickups)	Service Level (Dropoff)	Rank
2pm	500.748	0.948	4	2pm, 4pm	541.596	0.986	4
4pm	540.216	0.982	1	2pm, 6pm	551.985	0.984	2
6pm	511.278	0.960	2	2pm, 8pm	537.337	0.985	5
8pm	505.799	0.967	3	4pm, 6pm	554.641	0.987	1
				4pm, 8pm	543.524	0.988	3
				8pm, 10pm	531.967	0.983	6

Table 2.15: Triple & Quadruple Re-allocation - Evening

Triple				Quadruple		
Reset time	Ridership (Pickups)	Service Level (Dropoff)	Rank	Reset time	Ridership (Pickups)	Service Level (Dropoff)
2pm, 4pm, 6pm	556.021	0.990	3	2pm, 4pm, 6pm, 8pm	500.926	0.993
2pm, 4pm, 8pm	544.904	0.991	1			
2pm, 6pm, 8pm	552.369	0.990	4			
4pm, 6pm, 8pm	555.025	0.991	2			

2.8.2 Sources of Improvement

I analyze the sources of improvement in ridership and service level across stations in the network. As the number of stations is fairly large, I aim to identify any common existing demand patterns across stations. Upon close analysis of the pickup rate at stations 69 and 65 in Figure 2.4 and 2.5, I find two distinct pattern of pickups over the operational day. Station 69 experiences higher and lower pickups in morning and evening, respectively. I find the opposite pattern at station 65. To identify the common patterns, I perform cluster analysis with pickup rates across all stations.

I develop a dataset where each row represents station and columns represent pickup rates estimated for each 2 hour time slot. For each station, I normalize my variables by dividing pickup rate in each time slot by the total pickup rate for the day to retain the pattern. I apply a K-means clustering algorithm to the normalized data to identify groups of stations. Figure 2.9 shows the relationship between choice of clusters, within-cluster sum of squares and cluster groups by the first two principal components. I found 21 stations are characterized by higher pickups in morning, unlike the remaining 14 stations. Hence I tag 21 stations to cluster group S_1 and the remaining 14 to S_2.

I analyzed the difference between the current inventory allocation policy and an optimal inventory allocation policy for reset times of 8am and 10am during the morning and for 4pm and 6pm in the evening. The results are provided in Table 2.16 & 2.17. During morning, I found 48.7% (17) stations were under-stocked and 51.3% (18) were over-stocked, and that the average inventory change (ratio, capacity) from the current policy to an optimal policy was 10.6% and 10.5%, respectively. Among the group S_1, 28.57% of stations were under-stocked and 71.43% over-stocked. In group S_2, 78.57% of stations were under-stocked and 21.43% over-stocked.

Table 2.16: Improvement - Morning

Cluster group	Demand	Stations	Stations (Proportion)		Inventory change/Capacity	
			Under-stock	Over-stock	Under-stock	Over-stock
S1	Low pickups	21	6(28.57%)	15(71.43%)	−5.1%	11.8%
S2	High pickups	14	11(78.57%)	3(21.43%)	−13.3%	4.1%
Total		35	17(48.57%)	18(51.43%)	−10.6%	10.5%

During the evening, I found 54.29% (19) of stations were under-stocked and 45.71% (16) over-stocked. I also determined that the average inventory change (ratio, capacity) from the current policy to an optimal policy was 12.4% and 13.1%, respectively. Among the group S_1, 85.71% of stations were under-stocked and 14.29% over-stocked. In group S_2, 7.14% of stations were under-stocked and 92.86% over-stocked.

Table 2.17: Improvement - Evening

Cluster group	Demand	Stations	Stations (Proportion)		Inventory change/Capacity	
			Under-stock	Over-stock	Under-stock	Over-stock
S1	High pickups	21	18(85.71%)	3(14.29%)	−12.8%	5.2%
S2	Low pickups	14	1(7.14%)	13(92.86%)	−6.6%	15.0%
Total		35	19(54.29%)	16(45.71%)	−12.4%	13.1%

2.9 Substitution vs No Substitution - Allocation and Ridership

I aimed to study the value of incorporating substitution in demand estimation, optimal inventory allocation and ridership. In Section 2.5.2, I performed a pickup demand estimation process under three cases (under the absence of inventory data, under the presence of inventory data but without substitution, under the presence of inventory data and with substitution). I found difference in estimates during 8am slot for stations 69 and 70 in Figure

2.4 and 2.5. If the operators aim to increase pickups across the two stations during 8am slot, using estimates based on the first case or the second case can result in sub-optimal levels of inventory allocation.

Scholars in OM have studied (Smith and Agrawal [2000], Mahajan and Van Ryzin [2001], Kök and Fisher [2007], Honhon et al. [2010]) similar questions in the context of retailing. The results of existing research shows that if substitution is significant, avoiding its effect on demand estimation and inventory planning leads to incorrect inventory policies and hence lower profits.

In my context, I study the following three questions - (i) Does inventory allocation change when optimizing the system with demand parameters estimated under the assumption of no substitution versus an assumption of substitution? If so, how many stations incur a change in allocation? (ii) In what quantity does inventory change as a percentage of station capacity? (iii) What is the change in ridership if the system is optimized under no substitution versus substitution?

2.9.1 Estimation - No substitution

I consider the pickup demand model give by equation (2.1) discussed in section 2.5.1. For the second case, I set ($p_{jit} = 0$). This modifies the demand model to

$$\overline{\mu_{it}} = \mu_{it} \cdot (1 - \tau_{it})$$

I then perform Maximum Likelihood estimation to estimate $\{\mu_{it}, \forall i \in S\}$ and subsequently obtain the dropoff demand parameters using

$$\lambda_{it} = \sum_{j \in S} q_{jit} \cdot \mu_{jt} \qquad \forall i \in S$$

q_{jit} is estimated from trip data and represents proportion of trips realized from source station

"j" to destination "i" during "t".

Let the set of parameters estimated under no substitution be identified by

$$\theta^{\text{no subs}} = \{\lambda_{it}^{\text{no subs}}, \mu_{it}^{\text{no subs}} : i \in S\}$$

2.9.2 Estimation - Substitution

The demand estimation process under substitution (case 3) is extensively discussed in section 2.5.2. Let the set of parameters estimated be identified by

$$\theta^{\text{subs}} = \{\lambda_{it}^{\text{subs}}, \mu_{it}^{\text{subs}}, \alpha_t, \beta_t : i \in S\}$$

2.9.3 Optimization Model

My objective is to verify if inventory allocations change when optimizing the system with demand parameters estimated under no substitution versus those estimated under substitution. I choose the optimization problem as maximizing the sum of pickups and dropoffs against pickups under constrained dropoff service level. My choice of this objective function is to ensure I have the same feasible set of solution spaces under no substitution and substitution. I perform the analysis and report the results for the morning (8am slot) and the evening (4pm slot) busy periods.

For the morning busy period, I chose to use the starting inventory (8am) at each station in the network which maximizes the sum of pickups and dropoffs from 8am to 10am. The formulation is given by

$$\max_{x_i} \sum_{i \in S} z\Big(\text{period}, x_i, \theta\Big)$$

subject to

$$\sum_{i \in S} x_i = nb$$

$$0 \leq x_i \leq C_i$$

"S" refers to set of all stations in the network, x_i starting inventory at station "i" during selected period (morning/evening), "nb" number of bikes in the network, C_i capacity of station "i", θ is setting specific demand parameters.

$z\Big(\text{morning}, x_{i1}, \theta\Big) = f(x_{i1}, \theta) + g(x_{i1}, \theta)$ which refer to sum of pickups and dropoffs with starting inventory x_{i1} during morning.

Let $x^{*(\text{no subs})}, x^{*(\text{subs})}$ be the optimal inventory levels in the network when demand parameters $\theta^{\text{no subs}}, \theta^{\text{subs}}$ are used in the objective function. We ran the optimization problem under each setting with "nb" varying 251 to 400. I summarize the following two quantities - change in inventory allocation and ridership when the system is system is optimized under no substitution to substitution.

$$\text{Change in inventory allocation} = \Delta x = x^{*(\text{subs})} - x^{*(\text{no subs})}$$

$$\text{Change in number of stations} = \Delta N = N - \sum_{i \in S} \mathbf{1}_{x_i^{*(\text{subs})} = x_i^{*(\text{no subs})}}$$

$$\%\text{Change in objective function} = \Delta z = \frac{z\Big(x^{*(\text{no subs})}, \theta^{\text{subs}}\Big) - z\Big(x^{*(\text{subs})}, \theta^{\text{subs}}\Big)}{z\Big(x^{*(\text{subs})}, \theta^{\text{subs}}\Big)} * 100$$

A summary of the change in inventory allocation, number of stations and objective function value are provided in Table 2.18 and 2.19. During the 8am slot, I found the inventory levels change for between 5 (14.29%) and 13 (37.14%) stations. I determined that the inventory change across the same stations varyied from -2 to 4 (-7.41% to 21.05% of station capacity). I obtained similar results for the evening busy period. Hence I conclude that avoiding commuter substitution under stock-out can lead to sub-optimal decisions.

Table 2.18: Change in inventory allocation

Period	ΔN (ΔN/N)		Δx (Δx/Capacity)	
	Min	Max	Min	Max
8am slot	5(14.29%)	13(37.14%)	$-2(-7.41\%)$	4(21.05%)
4pm slot	4(11.43%)	12(34.29%)	$-1(-6.67\%)$	2(10.53%)

Table 2.19: Change in objective function

Period	Min	P25	Mean	Median	P75	Max
8am slot	-4.20%	-3.16%	-2.34%	-2.13%	-1.62%	-1.31%
4pm slot	-2.82%	-1.78%	-1.42%	-1.05%	-0.5%	-0.2%

Similarly, the ridership is less by a median of 2.13% and 1.05% during the 8am slot and 4pm slot (respectively) when the system is optimized under condition of no substitution versus substitution.

2.10 Conclusion & Managerial Implications

Performance of networked self-service systems, such as bike sharing systems, is deeply impacted by consumers' spatial and temporal preferences as well as the allocation of resources over the network. Demand estimation and managing bike inventory in the network of stations is an important problem in Bike Sharing Programs (BSP). Allocating incorrect amounts of inventory across stations without proper demand estimation can lead to major operational challenges - unavailability of bikes for pickup and a lack of empty docks for drop-off. These inefficiencies on a regular basis can lead to lower satisfaction for commuters eventually deterring them from participating in BSP, which in turn can impede the goal of BSP to reduce vehicle travel demand and associated carbon emissions.

I address these challenges by utilizing tools in Operations Research, Statistics, and Machine Learning for better planning of BSP. My results constitute four primary contributions to the literature: my first research question addresses how to estimate intrinsic commuter demand at each station in the network under non-stationarity while incorporating consumer substitution across stations during stock-outs; I also answer the implied question whether commuters substitute across stations under realization of stock-outs, and, if so, what is the impact of station pair distance on the likelihood of substitution. I integrated stochastic models and large datasets including censored trips and minute-level inventory data to build models of and to predict both station pickup and drop-off demand. My demand model captures non-stationarity, station substitution, and interdependency across stations. To the best of my knowledge, mine is the first work to model substitution and address censoring issues in a robust manner for demand estimation in BSP. I find statistical evidence of commuter substitution under stock-out. I also find that commuters tendency to substitute decreases with increase in walking distance between stations over the entire operational day, however this was only statistically significant during the 8am, 10am, 4pm and 6pm slots.

Second, I address how BSP operators should allocate inventory of bikes at the beginning of the day to maximize ridership. I extended existing work (Raviv and Kolka [2013], Henderson et al. [2016]) on inventory allocation problems in BSP to examine multiple periods each of fixed length. My optimization model is built on a dynamic program with non-stationary demand and transient analysis of inventory transition between two subsequent periods. To the best of my knowledge, my work is the first to incorporate transient analysis in Bike Sharing Programs. My results demonstrate the value of data driven research to estimate the benefits of optimization. I found the optimal allocation policy can improve ridership and service level by 7.60% and 1.69% respectively compared to the current policy.

Third, I provide insights on the optimal timing and quantity of bike inventory reallocation to improve ridership. I provide managerial recommendations on exhaustive combinations of single, double, triple and quadruple re-allocation policies. For example, if operators wish to reset bike inventories twice during the morning, my analysis suggests that the best time slots are 8am and 10am. While I found the timing of the current reallocation policy matches with the optimal policy, the current quantity reallocation policy does not match the optimal policy.

Finally, I quantify the impact of ignoring station substitution on inventory allocation and ridership. I find statistically significant differences in demand parameters during busy periods (8am-10pm, 4pm-6pm) of the operational day between a model that ignores station substitution and my model which incorporates station substitution. I show that failing to incorporate demand substitution in operational systems can lead to incorrect inventory allocation policies and can result in errors in inventory allocation by as much as 21.05% for some stations in the network.

I provide a methodological framework for integrating large scale dataset based empirical analysis and stochastic modeling to design and manage networked self-service systems. Future work could include running simulation studies on my data to validate my suggestions on inventory policies and ridership as well as the optimal re-balancing of bikes. The demand model described in this chapter assumes a homogeneous population of customers in the substitution process. The demand model could be extended to incorporate different classes of consumers in substitution. For example, one could consider two class of consumers, each with different level of sensitivity to walking distance. Another suggestion for future work would be to validate my approach using data from other Bike Sharing programs such as Citi Bike.

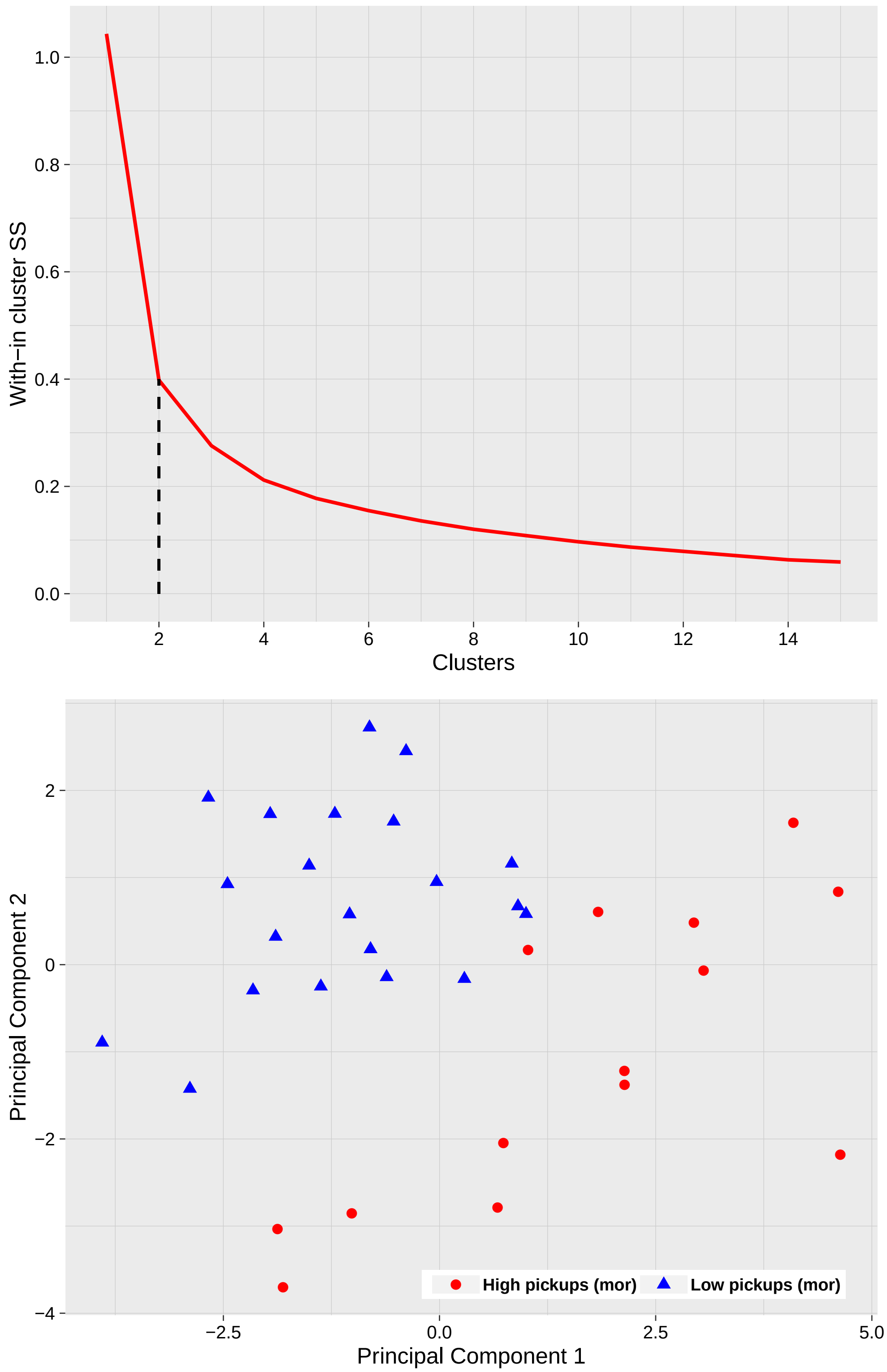


Figure 2.9: Cluster groups

CHAPTER 3: LOGISTICS PERFORMANCE, RATINGS, AND ITS IMPACT ON SALES IN E-COMMERCE PLATFORMS

3.1 Abstract

On-line product ratings on e-commerce platforms are key information for both sellers and customers. Sellers rely on the ratings to signal product or service quality whereas customers use this information for making the purchase decision. In this chapter, I examine the impact of logistics performance metrics such as delivery delays, customer's promised speed of delivery, order split, etc. on logistics service ratings of sellers on an e-commerce platform. Further, I analyze and quantify the impact of logistics service ratings and performance on customer purchasing behavior and sales. Prior work on online ratings in e-commerce platforms have largely analyzed customer response to product functional performance and biases that exist with-in. This study contributes to this stream by examining customer experience from a service quality perspective by analyzing logistics performance, logistics ratings and its impact on customer purchase behavior. The insights from this study are relevant to independent sellers as well as e-commerce platform managers who aim to improve long-term online traffic and sales. Using a large data set of customer orders ($\approx$ 21 million records) from the Tmall platform and the Cainiao network, I utilize ordered regression to understand the variation in logistics ratings and its drivers. I then use a customer utility model to quantify the impact of logistics ratings on customer purchasing behavior. I find, Logistics ratings are negatively impacted by delivery delays, but positively impacted by faster promised speed of delivery and total order amount paid. The impact of delivery delays on logistics ratings are moderated by the total order amount paid but not by faster promised speed of delivery.

For example, a customer who paid a higher order amount is likely to give a more negative rating to a delayed order compared to an on-time order than a customer who paid a low order amount. The results also show that splitting an order into multiple shipments, so that a part of the order is delivered on-time even if the overall order is delayed, does not improve logistics ratings. I also find that logistics ratings impact customer purchasing behavior positively. Lastly, I show that a reduction in delivery delay by one day can improve the average weekly sales by as much as 2.5%. This study emphasizes that logistics performance and ratings which measure service quality are important drivers of customer purchase behavior on e-commerce platforms. Hence, e-commerce platforms and sellers should pay attention to logistics performance, in addition to product performance, in driving traffic and sales online.

3.2 Introduction

E-commerce is one of the largest growing industries in the digital economy. Worldwide retail e-commerce sales have increased by 72% during 2014-17 and expected to increase by 112% during 2017-21[1]. Revenues of two large e-commerce retailers, Amazon and Alibaba, were $51.0 and $9.9 Billion respectively in the first quarter of 2018. During the period 2014-17, the industry has experienced a significant increase in the number of independent sellers and their sales on the digital platform. For example, as of 2018, Amazon has more than 2 million independent sellers around the world and their revenue share is more than 52%[2]. Independent sellers are immensely profitable to e-commerce platforms.

In the current digital age, with the low cost of sharing online information, independent sellers (hereafter "sellers") can now offer a wide variety of products and different versions (model, style, color etc.) of a single product. The rich information available about the products means that customers now have wide range of options (products and sellers) from

[1]https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/

[2]https://www.statista.com/statistics/259782/third-party-seller-share-of-amazon-platform/

which to choose. Occasionally, it can be perplexing for a customer to choose a specific product sold by multiple sellers on a platform at the same price. To ease the search cost, reduce the product uncertainty (Chen and Xie [2008]), and help customers in the purchase decision process, most platforms typically provide ratings and reviews of the products and sellers in addition to information such as price, discount, and available inventory. Typically platforms provide two forms of ratings or reviews: *Summary statistics* and *Full history* (Acemoglu et al. [2017]). *Summary statistics* display the volume of ratings by previous customers and their distribution typically in the range from 1 to 5 (1 being the lowest and 5 being the highest). *Full history* provides the entire verbal entry and rating by each previous customer in chronological order.

The cycle of an online purchase and rating entry process by a customer is typically as follows. An incoming customer visits the platform and the product page to obtain information on price, ratings or reviews, and other sources of data. Based on the rating or reviews, the customer builds an ex-ante valuation of the product and service quality. The customer then evaluates product price, ratings or reviews across multiple sellers and finally makes a decision on product purchase. After the purchase, the customer receives the product, evaluates it and realizes its true quality. Platforms typically send email or text messages with a list of questions requesting customers' feedback on the product or service on a discrete scale of 1 to 5. The customer's response typically describes their satisfaction level with respect to different measures: purchase process, product quality or functionality, and on-time delivery. In addition to their discrete response, customers are also requested to provide detailed text feedback (Chen and Xie [2008], Cui et al. [2012]) about their level of satisfaction with the entire process from purchase to delivery. After the customer provides the rating or review, the information is appended to the history and is made available to the next incoming customer. Higher ratings or great reviews about the seller and their products are likely to increase the likelihood of an actual purchase. Hence, it is no surprise that e-commerce platforms make enormous efforts regularly by sending email or text messages requesting feedback about their products and service from previous customers.

Sellers rely heavily on the online ratings to maintain market share and survive against fierce price competition on the e-commerce platform. Typically, sellers display ratings on two dimensions viz. product quality and logistics service quality. The product quality signifies the response to customers experience on the product performance as stated on their web page. On the other hand, the logistics service quality signifies the response to customer experience of timely order delivery. Customers are likely to give a lower rating when their experience on each of these dimensions does not meet their expectation. For example, consider a customer who places an order and requests for a one day delivery. When the seller fails to deliver the order by committed time, customer is unhappy with the delivery service and is likely to give a lower logistics rating. A cumulative volume of lower logistics ratings from customers over a period of time decreases the effective rating visible to an incoming customer. An overall lower rating is then likely to increase the likelihood of a “no purchase”, or switch the customer to a different seller (of potentially higher logistics rating) affecting long-term traffic and sales of the primary seller. In the above example, delivery delay is one potential metric which is likely to drive the logistics rating. However, it is unclear what other variables drive the logistics rating. As a result, it is vital from sellers’ point of view to identify exhaustive list of operational factors that drive the logistics ratings. Hence, the first research question is : How do the logistics performance metrics such as delivery delays, promised speed of delivery, order amount paid and other potential variables impact the logistics service ratings of sellers on an e-commerce platform (Q1). The insights from the anlaysis of this question helps sellers understand and uncover key measures, direction and magnitude of their impact on the logistics rating. Contingent on the magnitude of the impact, sellers can prioritize and take actionable steps to improve logistics performance, thereby improving logistics service rating from the customers.

When a customer places an order of multiple items (or multiple quantities of a single item), there are potentially two policies that sellers can choose for shipment of the order viz. bundled order delivery or split order delivery. The order is "bundled" if the entire order is shipped in a single shipment, while if the order is divided into multiple shipments each of which is delivered separately to the customer, the order is said to be "split". Each policy has different implications on shipping cost and delivery delays. Clearly, the shipping cost of bundled order is likely to be less than a split order. However, if the items are stored in different locations or if there is not sufficient inventory of all items, a seller may decide to split the order in multiple shipments. Splitting the order can sometimes enable the entire order to be delivered on-time. Although the implications on shipment costs are apparent, the impact of bundled versus split orders on customers' perception of service is unclear. Similarly, consider the alternative case where a seller receives a multi-item order and does not have sufficient inventory of all the items in their warehouse. If the seller implements a split order policy, the seller could potentially follow two paths - (i) ship the item whose inventory is available, accomplish a timely delivery for this item, and later ship the remaining item which is potentially delayed (defined as "partial delayed order"), or (ii) wait until the entire inventory of all items is available and realize a delayed delivery ("full delayed order"). However, it is not clear whether customers prefer a partially delayed order over a fully delayed order. These issues forms the basis for the second research question: How do customers ratings differ between bundled vs split orders for on-time delivery orders? If the orders are delayed, does the split moderate the impact of delay on rating? (Q2 (i)). Under split orders, do customers give higher rating to partial delayed orders compared to full delayed orders? (Q2 (ii)). The insights from the anlaysis of this question can help sellers determine when to split or not split multi-item orders by understanding customer preferences.

Prior work on online product quality ratings have shown evidence of their impact on customer purchase probability or sales (e.g, Chevalier and Mayzlin [2006], Chintagunta et al. [2010], Lin and Yang [2011]). As a result, a lower rating reduces the likelihood of purchase, subsequently affecting long-term traffic and sales of the seller. However, the answer to whether logistics performance impacts customer purchase decision is unclear. This forms basis for the third research question: Does the logistics service rating impact an incoming customer's purchase probability? (Q3). We also look at whether competition between sellers moderates or accentuates the impact of logitics rating on the customer purchase probability. The customers decision on purchasing an item depends on whether it is unique to a seller or it is sold by multiple sellers. If the item is sold by only one seller, the customer has only two choices - purchase or not to purchase. Whereas, if the item is sold by multiple sellers, customer can purchase one from among many sellers or decide not to purchase. The results from the analysis of this question helps sellers understand the value of logistics performance or rating in the e-commerce world in a competitive setting. If the logistics rating does not significantly impact the customer purchase probability, sellers can choose not to project it on their webpage. However, if the impact is significant, sellers should regularly make efforts to improve the rating as it affects their long-term traffic and sales. In addition, from a platform persepctive, they can potentially exclude sellers with inferior logistics ratings or negotiate stringent contracts which incentivize them to improve their logisitcs rating.

Finally, in our last research question we examine: how does logistics performance impact sales of a seller? We answer the question: what is the potential improvement in sales if sellers can reduce their delivery delays by one day? (Q4). The results from the analysis of this question has multiple benefits to the sellers. First, sellers can assess the economic value of improvement in logistic performance. Second, depending on the magnitude of the impact on sales, sellers can initiate process improvement efforts to improve the delivery performance of their customer orders.

The rest of the paper is organized as follows. In section 3.3, we provide details on prior work on online ratings and our contribution to the literature. In section 3.4, we describe the theory and propose the hypothesis to be tested. In, section 3.5, we describe the setting, data, summary statistics and variables for our model. In section 3.6, we provide the model, results, and analysis on the impact of order split on logistics rating, logistics rating on customer purchase probability and quantify the link between logistics performance and sales for a seller. Lastly, in section 3.7, we conclude the work with managerial insights and provide venues for future research.

3.3 Literature

Our work builds on the prior research into online ratings or reviews in e-commerce or social media platforms, which can be classified into the following three areas: (i) ratings as proxy or evidence for product functional quality; (ii) drivers of ratings; (iii) impact of rating on customer purchasing behavior or sales; and, (iv) distributional assumptions. We discuss the prior literature in each of these areas independently and identify the contribution of our paper.

First, the literature on ratings as a proxy or evidence for product functional quality reports mixed results. Most of the prior work has examined ratings on products such as books or movies. Scholars argue that the ratings on these kinds of products provided by individuals strongly depend on their personal taste and do not reflect an expert opinion. Hu et al. [2006] studied the customer ratings on books, DVDs and videos on the Amazon platform and demonstrated that the mean rating does not necessarily reveal true quality. The rating is posted only when a customer is extremely satisfied or unhappy with the product thus average rating is a biased signal. Dellarocas [2006] showed that ratings may not be completely trusted as they could be strategically manipulated by firms (potentially inflated) to remain competitive or maintain market share. In the healthcare sector, Lu and Rui [2017] investigated whether online physician ratings are informative of their medical quality and found that physician ratings could be a valuable information source for patients to learn about physician quality. Our work differs from previous research in that we study logistics service quality rather than product quality. Specifically, in the e-commerce world, the rating we study is on the customer experience of logistics service-fulfillment from the point of purchase to delivery, not the product performance quality.

Second, the prior work on ratings has primarily focused on different biases influencing the score. In a lab setting, Schlosser [2005] found "self-presentational behavior" where customers have adjusted their ratings downward after observing previous negative reviews. Li and Hitt [2008] found a "self-selection bias" and declining trend in book ratings over time. The bias is realized when early buyers tend to report a higher perceived quality, leading to better book ratings, than do later buyers. Similarly, Wu and Huberman [2008] argued that customers follow the history of ratings from previous opinions. They found a selection bias in the posting of ratings when the customers were required to pay to post their feedback. Hu et al. [2009] showed that the graphical distribution of ratings for three product categories (books, DVDs and videos) on the Amazon platform follow **J**-shaped distribution. The shape of the distribution is primarily driven by two self-selection biases: *purchasing bias* and *under-reporting bias*. Purchasing bias results when people with higher product valuations purchase a product and those with lower valuations are less likely to purchase the product. These customers are unlikely to write a negative review. Purchasing bias causes a positive skewness in the distribution. Second, people with higher valuation who purchased the product are more likely to write a review only when they are either extremely satisfied or unsatisfied. People who feel the product is average might not bother to write a review. Li and Xiao [2010] found bias in rating behavior, where customers are more sensitive to the high-rated sellers than the low-rated sellers.

More studies on biases in online ratings include factors like social identity (Wang [2010]), social comparisons (Chen et al. [2010]), rating environments like high valance and high volume (Moe and Schweidel [2012]), friends and strangers (Lee et al. [2015]), cultural influences (Koh et al. [2010]) etc. Our work does not address any of these biases but rather our contribution is primarily on examining the impact of measurable operational factors on logistics rating, which has not been studied previously.

Third, with respect to the distribution of realized ratings, Hu et al. [2009] showed that graphical distribution of realized ratings for three product categories (books, DVDs and videos) from the Amazon platform is distorted (**J**-shape) due to *purchasing bias* and *under-reporting bias*. They show people with higher valuation who purchased the product are more likely to write review only when they are either extremely satisfied or unsatisfied. People who feel the product is average may not bother to write a review. Ho et al. [2017] studied the effect of Disconfirmation-discrepancy between the assessment of the product's expected quality and an experienced assessment of the same product on the behavior of customers leaving online product ratings. They find that the individual decision of whether to post a rating and what rating to post are affected by Disconfirmation in two distinct manners. An individual is more likely to leave a review when the magnitude of Disconfirmation she encounters is larger. In addition, when the individual decides to review a product, the rating she chooses may not neutrally reflect her post purchase evaluation; the direction of such a bias is in accordance with the sign of Disconfirmation. Acemoglu et al. [2017] studied fast and slow learning analytically for different online rating systems (*Full history* and *Summary statistics*). The authors proposed a theoretical ordered response model where they placed the "no response" option in the median position of ordered ratings. Our work contributes towards addressing the distortion in distribution of ratings by empirically modeling the "no response" option which has not been studied in the prior literature.

Using the data from the same setting, Cui et al. [2018] study the removal and resumption of the high-quality logistics carrier option. The authors find removal leads to decrease in sales by 16.42%, resumption to increase in sales by 18.83%. However, these events did not impact variety and the logistics rating of sold products. On the contrary, our study quantifies the impact of logistics performance and ratings on customer purchasing behavior and sales.

3.4 Theory & Hypothesis

Customer ratings on e-commerce platform are an expression of satisfaction with the quality of their online transaction from purchase to delivery (Yi [1990]). Although scholars have debated whether ratings are an indicator of true quality (Moe and Trusov [2011]), they are largely considered as a manifestation of customers' perceived quality of the product or service (Kotler [1994], Koh et al. [2010]). High customer satisfaction ratings are believed to be the best indicator of a firm's future profits (Kotler [1994]). As a result, understanding the causes and effects of customer satisfaction has been an important area of research with a long history in Marketing or Information Systems (Oliver [1977], Oliver [1980], Oliver and Bearden [1985], Yi [1990]).

In the literature, customer satisfaction has been defined as post-purchase evaluation of product or service resulting from a comparison of the actual performance to pre-purchase expectation (Fornell et al. [1996], Churchill Jr and Surprenant [1982]). Different theories have placed emphasis on what drives customer satisfaction. For example, before initiating the purchase of product or service, a customer constructs a certain level of expectation of the quality. Assimilation theory (Anderson [1973]) states that satisfaction primarily depends on expectation and not the perceived performance after consumption. On the other hand, contrast theory (Sherif and Hovland [1961]) states that performance rating depends only on disconfirmation and not on expectation. Disconfirmation is the discrepancy between pre-purchase expectation and post consumption perception of product quality (Oliver [1977]). Although these are competing theories, research has largely shown that both expectation and disconfirmation affect customer satisfaction. This concept is derived from the Expectation-Disconfirmation theory (Oliver [1980]). This theory states customer satisfaction decision is determined by two factors: creation of expectation and expectancy disconfirmation (Parasuraman et al. [1988], Anderson and Sullivan [1993]). A customer builds a certain level of expectation for the quality before initiating the purchase of the product or service. After purchasing and receiving the product, customer consumes it and experiences its quality. Disconfirmation is the discrepancy between the expected quality and the experienced assessment of the product quality (Ho et al. [2017]). If the product performs more poorly than expected, customers experience negative discomfort, which leads to dissatisfaction (the greater the level of negative discomfort, the greater the dissatisfaction level). On the other hand, if the product performs better than expected, customers experience positive discomfort and are satisfied.

In the Operations Management literature, Kumar et al. [1997] examined the impact of a time guarantee on customers'satisfaction with waiting. They found that customer satisfaction decreased with an increase in wait if they observed that service times were greater than expected. In the context of our work, the customer places an online order and requests a promised speed of delivery. The seller commits to the promise and delivers the product to the customer. The promised speed creates an expectation of delivery service. If the order is received after the promised speed, customer experiences a delay and negative discomfort. Hence anchoring on the Expectation-Disconfirmation theory and model evidence we hypothesize - H1 as

- *Logistics rating decreases with increase in delivery delay*

We now focus on the impact of expectation on customer satisfaction. The antecedent theories on satisfaction are drawn from the Expectancy-Disconfirmation paradigm (Oliver [1977], Oliver [1977]). Prior research has shown that expectation is used as the reference comparison and affects customer perception of service performance (Olshavsky and Miller [1972], Olson and Dover [1979], Yi [1990]).

An explanation for the underlying effect of pre-purchase expectations on online product ratings is provided by the belief-adjustment model (Hogarth and Einhorn [1992], Bolton [1998]) and Assimilation theory (Anderson [1973]). It describes the order of belief updating over time as a process of anchoring and adjustments. The central message of the belief adjustment model is that individuals do not directly react to a new stimulus but rather adjust their prior expectations on the specific topic to the new stimulus while sustaining their belief structure in the vicinity of original anchor (Oliver [1980]). Thus, pre-purchase expectations should have a positive impact on satisfaction. This theory can be applied in the context of online shopping and the pre-purchase evaluation of products. First, customers form an expectation what the product might be like on the basis of information found on the product website. Second, they adjust this anchor within a reference frame set by the initial judgment when confronted with the product's performance after the purchase and delivery.

Product price has been identified as the most important inherent quality indicator in the offline world (Zeithaml [1988]). It has been identified to influence the perceived quality of the product in offline and online shops (Dodds et al. [1991], Rao and Monroe [1989], Chen and Dubinsky [2003]). Extending this theory to service quality, we hypothesize H2 as

- *(a) Logistics ratings increases with faster promised speed of delivery*;
- *(b) Logistics ratings increases with higher order payment*

In the previous two hypotheses, we theorized the effects of expectation and disconfirmation on customer satisfaction. Under Hypothesis 1, if the product or service performs more poorly than expected, customers experience negative discomfort, which leads to dissatisfaction (the greater the level of negative discomfort, the greater the dissatisfaction level). Under Hypothesis 2, greater pre-purchase expectation leads to greater positive impact on satisfaction. Combining the two, for a given disconfirmation level, customers with a greater pre-purchase expectation are likely to experience more discomfort and hence more dissatisfaction. Hence, we hypothesize H3 as :

- *(a) Faster promised speed of delivery moderates the impact of delivery delay on logistics rating*;
- *(b) Order payment moderates the impact of delivery delay on logistics rating*

3.5 Study Setting & Data

We utilize data from the **Tmall** platform and the **Cainiao network** provided by the MSOM committee of INFORMS to study the research questions stated in the paper. A brief description of the study setting is as follows.

The Cainiao network is a consortium of warehouse and logistics companies founded by Alibaba. Alibaba owns approximately 48% of the Cainiao network. The network operates as an integrated warehousing and logistics platform linking logistics providers as well as warehousing and distribution centers. It is dedicated to meet the current and future logistics demands of China's online and mobile commerce sector. Cainiao helps Alibaba extend its e-commerce network across China.

Tmall is a B2C Chinese language online retail website operated by Alibaba since April 2008. The platform helps Chinese and international businesses to sell their goods to customers in mainland China, Hong Kong, Macau and Taiwan. As of February 2018, Tmall has over 500 million monthly active users. Sellers sell products on Tmall platform. Before the Cainiao network was formed in 2013, sellers managed the inventory and fulfilled orders by themselves or by using various third-party logistics service providers. The Cainiao network opened its service to all sellers. Now sellers have the option to have their inventory managed and customer orders fulfilled in a worry-free way although they might choose to operate independently.

We now describe the entire life cycle of a customer order from its inception to delivery. The entire process is pictorially represented in Figure 3.1.

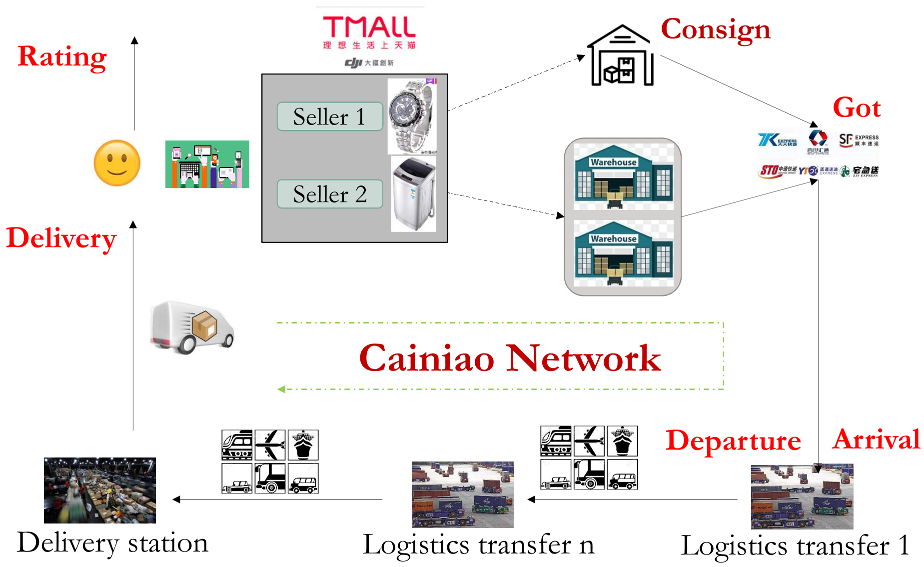


Figure 3.1: Life cycle of a customer order

A customer visits the Tmall platform on a personal computer or mobile app. Next, (s)he browses the products to purchase available (in identical form or with small variations) from different sellers. Sellers are classified into two types based on whether they depend on the Cainiao network for management of product inventory and fulfillment of order delivery. For example, in this figure, Seller 1 is independent of Cainiao, and manages the physical inventory of its products in their own warehouse and is responsible for delivery of orders. On the other hand, Seller 2 is dependent on Cainiao for managing the inventory and delivery of orders.

The sequence of events in the fulfillment process, as captured in the transactional data made available to us, is as follows: First, the shipment of the product from the warehouse to carrier is defined as a“Consign” event. This is followed by the carrier receiveing the package; this event is defined as “Got”. The carrier ships package and goes through multiple facility centers or logistics transfers before it reaches the delivery station nearest to the customer location. Two events are defined while the package moves through multiple transfers, "Arrival" and "Departure". These refer to the events when the package arrived and left at each facility. The package is then finally shipped from delivery station to the customer's location by the carrier. The customer acknowledges receipt of the package by providing a signature to the delivery person. Later, the customer may decide to leave an individual rating for each of the experiences on order quality, online purchase, and the logistics service quality.

The data for the entire life-cycle of each order comes from five different sources during the period 1^{st} January 2017 - 31^{st} July 2017. First, the **customer order** data, which describes the granular details of each order placed through the channel (personal computer or a mobile app). This data comprises information on: *day* (the date the customer placed the order); *order_id* (each order placed by the customer is uniquely identified by a number); *item_info* (each item id, quantity ordered, payment in Chinese Yuan); *pay_time_stamp* (time of payment rounded to 1 minute interval); *buyer_id* (the equivalent customer identifier); *promise_speed* of delivery (customer requested speed of delivery, takes 3 unique values, 1 : same day, 2 : next day, Unavailable : no promise); *if_cainiao* (service channel, takes two values 0, 1 ; 1 : if the order used Cainiao warehousing and logistics service, 0 : otherwise); *seller_id* (equivalent seller identifier); and *logistics review score* (the user's review of the logistics service, uses a Likert scale from 1 to 5 with 1 being poor quality and 5 being high quality). Although the data is highly granular, we were not provided with any information on type of item, e.g. book or electronics, and customer demographics, e.g., location of the customer, their gender, or their age. Hence, our research cannot address any heterogeneity due to item type or customer type.

Second, **inventory history** data which describes inventory information of the items sold by sellers who use Cainiao warehousing and logistics service. The data comprises information on: *day* (date of inventory snapshot), *item_id* (each item is uniquely identified by a number), *warehouse_id* (each warehouse is uniquely identified by a number), *warehouse_city_id* (city identifier of the warehouse), *tot_begin_qty* (item quantity at beginning of the day, 12:00AM), *tot_end_qty* (item quantity at the end of the day, 11:59PM), *replenishment_qty* (replenishment of inventory received), *transfer_in_qty* (transfers of item quantity from other warehouses), *sale_out_qty* (inventory shipped out to meet demand) and *transfer_out_qty* (transfers of item quantity to other warehouses). Although the data is highly granular, we do not have information on the frequency of replenishments, transfer in and outs events. These events are likely to happen multiple times in a day (reported by Cainiao team), however, only an aggregated data point is reported. We do not have information on inventory of products sold by Cainiao independent sellers .

Third, **item view** data which contains granular details on the number of views for each item sold by the sellers. The data comprises information on: *day* (the date the item was viewed), *item_id*, *seller_id*, *brand_id*, *category_id*, *sub_category_id*, *pc_pv* (total number of page views on personal computer), *app_pv* (total number of page views on mobile app), *pc_uv* (total number of unique customer visits on personal computer), *app_uv* (total number of unique customer visits on mobile app). The data on unique visits in app and pc are the key information we use in the analysis of impact of logistics ratings on customer purchase probability.

Fourth, **seller category view** data describes the page or app view of all the category products sold by a seller. The data comprises information on: *day* (date of snapshot), *seller_id*, *brand_id*, *sub_category_id*, *pc_pv* (total number of page views on personal computer),*app_pv* (total number of page views on mobile app), *pc_uv* (total number of unique customer visits on personal computer), *app_uv* (total number of unique customer visits on mobile app), *avg_logistics_review_score* (quality of logistics service), *avg_order_quality_score* (quality of products in the order), *avg_service_quality_score* (quality of online purchase).

Lastly, **logistics** data describes the event level and time stamp information from consignment to final delivery. The data comprises information on: *order_id*, *logistics_order_id*, *action* (events during the multi-transfer of order package to the customer), *action_time_stamp* (time of action). The *action* variable takes the following categories - (i) Consign (dispatch of the package from Cainiao or seller operated warehouse), (ii) Got (package received by the carrier), (iii) Arrival (Arrive at an intermediate facility), (iv) Departure (Leave at the intermediate facility), (v) Sent scan (package leaving the delivery station), (vi) Signed (customer receiving the package), (vii) Trade success (confirmation of package delivery), (viii) Failure (failure to deliver the package). Lastly, if an order has multiple *logistics_order_id* values, it is likely the order was split into multiple packages and shipped to the customer.

The total size of the entire dataset is close to 120 gigabytes. We utilized four of the above five datasets viz. **customer order**, **item view**, **seller category view** and **logistics** to address the research questions in the paper. We now provide an overview of the entire dataset in Table 3.1 before summarizing each dataset in detail.

Table 3.1: Overview of Cainiao network data

Variable	Value
Unique customers	84,357,531
Orders	137,131,394
Unique items	272,339
Sellers	527
Total Quantity ordered	262,347,425
Total Revenue	21,984,551,201 Yuan (approx $3.4B)
Cainiao warehouses	130
Logistics data records	15.2 Billion
Carrier companies	674

3.5.1 Customer Orders

The total number of customer orders realized during the period 1^{st} January 2017 - 31^{st} July 2017 is 137,131,394 (153,456,090 records). If a customer order has two unique items, it is represented as two records in the dataset. In Table 3.2, we summarize the variation in unique items, quantity and total paid for each customer order. We found that customer order had a median of one unique item, quantity one and the amount paid was 47.1 Chinese Yuan. The largest orders comprised 50 unique items, in total quantity of 241,162 and the amount paid was 1,389,199 Yuan.

Table 3.2: Summary of customer orders

	Min	P25	Mean	Median	P75	Max
Items	1	1	1.12	1	1	50
Total quantity	1	1	1.91	1	1	241,162
Total pay	0.01	23.1	160	47.1	105	1,389,199

Figure 3.2 displays the daily trend of orders and the quantity demand. We found evidence of seasonality after excluding days with abnormally high or low demand. We observed the lowest demand on 28th January 2017 (Chinese New Year) and the highest demand on 18th June 2017 (half year promotions). The remaining days of abnormally high demand included (in descending order) 8th January 2017, 8th March 2017 (International womens day), and 1st June 2017 (Children's day). The high demand on 8th January 2017 is due to customer orders placed in advance for Chinese New Year.

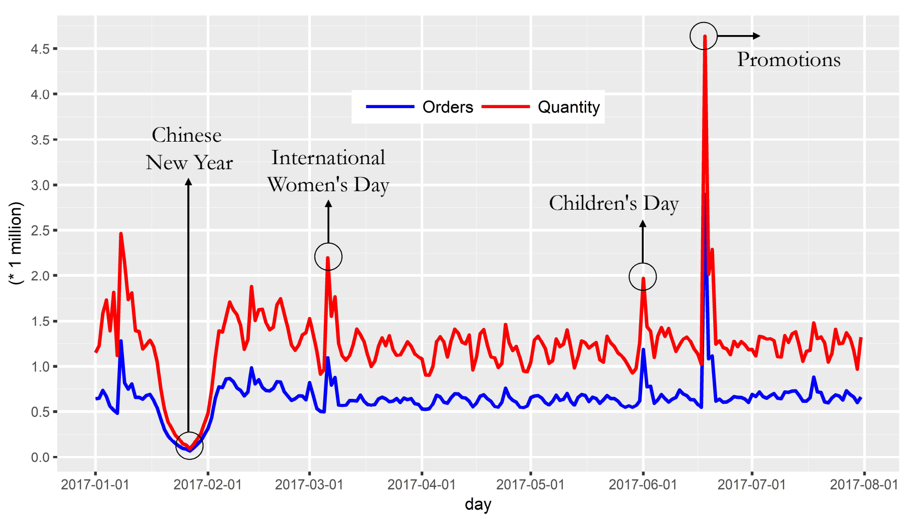


Figure 3.2: Daily demand of orders, quantity

Figure 3.3 shows the median orders for each hour and day type. We found clear evidence of non-stationarity in demand. We observed high demand on weekdays and low demand on weekends. A close analysis of the plot reveals that demand is higher during 10AM-11AM compared to the other time slots. This is probably due to customers placing orders from their personal computers at their workplace.

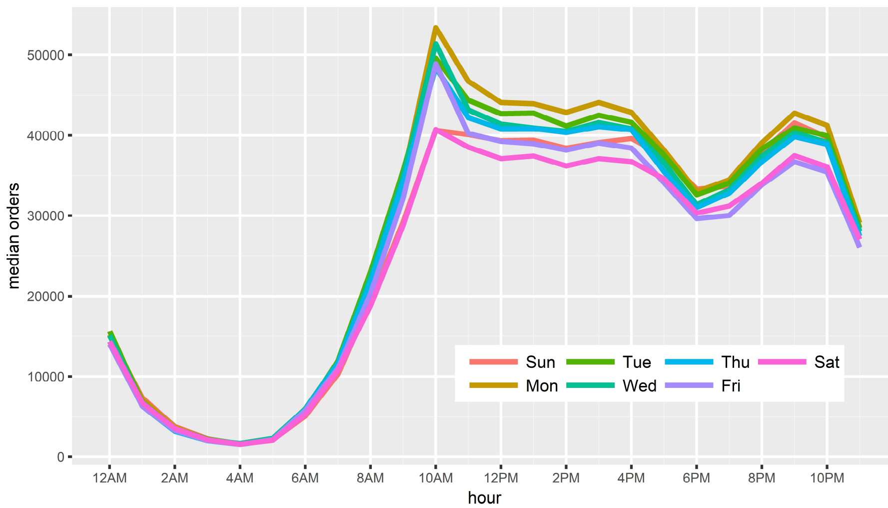


Figure 3.3: Hourly demand of orders

The customers utility from logistics service is likely to depend on abnormal days and hour of the day. Hence, we account for these variables later in our econometric model.

3.5.2 Items

The total number of unique items sold is 272,339. Of these, 99.3% (270,597) are unique to a specific seller. Unique items represent 94% of the total quantity order and 67% of total amount paid in the data. The remaining 0.7% (1,742) are competing items sold by multiple sellers. Competing items represent 6% of the total quantity order and 33% of total amount paid. Hence a third of revenue coming in to the platform results from the competing items available from multiple sellers. The distribution of competing items across multiple sellers is given below in Table 3.3; an example which appears in the first row is 814 items (of 1,742) which are sold by two sellers.

Table 3.3: Competing items and sellers

Number of items	Sellers
814	2
535	3
151	4
104	5
114	6
15	7
9	8

3.5.3 Sellers

The total number of unique sellers in our data are 527 and they are classified into two types as observed in the data. The segregation is shown in Figure 3.4. The first category (270) is Cainiao independent, meaning that these sellers managed their own inventory of all their items at their own warehouses and are accountable for the fulfillment of orders. These sellers account for 49% of the total number of orders, 62% of total quantity ordered and 15% of the total amount paid. The second category (257) are partially Cainiao dependent, meaning these sellers managed a subset of their items (26% of total orders, 23% of total quantity ordered and 23% of total amount paid) themselves but had the Cainiao network manage the remainder (25% of total orders, 15% of total quantity ordered and 62% of total amount paid). Although the order share is mostly similar, the quantity ordered and the share of the amount paid are significantly different. It is clearly evident that the second category of sellers are managing low revenue items in their own warehouses and delegating the responsibility for expensive items to Cainiao.

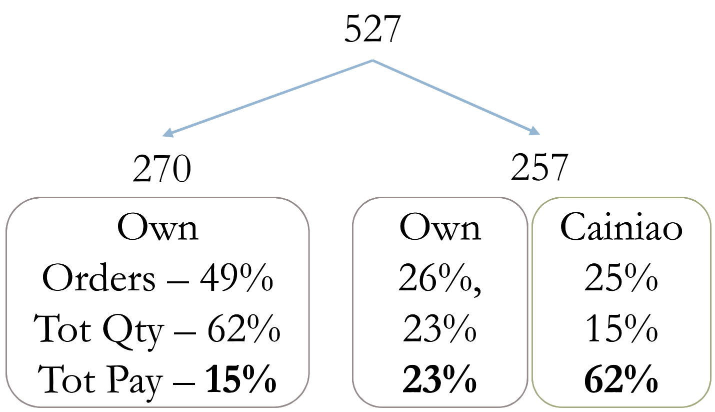


Figure 3.4: Seller type

3.5.4 Logistics

We next summarize the logistics data. The total number of unique carrier companies in the data are 674. Figure 3.5 displays the distribution of orders delivered by the top 14 companies, who delivered more than 95% of the orders. One carrier company with identifier "184" delivered more than 25% of total orders. It is clearly evident from the graph that there is a wide variation in the share of the seller or Cainiao orders handled by each carrier company.

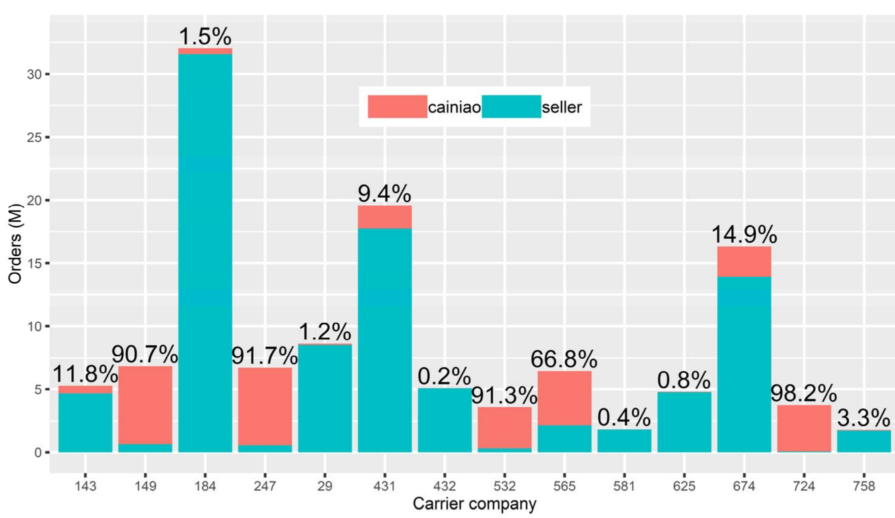


Figure 3.5: Median orders delivered by carrier company

Table 3.4 shows the distribution of promised speed requested by the customer for their orders: 91.0% of the orders had no promise meaning there was no time restriction on when the customer can receive the package. Remaining 9.0% of the orders account for same day and next day requests.

Table 3.4: Distribution of promise speed

Promise speed	Definition	Order count	Percentage
Unavailable	No promise	117,869,660	91.0%
1	Same day	848,385	0.7%
2	Next day	10,692,859	8.3%

Table 3.5 provides the distribution of customer-provided logistics ratings in the data (62.7% of orders do not have a customer review).

Table 3.5: Distribution of logistics rating

Logistics rating	Count	Percentage
No response	81,192,551	62.7%
1	859,570	0.7%
2	296,870	0.2%
3	1,123,670	0.9%
4	2,893,972	2.2%
5	43,044,018	33.3%

Figure 3.6 shows the timeline of an order from arrival to final delivery. The time difference between the "Consignment" event and the "Order arrival" event is defined as *Consignment time*. Similarly, the time difference between the "Order arrival" event and the "Signed" event is defined as *Order delivery time*. If the delivery time is greater than the promised speed, the order is delayed.

We now summarize the distribution of delivery times by promise speed and service (Cainiao or seller). Figure 3.7 shows the density plot of delivery time by service. Cainiao manages more same-day and next-day customer orders than Cainiao independent sellers. Hence, the distribution for Cainiao is more skewed than for the seller.

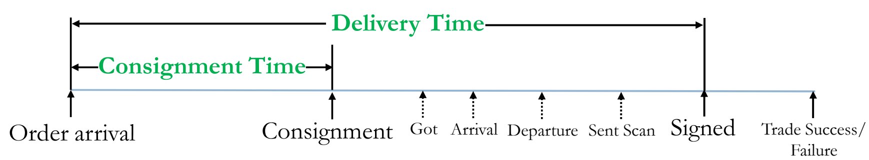


Figure 3.6: Consignment and Delivery time

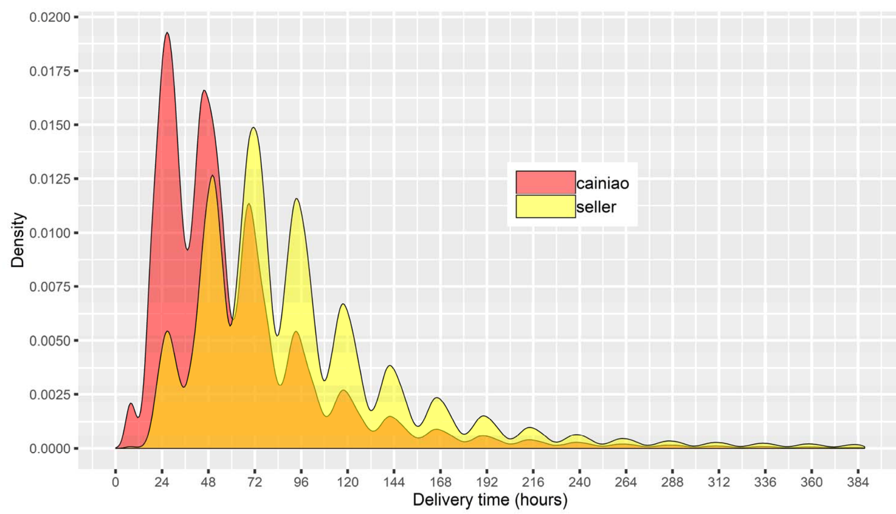


Figure 3.7: Density plot of delivery time by service

Similarly, Figure 3.8 shows the distribution of delivery time by promise speed. The average delivery time is correlated with the promised speed: most of the orders with "same day" promised speed were delivered within 24 hours, and a similar result holds true for "next day" promised speed.

We now compute an important operational performance measure, the delivery delay (in days). We defined the time difference between delivery time (signed by customer) and promised speed as the delivery delay (Figure 3.6). If the measure is negative, the order is delivered in time. We then forced the value to be **"0"** so that we could categorize it as on-time. Figure 3.9 shows the measure for delivery delays for same day and next day order requests received on Day 1. If the same day order is delivered on Day 2/ Day 3/ Day 4, the delay corresponds to one day / two day / three day, respectively. Similarly, if the next day

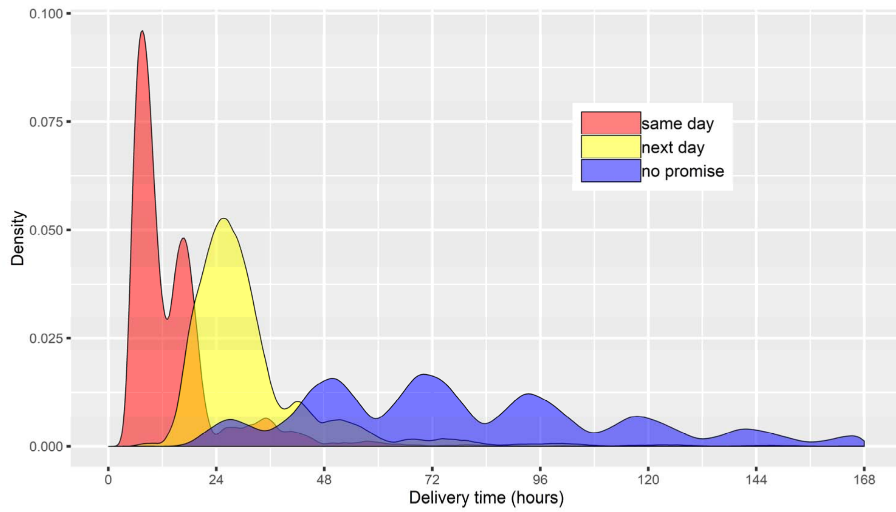


Figure 3.8: Density plot of delivery time by promise speed

order is delivered on Day 3/ Day 4, the delay corresponds to one day / two day, respectively.

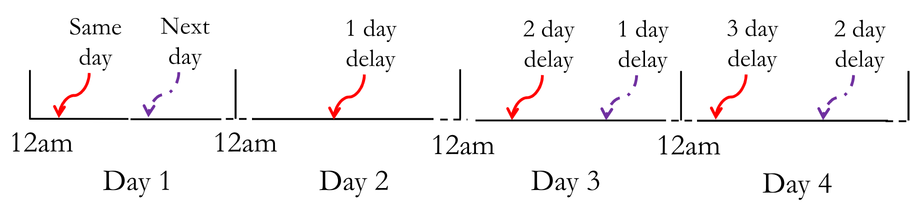


Figure 3.9: Delivery delay for same day and next day request

We now present the key variables used in the econometric model.

3.5.5 Measures

Using each customer order as the unit of analysis, we aim to understand the impact of different operational variables on the customer's provided logistics rating. Below we provide definitions of the measures which are specified in the econometric model.

Logistics rating: Logistics rating is the dependent variable provided by the customer for their experience with the logistics service. The variable uses a Likert scale with values ranging from 1 (poor quality) to 5 (high quality). The prior literature (Hu et al. [2009]) on online ratings indicates that ratings typically follow a **J**-shaped distribution with significant numbers of missing values (no response). Rather than excluding these observations, we model the option of "no response" in the econometric model as described later.

The independent variables are provided below:

Delay days: The time difference in days between the actual delivery time and promised speed for each customer order. If the measure is negative (meaning the order was delivered before the promise speed), we assign the variable's value to "**0**". In addition, we combine delay days greater than or equal to 4 into one category to keep as few categories as possible. Hence, this variable has 4 categories : "**0**" (delivered in time), "**1**" (one day delay), "**2**" (two day delay), "**3**" (three day delay), "**≥4**" (greater or equal to four day delay). Delivery delays can be computed only for orders with available promise speed. As a result, we focus on the subset of orders with same day and next day promise. Table 3.6 provides the distribution of delivery delays in the data. We found 6.7% (765,292) of the same-day and next-day customer orders were delayed in the data.

Table 3.6: Distribution of delivery delays

Delays	Count	Percentage
No delay	10,640,806	93.3%
One day	562,284	4.9%
Two day	121,362	1.1%
Three day	38,857	0.3%
≥ Four day	42,789	0.4%
Total	11,406,098	

Delay Ind: Delay Ind is a binary version of delay days which is defined as follows.

$$\text{Delay Ind} = \begin{cases} 1 & \text{Delay days} = 1, 2, 3, \geq 4 \\ 0 & \text{Delay days} = 0 \end{cases}$$

Promised Speed: Promised speed is the requested speed of delivery by the customers.

The measure takes two values: **1** (same day), **2** (next day). Promised speed can be interpreted along the dimension of willingness to wait for service completion. For example, a customer who specifies speed "**1**" is less willing to wait whereas one who requests promised speed "**2**" is more willing to wait.

Percentage Pay: Percentage Pay is the ratio of the total amount paid for each customer order to the mean of total amount paid for all orders in the data. We use this scaled measure of total amount paid rather than the actual value for two reasons: 1) the variation in total amount paid in the data is fairly large (minimum 0.01 Yuan and maximum 1,389,199 Yuan); and, 2) to prevent having extremely low value for the coefficient in the regression results. Due to the scaled measure, the interpretation of the coefficient becomes "one percentage increase in total amount paid for an order" on the dependent variable rather than increase in 1 Yuan increase.

Cainiao: Cainiao is a binary variable which takes the value of 1 if the customer order was managed by the Cainiao network, 0 if managed by an Cainiao independent seller.

Controls: The customer's utility from logistics service or the ability to provide a rating or no rating is likely to depend on additional factors such as time (e.g., month, week of the month, day of week, hour of day and holidays). We include an exhaustive list of controls for holidays. Holidays in our data include New Year's day, the day before the Spring Festival, Spring Festival, Chinese New Year, Lantern Festival, Zhonghe Festival, International Women's Day, Arbor Day, March equinox, Qing Ming Jie, Labour Day, Youth Day, Children's Day, Promotion, the day after promotion, two days after promotion, CPC Founding Day, and Maritime Day.

3.6 Model & Analysis

We explain the underlying data generating process of logistics ratings. The dependent variable in our data, logistics rating, is a discrete ordered outcome. Hence we specify a ordered regression model to examine the impact of different operational drivers. The latent continuous response, where observed ordinal responses Y_{ijklt} are generated from is given by

$$
\begin{aligned}
Y^*_{ijklt} = {} & \beta_1 \cdot \text{Delay days}_{ijklt} + \beta_2 \cdot \text{Promise Speed}_{ijklt} + \beta_3 \cdot \%\ \text{Pay}_{ijklt} + \\
& \beta_4 \cdot \text{Delay Ind}_{ijklt} \cdot \text{Promise Speed}_{ijklt} + \beta_5 \cdot \text{Delay Ind}_{ijklt} \cdot \%\ \text{Pay}_{ijklt} + \\
& \beta_5 \cdot \text{Cainiao}_{ijklt} + \beta_6 \cdot \text{Month}_j + \beta_7 \cdot \text{Week}_k + \beta_8 \cdot \text{Day}_l + \beta_9 \cdot \text{Hour}_t + \\
& \beta_{10} \cdot \text{Holidays} + \\
& \epsilon_{ijklt}
\end{aligned}
$$

where i, j, k, l, t represent customer order, month, week, day and hour of the day respectively.

As we mentioned previously, Hu et al. [2009] showed that the graphical distribution of ratings for three product categories from the Amazon platform follow **J**-shaped distribution, primarily due to two self-selection biases: *purchasing bias* and *under-reporting bias*. Purchasing bias causes the positive skewness in the distribution. Acemoglu et al. [2017] propose a theoretical ordered response model where they would place the "no response" option in the median position of ordered ratings. We follow the similar framework and include "no response" option in our econometric model. However, in our context the rating scale ranges from 1 to 5, thus "no response" can be either placed between 2 and 3 or 3 and 4. Hence, we analyze two versions of our model. The version 1 is provided below

$$
Y_{ijklt} = \begin{cases} 1 & Y^*_{ijklt} \leq K_1 \\ 2 & K_1 < Y^*_{ijklt} \leq K_2 \\ 3 & K_2 < Y^*_{ijklt} \leq K_3 \\ \text{No response} & K_3 < Y^*_{ijklt} \leq K_3 \\ 4 & K_4 < Y^*_{ijklt} \leq K_5 \\ 5 & K_5 < Y^*_{ijklt} \end{cases}
$$

Similarly, version 2 is as follows

$$
Y_{ijklt} = \begin{cases} 1 & Y^*_{ijklt} \leq K_1 \\ 2 & K_1 < Y^*_{ijklt} \leq K_2 \\ \text{No response} & K_2 < Y^*_{ijklt} \leq K_3 \\ 3 & K_3 < Y^*_{ijklt} \leq K_4 \\ 4 & K_4 < Y^*_{ijklt} \leq K_5 \\ 5 & K_5 < Y^*_{ijklt} \end{cases}
$$

where K_1 to K_5 are cut-off points on the underlying customer latent response curve.

The numerical results of the ordered regression model are presented in the Results section.

3.6.1 Results

We ran the econometric model (Ordered Logistic Regression[3]) for both the versions. The results are summarized in Table 3.7. As the econometric model has independent variables (Delay days, Promise Speed) each with different categories, there is a need to define the base

[3]We found the similar results with Ordered Probit model

category to interpret the results. We choose the base category as Delay days = "**0**" with Promise speed "**2**" meaning an on-time delivery for a next day order request (or customer who was more willing to wait).

Column 1 lists the independent and control variables. Columns 2 and 4 show the results for model version 1 and model version 2 (respectively) when Delay days is specified at four levels (one day delay, two day delay, three day delay, $\geq$ four day delay) while Columns 3 and 5 show similar results when Delay days is specified as a binary indicator variable. We find (for the first specification) that the co-efficient decreases with increase in Delay days, i.e, the logitics rating monotonically decreases with the increase in Delay days . Our finding that the logistics rating monotonically decreases with increase in Delay days provides support for Hypothesis 1.

We find the coefficient for Promised speed (same day) is positive and statistically significant, which implies that a same-day request customer (or customer with less willingness to wait) provides a higher rating than a next-day request customer (or customer with more willingness to wait). Hence, the logistics rating increases with faster promised speed of delivery supporting our Hypothesis 2(a). We find that the coefficient for the interaction between Delay Ind and Promise speed (same day) is not negative and statistically insignificant. Thus, a same-day request customer does not provide a lower rating for a delayed order compared to an on-time order than a next-day request customer. Thus, customer type, as measured by Promise speed, does not moderate the effect of a delivery delay on the logistics rating, which does not support the Hypothesis 3(a).

We find the coefficient for % Pay is positive and statistically significant, which suggests that a customer provides a higher rating for an order of higher Yuan value, thus, supporting Hypothesis 2(b). The coefficient for the interaction between Delay Ind and % Pay is negative and statistically significant, implying that a customer who paid a higher order amount gives a more negative rating to a delayed order compared to an on-time order than a customer

Table 3.7: Ordered Logit model estimates

Variable	Model 1 (Version 1)	Model 2 (Version 1)	Model 3 (Version 2)	Model 4 (Version 2)
One day delay	-0.135***		-0.107***	
	(0.003)		(0.003)	
Two day delay	-0.272***		-0.222***	
	(0.007)		(0.006)	
Three day delay	-0.376***		-0.307***	
	(0.007)		(0.007)	
$\geq$ Four day delay	-0.568***		-0.488***	
	(0.003)		(0.003)	
Delay Ind		-0.179***		-0.142***
		(0.004)		(0.004)
Promise speed (same day)	0.026***	0.027***	0.021***	0.023***
	(0.002)	(0.002)	(0.002)	(0.002)
Delay Ind * Promise speed (same day)		0.003		0.005
		(0.011)		(0.011)
% Pay	0.038***	0.038***	0.037***	0.038***
	(0.002)	(0.002)	(0.002)	(0.002)
Delay Ind * % Pay		-0.009***		-0.01***
		(0.001)		(0.001)
Cainiao	-0.007**	-0.007**	-0.007**	-0.007*
	(0.005)	(0.005)	(0.005)	(0.005)
Month	Yes	Yes	Yes	Yes
Week	Yes	Yes	Yes	Yes
Day	Yes	Yes	Yes	Yes
Hour	Yes	Yes	Yes	Yes
Holidays	Yes	Yes	Yes	Yes
K1	-5.433	-5.433	-5.436	-5.437
K2	-5.175	-5.175	-5.179	-5.179
K3	-4.527	-4.527	0.493	0.493
K4	0.520	0.517	0.516	0.516
K5	0.595	0.595	0.591	0.591
LR Chi^2	44,013	42,036	41,864	40,384
Observations	11,406,098	11,406,098	11,406,098	11,406,098
LL	-8,899,217	-8,900,205	-8,900,291	-8,901,031

*p<0.1; **p<0.05; ***p<0.01

who paid a low order amount, thus, providing support for Hypothesis 3(b).

3.6.2 Order Split vs Bundling and its impact on Logistics Rating

We now study the impact of order split on logistics rating. Consider a customer who places an order for two unique items (or multiple quantity of a single item). The order is "bundled" if the entire order is shipped in a single shipment, while if the order is divided into multiple shipments each of which is delivered separately to the customer, the order is said to be "split". We extracted the order split information from the **logistics** data. If an order has multiple unique *logistics_order_id* values, the order was split into multiple packages and shipped to the customer. We can also identify orders that could have been potentially split but were not, e.g., those with at least two unique items or one unique item with an order quantity greater than 1. We summarize below both the orders which were split and those which could have been split in Figure 3.10.

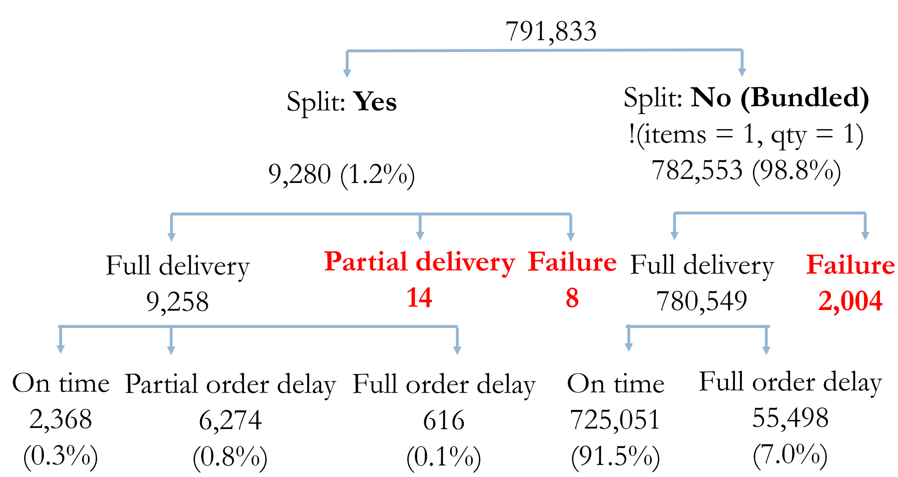


Figure 3.10: Order split

The total number of orders for the split analysis is 791,833, of which 9,280 (1.2%) orders were split and 782,533 (98.8%) could potentially have been split but were not. Of the 9,280 split orders, 9,258 were fully delivered, 14 were partially delivered (the customer only received a fraction of the split order) and 8 were not delivered (Failure). Of 9,258 fully delivered orders, 2,368 were delivered on or before the promised date, 6,274 were partially delayed (part of the order arrived on-time but the remainder was delayed), and 616 were fully delayed (all shipments were delivered after promise date).

Among the 782,533 "no" split or bundled orders, 780,549 were delivered and 2,004 were not delivered. Of the 780,549 fully delivered orders, 725,051 were delivered on or before the promised speed and 55,498 were fully delayed. We excluded both partial delivery and failed orders from the analysis, as we did not have a "signed" time for these records and, thus, could not compute Delay days. Hence, we focus our analysis only on fully delivered orders.

We study the following two questions:

- Q 2a : How does customers logistics rating vary between bundled vs split orders which were delivered on-time? Does the split moderate the impact of delay on logistics rating?
- Q 2b : For split orders, do customers give a higher rating to partially delayed orders compared to fully delayed orders?

We ran the econometric model for both the versions specified in Section 3.6 to analyze each question. The results are summarized in Table 3.8. Column 1 lists the independent and control variables; columns 2 and 3 provide model version 1 and model version 2 estimates for Q 2a (respectively), columns 4 and 5 provide similar estimates for Q 2b. As our specifications have multiple independent variables each with different categories, we must define the base category, which we do individually for Q 2a and Q 2b.

For Q 2a, we choose the base category as Delay Ind = 0, "bundled" order and "next day" Promise speed, i.e., a bundled order with on-time delivery for a customer more willing to wait. In both columns 2 and 3, we find that the coefficient for order split (Yes) is positive and significant, suggesting that customers provide a higher rating for split orders than bundled orders. However, the coefficients for the interaction between Delay Ind and order split (Yes) are negative and statistically significant. Thus, a customer who receives a split order gives a more negative rating for a delayed order compared to an on-time order than a customer

Table 3.8: Order split model estimates

Variable	Q1 (Version 1)	Q1 (Version 2)	Q2 (Version 1)	Q2 (Version 2)
Delay Ind	-0.165***	-0.145***		
	(0.012)	(0.012)		
Order split (Yes)	0.317***	0.312***		
	(0.026)	(0.026)		
Delay Ind * Order split (Yes)	-1.039***	-1.064***		
	(0.122)	(0.121)		
Partial order delay			-0.112	-0.124
			(0.136)	(0.132)
Promise speed (same day)	0.038***	0.036***	0.141	0.142
	(0.008)	(0.008)	(0.125)	(0.125)
Delay Ind * Promise speed (same day)	-0.007	-0.003		
	(0.041)	(0.041)		
% Pay	0.002***	0.001**	0.081***	0.082***
	(0.000)	(0.000)	(0.008)	(0.008)
Delay Ind * % Pay	-0.0001	-0.0001		
	(0.002)	(0.002)		
Cainiao	-0.054***	-0.056**	0.256*	0.245*
Month	Yes	Yes	Yes	Yes
Week	Yes	Yes	Yes	Yes
Day	Yes	Yes	Yes	Yes
Hour	Yes	Yes	Yes	Yes
Holidays	Yes	Yes	Yes	Yes
K1	-5.904	-5.903	-5.35	-5.369
K2	-5.69	-5.689	-5.232	-5.251
K3	-5.11	0.503	-4.922	1.903
K4	0.515	0.516	1.938	1.919
K5	0.567	0.568	1.987	1.968
LR Chi^2	1,907.460	1,850.160	173.470	174.670
Observations	783,533	783,533	6,890	6,890
LL	-580,085.890	-580,114.550	-3,210.604	-3,210.003

*p<0.1; **p<0.05; ***p<0.01

who receives a bundled order. Two key managerial insights can be drawn from these results. First, the carrier company can safely split an order if they have confidence that the individual shipments will reach the customer by the promised delivery date. Second, it is advisable not to split an order if there could be a delay in delivery, as customers give a more negative rating for a split delayed order compared to a bundled delayed order.

For Q 2b, we focus on partially delayed and fully delayed split orders. Here, we choose the base category as "next day" Promise speed and fully delayed orders, i.e., a fully delayed order for a customer who is more willing to wait. In columns 4 and 5, we find the coefficient for partial order delay is negative but statistically insignificant, i.e., there is no evidence that customers give a higher rating to partial delayed orders compared to full delayed orders. A key managerial insight from this is that if the seller believes the complete order delivery will be delayed, there is no need to split the order (thereby incurring additional shipment costs) since the customer treats a split order as delayed even if only a part of the order is delayed.

3.6.3 Logistics Rating and Customer Purchase Probability

In the context of our study, sellers display cumulative information on the item quality rating, online purchase (termed as "Service quality" in the data) rating, and logistics rating, which raises the question that do customers really value and consider these ratings before making a decision to purchase. The answers to these questions lie in understanding if the logistic rating (and other ratings) impacts the customer's purchase intention or probability. We examine the impact of logistics rating on purchase probability in two different ways: 1) across items which are unique to a seller, and 2) for non-unique or competing items available from multiple sellers.

For unique items, consider the example of a Seller 1 in Figure 3.11 who is selling a watch which is not available from any other seller. A visiting customer either chooses to purchase the watch, or not purchase the watch (outside option).

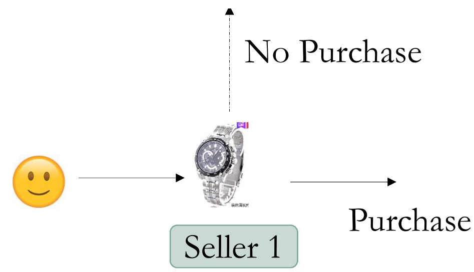


Figure 3.11: Purchase choice options for a unique item

More generally, consider a customer "c" visiting the web-page of item "j" is confronted with a choice of either to purchase, or not to purchase. The customer's preference or utility for item "j" on day "t" given by

$$U_{cjt} = X_{jt} \cdot \beta + \alpha \cdot p_{jt} + \zeta_j + \epsilon_{cjt} \tag{3.1}$$

where β and α represent the customer's sensitivity to variables in X_{jt}, unit price p_{jt}, ζ_j is the customer's perception (known only to them) of the quality of item "j", ϵ_{cjt} is an idiosyncratic or random component capturing $c^{'}s$ utility, which varies between the two choices. Under the assumption of a Gumbel distribution for the random component, the customer $c^{'}s$ probability of purchasing "j" is given by

$$p_{cjt} = \frac{e^{X_{jt} \cdot \beta + \alpha \cdot p_{jt} + \zeta_j}}{1 + e^{X_{jt} \cdot \beta + \alpha \cdot p_{jt} + \zeta_j}} \tag{3.2}$$

We need data on the market size to estimate coefficients in the choice model. The market size is defined by potential customers who visit the item's web-page. The **item view** data provides granular details on each item, including the number of views of each item sold by all the sellers at daily level. As a result, the different variables in X_{jt} and p_{jt} are averaged at daily level. The variables in X_{jt} include average volume of logistics rating available to the customer at the time of purchase, average logistics score, average order quality score, and the average service quality score. We used the total number of unique customer visits on personal computer and app as our measure for market size (M_{jt}). The **customer order** data has buyer id, which gives us the unique number of customers who purchased the item (q_{jt}). The market share for item "j" can be computed as $\frac{q_{jt}}{M_{jt}}$.

Following the methodological approach of Berry [1994], we equate the purchase probability function and market share value given by

$$s_{jt} = \frac{q_{jt}}{M_{jt}} = \frac{e^{X_{jt}\cdot\beta+\alpha\cdot p_{jt}+\zeta_j}}{1+e^{X_{jt}\cdot\beta+\alpha\cdot p_{jt}+\zeta_j}}$$

The above equation can be written as

$$\frac{s_{jt}}{s_{0t}} = e^{X_{jt}\cdot\beta+\alpha\cdot p_{jt}+\zeta_j}$$

where $s_{0t} = 1 - s_{jt}$ represents outside option share (no purchase). Applying natural logarithm on both sides further simplifies to

$$log(\frac{s_{jt}}{s_{0t}}) = X_{jt}\cdot\beta + \alpha\cdot p_{jt} + \zeta_j$$

The left side of the equation for each item "j" can be computed at daily level for the dataset. The parameters in the model can be estimated by running an item fixed effect regression. The total number of unique items sold by all the sellers is 272,339. Of these, 99.3% (270,597) are unique to a seller. As a result, the regression estimation includes 270,597 fixed effects. Our goal is to estimate the coefficient of the average logistics score (not the fixed effects) and the data mimics a panel data structure with item as cross-section and day as time series, we use a within-groups estimator (Greene [2003]) with robust standard errors corrected for heteroscedasticity and autocorrelation thereby reducing computational complexity.

Similarly, we study the impact of logistics rating on customer purchase probability for non-unique sold by multiple sellers (competing items). Consider the example in which two sellers are selling the same watch as shown in Figure 3.12. A visiting customer has three choices: purchase the watch from Seller 1, or from Seller 2, or choose to not purchase (outside option).

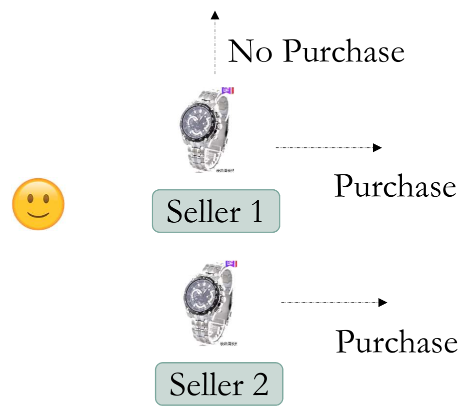


Figure 3.12: Purchase choice options for competing items

More generally, consider a customer "c" visiting the item "j" on seller 1 or seller 2 is confronted with a choice either to purchase or not to purchase. The customer preference or utility for purchase from seller 1 on day "t" is given by

$$U_{cj1t} = X_{j1t} \cdot \beta + \alpha \cdot p_{j1t} + \zeta_{j1} + \epsilon_{cj1t} \tag{3.3}$$

where β, α represents customers sensitivity to variables in X_{j1t}, unit price p_{j1t}, ζ_{j1} is unobserved (to the researcher but observable to the customer) quality of item from seller 1, ϵ_{cj1t} is a is a idiosyncratic or random component. Under the assumption of Gumbel distribution for random component, the customer $c^{'}s$ probability of purchasing from seller 1 is given by

$$p_{cj1t} = \frac{e^{X_{j1t} \cdot \beta + \alpha \cdot p_{j1t} + \zeta_{j1}}}{1 + \sum\limits_{l=1,2} e^{X_{jlt} \cdot \beta + \alpha \cdot p_{jlt} + \zeta_{jl}}} \tag{3.4}$$

Contrary to the unique items case, the measure for market size (M_{jt}) is the total number of unique customer visits for the item on **both** the sellers page. The regression equation with market share for the item on seller 1 and 2 is given by

$$log(\frac{s_{j1t}}{s_{0t}}) = X_{j1t} \cdot \beta + \alpha \cdot p_{j1t} + \zeta_{j1}$$

$$log(\frac{s_{j2t}}{s_{0t}}) = X_{j2t} \cdot \beta + \alpha \cdot p_{j2t} + \zeta_{j2}$$

where $s_{0t} = 1 - s_{j1t} - s_{j2t}$ represents outside option share (no purchase).

The parameters in the model can be estimated by running a item fixed effect (combination of seller and item) regression. Of the 272,339 items, 0.7% (1,742) are competing items sold by multiple sellers. We found a total of 4,769 combination of item id and seller id. Hence, the regression estimation includes 4,769 fixed effects. The parameter estimates of regression models for both unique and competing items case are given in Table 3.9.

Table 3.9: Impact of logistics rating on customer purchase probability

	Unique items:		**Competing items**:	
	(Model 1)	(Model 2)	(Model 3)	(Model 4)
% Pay	−0.206***	−0.206***	−2.486***	−2.486***
	(0.00001)	(0.00001)	(0.018)	(0.018)
% Volume of logistic ratings	0.082***	0.082***	0.004**	0.004**
	(0.0004)	(0.0004)	(0.002)	(0.002)
Average logistics rating	0.695***	0.695***	1.387***	1.387***
	(0.009)	(0.009)	(0.058)	(0.058)
Average order quality rating	−0.00002		−0.00002	
	(0.00004)		(0.0002)	
Average service quality rating		0.00002		0.00002
		(0.00004)		(0.0002)
Observations	9,411,630	9,411,630	372,384	372,384
R^2	0.009	0.009	0.054	0.054
Adjusted R^2	0.009	0.009	0.054	0.054
Residual Std. Error	0.606	0.606	0.707	0.707
F Statistic	22,360.270	22,360.270	5,323.135	5,323.135

Note: *p<0.1; **p<0.05; ***p<0.01

In Table 3.9, column 1 refers to all the independent variables. We use the scaled measure for price and volume of logistic ratings. Percentage Pay (% Pay) is the ratio of the unit price for each item order to the mean of unit price for all item orders in the data. Similarly, % Volume of Logistic ratings is the ratio of count of logistics ratings to its mean in the data. We use this scaled measure than the actual value to prevent having extremely low value for the coefficient in the regression results. Columns 2 and 3 show the results for unique items while columns 4 and 5 show similar results for competing items. We find in both unique and competing item cases, the correlation between average order quality and average service quality rating to be > 0.8. Hence we include only one variable to avoid multi-collinearity. We run different combinations of models, to check robustness of the results. Across models, we find coefficient of % Pay negative and statistically significant. This implies that higher the item price, less is the utility for customer to purchase the item which is a trivial result. We find coefficient of % Volume of logistics rating count and average logistics rating positive and statistically significant, implying that higher the number of ratings available to the customer, higher is the utility for the customer to purchase the item. We do not find any significance for remaining quality scores : average order and service quality. Hence, we find statistical evidence that logistics rating impacts customer purchase probability.

3.6.4 Logistics Performance and Sales

In this section, we aim to quantify the link between logistics performance and improvement in sales. For example, we ask the question : what is the improvement in a seller sales for a reduction in delivery delays by one day?

Our primary goal here is to provide a framework in quantifying the relation between logistics performance and seller sales. Rather than focusing on all sellers in the Cainiao network, we choose one seller and analyzed their customer orders. We choose the seller with id "225" as they constitute highest volume of orders in the data. Of the total 11,406,098 customer orders (final data used in analyzing the main econometric specification), seller "225" accounts for 915,248 orders (8.0%). The relationship between logistics performance and sales is established using a two step process. First, an improvement in logistics performance increases the expected logistics rating. Second, an increase in expected logistics rating increases the customer purchase probability, and subsequently the sales.

In the first stage, we initially evaluate the change in probability for each of the 6 categories of logistics rating (1 through 5 and "no response" categories) as a result of reduction delivery delay. We run both the versions of our main econometric specification (section 3.6) on the seller "225" data of 915,248 orders and generate the average marginal effects of delivery delays. The average marginal effects are given in Table 3.10.

Table 3.10: Average Marginal effects

Logistics rating	Model	One day delay	Two day delay	Three day delay	$\geq$ Four day delay
1	Version 1	-0.06%	-0.15%	-0.19%	-0.28%
2	Version 1	-0.01%	-0.04%	-0.05%	-0.07%
3	Version 1	-0.05%	-0.14%	-0.18%	-0.27%
No response	Version 1	-2.49%	-6.60%	-8.30%	-12.23%
4	Version 1	+0.05%	+0.13%	+0.17%	+0.25%
5	Version 1	+2.56%	+6.80%	+8.55%	+12.60%
1	Version 2	-0.04%	-0.13%	-0.17%	-0.23%
2	Version 2	-0.01%	-0.03%	-0.04%	-0.06%
No response	Version 2	-1.99%	-5.60%	-7.60%	-10.22%
3	Version 2	+0.01%	+0.03%	+0.05%	+0.06%
4	Version 2	+0.04%	+0.11%	+0.15%	+0.20%
5	Version 2	+2.00%	+5.61%	+7.62%	+10.25%

The numbers in the table are inferred as follows. For example, in Model Version 1, a reduction in delivery delay by one day decreases the probability of logistics rating categories 1, 2, 3, and "no response" by 0.06%, 0.01%, 0.05%, and 2.49%respectively, while it increases the probability of categories 4, and 5 by 0.05%, and 2.56% respectively. The current distribution of logistics ratings for seller "225" across all categories is : 1 (count: 25,691, proportion: 0.59%), 2 (6,637: 0.15%), 3 (26,331: 0.61%), no response (2,770,873 : 63.77%), 4 (70,373: 1.62%), 5 (1,445,519: 33.27%). Setting the proportion of "no response" to 0.00%, the final proportions of remaining 1 to 5 ratings are normalized to 100% as customers only see the volume of available ratings at the time of purchase. The initial expected rating is given by 4.8439 (= 1 * 1.63% + 2 * 0.42% + 3 * 1.67% + 4 * 4.47% + 5 * 91.81%). We add the average marginal effects to the current distribution of rating to arrive at final distribution. Using the normalized proportions of the new distribution, the final expected rating under Model Version 1 and Version 2 are given by 4.8622 (+0.0183 increase from initial expected rating 4.8439) and 4.8555 (+0.0116). A unique contribution of our approach above is that our method accounts for customers who do not provide any ratings (and, hence, whose valuation is invisible to potential customers) and is able to estimate the impact of reduction in delivery delay on them.

Table 3.11: Change in Logistics rating for a reduction in delivery delay by one day

Logistics rating	Model	Prop. (inital)	Norm. (inital)	Exp.rating (initial)	Marg. effects	Prop. (final)	Norm. (final)	Exp.rating (final)
1	Version 1	0.59%	1.63%		-0.06%	0.53%	1.37%	
2	Version 1	0.15%	0.42%		-0.01%	0.14%	0.37%	
3	Version 1	0.61%	1.67%		-0.05%	0.56%	1.44%	
No response	Version 1	63.77%	0.00%	**4.8439**	-2.49%	61.28%	0.00%	**4.8622**
4	Version 1	1.62%	4.47%		0.05%	1.67%	4.31%	
5	Version 1	33.27%	91.81%		2.56%	35.83%	92.51%	
1	Version 2	0.59%	1.63%		-0.04%	0.55%	1.44%	
2	Version 2	0.15%	0.42%		-0.01%	0.14%	0.37%	
No response	Version 2	63.77%	0.00%	**4.8439**	-1.99%	61.78%	0.00%	**4.8555**
3	Version 2	0.61%	1.67%		0.01%	0.62%	1.61%	
4	Version 2	1.62%	4.47%		0.04%	1.66%	4.34%	
5	Version 2	33.27%	91.81%		2.00%	35.27%	92.23%	

In the second stage, we first create a new dataset to predict customer purchase probability. The new dataset is a summarized version of seller "225" order data. For each combination of *item_id* and *day_type* (e.g., Monday) in the data, we calculate mean of average unit price, volume of logistics ratings, unique customer visits, and fix the average logistic score to 4.8439 which is the initial expected rating. We then compute the customer purchase probability using the choice model in section 3.6.3 for each row in the new dataset. The average logistic rating is then changed to 4.8622 and customer purchase probabilities are recomputed. The change in probability is the incremental chance of a customer purchasing the product due to a change in rating from 4.8439 to 4.8622 while keeping remaining variables constant. The increase in sales is then given by the product of change in purchase probability, number of unique customer visits, and average unit price. We compute the average weekly increment in sales and compare with actual average weekly sales. We find % increase in weekly sales to be 2.54% for a reduction in delivery delay by one day. Similarly, we repeat the above procedure for reduction in two, three, $\geq$ four day delay under Model Version 1 and Version

2. The results are provided in Table 3.12. We find that the reduction of two, three and $\geq$ four day delay leads to improvement is average weekly sales as much as by 5.34%, 6.36% and 8.56% respectively. Hence, we conclude that reduction in delivery delays can significantly improve sales because of the improvement in logistics ratings.

Table 3.12: Improvement in sales for reduction in delivery delay

Delay	Model	Expected rating (inital)	Expected rating (final)	% δ Sales per week
One day	Version 1	4.8439	4.8622	2.54%
Two day	Version 1	4.8439	4.8883	5.34%
Three day	Version 1	4.8439	4.8977	6.36%
$\geq$ four day	Version 1	4.8439	4.9168	8.56%
One day	Version 2	4.8439	4.8555	1.43%
Two day	Version 2	4.8439	4.8753	3.75%
Three day	Version 2	4.8439	4.8836	4.75%
$\geq$ four day	Version 2	4.8439	4.8951	6.12%

3.7 Conclusion & Managerial Implications

Online ratings on e-commerce platforms are valuable information to both sellers and customers. Sellers communicate ratings to signal product or service quality and reduce any uncertainty in the mind of the customer. On the other hand, customers use ratings to compare sellers and in making their purchase decisions. In this paper, we aim to study the value of logistics performance, logistics rating and its impact of sales in an e-commerce platform. Using a large data-set on the entire life cycle of customer orders from an e-commerce platform, we examined four research questions.

Our first research question was: How do logistics performance metrics such as delivery delays, customer's requested speed of delivery, order amount paid and other potential variables impact the logistics service ratings of sellers on an e-commerce platform? We utilized an ordered regression model to analyze the variation in the data. The results from our analysis are as follows. First, logistics ratings are negatively impacted by delivery delays. Second, a same-day request customer (or customer with less willingness to wait) provides a higher rating than a next-day request customer (or customer with more willingness to wait). Third, a same-day request customer does not provide a lower rating for a delayed order compared to an on-time order than a next-day request customer. Fourth, the logistics ratings are positively impacted by orders of higher Yuan value. Lastly, a customer who paid a higher order amount gives a more negative rating to a delayed order compared to an on-time order than a customer who paid a lower order amount.

Our second research question asks: How do customer ratings vary between bundled vs split deliveries? Does the split moderate the impact of a delay on ratings? Under split orders, do customers give higher rating to partially delayed orders compared to fully delayed orders? The results from our analysis are as follows. First, customers provide a higher rating for split orders than bundled orders for an on-time delivery. Second, a customer who receives a split order gives a more negative rating for a delayed order compared to an on-time order than a customer who receives a bundled order. Third, no evidence was found that customers give a higher rating to partially delayed orders compared to fully delayed orders.

Our third research question looks at: Does the logistics rating impact an incoming customer's purchase probability for unique and competing products sold by the sellers? The results from our analysis are as follows. Controlling for order and service quality ratings, we find that the logistics rating positively impacts the customers' utility to purchase both unique and competing items.

Our final research question answers: How does logistics performance impact sales of a seller? Using the estimates from the analysis of first three research questions, we establish the link between logistics performance and sales. For one of the largest seller in the network, we find that reduction of delivery delays by one, two, three and $\geq$ four day leads to improvement in average weekly sales by as much as 2.54%, 5.34%, 6.36% and 8.56% respectively.

Our results provide several key managerial implications. First, to improve the logistics rating, sellers should focus on performance metrics and attributes such as delivery delays, promised speed of delivery and order amount paid. Second, sellers or Cainiao network should prioritize the delivery performance of same-day orders or orders with high amount paid because these customers provide a higher rating than their counterparts. Third, sellers should incentivize these same customers to participate in the rating process as these customers provide higher rating for a superior delivery performance. Fourth, sellers or Cainiao can safely split the order if they have confidence that the individual shipments will reach the customer by the promised delivery date. Fifth, it is not advisable to split an order if there could be a delay in delivery, as customers give a more negative rating for a split delayed order compared to a bundled delayed order. Finally, our results demonstrate that logistics ratings are important drivers of customer purchase in e-commerce platform and can significantly drive sales.

3.7.1 Venues for future Research

Prior work has shown that customers learn about product quality of books from online ratings (Zhao et al. [2013]). Future work on our dataset could include empirically testing if customers learn about service quality from the logistics rating. Similarly, past work has shown different biases on item quality ratings, including self-presentational behavior (Schlosser [2005]), self-selection bias (Li and Hitt [2008]), purchasing bias, under-reporting bias (Hu et al. [2009]) etc. Future work could include empirically verifying if these different biases exist in logistics rating.

The rich dataset from Cainiao network prompts a wealth of more interesting research questions. First, the Cainiao network is an integrated warehouse and logistics organization equipped with big IT and decision support systems to manage a large inventory of items and their operations. Cainiao is likely to manage inventory efficiently such that they ship the item upon arrival of order with minimal delay and less variation. On the other hand, monitoring item inventory is likely to be difficult for independent sellers who manage more unique items. As a result, it is possible that independent sellers ship products late with more variation. Hence it would be interesting to study whether the mean and variability in consignment times is less for Cainiao managed orders as compared to seller managed orders.

CHAPTER 4: DEMAND ESTIMATION AND OPTIMAL SHIFT TIMING IN STREET-HAIL TAXI SERVICES

4.1 Abstract

Improving taxi services efficiency is an important problem as it affects drivers income, passenger service and importantly transportation revenue. Taxi regulations limit the number of hours a driver can operate. Driver service intensity is endogenous as they choose the location, start time and end time of their shift. An absence of centralized control under regulation and endogenous service intensity can lead to a sizable number of drivers starting or ending their shift during the same time period. Increased delay in shift change time when drivers share a common resource (taxi) can lead to low availability affecting service level and revenue. The impact can be significant if the shift changes overlap with peak demand period. I utilize large-scale datasets of the GPS information of pick-ups and drop-offs ($\approx$ 14 million records) from New York Yellow Taxi services during June 2013 for this study. I first develop a stochastic matching model (double-ended queue) to predict passenger demand in location and time. The stochastic model allows for non-stationarity, randomness in arrivals and relocations of both drivers and passengers. Using sample path information along with Maximum Likelihood Estimation, I estimate potential passenger demand as well as drivers and passengers relocate rates. The predicted demand is then used to analyze the optimal timing of drivers changing their shift to maximize revenue under current status quo of delay in shift changeover. Also, I quantify the amount by which revenue could be improved if the shift changeover is instantaneous (*ongoing*).

4.2 Introduction

Taxi services play an important role in urban mobility service systems as they affect customers requirements, drivers income, fuel consumption and most importantly overall transportation revenue. Taxi and limousine services are one of the largest components of the United States transportation industry, with revenue of $18.9 billion in 2016. This revenue has increased by 6% during 2011-16 and expected to increase by 1.2% during 2017-21(Brennan [2014]).

The potential sources of increase in demand are due to multiple factors. First, the growth in the global economy, coupled with an increase in income levels has allowed consumers to choose the more convenient mode of taxi services over public transport. Specifically, the per capita of disposable income has increased at an annual rate of 1.6% during 2011-16, thus an increase in average consumer spending can potentially increase future demand for taxi services. For example, during 2011-16, the number of domestic and international tourist visit trips has increased by an annual rate of 2.6% and 4.1%, respectively. Second, growing corporate profits provides executives in the private sector with more spending power thus they are more likely to choose convenient taxi services. Corporate profit earned by millions of businesses increased at an annual rate of 2.3% during 2011-2016. In summary, the demand for taxi services is likely to increase in the near future albeit at lower rate due to increased competition from ride-sharing companies.

Two factors are likely to provide strong competition to growth of traditional street-hail taxi services. First, transport network companies such as Uber and Lyft have enabled customers to request a taxi via the use of an app. These technologically-driven organizations have attracted a large pool of independent contractors who provide affordable on-demand service instead of street-hail taxi service. These alternate services can increase the competition level for traditional street-hail taxi services thereby slowing their growth rate. Second, increase in federal funding for transportation could initiate projects improving the accessibility of other public transit services (e.g., buses, subways and trains). Thus, more consumers are likely to choose the latter modes of transport instead of taxi services. Hence to maintain sustainable growth in the taxi services industry it is important for government agencies to identify any existing operational inefficiencies and implement policies to increase ridership or revenue.

Taxi services in cities are typically managed by a government agency. For example, the New York City Taxi and Limousine Commission manages taxis in New York city. The agency regulates licenses to operate (entry restriction), trip prices and shift schedules. A lack of timely policy implementations necessary to meet a growing demand can often lead to operational inefficiencies. First, governments regulate the maximum number of taxi licenses (medallions) that can be operated in a city. This limits the number of taxi drivers that can operate across different geographical locations in a city. As a result, some customers choose a different mode of transport due to the unavailability of a service. Second, taxi services operate under a two-part tariff pricing scheme which comprises both fixed and distance based fare. Although drivers are not allowed to refuse customer requests, they are likely to select longer trip time request and choose locations with high demand intensity to maximize their revenue. In addition, drivers are likely to relocate to a new location after completing a previous trip in search of the next profitable trip. The simultaneous effects of both drivers and customers changing location can cause a spatial and temporal imbalance of supply and demand. The imbalance potentially has two significant costs − an increased operational (e.g., fuel) cost for drivers, reducing their profit, and lost revenue from customers, impacting overall taxi revenue. Third, due to regulation of 24-hour availability of taxis, authorities mandate taxi drivers operate for stated shift hours (e.g., 9 hours for NYC yellow taxis). Typically taxi service intensity is endogenous as the drivers choose when to start and end their shift as well as the location at which to provide service during the operational day.

The absence of a centralized control combined with drivers endogenous service can lead to a sizable number of drivers starting or ending their shift during a specific time period. A longer shift change time between drivers can lead to less availability of taxis and subsequently lower revenue. The drop in revenue can be significant if sizable numbers of shift changes occur during a peak period.

Figure 4.1 shows the temporal variation of median taxi revenue, fare and active drivers for each hour on a weekday for New York Yellow taxi cabs. Three key insights can be derived from the plot. First, total revenue or fare is correlated with the number of active drivers in the taxi system.

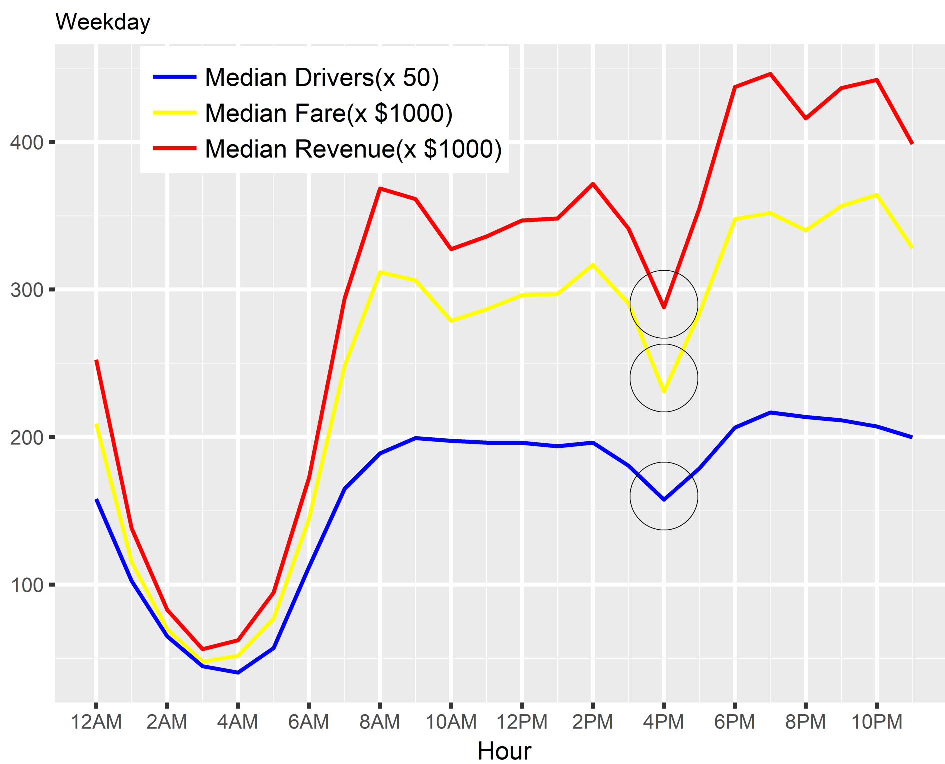


Figure 4.1: Median revenue, fare and active drivers by hour

Second, revenue drops during 12AM - 5AM, rises quickly until 9AM and stays almost flat until 2PM. The evidence can be explained by Figure 4.2 which provides the temporal variation in number of drivers starting or ending their shifts in each hour. The sharp rise in number of driver starting their shifts during 7AM-8AM and 5PM-6PM leads to rise in revenue.

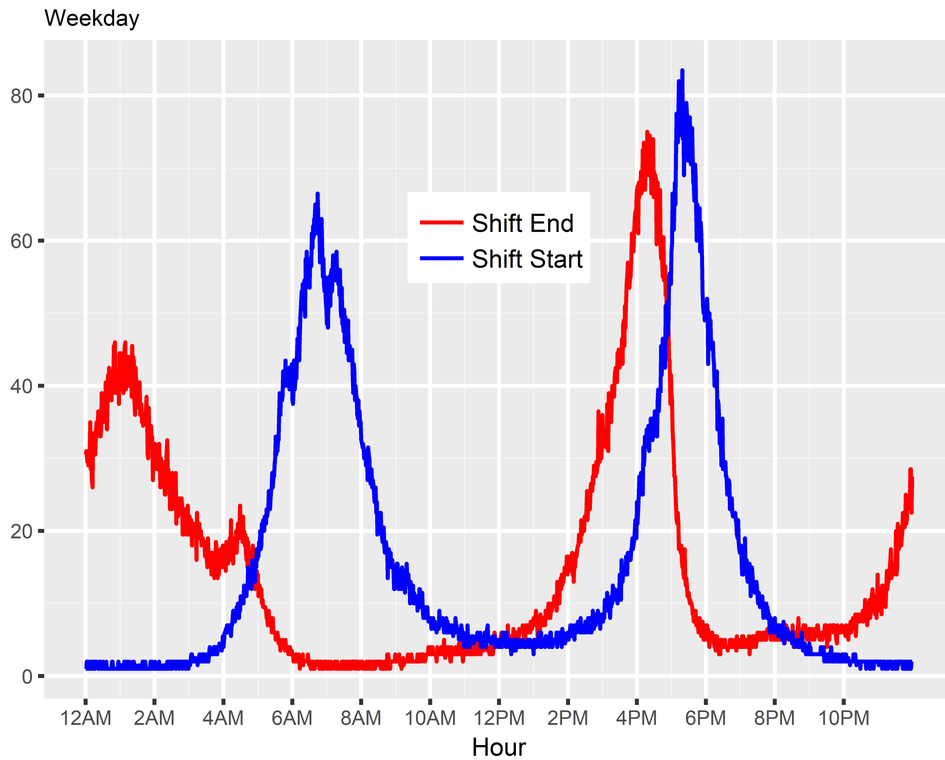


Figure 4.2: Driver shifts - start and end

Third, revenue drops significantly during 3AM - 5PM. Figure 4.3 shows a magnified version of the plot of driver shifts starting and ending during 2PM-8PM. The drop in revenue is due to two factors; (i) a sizable number of drivers end their shift around 4PM which leads to a sudden drop in the availability of taxis in the system; and, (ii) a significant delay in the evening drivers start-of-shift of approximately 80 minutes leads to a sustained unavailability of taxis. This pattern is partly due to the regulations that NYC TLC mandates to manage taxi operations in the city. This abnormal behavior in the taxi system is the seed of my research work here. In Section 4.3, I provide detailed description on the New York Yellow taxi industry and its operations.

Here I investigate how to improve the revenue of the taxi service system by identifying the optimal time for shift changeover. Specifically, I address the following three research

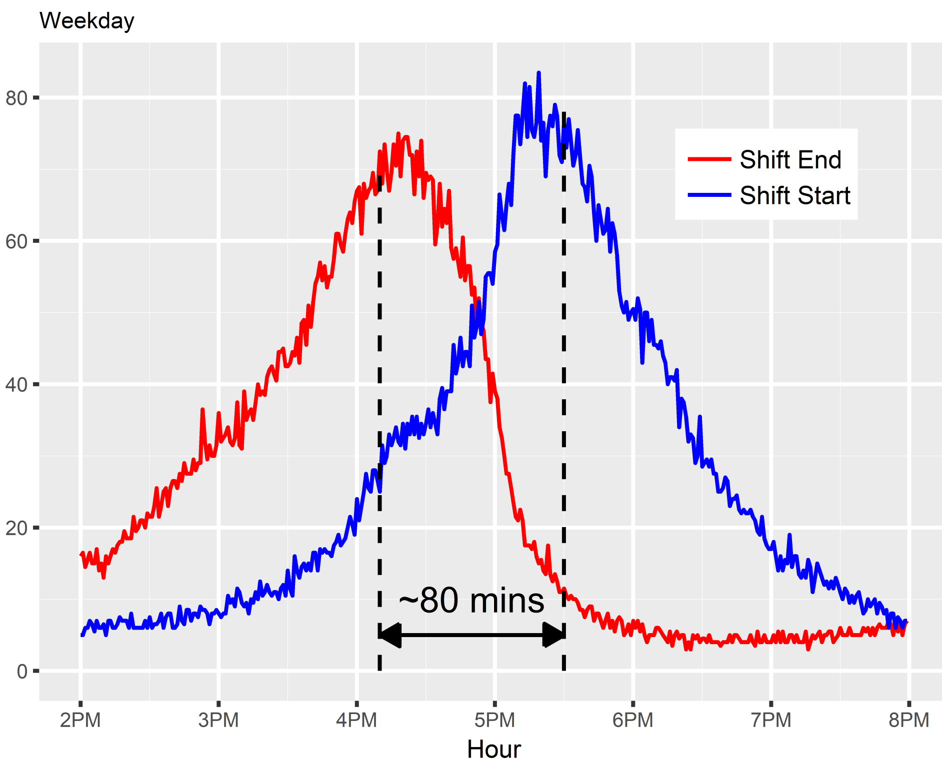


Figure 4.3: Evening driver shifts - start and end

questions: (i) How to estimate true passenger demand, capturing spatial and temporal distribution as well as driver and customer location changes? (ii) When should the drivers change their shift to maximize revenue under the current status quo of delay in shift changeover? (iii) How much revenue could be gained if the shift changeover is instantaneous?

The rest of the chapter is organized as follows. Section 4.3 provides details on New York City yellow taxi cab service operations. Section 4.4 provides details on prior work and the contribution of my results to the literature. Section 4.5 describes my data in detail and provides necessary summary statistics. In Section 4.6, I develop a demand model integrating both stochastic models and data to empirically estimate true passenger demand.

4.3 New York city Yellow Taxi market

New York city yellow taxi cabs provide street-hail services in the five boroughs, including Manhattan, Brooklyn, Staten Island, the Bronx and Queens, and at two airports (JFK and La Guardia). The taxi drivers require medallion licenses to operate their services and licenses fall into one of two categories - Fleet and Independent. Fleet licenses, also known as corporate medallions, are owned by multi-taxi companies or investors. Drivers qualified under the fleet model are required to operate for two 9-hour shifts each day. An independent medallion is owned by an individual who cannot own more than one license. Regulations require drivers with independent medallions to complete a minimum of 180 nine-hour shifts per year. In 2013, independent medallions were sold at prices between $800,000 and $1,050,000 and fleet medallions were sold at prices between $1,000,000 and $1,320,000.

The taxis operate in three distinct business models including Fleet, Driver Owned Vehicle (DOV), and Owner-Driver. First, under the fleet operation model, a corporation owns both the medallion and the vehicle to which it is affixed. Drivers operating in fleet lease both vehicle and the medallion, and work for 9-hour shifts regulated by the commission. Fleets operate about one third of taxis. Second, under the DOV model, the driver owns the taxi but not the medallion thus leases the medallion from an owner or agent to be able to operate his vehicle as a medallion taxi. About one third of taxis are operated as DOV's. Third, under the owner-driver model, the individual owns both the medallion license and the vehicle to which it is affixed. Regulations require an individual owning an independent medallion to complete a minimum of 180 nine-hour shifts per year. Owner-drivers operate about one third of taxis.

TLC regulation limits the number of shifts performed in a day ut does not control exactly when shift change will take place. Many fleet garage shift changes occur during 5AM-6AM and 5PM-6PM. The morning shift change tends to be less rushed, possibly because there is less demand for taxis in these early morning hours. The afternoon shift changes are more rushed than morning shift changes because they occur during a period of peak demand. However, this rush does not translate into very short times between shifts. This is probably due to rush hour traffic and the time drivers need to travel from their final fares to their shift change locations and from their shift-change locations to fare-rich areas of the city.

There are several ways drivers and fleets manage the transition between shifts. Some fleets require drivers to return their vehicles to their garages at the end of each shift to settle up finances and hand the vehicle over to the driver working the next shift. Other garages allow drivers to perform off-site shift changes, handing over the vehicle in person at a gas station or at a location near the drivers homes. Sometimes drivers park the taxi at a designated swap location. At this location, the next driver picks the vehicle up whenever he is ready to begin his shift, eliminating the in-person rendezvous and saving time for both drivers.

In December 2008, TLC initiated the Taxi Customer Enhancement Program, which enabled the organization to record detailed trip, fare and GPS information of taxi trips. The TLC commission follows a two-part tariff price structure which consists of fixed and distance based fare ($2.50 per ride plus $2.50 per mile or $0.5 per 60 seconds in slow traffic or when the vehicle is stopped).

For all the trips that end in New York city, there is a $0.5 state surcharge; there is also a daily $0.5 surcharge for 8PM to 6AM trips and $ 1 surcharge from 4PM to 8PM on weekdays, excluding holidays. Customers pay for bridge and tunnel tolls. For trips between JFK airport and any location in Manhattan, TLC follows a flat fare of $52 plus tolls, the $0.5 state surcharge, and $4.50 rush hour surcharge (4PM to 8PM weekdays, excluding legal holidays).

4.4 Literature

My work builds on the prior literature in customer demand estimation and revenue improving recommender systems in taxi services. I discuss prior research in each of these areas and describe the contribution of my paper.

First, most of the prior work in customer demand models utilized aggregate models without considering the spatial and temporal structure of the market (Schroeter [1983], Frankena and Pautler [1986], Arnott [1996], Yang et al. [2002]). Yang et al. [2005] extended the aggregated models by relaxing the time dimension under endogenous service intensity. For each period, the authors specified a parametric customer demand function as a function of fare or price of a taxi ride, and customer waiting time. Although the demand function is easy to interpret, it did not incorporate randomness in waiting time and fare. Previous work of taxi services is built on similar demand models under spatial aggregation. Buchholz [2015] built a time and location specific matching model to estimate customer demand. The matching function specification was parametric and allowed for randomness in customer arrivals and taxi arrivals. I build a stochastic model based on a double-ended queue and do not assume any pre-specified matching function. The queue allows randomness in customer and taxi arrivals and their location changes, which has not been studied in the previous literature. Also, I propose a technique to estimate customer relocation rate using a sample path approach from the real-time taxi trip data.

4.5 Data & Summary Statistics

I obtained taxi trip and fare data for yellow taxis in the New York City for the month of June 2013. The data is publicly available and released by New York City TLC through Freedom of Information Law request (FOIL).

The trip and fare data collected for Jun 2013 contains a total of 14.4 million records.

Each record in the data provides granular details of the trip and fare information. The trip data contains information on medallion id (equivalent to taxi id), hack license (driver id), date timestamp of pickup, drop-off, geographical coordinates (latitude, longitude) of exact pickup and drop-off location, trip distance (miles) and total number of customers driven on the trip. The fare data includes information on nominal fare, surcharge, MTA tax, tip and toll with all the units in dollars ($). I excluded any erroneous records present in the data. The abnormal records broadly comprise trips with customer count above 10, trip duration values either negative or more than 5 hours, speed of the trip more than 150 miles per hour, or geographical coordinates beyond the NYC land area. All the erroneous records in total combine to 2.72% of the data, leaving a cleaned 13.9 million records.

An overview of the cleaned data is as follows. I observed 13,360 unique taxi ids and 32,999 drivers. The total trip distance is approximately 41 million miles and the total trip duration is 3 million hours. The total revenue earned by the drivers is $209 million, which can be broken down into the nominal fare (175 million, 83.7%), surcharge (4.5 million, 2.2%), MTA tax (6.9 million, 3.3%), tip (19 million, 9.1%) and tolls (3.6 million, 1.7%). Figure 4.4 provides a matrix plot of trip distribution between and within the five boroughs. I find over 80% trips are realized within Manhattan, with the remainder distributed between Queens (primarily La Guardia and JFK airports) and Manhattan.

My granular data comprises driver id information for each record, which allows me to track the successive trips and duration of a driver shift. Using each individual record and driver shift information, I summarize my data. Table 4.1 provides summary statistics for the records in the data.

Table 4.1: Summary Statistics of key variables

Variable	Mean	Stdev	Min	P25	Median	P75	Max
Customer count	1.73	1.39	1.00	1.00	1.00	2.00	8.00
Trip distance (miles)	2.95	3.36	0.01	1.10	1.80	3.30	35.00
Trip duration (secs)	786.95	587.46	60.00	420.00	660.00	1020.00	10800.00
Speed (miles/hour)	12.89	6.91	0.01	8.40	11.60	15.75	81.30
Trip fare ($)	14.96	11.75	3.00	8.00	11.40	16.87	500.00

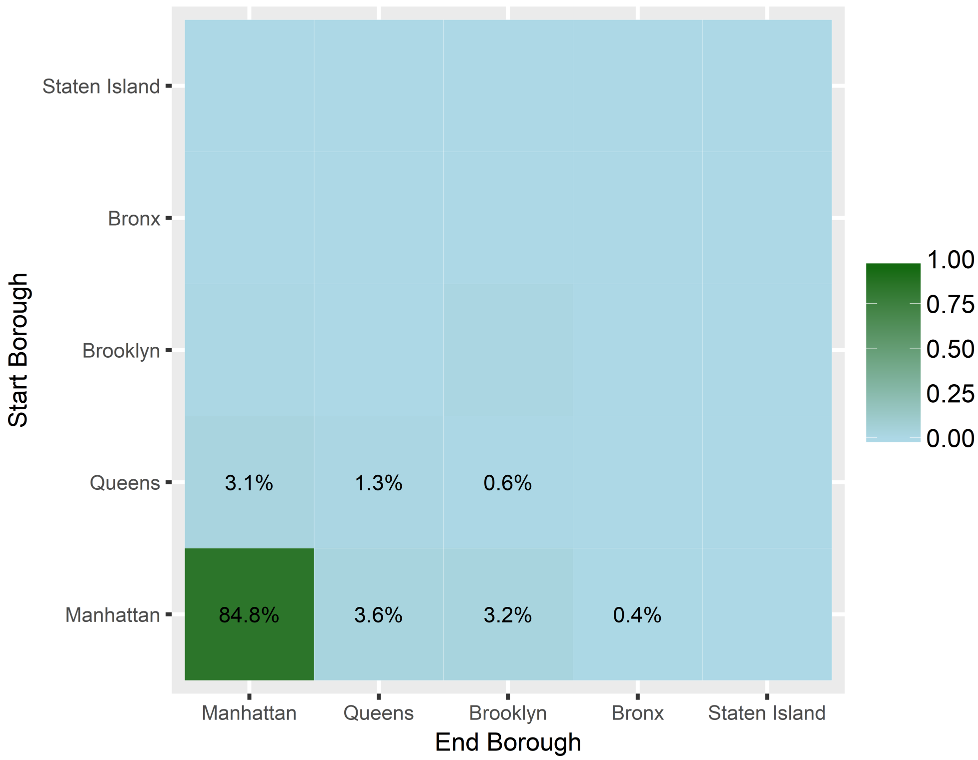


Figure 4.4: Trip distribution across five boroughs

I find median number of customers for a trip is 1, and the median trip distance, duration and speed are 1.8 miles, 660 seconds and 11.6 miles per hour, respectively. The median fare realized for a trip is \$11.4. Table 4.2 provides the summary of each drivers shifts for the month.

Table 4.2: Summary Statistics of driver information by shift and month

Level	Variable	Mean	Stdev	Min	P25	Median	P75	Max
Shift	Trips	20.72	9.38	1	15	21	27	62
	Revenue (\$)	309.89	122.95	3.00	238.61	311.87	380.51	886.47
	Shift (hours)	8.28	2.68	0.017	6.983	8.7	10.067	14.85
Month	Trips	424.10	200.73	1	284	441	568	1077
	Revenue (\$)	6,343.79	2,807.37	3.50	4,583.00	6,650.08	8,299.15	15,957.40

I find the median number of trips a driver takes during each shift is 21 and the revenue earned is \$311.87. The median shift duration is 8.7 hours, which is close to 9 hours as mandated by TLC. I find significant variation in the number of trips taken by a driver for the month with median of 441 and the median revenue earned is \$6,650. Interestingly I find one driver earned \$15,957 in the month.

I now summarize the spatial and temporal variation of trips.

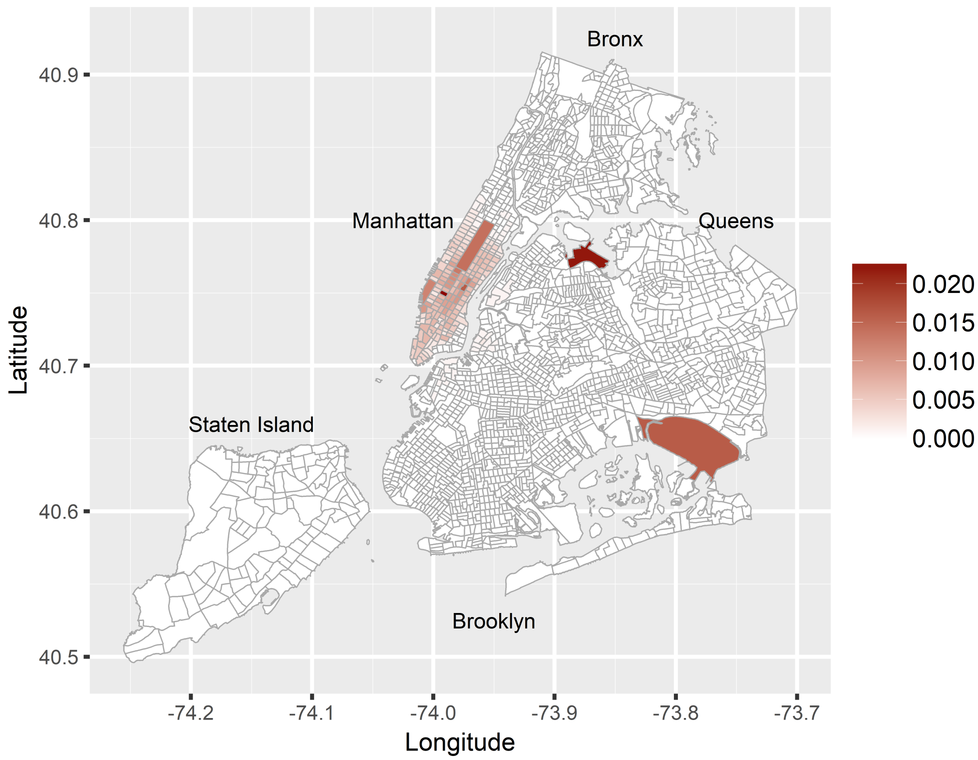


Figure 4.5: Spatial variation of trips in New York city

Figure 4.5 shows a heat map of trips across New York City for each borough. More than 92% of trips start in Manhattan, 85% of which end in Manhattan and the remainder end in Queens. Among the trips from Queens, 75% of them start either from JFK or La Gaurdia airport.

Figure 4.6 shows a heat map of trips across the neighborhoods in Manhattan. The following neighborhoods contribute to 50% of trips in Manhattan: Midtown-Midtown South, Hudson Yards-Chelsea-Flatiron-Union Square, West Village, Turtle Bay-East Midtown, and Upper East Side-Carnegie Hill. Second, Figure 4.7 displays inter-day variation of the trips. I observe that the largest number of trips occurs on Thursday; in descending order, the remaining days are Friday, Saturday, Wednesday, Tuesday and Monday, with the smallest

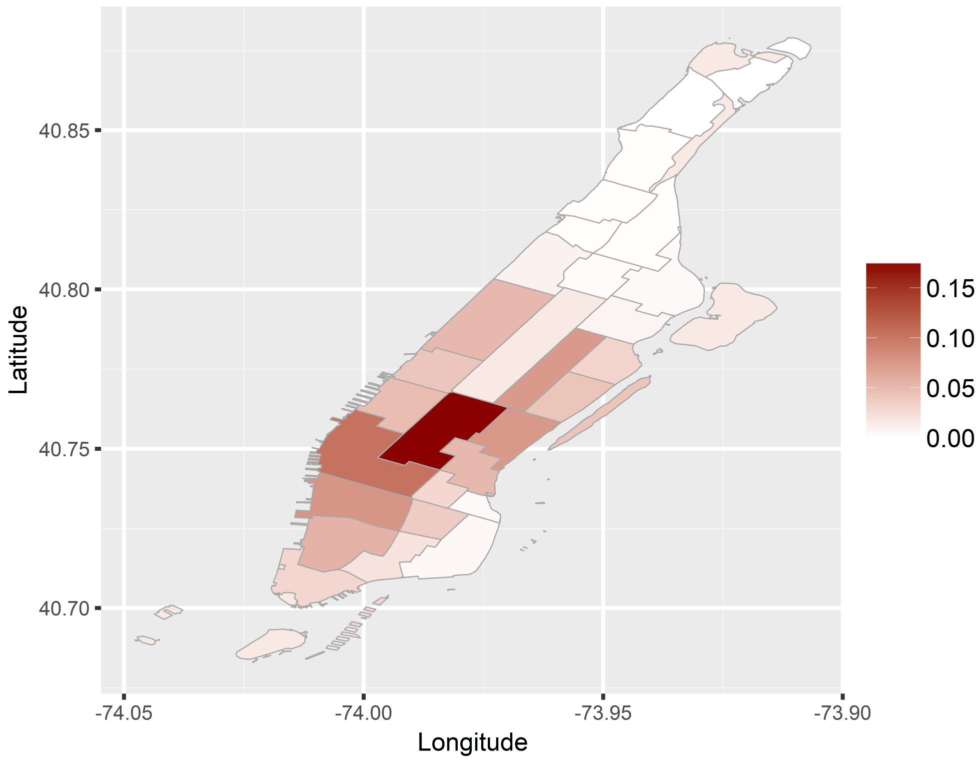


Figure 4.6: Spatial variation of trips in Manhattan

number occurring on Sunday.

I now analyze intra-day variation of trips. Figure 4.8 displays variation of trips by hour. I find distinct pattern for weekdays versus weekends. On weekdays, I find demand is low during 12AM - 5AM. Trip demand rises quickly after 5AM as most of the driver shifts start at the same time period. The peak demand is observed during 8AM-9AM. As discussed previously, a significant drop in number of trips at 5PM occurs due to the unavailability of significant numbers of drivers across the geographical locations, some of whom are ending their shift and others have not yet begun their shift.

4.6 Stochastic Matching Model

I now develop a stochastic matching model between taxis, drivers and customers in location and time using a double ended queue under relocating behavior. I consider the

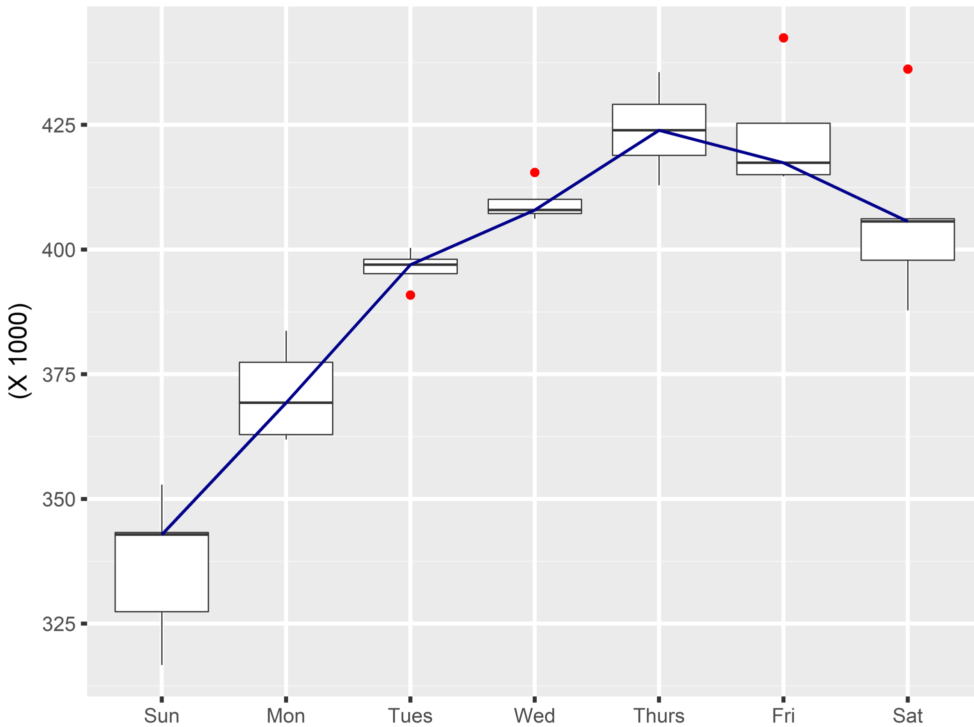


Figure 4.7: Inter-day trips by day of the week

following two assumptions: (i) the arrival rates of taxis and customers arrival rates follow a Poisson process; and, (ii) both leave their respective queues after waiting for an individual exponential amount of time. A schematic representation of the double ended queue for a given location and time is provided in Figure 4.9.

The left side of the queue is for taxis, while the right is for customers. The parameters λ_1, μ_1 represent taxi arrival and relocation rate; λ_2, μ_2 represent customer arrival and relocation rate. Under the stated assumptions and parameters, the dynamics in a double ended queue can have either taxis or customers or neither present in the system. Hence, I define a stochastic process $X(t)$ which represents the number of people in the system. If $X(t) > 0$ (< 0), taxis (customers) are waiting in the queue, otherwise the queue is empty. The generator or state transition diagram of the queue is given in Figure 4.10.

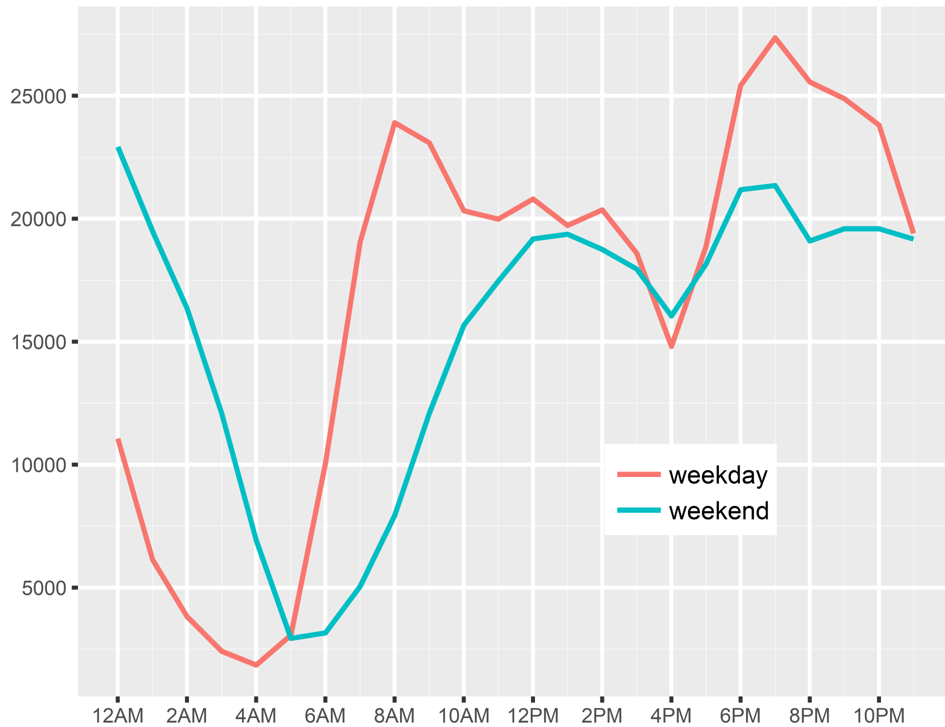


Figure 4.8: Intra-day trips by time slot (weekdays & weekends)

The steady state distribution of states is given by

$$\pi_i = \begin{cases} \frac{\lambda_1^i}{\prod_{j=1}^{i}(\lambda_2+j\mu_1)} \cdot \pi_0 & \text{if } i > 0 \\ \frac{\lambda_2^i}{\prod_{j=1}^{i}(\lambda_1+j\mu_2)} \cdot \pi_0 & \text{if } i < 0 \\ \frac{1}{\sum_{i=1}^{\infty}\left(\prod_{j=1}^{i}(\lambda_2+j\mu_1)+\prod_{j=1}^{i}(\lambda_1+j\mu_2)\right)} & \text{if } i = 0 \end{cases}$$

To estimate the set of parameters for a specified location and time, I need to know the distribution of matches. However, due to the unavailability of closed form expression for the number of matches, I resort to a sample path technique to estimate the underlying parameters. In Figure 4.11, I provide an example sample path in the event transition of number of taxi drivers for a given location and time.

Let $X(0)$ be number of taxi drivers available at time 0 ready to accept a random customer

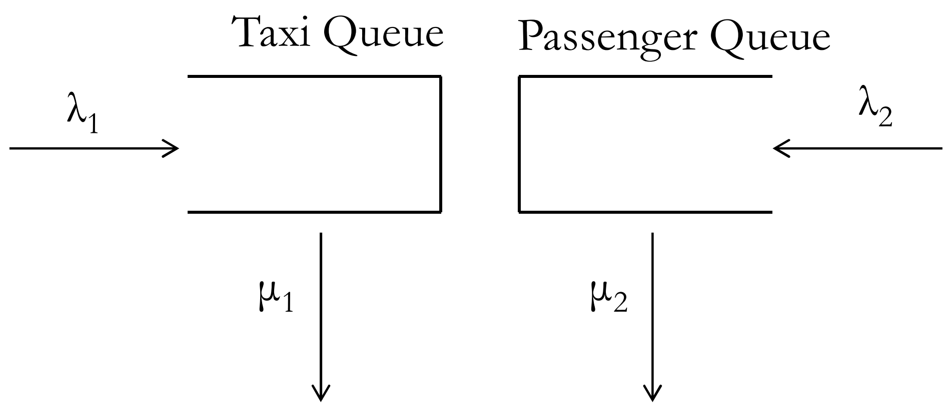


Figure 4.9: Taxi, customer matches

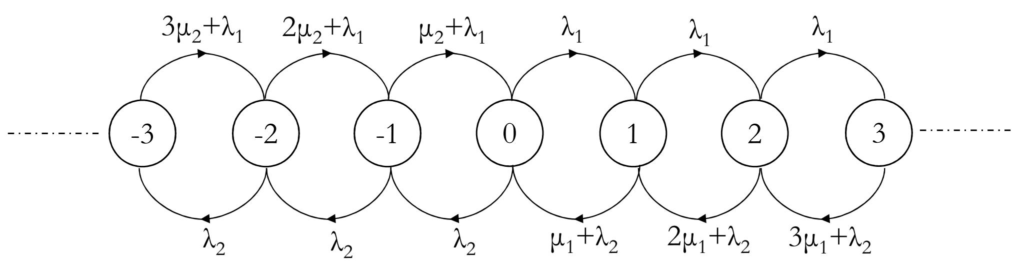


Figure 4.10: State transition in double ended queue

request. At time instant τ_1, a taxi relocation occurs, reducing the count of taxi drivers by 1. At time instant τ_2, an available taxi picks a customer, reducing the count of taxi drivers by 1. Similarly, at time instant τ_3, a new taxi driver finished a trip and is available for the next pickup, increasing the count by 1. The remaining events can be interpreted similarly. Under the assumption of event transition independence, I can build a likelihood function for the example sample path.

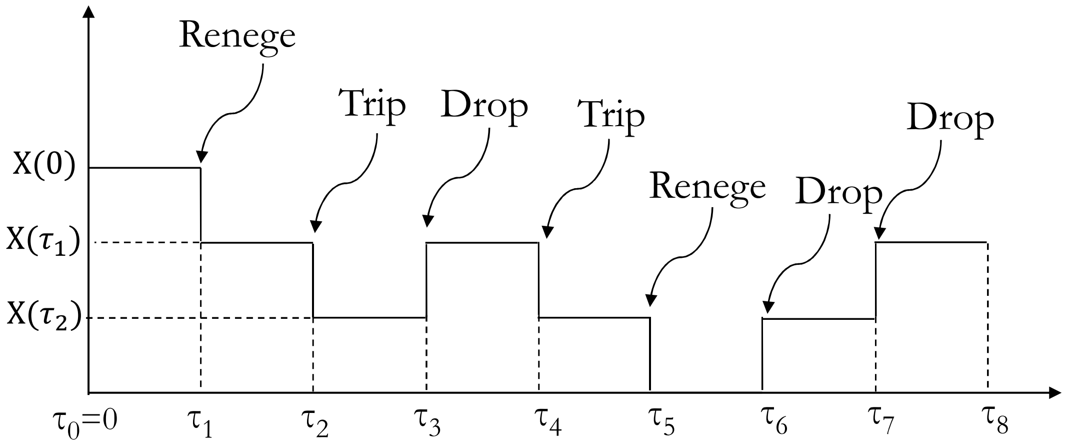


Figure 4.11: Example sample path

$$f_1(\tau_1 - \tau_0) = \Big(X(\tau_0)\mu_1 + \lambda_2 + \lambda_1\Big) \cdot e^{-\{X(\tau_0)\mu_1+\lambda_2+\lambda_1\}\cdot\tau_1} \cdot \frac{X(\tau_0)\mu_1}{X(\tau_0)\mu_1 + \lambda_2 + \lambda_1}$$

$$f_2(\tau_2 - \tau_1) = \lambda_2 \cdot e^{-\{X(\tau_1)\mu_1+\lambda_2+\lambda_1\}\cdot(\tau_2-\tau_1)} \qquad X(\tau_1) = X(\tau_0) - 1$$

$$f_2(\tau_3 - \tau_2) = \lambda_1 \cdot e^{-\{X(\tau_2)\mu_1+\lambda_2+\lambda_1\}\cdot(\tau_3-\tau_3)} \qquad X(\tau_2) = X(\tau_1) - 1$$

$$\vdots$$

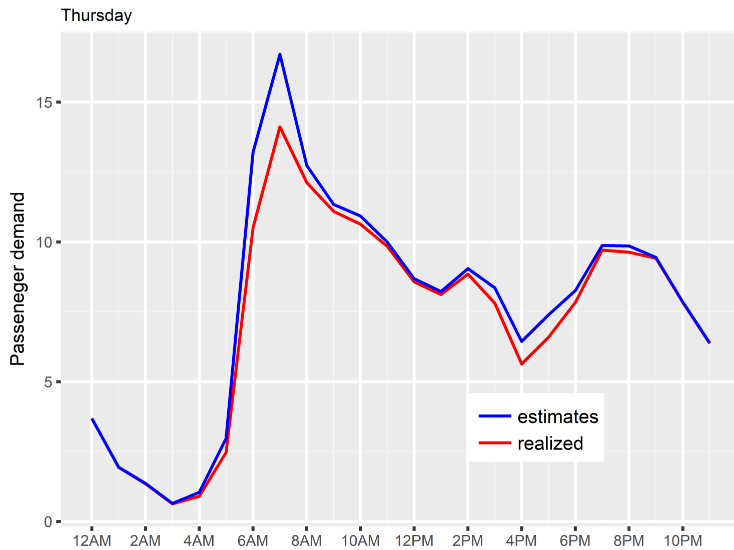
$$\vdots$$

The likelihood and optimization formulation to estimate the parameters is given by

$$\max_{\lambda 1,\, \mu_1,\, \lambda_2} \prod_{i\geq 1} f_i(\tau_i - \tau_{i-1})$$

I differentiate the parameters for each location, day type (weekday or weekend) and hour of the day. The entire geographical area where the taxi services are spread comprises 2,164 tracks. Hence, I ran the estimation for 2164 * 2 * 24 times.

Figure 4.12 display the estimated passenger demand rate (λ_2) during weekday and weekend for the busiest tract (Midtown) in Manhattan. The difference in estimated and actual



values is due to the parameter for taxi driver relocation included in the model. I find the difference is significant during the peak demand period and shift change period.

4.7 Conclusion

Improving taxi services efficiency is an important problem as it affects drivers income, passenger service and transportation revenue. Taxi regulations such as the limit on issued licenses, the two-part tariff pricing policy and 24 hour availability of services can often lead to a spatial and temporal imbalance of supply and demand in the service area. Drivers service intensity is endogenous as they choose location as well as the time to start and end their shift. An absence of centralized control under regulation and endogenous service intensity can lead to a sizable number of drivers starting or ending their shift during the same time period. Increased delay in shift change time when drivers share a common resource (taxi) can lead to low availability affecting service level and revenue. In this chapter, I studied passenger demand estimation and taxi driver optimal shift timing problem. I first developed a stochastic matching model (double-ended queue) to predict passenger demand in location and time. The stochastic model allows for non-stationarity, and randomness in arrivals and

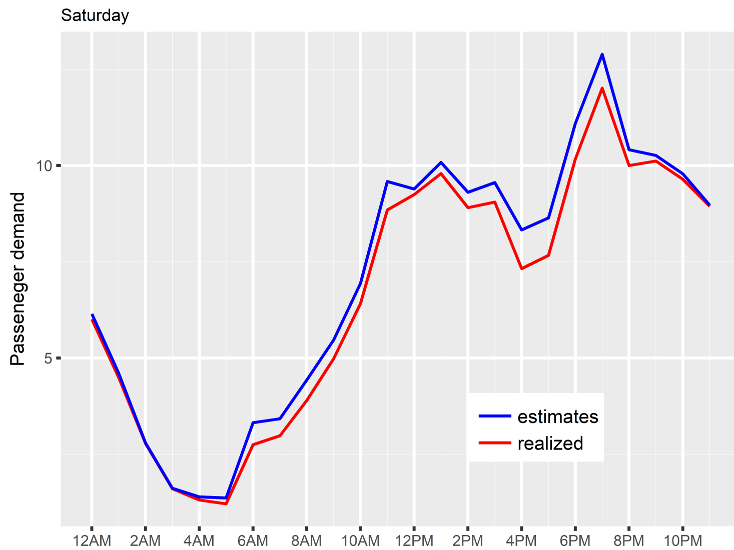


Figure 4.12: Estimated and Realized passenger demand rate in Midtown tract

relocations of both drivers and passengers. I estimate potential passenger demand, drivers and passenger relocation rate. The predicted demand is then used to analyze the optimal time for drivers to change their shift to maximize revenue under the current status quo of delay in shift changeover. I find that avoiding the relocating behavior of taxi drivers at each location after finishing a previous trip can lead to under-estimation of passenger demand at each location. In addition, I propose a framework to estimate passenger relocation rate under no availability of taxis in any given location.

4.8 Venues for future Research

My work in this chapter provides a methodological framework for integrating large scale dataset based empirical analysis and stochastic modeling to estimate passenger demand in Street-Hail Taxi services. The stochastic demand model assumes an exogenous specification of the taxi driver relocation rate, however in reality taxi drivers relocating behavior from a location depends on how late they are in the shift. Future work could include a specification of endogenous relocation behavior. Currently the double-ended queue framework of taxis and passengers assumes perfect matching meaning the queue contains a non-negative number of taxis or passengers but not both. In reality, it is possible that in a geographical location a positive number of both taxis and passengers could exist. My demand model could be improved with parametric specification of imperfect matching in the queuing framework.

CHAPTER 5: CONCLUSION

By the end of 2015, the Service industry accounted for 66% of world GDP and over half of the working population. Given its magnitude and share, Services is an extremely important industry to study in the present day and in the future. Improving services at each firm level eventually can add benefits and lead to the growth and creation of more jobs in the economy. Motivated by its importance and impact, I have chosen to study a few operations problems in this industry during my Ph.D. program. Customers and providers are key players in any service transaction. A transaction often involves the transfer or exchange of information between customers and providers. An outcome of a successful transaction is the provision of quality service and high satisfaction to the customer, which is the primary objective of providers.

In my thesis, I examined two types of operational problems. First, I utilized large scale datasets and stochastic models to predict true demand under customer behavior, time and spatial dynamics. After predicting demand, I built models to optimally allocate resources for improved performance. The settings I focused on for this topic are Bike Sharing systems and Street-Hail Taxi services. Second, I developed micro models to determine the drivers of customer provided satisfaction and its impact on operational performance measures.

The key generalized findings from the three chapters in my thesis are as follows. First, I find evidence of non-stationarity and spatial variation in demand. In the Bike Sharing and Taxi services chapters, I find proof of pickups and drop-offs (taxi trips) varying at each station (tract location) and by time of the day. Second, I find evidence of customer behavior in operational systems. In the Bike Sharing chapter, I find statistical evidence of commuter substitution under stock-out; specifically, commuters likelihood of substitution decreases with an increase in the walking distance between stations over the entire operational day. In the Taxi services chapter, I observe drivers relocate from one location to another in the search for a new trip after the previous drop-off. In the e-commerce study, I find statistical evidence of consumers reacting to logistic delivery performance through their ratings of the service. Third, micro models built using data-driven techniques can help predict true customer demand. In Bike Sharing, I find that, under commuter substitution, predicted true demand is different (more or less) from realized demand at each station. Similarly, in Taxi services, I find true passenger demand is different from realized demand under passenger and driver relocating behavior. Lastly, incorporating customer behavior in operational planning is essential to provide better service and improved performance. In Bike Sharing, I show modeling commuter substitution improves demand estimates and subsequently allocation policy, which can improve ridership and service level by 7.60% and 1.69%, respectively.

In this thesis, I conducted rigorous analyses using large datasets to study three different service systems. Similar studies can be executed on other service systems such as call centers and healthcare. The tools were primarily drawn from Stochastic Models, Optimization, Econometrics and Data Visualization. My work can further be extended in building more realistic demand models by incorporating finer details of customer behavior. For example, in Bike Sharing study, the demand model assumes a homogeneous population of customers in the substitution process. The model can further be improved by incorporating different classes of consumers in substitution. For example, one could consider two classes of consumers, each with different levels of sensitivity to walking distance. Similarly, in the Street-Hail Taxi services chapter, my stochastic demand model assumes an exogenous specification of taxi driver relocate rate, however, in reality drivers relocating behavior from a location depends on how late they are in the shift. Future work could include a specification of endogenous relocating behavior. Currently the double-ended queue framework of taxis and passengers assumes perfect matching meaning the queue contains non-negative number of taxis or passengers but not both. In reality, it is possible that in a geographical location a positive number of both taxis and passengers can happen. The model could be further improved with parametric specification of imperfect matching in the queuing framework.

Lastly, in the e-commerce chapter, prior work has shown that customers learn about the product quality of books from the online ratings (Zhao et al. [2013]). Future work on this dataset could include empirically testing whether customers learn about service quality from the logistic rating. Similarly, past work has shown different biases in item quality ratings such as self-presentational behavior (Schlosser [2005]), self-selection bias (Li and Hitt [2008]), purchasing bias, and under-reporting bias (Hu et al. [2009]). Future work could include empirically verifying if these different biases exist in logistic rating.

The rich dataset from the Cainiao network prompts a wealth of more interesting research questions. First, the Cainiao network is an integrated warehouse and logistics organization equipped with big IT and decision support systems to manage a large inventory of items and their operations. Cainiao is likely to manage inventory efficiently such that they ship the item upon arrival of order with minimal delay and less variation. On the other hand, monitoring item inventory is likely to be difficult for independent sellers who manage more unique items. As a result, it is possible that independent sellers ship products late with more variation. Hence it would be interesting to study whether the mean and variability in consignment times is less for Cainiao managed orders compared to seller managed orders. Second, Sellers who transfer most of their warehousing and logistics responsibility to Cainiao have the time, energy, and resources to devote to their primary areas of business or competency. As a result, they are likely to offer more unique products over time. Hence, it would be interesting to study if sellers scale up or offer more unique items when there is an increase in items managed by Cainiao.

CHAPTER 6: APPENDIX

In this chapter, I provide details of proof and additional results of Chapter 2.

6.1 Alternative test for Poisson assumption

In my paper, I use the familiar Kolmogorov-Smirnov one-sample test frequently applied in the call center context by Brown et al. [2005] to verify my assumption that inter-arrival time follows an exponential distribution. To ensure robustness, I conduct my tests using a Chi-Square goodness of fit test for Poisson distribution. I tested pickup and dropoff events individually on observations for each group formed by the combination of station, day type(Monday - Friday), 2 hour slot (6am, 8am, 10am, 12pm, 2pm, 4pm, 6pm & 8pm). For 35 stations, 5 day types and 8 slots, the total number of groups was 1400. Sample tests are provided for pickups at station SFO Caltrain 2-330 Townsend on Tuesday during 10am-12pm and dropoffs at SFO Caltrain-Townsend at 4th on Monday during 2pm-4pm.

H_0 : Pickups/Dropoffs fit Poisson distribution (Null)

H_1 : Pickups/Dropoffs does not fit Poisson distribution (Alternative)

Table 6.1: Goodness-of-fit test for Poisson distribution

	χ^2	df	p-value
Pickups	9.275	10	0.506
Dropoffs	14.589	13	0.334

Among the 1400 groups, we find 78% pickup and 71% dropoff groups follow Poisson process.

6.2 Multinomial Logit model - Robustness

In section 2.5.1, we provided argument for the choice of commuters walking distance to substitute across stations. Though our options are 5 mins, 10 mins, we estimated demand parameters with later choice as it includes larger set of stations for substitution. The tables below provide Multinomial Logit model estimates under 5 mins, additionally with quadratic walking distance.

Table 6.2: Multinomial logit model estimates (5 mins of walking time)

Variable	6am	8am	10am	12pm	2pm	4pm	6pm	8pm
constant	-0.79***	0.3289***	0.633***	-0.0104	0.0205	-0.5589***	-0.0809***	-0.0158
	(0.052)	(0.019)	(0.032)	(0.08)	(0.05)	(0.025)	(0.024)	(0.061)
walk_dist	-0.177*	-1.055***	-0.097**	-0.009	0.002	-0.021	0.067	-0.005
	(0.105)	(0.022)	(0.039)	(0.084)	(0.065)	(0.052)	(0.068)	(0.004)
-logLL	-18822.55	-109287.12	-720.78	751.73	700.25	-69787.14	-24249.11	5913.53

*p<0.1; **p<0.05; ***p<0.01

Table 6.3: Multinomial logit model estimates (5 mins of walking time)

Variable	6am	8am	10am	12pm	2pm	4pm	6pm	8pm
constant	-0.681***	0.278***	0.7	-0.01	0.019	-0.564***	-0.095***	-0.015
	(0.023)	(0.044)	(0.877)	(0.108)	(0.055)	(0.04)	(0.021)	(0.013)
walk_dist	-0.161***	-0.671***	-0.093	-0.011	0.002	-0.027	0.08**	-0.001
	(0.031)	(0.008)	(0.075)	(0.111)	(0.072)	(0.05)	(0.032)	(0.055)
walk_dist^2	-0.045	-0.274***	-0.003	0	0	0.018	0.048	0.002
	(0.05)	(0.019)	(0.301)	(0.088)	(0.091)	(0.027)	(0.031)	(0.105)
-logLL	-18822.48	-109286.61	-720.82	751.73	700.25	-69787.13	-24249.12	5913.53

*p<0.1; **p<0.05; ***p<0.01

Similar to my results in Table 2.6, I find that the sign of the coefficient of distance is negative which indicates an inverse relationship, i.e., consumers are less likely to substitute (less utility) as distance increases. In most of the slots, the coefficient of the quadratic distance is also negative, indicating that the rate of utility drop is higher at higher walking distance. In addition I find the inverse relationship occurs more during "busy" (8am, 10am, 4pm and 6pm slots) than "non busy" periods.

Table 6.4: Multinomial logit model estimates (10 mins of walking time)

Variable	6am	8am	10am	12pm	2pm	4pm	6pm	8pm
constant	-0.252***	2.761***	1.181***	-0.002	0	-0.516	0.66***	-0.016
	(0.019)	(0.013)	(0.025)	(0.039)	(0.017)	(1.37)	(0.01)	(0.059)
walk_dist	-0.361***	-3.818***	-6.024***	-0.016	-0.074**	-1.396	-4.188***	-0.011
	(0.011)	(0.067)	(0.034)	(0.023)	(0.034)	(1.471)	(0.017)	(0.096)
walk_dist^2	-0.169***	-5.65***	-3.286***	-0.008	-0.03	-0.726	-2.712***	-0.003
	(0.048)	(0.009)	(0.041)	(0.034)	(0.043)	(0.732)	(0.019)	(0.015)
-logLL	-18828.07	-109427.35	-722.11	749.82	701.46	-69794.10	-24236.88	5911.23

*p<0.1; **p<0.05; ***p<0.01

6.3 Model Validation

I validate the fit of my stochastic model built under the setting - Demand substitution under stock-outs (Case 3). I choose this setting as it likely represents real world behavior. A sample run of my model validation process is as follows. I separate both the *trip* and *inventory* data into two parts - training and validation. The training data is obtained by randomly selecting 180 (80%) out of 223 busy days. The remaining 43 (20%) busy days is referred as validation data. I estimate the parameter set - intrinsic pickup, dropoff rates $\theta^{\text{training}} = \mu_{it}, \lambda_{it} \mid i \in S$ using my stochastic model and the training data (2.5.2). I next obtain actual and predicted ridership for each of the 43 days in the validation data. The actual ridership is given by calculating number of pickups during 6am-10pm for each of the busy days.

After analyzing the *inventory* data, I found managers were re-balancing bikes during 8am, 10am, 4pm and 6pm slots. I extracted the starting distribution of bikes from *inventory* data for each station during the re-balancing slots. To ensure robustness in the values of the starting distribution, I average the inventory values during first 30 minutes for each of 8am, 10am, 4pm and 6pm slots rather than using the values at exact time stamp. The distribution of each busy day in validation data is applied to the stochastic model with parameter set θ^{training} to obtain predicted ridership. Comparing the actual and predicted ridership for the 43 days, I calculate Mean Absolute Percentage Error (MAPE). To assess the range of MAPE values, I execute the above sample run 25 times. The summary of MAPE for all the 25 runs is as follows. I find MAPE varies from 3.97% to 6.313%.

Table 6.5: Summary of MAPE for the 25 runs

Variable	Min	P25	Median	Mean	P75	Max
MAPE	3.970%	5.103%	5.387%	5.409%	5.886%	6.313%

6.4 Time dependent equations - M/M/1/C queuing system

The forward time dependent equations of $M/M/1/C$ queuing system are given by

$$\frac{dP(t)}{dt} = P^{'}(t) = P(t) \cdot Q$$

where $P(t)$ is time dependent transition probability and Q is generator matrix.

For $n \in \{0, 1, 2, 3, \dots, C\}$, individual equations are given by

$$\frac{dp_0(t)}{dt} = \mu \cdot p_1(t) - \lambda \cdot p_0(t)$$

$$\frac{dp_n(t)}{dt} = \mu \cdot p_{n+1}(t) + \lambda \cdot p_{n-1}(t) - (\lambda + \mu) \cdot p_n(t)$$

$$\frac{dp_C(t)}{dt} = \lambda \cdot p_{C-1}(t) - \mu \cdot p_C(t)$$

Setting $p_n(t) = \rho^{\frac{n}{2}} \cdot B_{n,s} \cdot e^{-k_s t}$ where $k_s = \mu \cdot x_s$ reduces the above 3 equations as

$$\sqrt{\rho} \cdot B_{1,s} + (x_s - \rho) \cdot B_{0,s} = 0$$

$$\sqrt{\rho} \cdot (B_{n+1,s} + B_{n-1,s}) + (x_s - 1 - \rho) \cdot B_{n,s} = 0$$

$$\sqrt{\rho} \cdot B_{C-1,s} - (x_s - \rho) \cdot B_{C,s} = 0$$

Considering $B_{n,s} = \sin(ny) - \sqrt{\rho} \cdot \sin(n+1)y$ will reduce the above three equations to a single solution

$$x_s = \rho + 1 - 2 \cdot \sqrt{\rho} \cdot \cos(y)$$

which is equivalent to $k_s = \lambda + \mu - 2 \cdot \sqrt{\lambda \cdot \mu} \cdot \cos(y)$

If $\sin(C+1)y = 0$, then $y = \frac{s\pi}{C+1}$ where $s = \{1, 2, 3, \ldots C\}$

The solution to forward equations is be given by

$$p_n(t) = \pi_n + \rho^{\frac{n}{2}} \sum_{s=1}^{C} \alpha_s \cdot \left(\sin \frac{sn\pi}{C+1} - \sqrt{\rho} \sin \frac{s(n+1)\pi}{C+1} \right) \cdot e^{-k_s t}$$

$$k_s = \lambda + \mu - 2 \cdot \sqrt{\lambda \cdot \mu} \cdot \cos\left(\frac{s\pi}{C+1}\right)$$

where $s = \{1, 2, 3, \ldots C\}$ and $j = \{0, 1, 2, 3, \ldots C\}$

The coefficients α_s in the above equation can be found by initial state "m". This implies $p_m(0) = 1$. Then the above solutions can be represented as

$$p_{mn}(u) = \pi_n + \frac{2 \cdot \rho^{\frac{n-m}{2}}}{C+1} \cdot \sum_{s=1}^{C} \frac{\mu}{k_s} \cdot K_s \cdot e^{-k_s u}$$

where

$$K_s = \left(\sin \frac{sm\pi}{C+1} - \sqrt{\rho} \cdot \sin \frac{s(m+1)\pi}{C+1} \right) \cdot \left(\sin \frac{sn\pi}{C+1} - \sqrt{\rho} \cdot \sin \frac{s(n+1)\pi}{C+1} \right)$$

$$k_s = \lambda + \mu - 2 \cdot \sqrt{\lambda \cdot \mu} \cdot \cos\left(\frac{s\pi}{C+1}\right)$$

$$\rho = \frac{\lambda}{\mu}$$

$$\pi_n = \begin{cases} \frac{1}{C+1} & \text{if } \rho = 1 \\ \frac{1-\rho}{1-\rho^{C+1}} \cdot \rho^n & \textit{otherwise} \end{cases}$$

BIBLIOGRAPHY

D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar. Fast and slow learning from reviews. Technical report, National Bureau of Economic Research, 2017.

Z. Aksin, B. Ata, S. M. Emadi, and C.-L. Su. Structural estimation of callers' delay sensitivity in call centers. *Management Science*, 59(12):2727–2746, 2013.

E. W. Anderson and M. W. Sullivan. The antecedents and consequences of customer satisfaction for firms. *Marketing science*, 12(2):125–143, 1993.

R. E. Anderson. Consumer dissatisfaction: The effect of disconfirmed expectancy on perceived product performance. *Journal of marketing research*, pages 38–44, 1973.

R. Anupindi, M. Dada, and S. Gupta. Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Science*, 17(4): 406–423, 1998.

R. Arnott. Taxi travel should be subsidized. *Journal of Urban Economics*, 40(3):316–333, 1996.

S. T. Berry. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262, 1994.

R. N. Bolton. A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction. *Marketing science*, 17(1):45–65, 1998.

A. Brennan. Ibisworld industry report 48533 taxi & limousine services in the us. Technical report, Tech. rep., IBIS World, 2014.

L. Brown, N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, and L. Zhao. Statistical analysis of a telephone call center: A queueing-science perspective. *Journal of the American statistical association*, 100(469):36–50, 2005.

P. Brown, M. Broomfield, G. Buys, L. Cardenas, E. Kilroy, J. MacCarthy, T. Murrells, Y. Pang, N. Passant, J. Ramirez Garcia, et al. *UK greenhouse gas inventory, 1990 to 2014: annual report for submission under the Framework Convention on Climate Change*. Ricardo Energy & Environment, 2016.

N. Buchholz. Spatial equilibrium, search frictions and efficient regulation in the taxi industry. Technical report, Technical report, University of Texas at Austin, 2015.

Y. Chen and J. Xie. Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management science*, 54(3):477–491, 2008.

Y. Chen, F. M. Harper, J. Konstan, and S. X. Li. Social comparisons and contributions to online communities: A field experiment on movielens. *American Economic Review*, 100 (4):1358–98, 2010.

Z. Chen and A. J. Dubinsky. A conceptual model of perceived customer value in e-commerce: A preliminary investigation. *Psychology & Marketing*, 20(4):323–347, 2003.

J. A. Chevalier and D. Mayzlin. The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, 43(3):345–354, 2006.

P. K. Chintagunta, S. Gopinath, and S. Venkataraman. The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science*, 29(5):944–957, 2010.

G. A. Churchill Jr and C. Surprenant. An investigation into the determinants of customer satisfaction. *Journal of marketing research*, pages 491–504, 1982.

M. C. Cohen. Big data and service operations. *Production and Operations Management*, 2017.

C. T. Conlon and J. H. Mortimer. Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics*, 5(4):1–30, 2013.

G. Cui, H.-K. Lui, and X. Guo. The effect of online consumer reviews on new product sales. *International Journal of Electronic Commerce*, 17(1):39–58, 2012.

R. Cui, M. Li, and Q. Li. Value of high-quality logistics: Evidence from a clash between sf express and alibaba. *SSRN*, 2018.

K. Cukier. *Data, data everywhere: A special report on managing information*. Economist Newspaper, 2010.

C. Dellarocas. Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management science*, 52(10):1577–1593, 2006.

L. Dell'Olio, A. Ibeas, and J. L. Moura. Implementing bike-sharing systems. In *Proceedings of the Institution of Civil Engineers-Municipal Engineer*, volume 164, pages 89–101. Thomas Telford Ltd, 2011.

W. B. Dodds, K. B. Monroe, and D. Grewal. Effects of price, brand, and store information on buyers' product evaluations. *Journal of marketing research*, pages 307–319, 1991.

A. Faghih-Imani, R. C. Hampshire, L. Marla, and N. Eluru. An empirical analysis of bike sharing usage and rebalancing: Evidence from barcelona and seville. *Available at SSRN 2657197*, 2015.

C. Fornell, M. D. Johnson, E. W. Anderson, J. Cha, and B. E. Bryant. The american customer satisfaction index: nature, purpose, and findings. *the Journal of Marketing*, pages 7–18, 1996.

M. W. Frankena and P. A. Pautler. Taxicab regulation: an economic analysis. *Research in Law and Economics*, 9:129–165, 1986.

N. Gans, G. Koole, and A. Mandelbaum. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management*, 5(2):79–141, 2003.

J. C. García-Palomares, J. Gutiérrez, and M. Latorre. Optimizing the location of stations in bike-sharing programs: a gis approach. *Applied Geography*, 35(1):235–246, 2012.

W. H. Greene. *Econometric analysis*. Pearson Education India, 2003.

S. G. Henderson, E. O. Mahony, and D. B. Shmoys. (citi)bike sharing. Submitted, 2016.

Y.-C. Ho, J. Wu, and Y. Tan. Disconfirmation effect on online rating behavior: A structural model. *Information Systems Research*, 28(3):626–642, 2017.

R. M. Hogarth and H. J. Einhorn. Order effects in belief updating: The belief-adjustment model. *Cognitive psychology*, 24(1):1–55, 1992.

D. Honhon, V. Gaur, and S. Seshadri. Assortment planning and inventory decisions under stockout-based substitution. *Operations research*, 58(5):1364–1379, 2010.

N. Hu, P. A. Pavlou, and J. Zhang. Can online reviews reveal a product's true quality?: empirical findings and analytical modeling of online word-of-mouth communication. In *Proceedings of the 7th ACM conference on Electronic commerce*, pages 324–330. ACM, 2006.

N. Hu, J. Zhang, and P. A. Pavlou. Overcoming the j-shaped distribution of product reviews. *Communications of the ACM*, 52(10):144–147, 2009.

A. Kabra, E. Belavina, and K. Girotra. Bike-share systems: Accessibility and availability. 2016.

A. Kaltenbrunner, R. Meza, J. Grivolla, J. Codina, and R. Banchs. Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing*, 6(4):455–466, 2010.

N. S. Koh, N. Hu, and E. K. Clemons. Do online reviews reflect a products true perceived quality? an investigation of online movie reviews across cultures. *Electronic Commerce Research and Applications*, 9(5):374–385, 2010.

A. G. Kök and M. L. Fisher. Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research*, 55(6):1001–1021, 2007.

A. G. Kök, M. L. Fisher, and R. Vaidyanathan. Assortment planning: Review of literature and industry practice. In *Retail supply chain management*, pages 175–236. Springer, 2015.

P. Kotler. *Marketing management, analysis, planning, implementation, and control, Philip*

Kotler. London: Prentice-Hall International, 1994.

P. Kumar, M. U. Kalwani, and M. Dada. The impact of waiting time guarantees on customers' waiting experiences. *Marketing science*, 16(4):295–314, 1997.

N. Lathia, S. Ahmed, and L. Capra. Measuring the impact of opening the london shared bicycle scheme to casual users. *Transportation research part C: emerging technologies*, 22: 88–102, 2012.

Y.-J. Lee, K. Hosanagar, and Y. Tan. Do i follow my friends or the crowd? information cascades in online movie ratings. *Management Science*, 61(9):2241–2258, 2015.

L. I. Li and E. Xiao. Money talks? an experimental study of rebate in reputation system design. 2010.

X. Li and L. M. Hitt. Self-selection and information role of online product reviews. *Information Systems Research*, 19(4):456–474, 2008.

J.-R. Lin and T.-H. Yang. Strategic design of public bicycle sharing systems with service level constraints. *Transportation research part E: logistics and transportation review*, 47 (2):284–294, 2011.

Z. Liu, X. Jia, and W. Cheng. Solving the last mile problem: Ensure the success of public bicycle system in beijing. *Procedia-Social and Behavioral Sciences*, 43:73–78, 2012.

S. F. Lu and H. Rui. Can we trust online physician ratings? evidence from cardiac surgeons in florida. *Management Science*, 2017.

B. T. Ma, C. Liu, and S. Erdoğan. Bicycle sharing and transit: Does capital bike-share affect metrorail ridership in washington, dc? In *83rd Annual Meeting of the Transportation Research Board*, 2015.

S. Mahajan and G. Van Ryzin. Stocking retail assortments under dynamic consumer substitution. *Operations Research*, 49(3):334–351, 2001.

E. Martin, A. Cohen, J. L. Botha, and S. Shaheen. Bikesharing and bicycle safety. 2016.

W. W. Moe and D. A. Schweidel. Online product opinions: Incidence, evaluation, and evolution. *Marketing Science*, 31(3):372–386, 2012.

W. W. Moe and M. Trusov. The value of social dynamics in online product ratings forums. *Journal of Marketing Research*, 48(3):444–456, 2011.

P. M. Morse. *Queues, inventories and maintenance: the analysis of operational systems with variable demand and supply*. Courier Corporation, 2004.

A. Musalem, M. Olivares, E. T. Bradlow, C. Terwiesch, and D. Corsten. Structural estimation of the effect of out-of-stocks. *Management Science*, 56(7):1180–1197, 2010.

R. Nair and E. Miller-Hooks. Fleet management for vehicle sharing operations. *Transportation Science*, 45(4):524–540, 2011.

R. Nair, E. Miller-Hooks, R. C. Hampshire, and A. Bušić. Large-scale vehicle sharing systems: analysis of vélib'. *International Journal of Sustainable Transportation*, 7(1):85–106, 2013.

R. Oliver. Effects of expectations and disconfirmation on postexposure product evaluations. *Journal of Applied Psychology*, 62:246–50, 1977.

R. L. Oliver. A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of marketing research*, pages 460–469, 1980.

R. L. Oliver and W. O. Bearden. Disconfirmation processes and consumer evaluations in product usage. *Journal of Business Research*, 13(3):235–246, 1985.

R. W. Olshavsky and J. A. Miller. Consumer expectations, product performance, and perceived product quality. *Journal of marketing research*, pages 19–21, 1972.

J. C. Olson and P. A. Dover. Disconfirmation of consumer expectations through product trial. *Journal of Applied psychology*, 64(2):179, 1979.

S. O'Sullivan and J. Morrall. Walking distances to and from light-rail transit stations. *Transportation research record: journal of the transportation research board*, (1538):19–26, 1996.

A. Parasuraman, V. A. Zeithaml, and L. L. Berry. Servqual: A multiple-item scale for measuring consumer perc. *Journal of retailing*, 64(1):12, 1988.

O. Perdikaki, S. Kesavan, and J. M. Swaminathan. Effect of traffic on sales and conversion rates of retail stores. *Manufacturing & Service Operations Management*, 14(1):145–162, 2012.

A. R. Rao and K. B. Monroe. The effect of price, brand name, and store name on buyers' perceptions of product quality: An integrative review. *Journal of marketing Research*, pages 351–357, 1989.

T. Raviv and O. Kolka. Optimal inventory management of a bike-sharing station. *IIE Transactions*, 45(10):1077–1093, 2013.

T. Raviv, M. Tzur, and I. A. Forma. Static repositioning in a bike-sharing system: models and solution approaches. *EURO Journal on Transportation and Logistics*, 2(3):187–229, 2013.

A. E. Schlosser. Posting versus lurking: Communicating in a multiple audience context. *Journal of Consumer Research*, 32(2):260–265, 2005.

J. R. Schroeter. A model of taxi service under fare structure and fleet size regulation. *The Bell Journal of Economics*, pages 81–96, 1983.

J. Schuijbroek, R. Hampshire, and W.-J. van Hoeve. Inventory rebalancing and vehicle routing in bike sharing systems. 2013.

S. Shaheen and S. Guzman. Worldwide bikesharing. *Access Magazine*, 1(39), 2011.

M. Sherif and C. I. Hovland. Social judgment: Assimilation and contrast effects in communication and attitude change. 1961.

J. Shu, M. C. Chou, Q. Liu, C.-P. Teo, and I.-L. Wang. Models for effective deployment and redistribution of bicycles within public bicycle-sharing systems. *Operations Research*, 61 (6):1346–1359, 2013.

D. Singhvi, S. Singhvi, P. I. Frazier, S. G. Henderson, E. OMahony, D. B. Shmoys, and D. B. Woodard. Predicting bike usage for new york citys bike sharing system. In *AAAI 2015 Workshop on Computational Sustainability*. Citeseer, 2015.

S. A. Smith and N. Agrawal. Management of multi-item retail inventory systems with demand substitution. *Operations Research*, 48(1):50–64, 2000.

G. van Ryzin and G. Vulcano. A market discovery algorithm to estimate a general class of nonparametric choice models. *Management Science*, 61(2):281–300, 2014.

G. Vulcano, G. Van Ryzin, and R. Ratliff. Estimating primary demand for substitutable products from sales transaction data. *Operations Research*, 60(2):313–334, 2012.

Z. Wang. Anonymity, social image, and the competition for volunteers: a case study of the online market for reviews. *The BE Journal of Economic Analysis & Policy*, 10(1), 2010.

J. Woodcock, M. Tainio, J. Cheshire, O. OBrien, and A. Goodman. Health effects of the london bicycle sharing system: health impact modelling study. *Bmj*, 348:g425, 2014.

F. Wu and B. A. Huberman. How public opinion forms. In *International Workshop on Internet and Network Economics*, pages 334–341. Springer, 2008.

H. Yang, M. Ye, W. H. Tang, and S. C. Wong. Modeling urben taxi services: A literature survey and an analytical example. In *Advanced Modeling for Transit Operations and Service Planning*, pages 256–286. Emerald Group Publishing Limited, 2002.

H. Yang, M. Ye, W. H.-C. Tang, and S. C. Wong. A multiperiod dynamic model of taxi services with endogenous service intensity. *Operations research*, 53(3):501–515, 2005.

Y. Yi. A critical review of consumer satisfaction. *Review of marketing*, 4(1):68–123, 1990.

V. A. Zeithaml. Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. *The Journal of marketing*, pages 2–22, 1988.

F. Zhao, L.-F. Chow, M.-T. Li, I. Ubaka, and A. Gan. Forecasting transit walk accessibility: regression model alternative to buffer method. *Transportation Research Record: Journal* *of the Transportation Research Board*, (1835):34–41, 2003.

Y. Zhao, S. Yang, V. Narayan, and Y. Zhao. Modeling consumer learning from online product reviews. *Marketing Science*, 32(1):153–169, 2013.

F. Zheng, P. He, E. Belavina, and K. Girotra. Customer preference and station network in the london bike share system. *working paper*, 2018.