8-10-2018

Understanding Shared Demand: A Quantitative Investigation into the Changing Modal Landscape of New York City

Raymond Gerte
raymond.gerte@gmail.com

Recommended Citation
https://opencommons.uconn.edu/gs_theses/1265

This work is brought to you for free and open access by the University of Connecticut Graduate School at OpenCommons@UConn. It has been accepted for inclusion in Master's Theses by an authorized administrator of OpenCommons@UConn. For more information, please contact opencommons@uconn.edu.
Understanding Shared Demand: A Quantitative Investigation into the Changing Modal Landscape of New York City

Raymond J Gerte

B.S., University of Connecticut, 2016

A Thesis

Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science

At the University of Connecticut

2018
Copyright by
Raymond J Gerte

Note:

2018
Masters of Science Thesis
Understanding Shared Demand: A Quantitative Investigation into the Changing Modal Landscape of New York City

Presented by
Raymond J Gerte, B.S.

Major Advisor________________________________
Karthik C. Konduri

Associate Advisor________________________________
Nicholas E. Lownes

Associate Advisor________________________________
Nalini Ravishanker

University of Connecticut
2018
ACKNOWLEDGMENTS

Before presentation of the analysis or findings completed in support of this thesis, I would like to take a moment to thank those that have helped me arrive at this point.

First and foremost I would like to thank my parents, Amy and Ray Gerte, and my grandparents, Patricia and John Cervone, for always encouraging my curiosity and desire to learn. Without that thirst for knowledge I do not think I would have made it this far in my academic journey.

I would also like to thank my advisor and academic mentor, Dr. Karthik Konduri for his guidance and support throughout my graduate tenure. From when I started in the summer of 2016, you always encouraged me to learn something new, expand my technical skillset, and push myself as a researcher, and for that I am grateful. Special thanks are also due to my associate advisors for their support. Thank you to Dr. Nicholas Lownes for his feedback on this work, and his engaging instruction throughout my time here at UConn. Thank you to Dr. Nalini Ravishanker for her technical expertise and patience in answering all my questions. Thank you to Dr. Naveen Eluru for his guidance and helpful comments.

In addition to those mentioned above, I would also like to acknowledge my friends and colleagues within the Transportation and Urban Engineering group, and without, for their support, conversation, and ideas that made even the toughest days seem bright.

Thanks also is in order for the Federal Highway Administration and the Dwight D. Eisenhower Transportation Fellowship for their financial support.

Finally, I want to thank the unsung heroes of the Civil Engineering Department, Kristina Ashley and Jessica Zoldak for all of their work behind the scenes.
TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................ iv

TABLE OF CONTENTS ....................................................................................................... v

LIST OF TABLES .................................................................................................................. vii

LIST OF FIGURES ............................................................................................................... viii

CHAPTER 1. INTRODUCTION ............................................................................................... 1

1.1 Background Information ............................................................................................... 1

1.2 Motivation ....................................................................................................................... 4

1.3 Research Objectives ....................................................................................................... 5

1.4 Thesis Structure ............................................................................................................. 5

CHAPTER 2. IS THERE A LIMIT TO ADOPTION OF DYNAMIC RIDESHARING SYSTEMS? . 6

2.1 Introduction ..................................................................................................................... 6

2.2 Literature Review .......................................................................................................... 6

2.3 Data ................................................................................................................................ 9

2.4 Descriptive Analysis ..................................................................................................... 11

2.5 Panel Model .................................................................................................................. 16

2.6 Discussion of Results ................................................................................................. 17

CHAPTER 3. UNDERSTANDING THE RELATIONSHIPS IN DEMAND FOR SHARED RIDE

MODES ................................................................................................................................. 26

3.1 Introduction ..................................................................................................................... 26
3.2 Literature Review ........................................................................................................... 26

3.3 Data and Descriptive Analysis ....................................................................................... 30

3.4 Methodology .................................................................................................................. 38

3.5 Results ........................................................................................................................... 41

3.6 Discussion ....................................................................................................................... 48

3.7 Forecasting ..................................................................................................................... 51

CHAPTER 4. CONCLUSIONS, LIMITATIONS, AND FUTURE WORK ............................................. 53

4.1 Conclusions ..................................................................................................................... 53

4.2 Limitations and Future Work .......................................................................................... 57

REFERENCES ....................................................................................................................... 61

APPENDIX ............................................................................................................................. 65
LIST OF TABLES

Table 1. GLS Random Effects Panel Model Results................................................................. 18

Table 2. Average Weekday and Weekend Demand by Shared Mode........................................ 37

Table 3. Univariate DLM Results for TNC............................................................................ 42

Table 4. Univariate DLM results for Taxi............................................................................... 43

Table 5. Univariate DLM Results for Bikeshare and Subway .................................................. 44

Table 6. Univariate DLM Results for TNC with Modal Demands .......................................... 47

Table 7. DLM results for Taxi with Other Modal Demands...................................................... 58

Table 8. DLM Results for Bikeshare and Subway with Other Modal Demand ....................... 58
LIST OF FIGURES

Figure 1. Manhattan Taxi Zone Land Use Distribution ................................................................. 12

Figure 2. Weekly Pickup Demand in Commercial Taxi Zones (N=23) .............................................. 13

Figure 3. Weekly Pickup Demand in Commercial Taxi Zones, N. Manhattan (N=1) ................. 14

Figure 4. Weekly Pickup Demand in Residential Taxi Zones (N=27) .............................................. 15

Figure 5. Weekly Pickup Demand in Residential Taxi Zones, N. Manhattan (N=11) .............. 16

Figure 6. Daily NYC Shared Mode Demand, January 2015 - June 2017 ....................................... 34

Figure 7. Daily NYC Shared Mode Demand, Excluding Subway .................................................. 35

Figure 8. Daily Demand Patterns, Excluding Subway, January 2015 .............................................. 36

Figure 9. CCF of residuals from TNC-S and Model 1 Taxi............................................................... 45

Figure 10. CCF of residuals from TNC-S and Model 1 Bikeshare .................................................. 46

Figure 11. CCF of residuals from TNC-S and Model 1 Subway ...................................................... 46

Figure 12. TNC-D Latent State Components .................................................................................... 50

Figure 13. TNC Demand Forecast Using Predicted Demand for Taxi, Bikeshare, & Subway ....... 52
CHAPTER 1. INTRODUCTION

1.1 Background Information

Technology has long been the driver for advances in our everyday means of transport. The development of data-enabled and location aware mobile phones, combined with adoption of sharing economies (also referred to as collaborative economies or peer-to-peer networks), has opened the door for novel modal offerings like Uber, Lyft, and citywide bike sharing systems. These specific modal offerings are referred to generally as transportation networking, dynamic ridesharing, or ridesourcing and are provided by transportation networking companies (TNCs). Recent consensus in the literature appears to lean towards the use of TNC and as such, that terminology will be used in this work. TNCs capitalize on the growth of mobile phone adoption and integrated global positioning (GPS) capabilities to actively link riders to drivers in an on-demand setting, as of 2016 77% of US adults are smartphone owners (1). One of the suggested benefits of TNCs is the potential boom in ridesharing. Historically, ridesharing has filled unique niches in the transportation landscape (carpooling, preplanned trips, etc.) and has consistently been encouraged for its ability to minimize user costs and reduce congestion. However, the widespread adoption of conventional ridesharing has been linked with an inability to capture impromptu trips (2). Due to the fact that almost all trips need to be prearranged, conventional ridesharing has substantial limitations and users that recognize its benefits are most often unwilling to make the switch because of inconvenience. One aspect important to note is, that while TNCs have been linked with ridesharing; the rise of services like UberPool and Lyft Line constitute a move towards conventional ridesharing, the majority of TNC operations are hailed in a taxi like manner and should not necessarily be considered conventional ridesharing. This one
vehicle per trip model can be problematic given the rapid explosion in TNC trips presented below.

As the current industry leader, Uber’s growth is readily apparent. At the close of 2014, Uber operated in 230 cities worldwide, and by June 2015, Uber grew to operate in 300 cities (3, 4). In only six years of existence (2009 to 2015), Uber reported facilitating over one billion trips. As of June 2016, Uber reported surpassing two billion total trips in 450 cities (5). In a report published in June of 2017, Uber has facilitated over 5 billion trips worldwide (6). This accelerating growth is impressive, and the popularity of Uber can be associated with its inherent flexibility and user-perceived efficiency. When using Uber or any other TNC provider’s app, individuals are able to hail a ride remotely from anywhere within the service city, given that a driver is available. This is an improvement over the call ahead or on street hailing methods of its most similar competitor, the taxi, while also offering service in locations that may not have many viable alternatives. In a major metropolitan hub like New York City, the growth of Uber is extensive. From June 2014 to June 2015, Uber’s Manhattan monthly June pickups increased by over 250% with over one million more trips made using the dynamic ridesharing app (7). As the first major TNC to take off, Uber has been afforded a leg up over other TNCs, however other providers are catching up. Lyft, the second largest TNC provider in the US, announced in July 2017 that it was completing over 1 million trips per day in the US (8). According to Second Measure, these two industry giants encompass almost 98 percent of the TNC market in the US, with the remaining share coming from smaller companies like Gett, Juno, Sidecar, and Via (9).

While Uber’s growth is noteworthy, there has also been substantial controversy in regards to the transportation networking giant. Established taxi groups and regulators have filed numerous lawsuits against Uber and TNCs for infringing on market share without the burden of
for hire vehicle (FHV) licensure and regulation (10, 11). There has also been considerable
debate on the ethics of the dynamic pricing incorporated into Uber’s app (12). Furthermore, there
is ongoing discussion regarding Uber’s safety, its policies concerning drivers, and how these
policies compare to similar regulations governing other FHV companies (13).

This growing impact of TNCs is becoming an important piece in the planning of cities
and regions around the U.S. and elsewhere. The controversies ignited by TNC growth prompt
additional questions about how TNCs are impacting mobility, congestion, and the environment.
The sharing of transportation resources is nothing new, but TNC growth has induced a
heightened interest in shared mobility. In a recent FHWA report on the topic, Shaheen, Cohen,
and Zohdy define shared mobility to be “the shared use of a vehicle, bicycle, or other travel
mode” (14). Using this definition, much of the modern transportation system can be classified as
shared mobility, including traditional services like transit, taxi, and carpooling, but also newer
offerings like TNCs and bikesharing. In this way, understanding the impacts of shared demand
plays a vital role in the planning, policy, and operation of not only TNCs, but also for
longstanding modes like transit and potential future offerings like autonomous vehicles. The
importance of this matter is evident in the rapidly expanding body of literature on TNCs and
shared mobility. From a positive perspective, it appears growing demand for shared modes may
influence individuals to use transit, own fewer cars, and spend less on transportation overall (15).
However, these potential impacts also may not be as positive. Schaller points out in his review of
TNCs in NYC that they have negative implications like increased congestion, increased vehicle
miles traveled (VMT), and increased green-house gas emissions under current the current
operational model (16). While not linked directly to TNC growth, many municipalities in recent
years have taken strides to improve multimodality and it is important to understand how TNCs may be impacting mode choice.

These systems are not only expanding and growing in the short term, but the on-demand operational model they employ has implications for the future, namely in regards to the deployment of autonomous vehicles (AVs). AV offerings are projected to perform in a shared, on-demand manner, similar to modern TNC offerings. In this sense, studying demand for TNCs and how their growth is impacting existing transportation networks is important. As has been noted with some of the current negative implications of TNC growth, true sharing of the vehicles is an important factor in realizing the congestion and environmental benefits associated with AVs (17).

1.2 Motivation
Transportation networking is rapidly reshaping shared mobility in cities throughout the US and around the globe. This work is motivated by the fact that to date there are relatively few data-driven explorations into what factors impact demand for TNCs and how growing demand for TNCs impacts the demand for other shared modes. By fostering understanding at both a modal level and across modes, the hope is that planners and policy makers can use the findings of this work to continue to promote equitable, sustainable, and useful transportation systems. In conjunction with promoting planning initiatives in the near term this work is also motivated by the future of autonomous vehicles. By fostering understanding in TNC operations now it is assumed that at least high level implications can be drawn concerning the potential future impacts of AVs.
1.3 Research Objectives
The objectives of this thesis can be summarized in 2 parts:

1. Quantify the demand for a transportation networking company in a major urban context as it relates to land use, demographic, and environmental factors. Essentially, explore demand at a disaggregate spatial level to understand sub-regional adoption of transportation networking.

2. Holistically explore how the growing TNC demand is impacting the shared mode landscape in a city; in particular explore how TNC demand is impacting demand for prominent shared modes like taxis, bikesharing, and transit.

1.4 Thesis Structure
This work is comprised of two studies completed in 2017 and 2018 using partially overlapping data. Each of the following two chapters highlights one study. Chapter 2 focuses on a single TNC (Uber) to understand demand and growth in demand for that service in the core of NYC in select months in 2014 and 2015. This investigation is done at a neighborhood level and incorporates land use and demographic information to understand their relationship to neighborhood level Uber demand. The third chapter takes a more broad approach and investigates TNC’s impact on other shared modes, including taxi, bikeshare, and subways, at a citywide scale from 2015-2017. The final chapter presents the conclusions, discusses key study limitations, and outlines future work for both endeavors.
CHAPTER 2. IS THERE A LIMIT TO ADOPTION OF DYNAMIC RIDESHARING SYSTEMS?

2.1 Introduction
As noted in chapter 1, transportation networking is rapidly growing and impacting how people traverse our cities. Understanding TNC demand can provide insight into how these dynamic systems are currently performing and being received by the public, while also offering a possible window into the impact of future modal offerings employing a similar on-demand operational concept. Substantial investigation into how transportation networking impacts travel behaviors and integrates into existing networks is lacking. To address these limitations, this study focuses on quantifying and understanding temporal trends in Uber demand within a major metropolitan center. In particular, the study explores if growth in the adoption and usage of TNCs is unbounded or if there is a limit to their adoption. In order to pursue this research, trip pickup data made available by Uber through a Freedom of Information Law (FOIL) request by the analytics group FiveThirtyEight, is utilized. Weekly demand, aggregated spatially to a taxi zone level, is analyzed using a panel based random effects model while also accounting for potential heteroscedasticity and autocorrelation issues. The demand is expressed as a function of various demographic, land use, and environmental factors to further explore the trends.

2.2 Literature Review
The existing research on TNC demand and analysis of demand trends in particular, is quite limited. These limitations can be attributed to the lack of readily available data from the providers of these modal offerings, and limitations in data that is publicly available. The first vein of research aims to offer a more complete understanding of dynamic ridesharing system performance in relation to available supply and demand. Chen et al. and Guo
et al. (11, 12, 18) explore the characteristics of Uber’s surge pricing and assess its impact on
supply and demand. The authors collected four weeks of comprehensive data on all Uber vehicle
types using the Uber App API and Client App API. The data gathered included information on
the supply of vehicles, the demand for vehicles, and the associated surge price multiplier in both
downtown San Francisco and Midtown Manhattan. They found that the majority of vehicles in
their study areas were concentrated in commercial and tourist areas within each study zone.
Further, they observed that surge pricing did in fact negatively impact demand while
encouraging supply at the study scale. Correa et al. (19) used the same data explored in this study
to quantify changes in Uber and taxi demand over space and time and found substantial
significant growth in demand over time with high correlation between taxi and Uber modes in
the core of the city. Increases in roadway length, income, and job opportunities, as well as
decreases in transit access time and vehicle ownership, were observed with increased demand for
Uber.

Another major line of research is the comparison between dynamic ridesharing and
traditional mode offerings such as taxi and transit. Rayle et al. (20) found that dynamic
ridesharing users are typically younger and more educated than those that use taxis. Also, while
ridesharing does remove potential trips from taxis, there is substantial evidence that users opted
to switch from available transit modes citing travel time as the reason. Survey respondents using
dynamic ridesharing were skewed towards no vehicle ownership with multiple users traveling
together. Researchers Cramer and Krueger (21) compiled a direct comparison of Uber’s
utilization rate (percentage of time drivers are serving passengers) against that of taxis in five
major U.S. cities. Their results showed that as a whole, Uber is the more efficient mode of
transportation compared to taxis. Poulsen et al. (22) explored the relationship between Uber and
green cabs in New York City and found that while demand for both modes is growing, Uber’s is growing at a much faster rate and green cabs perform better in poor neighborhoods. These studies highlight the fact that Uber is similar to existing taxi infrastructure, however, they highlight key differences that warrant further investigation into how people are using dynamic ridesharing on a citywide scale.

The final set of literature pertaining to dynamic ridesharing and demand is informal in the form of blog posts and news articles. These are mostly based on Uber demand information for all of New York City made openly available by researchers at FiveThirtyEight; the same data was also used in this paper. FiveThirtyEight researchers have published numerous articles on the topic. Fischer-Baum and Bialik (23) address how Uber’s growth should be classified. They found that in the urban core of Manhattan, Uber trips seem to come almost entirely from taxis, with the sum of total trips across both modes remaining the same across study years. Schneider (7) explored the trends in Uber demand and found that Uber pickups in the urban core are growing, and notes that these pickups increase more substantially in times of inclement weather compared to increases in taxi pickups.

The research highlighted above provides interesting insights into the demand for dynamic ridesharing, but there are substantial gaps. A key question that this research attempts to answer is if growth in demand for these systems is subject to saturation, or if it is unbounded. To this end, the study explores demand over time, while also capturing the impact of demographic, built environment, and environmental variables. In order to explore this idea quantitatively, this study develops a panel-based random effects model that is similar in nature to the works by Faghih-Imani et al. (24). The authors study the relationship between NYC Citi Bike ridership, land use, and urban form amongst other studies focused on Citi Bike and Taxi travel time. Their research
focuses on three main types of variables: those related to the environment, temporal variables, and land use variables. A similar structure is also observed in an analysis of bus ridership with the incorporation of real-time location information by Brakewood et al. (25), and work by Campbell and Brakewood (26). These studies, along with some of the variables described in the work above, served to motivate the modeling aspect of this research and helped guide selection of potential demographic, land use, and environmental factors that may influence Uber demand.

2.3 Data
As noted in the limitations regarding dynamic ridesharing research, restrictions in data availability and quality have had a substantial impact on the volume of research. Recent adoption of open-data initiatives (wherein data is made publicly available) and the practice of making data available in response to public requests (through FOIL and other similar mechanisms) have made much of this work possible. The Uber trip information was made available by Uber after a FOIL request and was gathered from FiveThirtyEight’s GitHub repository (27). While these data are the only source of information on Uber demand, they also have several limitations that dictated the data preparation and approach to analysis. First, Uber data is only available for six months in 2014 (April through September) and six months in 2015 (January through June). Within this, trip pickup location is recorded at a latitude/longitude level in 2014 and aggregated up to predefined taxi zone level in 2015, including pickup date and time information. Due to the spatially aggregated nature of 2015 data, taxi zone was chosen as the spatial unit of analysis. Subsequently, all remaining data (demographic, land use, and environmental factors) were spatially aggregated to the taxi zone level. Taxi zone delineations are defined by the NYC Taxi and Limousine Commission and line up with neighborhood boundaries in New York City. This analysis focused only on the taxi zones in the Borough of Manhattan. There are 69 unique taxi
zones in Manhattan which includes three zones for ferries to Ellis Island, Liberty Island, and Governors Island, and four zones for major parks, namely, Central Park, Battery Park, Inwood Hill Park, and Highbridge Park. To add context, there are 288 census tracts located in the same study area. The boundaries of census tracts and taxi zones are coincident. However, taxi zones are relatively larger in spatial extent compared to tracts and on average each taxi zone is comprised of about four census tracts. Built environment variables were collected at the tax lot level and gathered from the NYC Planning Departments website (28). Information on transit stations/stops and bike share infrastructure was accessed through NYC open data hub (29). Environmental variables like temperature and precipitation were gathered at the day level from 2014 through 2015 in Central Park from the National Climate Data Center hosted by the National Oceanic and Atmospheric Administration (30). All demographic data was collected at the census tract level from the American Community Survey’s five year estimates for 2014 and 2015 and aggregated to the taxi zone level.

Substantial data processing and data preparation were performed prior to the analysis. First, the taxi zone aggregated Uber pickups were summed to the week level because of the high level of variability at the day level. From the 12 months of initial data, there were 54 total weeks, however, weeks that fell at the beginning and end of the dataset were incomplete. As a result, these incomplete weeks were dropped leaving 49 total complete weeks of demand data. For the remainder of the paper, the weekly pickup counts will be used to constitute the demand and will be referred to as such. Daily weather information was treated similarly. Built environment variables were assumed to not vary across the investigation period and the demographics of each taxi zone were assumed to be represented by the 2014 and 2015 ACS estimates, respectively.
2.4 Descriptive Analysis

The growth in Uber overall has been well documented, in both the literature presented above and in its emergent prevalence in the societal lexicon. When analyzing the data at the borough level, substantial growth in Uber demand over time is easily observed. The total number of trips made in Manhattan from April 2014 through September 2014 was just under 140 thousand per week. In 2015, that number rose to just over 430 thousand trips per week from January through June. However, it is important to note that this growth is not evenly distributed. At the taxi zone level, the top five taxi zones in both 2014 and 2015 are as follows: Midtown Center, Union Square, Tribeca/Civic Center, and both East and West Village. All of these zones are located either in midtown or lower Manhattan and together amount for about 20 percent of the total trips in 2014 and 2015, respectively. In reviewing the shifts in demand by each taxi zone individually, differences were noted in zones that were more residential. These differences were typically manifested as lower overall demand each week, but the growth of that demand appeared lower than that in commercial zones. This observation is reasonable considering residential areas typically have less tourism or visitor traffic.

To better understand overall demand and the temporal trends, the taxi zones were classified into five groups based on a residential built environment measure derived from the land use dataset, seen in Figure 1. The first group, residential taxi zones, is comprised of those zones that have a higher ratio of residential land use than commercial land use. The second group, commercial zones, is comprised of those zones that have a higher ratio of commercial land use than residential. These two groups were then further divided into those found in Northern Manhattan (north of 86th street on the east side of the island and north of Central Park otherwise) and those not which is noted by the red line in Figure 1. Taxi zones that were exclusively parks were treated as the fifth group. These trends with respect to four of the land use
areas classified in the following figures, excluding parks. From Figure 1, it is shown that residential zones are located around Central Park, the northern portion of the island, as well as pocket in the south east (or the Lower East Side).

Figure 1. Manhattan Taxi Zone Land Use Distribution

The line shading for the demand plots in Figures 2, 3, 4, and 5 range from dark (less residential/commercial) to light (heavily residential/commercial). Additionally, residential and commercial dominant zones are shaded in blue and orange hues respectively. Therefore, more
mixed-use areas i.e. places with an even split of residential and commercial land, are represented by the darker hues. Weekly trends show that there is general increase in demand for all zones from 2014 to 2015.

Looking at Figure 2, the commercial dominant taxi zones, a few notable trends are apparent. First, a banding effect can be observed wherein adjacent zones experience similar levels of demand. The upper bands with the most demand correspond to Midtown and surrounding taxi zones, while the remaining zones are distributed further downtown. Almost all commercial zones located in central and southern Manhattan have an upward trend even after noting seasonal decreases.

Figure 3 highlights the only commercial dominant taxi zone found in the northern part of the island, Morningside Heights. Home to Colombia University, this zone exhibits atypical behavior when compared to the other commercial and residential zones. Interestingly, the
substantial increases and decreases observed in 2015 line up with typical semester beginning and ending points.

![Figure 3. Weekly Pickup Demand in Commercial Taxi Zones, N. Manhattan (N=1)](image)

On the other hand, focusing in on the residential areas in central and lower Manhattan highlighted in Figures 4, darker lines indicate the more mixed-use residential areas. These zones have the largest increase in demand of the residential zones and have higher volatility and clear seasonal shifts. The lighter color lines indicate heavily residential zones and have fewer weekly trips and less overall variability at the week level. Perhaps the most interesting observed trend within these zones is that demand appears to stagnate over time in the heavily residential areas. This stagnation is not meant in the sense that demand shows no signs of growth, it is taken to
mean relative to the other residential zones in central and lower Manhattan, these zones show a much flatter change in demand over time.

The results are similar in the residential zones in northern Manhattan, Figure 5. The more mixed zones have the highest demand and growth while the most residential of the group fall within the middle of the graph. Here too, the lines below the 1000 trip mark appear to decrease. The observations of flattening demand lend credence to the hypothesis that growth in Uber demand may not be unbounded and that it may be subject to potential saturation level. We further explore this hypothesis using a random effects panel model below.

Figure 4. Weekly Pickup Demand in Residential Taxi Zones (N=27)
The nature of the weekly Uber demand data (i.e. panel) lends itself to a linear panel based analysis (31). The demand data are available temporally by week for 25 complete weeks in 2014, and 24 complete weeks in 2015. However, not all elements of the demographic, land use, and environmental data used to explain the demand are time variant. As a result, a random effects structure was employed that allows consistent estimation of coefficients associated with time invariant variables.

Following the exposition of linear panel models in Cameron and Trivedi (31), an outline of the model structure and assumptions are presented below. Let $z = 1, 2, ..., Z$ represent an index for the unique taxi zones and $t = 1, 2, ..., T$ represent an index for the week. The dependent variable, weekly pickup demand, is then modeled using a linear panel structure similar to the one below:
\[ y_{zt} = \alpha_z + \beta' X_{zt} + \epsilon_{zt} \]

Where \( y_{zt} \) represents the weekly pickup demand in taxi zone \( z \) at week \( t \), \( \alpha_z \) captures the zone specific unobserved heterogeneity, \( X_{zt} \) represents the vector of independent variables and time variables, and \( \beta \) represents the regression coefficients associated with the independent variables, \( \epsilon_{zt} \) represents the errors within each taxi zone \( z \) for time period \( t \). \( \epsilon_{zt} \) is assumed to be strictly exogenous and normally distributed as \( N(0, \sigma^2_{\epsilon}) \). Further, specific assumptions on \( \alpha_z \) result in the fixed effect and random effect variants. In a random effects model, used in the analysis, \( \alpha_z \) is assumed to be random and distributed normally as \( N(\alpha, \sigma^2_{\alpha}) \). Also, \( \epsilon_{zt} \) and \( \alpha_z \) are assumed to be independent. Further, heteroscedasticity and autocorrelation are allowed in specification of \( \epsilon_{zt} \).

The model was estimated using Stata software using the feasible generalized least squared routine (i.e. xtgls). In order to test the autocorrelation and heteroscedasticity specification of the error term, separate hypothesis tests were run. First, to test heteroscedasticity, a likelihood ratio test proposed by Poi and Wiggins (32) was used that compares heteroskedastic treated panel model against the same model assumed to be homoscedastic. Second, to test autocorrelation, the procedure outlined in Wooldridge (33) was adopted.

2.6 Discussion of Results
As noted earlier, a linear panel model was estimated using the generalized least squares approach in Stata software. The model allowed heteroscedasticity and autocorrelation in the specification of the disturbance term. Due to the presence of autocorrelation and gaps in the data (i.e. missing weekly data between 2014 and 2015), standard procedure of maximum likelihood estimation was not applicable. The final panel contained 61 taxi zones over 49 weeks for a total of 2,989
observations. Parks and zones with missing independent variables were dropped from the model, including Ellis Island, Liberty Island, Governors Island, Randall’s Island, Central Park, Battery Park, Inwood Hill Park, and Highbridge Park. Before running the panel model, the tests for heteroscedasticity and autocorrelation identified above were performed. The results indicated that both autocorrelation (with an F-test statistic of 331 with degrees of freedom of 1 and 60 and p-value of 0.00), and heteroscedasticity (with a chi squared statistic of 2242.47 with degrees of freedom of 60 a p-value of 0.00) were present and as such the error term specification needed to accommodate these issues. In the interest of space, the complete listing of autocorrelation and heteroscedasticity parameters is not presented here.

A variety of explanatory factors, including demographic, land use, and environmental, were explored. Additionally, temporal variables were explored to capture the impact of unobserved time varying factors. The final model is presented in Table 1. All the variables included are significant at least at the 10 percent level of significance. Below, the influence of the various explanatory variables on the weekly Uber pickup demand are discussed.

**Table 1. GLS Random Effects Panel Model Results**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Random Effects Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
</tr>
<tr>
<td>Constant</td>
<td>-19056.180</td>
</tr>
<tr>
<td>Time Variables</td>
<td></td>
</tr>
<tr>
<td>Summer Dummy Variable</td>
<td>-278.133</td>
</tr>
<tr>
<td>Winter Dummy Variable</td>
<td>89.647</td>
</tr>
<tr>
<td>Difference in weeks from initial week</td>
<td>93.323</td>
</tr>
</tbody>
</table>
### Environmental Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean 1</th>
<th>Mean 2</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly Precipitation (in.)</td>
<td>25.892</td>
<td>6.262</td>
<td>***</td>
</tr>
<tr>
<td>Weekly Average Max Temperature (°F)</td>
<td>-10.143</td>
<td>1.174</td>
<td>***</td>
</tr>
</tbody>
</table>

### Built Environment Variables (Taxi Zone Level)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean 1</th>
<th>Mean 2</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count of Citi Bike Stations</td>
<td>242.850</td>
<td>29.211</td>
<td>***</td>
</tr>
<tr>
<td>Total Built Area (10,000 sq. ft.)</td>
<td>0.936</td>
<td>0.071</td>
<td>***</td>
</tr>
<tr>
<td>Percent Residential by Floor Area</td>
<td>64.247</td>
<td>7.003</td>
<td>***</td>
</tr>
<tr>
<td>Percent Retail by Floor Area</td>
<td>151.426</td>
<td>33.043</td>
<td>***</td>
</tr>
</tbody>
</table>

### Demographic Variables (Taxi Zone Level)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean 1</th>
<th>Mean 2</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent African American</td>
<td>-58.925</td>
<td>7.517</td>
<td>***</td>
</tr>
<tr>
<td>Percent Male</td>
<td>60.847</td>
<td>17.042</td>
<td>***</td>
</tr>
<tr>
<td>Percent Age Under 19</td>
<td>220.657</td>
<td>23.752</td>
<td>***</td>
</tr>
<tr>
<td>Percent Age Over 65</td>
<td>-0.213</td>
<td>0.047</td>
<td>***</td>
</tr>
<tr>
<td>Percent of Households with No Vehicle</td>
<td>126.400</td>
<td>60.253</td>
<td>**</td>
</tr>
<tr>
<td>Percent of Households with One Vehicle</td>
<td>131.103</td>
<td>62.971</td>
<td>**</td>
</tr>
<tr>
<td>Average Household Size</td>
<td>-2991.058</td>
<td>363.236</td>
<td>***</td>
</tr>
<tr>
<td>Percent of Adults with at least Some College</td>
<td>17.922</td>
<td>9.481</td>
<td>*</td>
</tr>
</tbody>
</table>

### Interaction Variables

<table>
<thead>
<tr>
<th>Interaction</th>
<th>Mean 1</th>
<th>Mean 2</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Percent Residential by Floor Area) x (Difference in weeks)</td>
<td>-0.909</td>
<td>0.089</td>
<td>***</td>
</tr>
<tr>
<td>(Percent Retail by Floor Area) x (Difference in weeks)</td>
<td>9.581</td>
<td>0.653</td>
<td>***</td>
</tr>
</tbody>
</table>

(* p < 0.10, ** p < 0.05, *** p < 0.01)

#### 2.6.1 Influence of Temporal Variables

A variety of temporal variables were explored to capture the influence of unobserved time varying and seasonal factors. As the scope of the data was limited to only two six month periods,
the extent of seasonal coverage was also limited to only include winter, spring, and summer. In the model, seasonal effects are reported with reference to spring. Spring was chosen because it was the only season present in both years of the data used in the analysis. It appears that Uber demand at the taxi zone level decreases during the summer as noted by the negative coefficient. This observation is also confirmed by the descriptive analysis wherein a clear drop was observed during the summer months of 2014. This may be attributed to individuals’ preference to use non-motorized modes, such as walking or biking, when the weather is warmer compared to other seasons. While reviewing research on seasonal influences on motorized and non-motorized travel, Guo et al. (34) found that during the summer, people make more discretionary trips via non-motorized modes. On the other hand, an opposite reasoning can be offered to explain the positive effect of the winter dummy variable on Uber demand: given the extreme weather that New York City experiences during the winter months, people may seek the comfort of a climate controlled vehicle in order to escape cold conditions.

Additionally, the difference in week from initial week variable is defined as the number of weeks past the first complete week (i.e. week ending on April 13th, 2014) of demand data. Here, a positive influence is apparent and can be interpreted to mean that as time progresses, the pickup demand within taxi zones is increasing. This finding is in line with both the observed trends at the descriptive level, as well as research and reports presented earlier on the growth of Uber overall.

One of the primary hypotheses was to explore potential tapering off of Uber demand with time. To this end, the impact of higher order versions of the difference in week variable were explored after controlling for the other explanatory variables. For example, a negative coefficient associated with a second order influence will lend evidence in support of the hypothesis. It was
found that higher order versions of the difference in week variable were not significant. However, this contradicts the observations from the trend analysis by taxi zones that showed that there is evidence of a tapering effect for residential zones. To further explore the tapering off effect, interaction variables between difference in week and other explanatory factors, namely demographic and land use variables were explored. The results for the interaction variables are discussed at the end of this section.

2.6.2 Influence of Environmental Variables
The environmental variables are included to extend interpretation beyond just seasonality and quantify how weekly weather trends influence zonal level Uber demand. The positive influence related to total weekly precipitation was expected and is intuitive. As the overall total of weekly precipitation increases, demand for Uber also increases. This is plausible because rain/sleet/snow move people away from other non-automobile modes and towards trip making in the comfort of an on-demand vehicle. A substantial amount of users in Manhattan use non-automobile modes like Citi Bike, walk, and transit among others (25, 26). These riders may potentially opt for an Uber trip if it means avoiding exposure to the elements. As the average high temperature by week increases the demand for Uber also increases. Increased temperatures, especially those outside of the normal weather condition, may be prompting individuals to trend towards non-motorized travel in a similar manner to the summer influence.

2.6.3 Influence of Built Environment Variables
Unfortunately, in the final model formulation, the influence of number of bus stops and subway stops at the taxi zone level was found to be statistically insignificant. However, the large positive influence of Citi Bike stations at the taxi zone level is an interesting result. At a high level, the
positive influence makes sense as Citi Bike stations are primarily located in lower Manhattan and these taxi zones correspond to the areas with the highest trip attraction. However, upon closer inspection, it can be seen that even with the inclusion of other built factors like total floor area which are more direct determinants of trip attraction, Citi Bike stations have a significant and sizeable impact. One plausible reason for this is that the users of the on-demand shared Citi Bike system may also be inclined to use shared on-demand services provided by Uber. It is also plausible that Uber is used by such individuals to complete trips that Citi Bike could not. This is consistent with research by Faghih-Imani et al. (24) – in their analysis of Citi Bike and Taxi modes, the authors found that for trips with substantial distance taxi modes proved to be more suitable option – a similar scenario could be at work here.

The total built area is assumed to capture the overall intensity of each taxi zone, and the positive relation to Uber demand follows from our expected results. With more built area, there is more to see, do, and places to live and this falls within other findings that Uber demand is highest in the core of Manhattan and in tourism driven taxi zones. The percent of residential floor area has a positive influence on Uber pickup demand. Residents, just like visitors and tourist, make trips and there are substantial amounts of demand in these residential zones – referring to Figure 1, it can be seen that residential zones like East and West village have some of the highest demand of all observed taxi zones.

Variables representing other land use, namely, percentages of overall commercial, retail, garage (parking), and office areas were explored. These remaining land use percentages were also included in the model but the only other type that proved significant was the percentage of retail by floor area. This strong positive interaction shines light on previously noted trends
regarding Uber’s propensity to serve tourism and central business districts. Within Manhattan, these locations coincide with higher proportions of retail area and as such higher Uber demand.

2.6.4 Influence of Demographic Variables

Understanding how demographic makeup of taxi zones influence Uber demand can provide valuable insights into how different groups are adopting dynamic ridesharing services. When exploring the link between Uber demand and percentage of the population that is African American in a taxi zone, a negative relationship was observed. This observation raises questions on equity. A better understanding of the factors influencing (or lack thereof) African American users to Uber is needed.

The second demographic factor explored is that of gender on Uber demand. Here the percentage of males within the population was found to have a positive relationship with Uber pickup demand. A possible explanation for this can be safety. As mentioned previously (13), Uber has been dealing with issues regarding safety. As a result, male riders may feel more comfortable using the dynamic ridesharing system than their female counterparts.

Age has been an important topic in regard to Uber adoption, and previous research indicates that Uber users tend to skew younger (20). When holding the typical adult age range as the constant categorical factor and exploring the effects on individuals under 19 and over 65, our findings fall in line with existing research. The percentage of the population under 19 years of age is positively associated with demand for Uber than their older counterparts. On the other hand, senior individuals exhibit a lower demand for Uber. Propensity for technology and existing mode bias are potential explanations for this relationship.

Vehicle ownership was investigated over conventional income variables due to findings in the Uber literature. In San Francisco, survey respondents that utilized Uber identified as
owning few vehicles (20). As such, we investigated the influence of household vehicle ownership on Uber demand. It was observed that both no vehicle and one vehicle households exhibit heightened Uber demand compared to households with two or more vehicles. Overall vehicle deficiency can explain this, as households still need to make trips to fulfill their activity needs.

The modeled effects of household size in regard to demand indicate that larger households have a substantially lower demand for Uber than smaller ones. As household size increases, traveling in general typically becomes more difficult and potentially also expensive especially for a mode like Uber that has limited capacity per vehicle.

The final significant demographic variable explored was aimed at quantifying the relationship of education on Uber demand. While only marginally significant, the findings indicate that individuals with higher education exhibit higher demand for Uber. Once again in line with preliminary survey findings, this can be can be attributed to a higher propensity for adopting and using technology by those associated with higher levels of education.

2.6.5 Influence of Interaction Variables
As noted earlier, the main hypothesis was to explore if there is a potential tailing off effect in the growth of Uber demand with time. As noted before, higher order versions of the weekly difference variable were found to be insignificant despite the evidence from the descriptive analysis. In an effort to further explore the tailing off effect, interaction variables between the weekly difference variable and different demographic, and land use variables were investigated to see if there are any potential negative shifts in the demand for specific tax zones and/or population groups over time. This exploration provided substantial evidence in support of our proposed hypothesis. We found that the interaction between percent retail and weekly difference
had a substantial positive impact on Uber demand, suggesting that Uber demand in high retail areas is growing with time. As Uber becomes more readily adopted, a greater proportion of users are making trips from and to areas with higher retail using the service. However, when looking at the interaction of the percentage of residential area by weekly difference there is a clear negative impact. While the magnitude of the relationship is small, the negative coefficient indicates that residential areas over time have a waning or decline in overall Uber demand. There are a couple of possible reasons for this observation. First, some residents may have preference towards existing modes (and bias against Uber) and as such are not going to adopt new systems. Second, if Uber only replace SOV trips, and/or borrows from other non-automobile modes, it will not improve congestion on roadways. As a result, those who are conscious to travel time performance and are price conscious may prefer the cheaper alternative with similar travel time performance (e.g. subway, bus).
CHAPTER 3. UNDERSTANDING THE RELATIONSHIPS IN DEMAND FOR SHARED RIDE MODES

3.1 Introduction
Demand for transportation networking is irrefutably growing, and while the findings above indicate that growth is not uniform across a space and time, the magnitude of that growth at a citywide scale has substantial implications for shared mobility. Beyond mobility, this growth also has implications for congestion and sustainability in modern cities. As mentioned, the goal of this portion of the thesis is to further understand the impact TNCs have on the shared mode landscape, particularly in an urban context. This analysis takes a comprehensive approach by making use of openly available data for a variety of shared modes; the investigation is focused on New York City from January 2015 through June 2017. The modes investigated include four key shared mobility categories: transit, taxi, bikeshare, and transportation networking. Transit in this study is characterized solely by the NYC subway system, taxi is comprised of both Yellow and Green cabs, bikeshare is represented by NYC’s Citibike service, and the TNCs that included are Uber, Lyft and Via. In order to capture the time varying aspects inherent to this investigation, a Dynamic Linear Modeling (DLM) framework is employed. Along with the exploration into demand effects, this study also proposes and examines a forecasting methodology for TNC demand based on the demand for transit, taxi, and bikeshare.

3.2 Literature Review
Existing literature regarding transportation networking has been growing over the past few years as data becomes more openly available and the influence of these systems continues to grow. The majority of the literature can be categorized into two main pools: 1) explicit investigations into various TNCs and 2) investigations into the impacts TNCs have on other modes, traffic
operations, and the environment. This work falls squarely into the second category, and as such this literature review is focused on that vein of research. For readers interested in analyses aimed solely at TNCs, the authors suggest works by Alemi et al. (2017, 2018) and Guo et al. (2017) in addition to the findings presented above (35, 36, 18).

Focusing specifically on TNC impacts on other shared modes, a common theme in both the academic literature and popular media is the comparison between taxi modes and transportation networking. The fact that TNCs operate in a manner similar to taxis logically creates competition between the two modes. Informal research by FiveThiryEight reinforces this concept by highlighting the large decrease in taxi trips in NYC coupled with a large increase in Uber trips in from 2014-2015 (23). Also in NYC, researchers Correa et al present findings that demand for Uber and Taxi respond to similarly to the same land use and socio-demographic characteristics, but that Uber grew by at least 200% in each of the five study boroughs (19). Cramer and Kruger recognize similar findings in their study of Taxi and Uber utilization rates in the US, where they suggest Uber’s improved passenger matching, larger scale, lack of limiting taxi-like regulations, and dynamic supply and demand matching result in Uber’s higher utilization rates (21). In Los Angeles, Smart et al present findings that Uber out performed taxis in low income areas while providing faster and cheaper service (37). Similarly, Poulsen et al studied Green cabs and Uber in the outer boroughs of NYC and they conclude that while both services are capturing previously unmet demand, Uber is growing at a much faster rate relative to its green counterpart (22). In their recent review of the transportation networking literature, Jin et al present observed impacts on three levels: transportation efficiency, equity, and sustainability (38). The review confirms the negative impact of TNCs on taxi demand in a variety of contexts,
but it also discusses to the second major modal comparison found in the literature: TNC’s impact on public transit.

Unlike the influence on taxis, the literature appears split when it comes to TNC’s impact on transit demand, with context seemingly playing an important role (38). In one of the first studies to explore interactions across shared modes, Rayle et al used intercept surveys in San Francisco and find that the TNC users that had the option for a comparable transit trip opt for a ridesourced trip because of the improved travel time (20). This travel time dependence is also highlighted in a data-driven analysis of the city specific TNC, RideAustin. In their study, Komanduri et al compare the observed trips made via RideAustin with temporally comparable transit options and find that the majority of comparable transit trips would have been over 40% longer (39). In this sense, it appears TNCs may draw potential transit trips based on the travel time improvements. Presenting a slightly different relationship, Davidson et al suggest that TNCs appear to be complimenting public transit in NYC by ensuring mobility outside of service hours, or when transit service is down (40). In an investigation using the same data explored in this work, Hoffmann et al identify a positive correlation between subway turnstile entries and Uber, and also note that subway service outages have a significant positive influence on both Uber and Yellow cab demand (41). In his review of seven US cities, Murphy also concludes TNCs serve as a compliment to transit outside normal service, and that the use of shared modes leads to higher public transit utilization overall (15). This complimentary influence however is not ubiquitous. Research indicates that the type of transit plays an important factor in the direction of influence. In an Uber funded study in California, researchers find that when new suburban rail stations are opened, the number of Uber trips in the areas around the stations increases, suggesting a feeder-style relationship between Uber and commuter rail (42).
agency level analysis of operators, Barbar and Burtch explore transit ridership data before and after the introduction of TNCs in a region. They find a significant decrease in bus ridership but note increases in subway and commuter rail ridership, while simultaneously noting that these influences are also dependent on the quality of transit service before (43). Some transit agencies have already taken note of the negative impact on TNCs specifically on bus ridership, and some municipalities are subsidizing transportation networking trips in lieu of continued operation of existing bus service (44).

The last mode included in this study, bikeshare, sees little to no mention in the literature on TNC impacts. In their analysis of TNC and subway interaction, Hoffmann et al note a positive correlation between subway and Citi Bike, but make no mention of Citi Bike and Uber (41). In a study of travel time between taxi and bikeshare in NYC, Faghih-Imani et al find that in the dense urban core of Manhattan, Citi Bike trips are competitive in terms of travel time with taxi trips (24). This study highlights one of the few works exploring comparisons between bikeshare and other shared modes, noting the similarity between TNCs and taxi modes, a similar relationship may be at work with Citibike and TNCs.

The observed modal shifts linked to growing TNCs use highlighted above represents the primary focus of this study, however these shifts have environmental and operational impacts that are also present in the literature. Schaller explored the impact of this growth in the context of New York City and concludes that dynamic rideshare growth has detrimental impact on the sustainability goals of New York City with the potential to increase congestion and vehicle miles traveled (VMT) (16). On the other hand, Li et al determine that Uber’s entry into an urban area actually reduces congestion (45). In the sustainability section of their review, Jin et al note there
is no current consensus on TNC impact on energy consumption and greenhouse gas emission (38).

After reviewing the literature, it can be seen that while the relationships between different pairs of shared ride modes have been explored, a comprehensive exploration of all shared ride modes in a region is missing. Only Hoffmann et al study a variety of modes together, but ultimately the study is aimed at characterizing the pairwise relationship of each mode to public transit. A comprehensive understanding of all shared modes will allow cities that are aiming for effective and efficient multimodal systems, to make informed decisions that positively impact all offerings within the shared mode landscape. Therefore, there is a need to understand the impact of transportation networking on all other shared modes together in a comprehensive framework. Due to the lack of widely available demand data on TNCs, much of the literature is survey based and can be subject to respondent biases. This study makes use of openly available disaggregate demand data to characterize shifts in demand patterns over time. Another major aspect unique to this study is the inclusion of bikesharing. To the best of the authors’ knowledge, this study marks the only exploration into TNC’s relationship with bikeshare demand. Finally, the study develops a multivariate forecasting approach that can be used to forecast the demand for different shared modes over time. This framework can be used by transportation planning agencies for short-term and long-term planning of these different shared ride mode offerings.

3.3 Data and Descriptive Analysis
As in any data driven exploration, a solid understanding and characterization of the input data is of the utmost importance. This study makes use of demand data for seven distinct shared operators: Uber, Lyft, Via, Yellow Cabs, Green Cabs, Citi Bike, and Subway. Disaggregate trip data for all seven providers are made openly available by three distinct agencies. The first of
these, the NYC Taxi and Limousine Commission (TLC) governs taxis and other for-hire vehicles (FHV), which includes TNCs. The TLC openly publishes monthly installments of data on both taxi modes, as well as all the TNCs mentioned above (46). The FHV dataset also includes information on two additional TNCs, Juno and Gett; however due to the lack of complete information they are excluded from this investigation. Citi Bike, the predominant bikeshare service in the city, openly publishes their disaggregate trip data every month (47). The final agency, the Metropolitan Transit Authority (MTA), publishes the hourly cumulative turnstile entrance/exit count information from all subway turnstiles each week (48). Prior to discussing each dataset, it is important to note that this investigation focuses on only the pick-up trip-end due to limitations in the TNC and subway data. While full trip information (including start/end time and location) are available for the taxi and bikeshare modes, only the access end is available for subway and TNC data at the time of this investigation. To allow for uniform analysis, the trips ends are aggregated to a daily level and summed spatially to represent all daily trips started within NYC from January 2015 through the end of June 2017, 912 total days. Additional explanatory datasets on daily weather, city issued permits for events, and federal holidays are also assembled. The weather information comes originally from three observation stations around the city (Central Park, LaGuardia International Airport, and John F. Kennedy International Airport); the average across all stations was used due to the aggregate nature of this investigation. This information is gathered from the National Oceanic and Atmospheric Administration, and through NYC’s open data pool (49, 50). The following paragraphs discuss the data cleaning and aggregation processes for each mode.

The TNC data explored in this investigation is disseminated with a unique pickup timestamp (date & time), and a location that has been spatially aggregated by the TLC to protect
privacy. Lyft and Via are not present across the full temporal range and based on the data only begin making trips in April 2015. Data processing for each TNC was straight forward, with no observed data quality issues. Trip ends are simply summed to a daily level across all of NYC for each provider and then aggregated again to encompass total TNC demand across providers.

The Taxi data structure is very similar to that of the TNC data, and includes additional information on trip length, passengers, cost, etc.; however missing weeks of raw data are present for Green Cabs and required treatment. The last week of nearly every month of raw data in 2016 is missing, in total 72 days. To treat this, a local averaging method is employed where data from the same weekday two weeks prior and two weeks after are averaged and the missing value is replaced. Take for example Monday May 23rd 2016, a date missing in the Green data. To treat this, observed data on May 9th, May 16th, May 30th, and June 6th are averaged and the missing value on May 9th is replaced with the average value. After complete treatment, a similar aggregation procedure to TNCs is employed to arrive at the final daily observed pickup demand for NYC Taxis (both green and yellow).

The Citi Bike dataset contains information on the pickup/dropoff timestamp and docking station location, along with additional rider variables. Using the pickup time and location, the trips are treated in the same manner as those above to arrive at the citywide total of Citi Bike trips per day.

The Subway dataset is distributed as cumulative turnstile counts, with readings taken approximately every four hours. This presented unique aggregation challenges. First, the turnstiles must be grouped and ordered to determine the number of entries by differencing the cumulative counts. While direct in theory, issues like closed turnstiles, cumulative count resets,
phantom readings (additional readings taken a few seconds after a normal 4-hour reading), and decreasing cumulative counts are present in the data. Closed turnstiles are a valid observation, but to treat the other issues three treatments are employed:

1. **Drop the observation** – applied to the “phantom” readings where the cumulative count almost always demonstrates a cumulative decrease of one.

2. **Take the absolute value of the entry count** – applied when the cumulative counts for entries are decreasing but otherwise comparable with adjacent turnstiles, essentially these values are treated as valid readings just reported in the wrong direction.

3. **Impute the value as the average of adjacent counts before and after the reading** – applied when cumulative totals are reset resulting in very large positive or negative readings (cumulative counts are not always reset to 0).

After addressing the observed issues, the counts are aggregated to a daily level across all turnstiles. To verify the validity of the data treatment, the yearly totals are compared directly to MTA’s reported subway ridership data in 2015 and 2016. The observed totals are 1,762,565,419 riders in 2015 and 1,756,814,800 riders in 2016 (51). The yearly totals derived from the corrected turnstile data slightly overestimate these totals by 0.92% and 0.93% respectively, or by about 16 million trips. The cleaned data also maintains the change in ridership across the study period with a decrease of about 5.62 million from 2015 to 2016.
Figure 6. Daily NYC Shared Mode Demand, January 2015 - June 2017

Descriptive Analysis

This subsection is focused on highlighting observed trends and presenting a high level summary of the input data. Figure 6, above, presents the demand for all 4 modes. The most direct conclusion from Figure 6 is the sheer dominance of subway compared to the other shared modes. Considering the NYC subway system is one of the most utilized in the world, this vast gap is not surprising. Due to the dominance of subway, Figure 7 highlights the non-subway modes at a more appropriate scale.
Figure 7. Daily NYC Shared Mode Demand, Excluding Subway

From Figure 7, the growth in TNC and decline in taxi demand are readily apparent. By the end of 2016, taxis and TNCs in NYC are making about the same number of daily trips, and by June 2017 TNCs are experiencing substantially more demand than their medallion bearing counterparts. Also apparent at this scale is a clear seasonality in the bikeshare demand. Demand for bikeshare appears to peak in the Fall with the values logically dipping in the winter months. The adverse weather in the region during the winter months is one intuitive explanation for this.

Another important point to note is the noisiness of the observations in Figures 6 and 7. These fluctuations can be characterized by repetitive weekly patterns that are present in the data. For the sake of readability, Figure 8 excludes subway, but is included to highlight these weekly patterns. The weekly subway fluctuations are excluded, but the pattern in the weekly subway demand is the most uniform of the studied modes, and can even be observed in the rounded peaks in Figure 6. This is due to the fact that subway demand is relatively flat across weekdays.
and declines sharply on the weekends, a byproduct of fewer commuters and declined service. Shifting focus to the other modes, taxis and TNCs appear to follow a similar weekly pattern with demand at a minimum on Sundays, increasing steadily throughout the week, and peaking on Saturday. Citi Bike, like subway, appears to be more commuter dependent with the majority of the demand falling during the weekdays and declining on the weekends. These findings make sense; taxis and TNCs are inherently flexible and can be used more easily by those completing recreational trips on evenings or weekends. Bikeshare as a commuting mode, particularly in NYC, also has support as Faghih-Imani et al note competitive travel times with auto modes during peak commute hours (24).

Figure 8. Daily Demand Patterns, Excluding Subway, January 2015
With noted differences in the patterns of demand during the week and on the weekend for each mode, it is useful to look at these averages and how they have changed over time, see Table 2.

Table 2. Average Weekday and Weekend Demand by Shared Mode

<table>
<thead>
<tr>
<th></th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
<th>% Change 2015-2017</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TNC</strong></td>
<td>Avg. Weekday</td>
<td>112,845</td>
<td>233,006</td>
<td>343,089</td>
</tr>
<tr>
<td></td>
<td>Avg. Weekend</td>
<td>119,433</td>
<td>255,746</td>
<td>398,004</td>
</tr>
<tr>
<td><strong>Taxi</strong></td>
<td>Avg. Weekday</td>
<td>439,579</td>
<td>394,644</td>
<td>353,971</td>
</tr>
<tr>
<td></td>
<td>Avg. Weekend</td>
<td>462,770</td>
<td>399,937</td>
<td>358,331</td>
</tr>
<tr>
<td><strong>Citi Bike</strong></td>
<td>Avg. Weekday</td>
<td>29,147</td>
<td>40,702</td>
<td>40,210</td>
</tr>
<tr>
<td></td>
<td>Avg. Weekend</td>
<td>22,410</td>
<td>30,687</td>
<td>31,155</td>
</tr>
<tr>
<td><strong>Subway</strong></td>
<td>Avg. Weekday</td>
<td>5,576,099</td>
<td>5,578,196</td>
<td>5,571,128</td>
</tr>
<tr>
<td></td>
<td>Avg. Weekend</td>
<td>3,109,717</td>
<td>3,021,371</td>
<td>2,978,651</td>
</tr>
<tr>
<td><strong>All Modes</strong></td>
<td>Avg. Weekday</td>
<td>6,157,669</td>
<td>6,246,549</td>
<td>6,308,398</td>
</tr>
<tr>
<td></td>
<td>Avg. Weekend</td>
<td>3,714,330</td>
<td>3,707,743</td>
<td>3,766,141</td>
</tr>
</tbody>
</table>

Table 2 further reinforces the patterns observed in the demand for each mode. TNCs saw a huge leap in demand from 2015 to 2017, with a 204% growth in trips during an average week day and 233% growth in trips on an average weekend day. The observed decline in taxi is also reinforced with about a 20% decrease in average trips across both weekdays and weekends. Bikes share demonstrates substantial growth on both weekends and weekdays, while average weekday trips on the subway remain remarkably consistent. It is interesting to note that across all the shared modes analyzed demand appears to be increasing. Slightly more shared trips are being made on average both during the week and on the weekend at 2.4% and 1.4%, respectively. This growth may potentially bode well for multimodality as previous findings indicate those individuals that use any shared mode are more likely to use other shared modes and transit (15). At a high level, these observations present potential relationships across modes. It appears the growing demand
for TNCs is coming directly from that of taxis, but the magnitude of TNC growth, along with
growth in the overall demand for all shared modes, may indicate previously unmet demand is
now being filled. These potential relationships are explored in the following sections using a
preliminary Dynamic Linear Modeling Framework.

3.4 Methodology
With the goal of understanding tradeoffs in demand across shared modes in mind, there are a
wide array of techniques one could apply to quantify the desired relationships. In deciding an
applicable methodology, the inherent temporal nature of the data naturally lent itself to a time
series analysis. A Dynamic Linear Modeling (DLM) framework enables study of the changes in
demand patterns, and to understand the relationship to demand for shared modes (52). The DLM
is a special case of State Space Models. Durbin and Koopman (2012) offer a detailed exposition
of state-space modeling and its different variants (53). DLM is comprised of two parts: first, an
observation equation relates the observed demand on day $t$ to some underlying unobservable
random time-varying state(s) at time $t$ plus an unobserved observation (or measurement) error,
and second, a state equation that defines how the unobserved state value(s) on day $t$ are
generated as a function of the state(s) on day $t - 1$ and other exogenous inputs on day $t$ plus an
unobserved state(s) error. The states and the errors are assumed to be mutually independent,
Gaussian random variables. In terms of the relevance of this modeling approach to the study of
demand for different shared modes, the observation equation assumes that the observed demand
for a mode is a manifestation of true but latent state(s) (e.g. trend, seasonality) for a mode,
subject to some noise. On the other hand, the state equation represents a characterization of the
true underlying state(s) as a function of various influencing factors including trends (e.g. growth

38
in demand over time), seasonality (e.g. day of week effect), and exogenous variables (e.g. events, disruptions, holidays, weather) among others.

The exposition of the DLM formulation below is limited to the specific version used in this study. For a more general treatment of DLM, please review Shumway and Stoffer (2011) and Durbin and Koopman (2012) (53, 54). The observed scalar response $y_t$, which is the demand for TNC, for example, at time (day) $t$ is expressed via an additive structural time series model as:

$$ y_t = T_t + S_t + v_t $$  

(1)

where $T_t$ denotes the trend and $S_t$ denotes the seasonality at time $t$, while $v_t$ denotes random error. The trend component $T_t$ is modeled as an autoregressive process of order 1 (i.e. an AR(1) process) with drift $\delta$ and AR(1) coefficient $\varphi$ as shown in Equation 2.

$$ T_t = \delta + \varphi T_{t-1} + w_{1t} $$  

(2)

The seasonality component $S_t$ is the weekly periodic component such that:

$$ S_t + S_{t-1} + S_{t-2} + S_{t-3} + S_{t-4} + S_{t-5} + S_{t-6} = w_{2t} $$  

(3)

For each $t$, the errors $w_{1t}$ and $w_{2t}$ are assumed to be independent and distributed as $N\left(0, \sigma_{w_1}^2\right)$ and $N\left(0, \sigma_{w_2}^2\right)$ respectively and assumed to be independent of the observation error which is assumed to be $N\left(0, \sigma_v^2\right)$.

The Gaussian DLM formulation requires that the observation equation be linear and Gaussian, while the state equation must be linear, Gaussian, and Markovian. In Equations 2 and 3 above, while the model for $T_t$ only depends on $T_{t-1}$, it can be seen that $S_t$ depends on six past values. By expanding the dimension and denoting the unknown state vector as $x_t = (T_t,$
\( S_t, S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}, S_{t-5} \), the above equations can be recast as a Gaussian DLM, see Shumway and Stoffer (2011), Ch. 6 for details (54).

The observation equation for modeling the demand for TNC is:

\[
y_t = A_t x_t + v_t
\]  

(4)

where \( y_t \) is the scalar observed demand at time (day) \( t \). \( x_t \) is a 7-dimensional state vector, and \( A_t \) is a \( 1 \times 7 \) observation vector which defines how the observation \( y_t \) and state vector \( x_t \) are related. \( A_t \) is defined as \([1,1,0,0,0,0,0]\). The state equation is:

\[
x_t = \Phi x_{t-1} + Y u_t + w_t
\]  

(5)

where, \( \Phi \) is a 7x7 dimensional state transition matrix with rows given by \((\phi,0,0,0,0,0,0), (0,-1,-1,-1,-1,-1,-1), (0,1,0,0,0,0,0), (0,0,1,0,0,0,0), (0,0,0,1,0,0,0), (0,0,0,0,1,0,0), \) and \((0,0,0,0,0,1,0)\). The state error vector is \( w_t = (w_{1t}, w_{2t}) \). \( u_t \) is an \( r \)-dimensional vector of time varying exogenous predictors (e.g. precipitation, events) and \( Y \) is the associated coefficient vector.

The parameters to be estimated are the variances associated with the two state noise variance components: \( \sigma^2_w, \sigma^2_w \), the observation noise variance: \( \sigma^2_v \), the coefficient \( \phi \) and drift \( \delta \) associated with the trend state component, and the \( Y \) coefficient vector corresponding to the exogenous factors. The estimation process of the parameters is accomplished via numerical maximization of the innovations likelihood using the Newton-Raphson algorithm. This process was operationalized in R using the \texttt{astsa} package (54).
3.5 Results
To explore the demand and relationships between TNC and other shared modes, the DLM was estimated with demand for TNC as the response variable ($y_t$). The two separate models for TNC demand were estimated as described below along with additional models for the other shared modes:

1. First, a DLM was fit to TNC demand with the expanded state vector as discussed above and time varying exogenous variables (related to weather, events, and holidays). This model will be referred to as “TNC-S” moving forward. Similarly, additional models were fit for Taxi, Bikeshare, and Subway. After the best fitting models were determined, cross-correlation analysis between the residuals of the four models was done to identify any contemporaneous or lagged relationships with the demand for other shared modes.

2. Second, a DLM for TNC demand was estimated again with the same specification as before, but with the addition of appropriate concurrent or lagged demand for other shared modes determined from the correlation analysis included as additional exogenous predictors. This model will be referred to as “TNC-D” moving forward.

For all models, performance was estimated with a thirty day holdout sample to do out-of-sample predictive validation. The Mean Absolute Percent Error (MAPE) was used to quantify the validation performance and model fit. Additionally, diagnostic checks were made to assess normality and independence of residuals. These checks can be found in the Appendix.

In terms of the model specification, the trend portion of the latent state was specified as either AR(1), random walk with drift ($\phi$ is set as 1) or just random walk ($\phi$ is set as 1 and $\delta = 0$). The final choice was made based on overall model fit. Daily recurring patterns over the
course of a week were captured using a seasonality portion of the latent state. As noted earlier, exogenous factors related to weather, events and holidays were explored. For the TNC-D model, demand for other shared modes were included as predictors. Below, TNC-S model results are presented first followed by a the cross-correlation analysis and finally the TNC-D model results.

Table 3. Univariate DLM Results for TNC

<table>
<thead>
<tr>
<th></th>
<th>estimate</th>
<th>SE_tnc</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_{w1} )</td>
<td>1.16503E-02</td>
<td>1.05461E-03</td>
<td>11.05</td>
</tr>
<tr>
<td>( \sigma_{w2} )</td>
<td>2.12051E-03</td>
<td>1.30601E-04</td>
<td>16.24</td>
</tr>
<tr>
<td>( \sigma_{v} )</td>
<td>1.36043E-02</td>
<td>8.68196E-04</td>
<td>15.67</td>
</tr>
<tr>
<td>Precipitation (in.)</td>
<td>6.74644E-03</td>
<td>2.06970E-03</td>
<td>3.26</td>
</tr>
<tr>
<td>Holiday Indicator</td>
<td>-1.06592E-02</td>
<td>3.81212E-03</td>
<td>-2.80</td>
</tr>
<tr>
<td>NYC Peak Travel (Sep-Dec)</td>
<td>-3.33001E-04</td>
<td>8.42827E-04</td>
<td>-0.40</td>
</tr>
<tr>
<td>Number of Event Permits</td>
<td>-1.09861E-06</td>
<td>3.15818E-06</td>
<td>-0.35</td>
</tr>
<tr>
<td>MAPE</td>
<td>9.78%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Log-Likelihood</td>
<td>-2919.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3 presents results for the TNC-S model. Indicated by the MAPE, the fit for TNC is quite good with the out of sample predictions matching observed demand within 10%. The trend component was limited to a simple random walk process for TNC. Precipitation and federal holidays were found to significantly impact the demand for TNCs. While holidays substantially decrease demand for TNC, precipitation actually increases the demand. The peak holiday season indicator and count of events were also explored as potential exogenous factors. However, their impact was insignificant. The standard deviation estimates for the noise terms also indicates that the variation of the weekly periodicity, \( \sigma_{w2} \), is lower than those associated with the trend or observations. Also, based on the diagnostic checks, while the residuals are roughly normal, some
uncaptured temporal dependence still exists. This can potentially be addressed by including the demand for other shared modes.

Prior to the inclusion of the demand for other modes in the TNC model, it was important to explore the appropriate temporal relationships in the demand for each mode. What this exploration attempts to answer is, after accounting for the observed trend and periodic effects, does the demand for TNC today most depend on the demand for Taxi/bikeshare/subway today, yesterday, or some other temporal lag. The way that was done was to first fit DLMs for each of the other three modes (Taxi, Bikeshare, and Subway) to account for their own trend and seasonal components, and then explore any remaining correlations in the modeled residuals to identify any existing relationships between modes. To highlight this, Tables 4 and 5 present the model results and Figures 9, 10, and 11 present the cross correlation analysis.

Table 4. Univariate DLM results for Taxi

<table>
<thead>
<tr>
<th></th>
<th>Model 1. Taxi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>estimate</td>
</tr>
<tr>
<td>$\sigma_{w_1}$</td>
<td>2.68538E-02</td>
</tr>
<tr>
<td>$\sigma_{w_2}$</td>
<td>2.29988E-03</td>
</tr>
<tr>
<td>$\sigma_v$</td>
<td>1.11329E-02</td>
</tr>
<tr>
<td>Precipitation (in.)</td>
<td>-2.04801E-02</td>
</tr>
<tr>
<td>Holiday Indicator</td>
<td>-3.23995E-02</td>
</tr>
<tr>
<td>NYC Peak Travel (Sep-Dec)</td>
<td>-6.77869E-04</td>
</tr>
<tr>
<td>Number of Event Permits</td>
<td>-9.30607E-06</td>
</tr>
<tr>
<td>drift, $\delta$</td>
<td>4.27933E-03</td>
</tr>
<tr>
<td>MAPE</td>
<td></td>
</tr>
</tbody>
</table>

Much like the results for the TNC model, the fit for Taxi is quite good with a MAPE under 6%. Unlike the TNC model, the trend here is best characterized by a random walk with
drift process. The same explanatory variables proved significant, with the only major difference being the negative influence of precipitation on daily Taxi demand.

Moving on to the models for the next two modes, bikeshare presented a unique challenge because demand appears, at a high level, to be more vitally linked to weather and seasonality than the other modes. This effect is visible in the peaks during summer and fall months in Figure 7. To accommodate this, alternative exogenous predictors were applied in an attempt to improve the fit of the bikeshare model. These additional predictors include an alternative peak travel indicator for “fair-weather” months as well as the maximum observed daily temperature. The final model results for bikeshare and subway are presented in Table 5.

### Table 5. Univariate DLM Results for Bikeshare and Subway

<table>
<thead>
<tr>
<th></th>
<th>Model 1. Bikeshare</th>
<th></th>
<th>Model 1. Subway</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>estimate</td>
<td>SE_citi</td>
<td>t</td>
<td>estimate</td>
</tr>
<tr>
<td>Phi ((\varphi))</td>
<td>6.43253E-01</td>
<td>1.93718E-02</td>
<td>-18.42</td>
<td>1.00673E+00</td>
</tr>
<tr>
<td>(\sigma_w1)</td>
<td>6.08079E-03</td>
<td>1.39958E-04</td>
<td>43.45</td>
<td>2.54383E-01</td>
</tr>
<tr>
<td>(\sigma_w2)</td>
<td>2.14207E-04</td>
<td>1.46533E-04</td>
<td>1.46</td>
<td>1.22382E-01</td>
</tr>
<tr>
<td>(\sigma_v)</td>
<td>5.29217E-04</td>
<td>5.23303E-04</td>
<td>1.01</td>
<td>3.30243E-01</td>
</tr>
<tr>
<td>Precipitation (in.)</td>
<td>-1.49115E-02</td>
<td>6.90794E-04</td>
<td>-21.59</td>
<td>-1.61787E-01</td>
</tr>
<tr>
<td>Holiday Indicator</td>
<td>-5.57909E-03</td>
<td>1.24109E-03</td>
<td>-4.50</td>
<td>-8.14995E-01</td>
</tr>
<tr>
<td>NYC Peak Travel (Sep-Dec)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8.54187E-03</td>
</tr>
<tr>
<td>Peak Bike (May-Oct)</td>
<td>1.9985E-03</td>
<td>6.1825E-04</td>
<td>3.23</td>
<td>-</td>
</tr>
<tr>
<td>Number of Event Permits</td>
<td>-1.43690E-05</td>
<td>3.22039E-06</td>
<td>-4.46</td>
<td>-7.18904E-06</td>
</tr>
<tr>
<td>Max Daily Temp (F)</td>
<td>2.2854E-04</td>
<td>1.1672E-05</td>
<td>19.58</td>
<td>-</td>
</tr>
<tr>
<td>MAPE</td>
<td>28.42%</td>
<td>-</td>
<td>-</td>
<td>12.50%</td>
</tr>
</tbody>
</table>

The trend components in bikeshare and subway demand appear to be reasonable captured by an AR(1) process each with a Phi significantly different than 1, but the fit for these models is not as good as those for the auto based modes. Significant negative relationships appear to be present for both precipitation and holidays. Bikeshare demand also is positively impacted by higher daily
temperatures and the peak months. Again, validity checks for all these models are included in the Appendix. While the residuals are roughly normal, there does still appear to be some uncaptured temporal dependence.

The models above appear to fit reasonably well but exclude any interactions between the demands for shared modes. Using the residual errors of these models as a proxy for additional unexplained components, the cross-correlation effects are explored in an attempt to identify key lagged or contemporaneous relationships between these modes.

![Figure 9. CCF of residuals from TNC-S and Model 1 Taxi](image)
After treating the temporal dependence, it appears the same day demand (Lag 0) is the most impactful temporal piece between TNCs and the other shared modes. Using this as a guide, the model for TNC demand is re-evaluated with the contemporaneous demand for the other three respective shared modes included as state-level predictors. Tables 6 presents these results.
Table 6. Univariate DLM Results for TNC with Modal Demands

<table>
<thead>
<tr>
<th>Model 2. TNC-D</th>
<th>estimate</th>
<th>SE_tnc</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\sigma_{w1})</td>
<td>1.1586E-02</td>
<td>1.0438E-03</td>
<td>11.10</td>
</tr>
<tr>
<td>(\sigma_{w2})</td>
<td>-2.0840E-03</td>
<td>1.2924E-04</td>
<td>-16.13</td>
</tr>
<tr>
<td>(\sigma_v)</td>
<td>-1.3463E-02</td>
<td>8.7592E-04</td>
<td>-15.37</td>
</tr>
<tr>
<td>Precipitation (in.)</td>
<td>4.2111E-03</td>
<td>2.2577E-03</td>
<td>1.87</td>
</tr>
<tr>
<td>Holiday Indicator</td>
<td>-9.2408E-03</td>
<td>3.9433E-03</td>
<td>-2.34</td>
</tr>
<tr>
<td>NYC Peak Travel (Sep-Dec)</td>
<td>-2.1734E-04</td>
<td>1.0186E-03</td>
<td>-0.21</td>
</tr>
<tr>
<td>Number of Event Permits</td>
<td>3.6613E-06</td>
<td>6.3413E-06</td>
<td>0.58</td>
</tr>
<tr>
<td>TNC Demand (Mil. Trips)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Taxi Demand (Mil. Trips)</td>
<td>-2.0432E-02</td>
<td>1.0251E-02</td>
<td>-1.99</td>
</tr>
<tr>
<td>Bikeshare Demand (Mil. Trips)</td>
<td>-1.1103E-01</td>
<td>5.1766E-02</td>
<td>-2.14</td>
</tr>
<tr>
<td>Subway Demand (Mil. Trips)</td>
<td>2.4653E-03</td>
<td>1.0491E-03</td>
<td>2.35</td>
</tr>
<tr>
<td>MAPE</td>
<td></td>
<td></td>
<td>11.23%</td>
</tr>
<tr>
<td>Difference in MAPE</td>
<td></td>
<td></td>
<td>1.45%</td>
</tr>
<tr>
<td>Final Log-Likelihood</td>
<td></td>
<td></td>
<td>-2923.9945</td>
</tr>
</tbody>
</table>

The TNC-D model performed slightly worse in terms of MAPE when compared to the TNC-S model (9.78\% versus 11.22\%) but had a better overall fit based on the log-likelihood at convergence (-2919.8 versus -2923.9). The TNC-D model also offers a way to understand the interactions across different shared modes. The demand for the three other modes are significant and their direction of influence can be seen in Table 6. Again, the diagnostic checks for these models indicated that the model does not violate the underlying assumptions significantly. While reasonably fitting, the model even after including the demand variables for other modes does not fully account for all the variability in the data. Other factors and alternate model formulations may be necessary – these will be discussed in the conclusions section. The following section presents a brief discussion of these results.
3.6 Discussion
As indicated by the cross-correlation results, the demand for other TNC appears to be most directly linked to the demand for the other shared modes within the same time period (demand for one mode today is the most important influencer on demand for other shared modes today). Intuitively this makes sense, but even more so in the context of New York City where most individuals have access to most if not all of studied modes with no inherent access restrictions, outside of proximity and availability. Of course, some individuals may be cost sensitive or lack the required technology to access TNCs or bikeshare, but these aspects are not identifiable in this research. In this way, the demand for these modes may be reliably characterized together and encompass demand for shared systems in general. Existing research into shared mobility posits a similar idea, citing that individuals that use one shared mode tend to use others (15). With this holistic demand relationship in mind, the remainder of this discussion section is focused on the TNC-D model that includes the demand for the other shared modes. This is done to emphasize the cross demand influence and address the original goal of this work in assessing the impacts of TNCs on other shared systems.

A brief discussion is presented on the influence of precipitation, holidays, events, and the time of year indicators included in this investigation. As with the previous TNC-S model, precipitation appeared to increase demand for TNCs. As the only mode to provide true door-to-door service, this relationship is reasonable. Individuals that otherwise would need to walk in the rain to a bike docking station, hail a cab on the street, or walk to the nearest subway entrance can simply order a ride from in-doors and avoid the elements all together. This positive relationship between TNC demand and precipitation was also noted in a previous study of Uber and rain in NYC (55). The interpretation of holiday’s impact on demand is straightforward. An indicator for
federal holidays was associated with a decrease in trips for TNCs. This can be explained by the fact that people naturally travel less on these days. As for events, the number of permits issued (like street festivals or events in the park) did not appear to have any statistically significant relationship. Also, there did not seem to be any statistically significant time of the year effects on the demand for TNC.

Shifting the focus to how the other modes impact TNC demand can offer insights into the relationship between different shared modes. The model indicated that increased taxi and Citi Bike demand negatively influence TNC demand. Due to the direct competition between taxi and TNC modes, the negative relationship may be reasonable. The influence of Citi Bike on the other hand is not so straightforward. A simple explanation is that Citi Bike users and TNC users may be one and the same and depending on the circumstances (weather, TNC/bike availability, cost), users may be switching back and forth. Further work needs to be done to explore this relationship, but from a policy perspective this presents a potential avenue for reducing VMT caused by TNCs by further promoting bikeshare. From an industry perspective, the link between TNCs and bikeshare has also been acknowledged with both Uber and Lyft making significant investment in established bikeshare providers and pursuing efforts to offer bikeshare alongside their traditional TNC offering (56). The relationship between Subway and TNCs on the other hand seems to confirm the complementary aspects noted by other researchers (15, 40-42). In a location like NYC, where Subway use is entrenched especially for weekday commuters, a true complementary relationship may be at work. Others have noted that TNCs appear to fill the gaps in Subway service, which presents one explanation and is beneficial for the overall mobility options in a region. An additional aspect of the complementary relationship may be more multimodal trips are now made easier – a similar relationship is also confirmed by Davidson et
al. (2016) (40). Where a transit trip was once infeasible, individuals can now use transit and TNCs to quickly and easily arrive at their final destination, effectively solving first and last mile issues. While this requires further study, this potential relationship has substantial implications for the future of shared on-demand fleets, like TNCs, and public transit offerings.

In addition to the relationships in demand, the TNC-D model also provides insight into the changing demand for TNC over time. Figure 12 presents the predicted values for the two state components, trend in blue and seasonality in red.

![Figure 12. TNC-D Latent State Components](image)

As can be seen the trend closely mimics a random walk process. The seasonality (capturing the day of week effect) is more sinusoidal in shape oscillating around the zero value. This is a manifestation of how it was specified (see Equation 3). It is interesting to note that the
magnitude of the oscillations are increasing with time. This is reasonable because the demand for TNC is changing with time and as a result there are larger day of week fluctuations.

3.7 Forecasting

The nature of the concurrent temporal relationships observed in the demand for shared modes presents some challenges for forecasting. In order to use the TNC-D model presented above for prediction, one would need to know the demand for all other shared modes at future time points in order to predict TNC demand at future time points. To overcome this, a two part forecasting procedure is proposed to estimate the demand for TNCs. The procedure makes use of the 3 univariate DLMs: Model 1 Taxi, Model 1 Bikeshare, and Model 1 Subway modes, and the TNC-D model that includes the demand for other modes as predictors. This procedure requires the forecaster to have information on the future precipitation, number of city issued event permits, upcoming holidays, and time of year indicators as exogenous predictors. It is reasonable to assume that information on these variables is readily available for short-term forecasts of 7-10 days through local weather forecasts and access to upcoming permit request totals. The forecasting procedure is as follows:

1. Predict the demand for taxi, bikeshare, and subway for n future time points using their respective DLMs.
2. Use the n future demand predictions for taxi, bikeshare, and subway as input data to forecast the TNC demand using the TNC-D model.
3. As the observed data becomes available, update the models with the new information to actively reflect any changes in the demand for all modes.
This procedure was evaluated using the same 30 day holdout period used to assess fit. Figure 12 presents the fit, highlighted by the values to the right of the vertical line, with the black line representing forecasted demand and the red line representing observed demand.

![Figure 12. Fit of model with forecasted demand](image)

**Figure 13. TNC Demand Forecast Using Predicted Demand for Taxi, Bikeshare, & Subway**

The resulting MAPE using the forecasting procedure actually marginally outperforms the original model using the observed demand for taxi, bikeshare and subway by about 0.5%. The benefit of a forecasting framework like this lies in the ease of estimation and model updating. Regions with actively collected data, like NYC, can automatically update the models with new observed data via a script based approach. This method is limited by a confounding of errors. This means that any errors in the 3 preliminary demand models are inflated when used in the TNC model.
CHAPTER 4. CONCLUSIONS, LIMITATIONS, AND FUTURE WORK

The organization of this chapter follows the format of the thesis, with the conclusions from chapter 2 presented first followed by the conclusions from chapter 3. Overall limitations and future work for both studies are discussed at the end.

4.1 Conclusions

With the substantial growth in TNC offerings over the last few years, a comprehensive investigation into understanding the underlying factors impacting demand is of interest. Chapter 2 explores multiyear data from an existing TNC, namely, Uber in New York City, to understand temporal trends in demand for this service. Existing research supports the growing adoption of Uber, but here it is hypothesized that while macroscopic trends may suggest unbounded increases in the demand, meso/microscopic levels may offer a different picture. It is posited that the demand for Uber has a saturation limit, but what the limit may be is outside the scope of this work. To explore these underlying factors and test our hypothesis, a random effects panel model was developed.

The findings of this analysis indicate that there is a significant growth over time, but that seasonality in demand exists with an increase in demand in winter months and decrease in demand in summer months. Weather was also noted to have a statistically significant impact, with increased precipitation positively influencing demand and warmer temperatures decreasing the weekly demand. At a built environment level, the presence of transit was not determined to have significant impact, but bike share infrastructure had a positive influence. Similarly, the more built area a taxi zone had, as well as the higher percentage of the floor area that is residential and retail based land use, indicated a higher demand for Uber service. The influence of zonal level demographics was also investigated. The percentage of population that is male,
under 19 years of age, has some college education, and has one or no household vehicles showed positive impacts on weekly Uber demand. The percentage of the population that is African American, over 65, and larger households all had negative impacts on Uber demand. Within the model, the weekly difference had a positive effect on Uber demand, it was included at higher order to capture the observed tailing over time, however the results were insignificant. Interaction variables were then included to test our hypothesis for observed declines in demand over time with respect to various built environment variables. A positive influence of retail area across time was determined. Perhaps the most substantial finding, a negative relationship was observed when exploring the influence of percent residential area across time. This negative relationship lines up with the decreasing trends observed in the descriptive analysis. Together, these results offer substantial evidence in support of a trailing in Uber demand over time in residential dominant areas.

These last findings have substantial impact beyond the scope of just Uber demand analysis and have potentially important policy implications for the overall adoption of dynamic shared modes. The significant tapering off within residential zones indicates that even as the demand grows at a citywide scale, residents’ inertial modal biases and familiarity with existing modes may limit TNC adoption by all individuals within individual subareas (57). A second policy implication that can be drawn here is that users are typically motivated to change modes if the alternative provides improvement over existing offerings (e.g. lower costs, or shorter travel times). In the current operational framework, Uber and other TNCs most often replace single occupancy (SOV) auto-modes further contributing to the effects of congestion and hindering overall adoption (20, 23). Further investigation into these effects is warranted.
As one of the first works to explore demand across multiple shared modes in a collective way, Chapter 3 represents a substantial first pass in understanding the changes in shared mobility due to the expansion of TNCs in a real world context. By using observed data on pickup trips for TNCs, taxis, bikeshare and subway entrances counts in New York City this study takes a data driven approach to exploring the interactions in demand across shared modes. In addition to these influences, the effects of weather, federal holidays and publicly sanctioned events are also explored. To accommodate the temporal dependence of the data, the investigation is completed using a Dynamic Linear Modeling framework that structurally decomposes the observed demand into a trend and seasonal component.

At a high level, the demand for TNCs has grown substantially across the study period. The average number of TNC pickups increased by over 200% from 2015 to 2017 while simultaneously the average number of pickups by taxis fell by about 20% over the same period. The average number of daily trips for bikesharing has also grown over the same study period by about 40%, possibly linked for the continued commitment to biking infrastructure. As the clear shared mode king, the subway system moves substantially more individuals than all the other modes combined. Over the study period, average subway demand during the week was essentially constant, -0.1% change, but a slight decline of about 4% was noted on the weekends. These fluctuations in specific modes however do not entirely capture the overall growth in the total shared demand across all the studied modes. The average number of observed shared trips during a weekday increased by 2.4% while trips on an average weekend day increased by 1.4%. This overall growth may indicate previously unmet demand is now being satiated, but further work should be done to substantial this claim.
Simply noting the changes in average trip rates cannot capture how the demand is shifting over time. To uncover these relationships, a DLM explaining the demand for each of the TNC was built. The model was also developed to capture the temporal trends, and the influence of other exogenous factors along with the demand for the three other shared systems. The key findings indicate the demand for these systems to be largely contemporaneous, highlighting the unprecedented ease in which people can make use of any or all of these modes. The results indicate growing demand for taxi and bikeshare appears to lessen the demand for TNCs. Recent acquisitions of bikeshare providers by the major TNC operators may lend evidence to this relationship as they move to fully capture a segment of the shared mobility market (56). In the long run, a better integration of bikeshare and TNC services may promote better efficiency across both modes while simultaneously promoting more sustainable and active travel. When it comes to public transit and TNCs, the complementary relationship identified in the existing literature is further confirmed. Partnerships between TNCs and transit agencies have already been expanding to fill travel outside of service hours or where demand does not constitute a dedicated route (44). In the context of NYC, where the subway plays a vital role in the transportation network, it appears growing TNC demand may not mark its demise. Without available disaggregate data on the bus network, no comments can be said as to how growing TNC demand is impacting bus service. In addition to the policy insights, the model developed can also be used to carry out forecasting to support short- and long-term operations and planning applications.

It is important to note that these findings are indicative of a simple univariate approach with other modes induced as exogenous predictors. This formulation operates under the limited assumption of independence across variables. There are also potential issues with
multicollinearity and endogeneity within the data and as such the final conclusions may be subject to bias. This is discussed further in the following section, and future work is proposed to accommodate these issues.

4.2 Limitations and Future Work

Both studies represent substantial first steps in understanding dynamic ridesharing in a real world context, but it is important to recognize some of their limitations. First, holding TNC pickups as a stand in for the demand in both studies can be problematic as it inherently omits any demand by users who attempted to request rides but were not able to have those requests filled. It also essentially omits any user without a smartphone or knowledge of transportation networking companies, which may be disproportionately represented by elderly and low income groups. It should be noted that the continued growth of TNC popularity and smartphone adoption may eliminate this problem in the future. Second, with data only on the pickup event of each trip, a rich understanding of how these trips are integrating into the existing network is not possible. Third, the spatially aggregated nature of the data limited the analysis in chapter 2 to the neighborhood level offered by taxi zones, but was not a problem for the aggregate approach taken in chapter 3. The data quality and structure issues noted with the TNC and subway data motivates an important discussion. As open data becomes more readily available, it is vital for open data providers to adopt a standard of practice for quality and structure that can in turn promote additional future research endeavors.

From a methodological stand point, both studies face limitations as well. In Chapter 2, the panel model assumes strong exogeneity which may not be supported by the data and alternate model forms must be explored. In Chapter 3, by expressing demand for TNC as a univariate model with other demands as exogenous predictors, the covariance between modes cannot be
accurately captured. Also, because the demand for all modes appears to be related to common observed and unobserved factors, there may be endogeneity and multicollinearity issues at play. Early evidence to this effect was observed when similar models were developed to understand the impact of three other modes on the demand for taxi, bikeshare, and subway. Tables 7 and 8 present these modeled results.

Table 7. DLM results for Taxi with Other Modal Demands

<table>
<thead>
<tr>
<th></th>
<th>estimate</th>
<th>SE_taxi</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_w1 )</td>
<td>2.876E-02</td>
<td>1.471E-03</td>
<td>19.55</td>
</tr>
<tr>
<td>( \sigma_w2 )</td>
<td>-8.571E-07</td>
<td>4.281E-04</td>
<td>0.00</td>
</tr>
<tr>
<td>( \sigma_v )</td>
<td>4.888E-03</td>
<td>3.922E-03</td>
<td>1.25</td>
</tr>
<tr>
<td>Precipitation (in.)</td>
<td>-2.421E-02</td>
<td>3.676E-03</td>
<td>-6.58</td>
</tr>
<tr>
<td>Holiday Indicator</td>
<td>8.020E-03</td>
<td>6.700E-03</td>
<td>1.20</td>
</tr>
<tr>
<td>NYC Peak Travel (Sep-Dec)</td>
<td>-6.777E-04</td>
<td>2.420E-03</td>
<td>-0.28</td>
</tr>
<tr>
<td>Number of Event Permits</td>
<td>3.880E-05</td>
<td>1.243E-05</td>
<td>3.12</td>
</tr>
<tr>
<td>drift, ( \delta )</td>
<td>-1.033E-01</td>
<td>7.748E-03</td>
<td>-13.33</td>
</tr>
<tr>
<td>TNC Demand (Mil. Trips)</td>
<td>5.124E-02</td>
<td>1.178E-02</td>
<td>4.35</td>
</tr>
<tr>
<td>Taxi Demand (Mil. Trips)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Bikeshare Demand (Mil. Trips)</td>
<td>-5.656E-01</td>
<td>1.011E-01</td>
<td>-5.59</td>
</tr>
<tr>
<td>Subway Demand (Mil. Trips)</td>
<td>2.255E-02</td>
<td>1.547E-03</td>
<td>14.58</td>
</tr>
<tr>
<td>MAPE</td>
<td></td>
<td>20.23%</td>
<td></td>
</tr>
<tr>
<td>Difference in MAPE</td>
<td></td>
<td>14.63%</td>
<td></td>
</tr>
</tbody>
</table>

Table 8. DLM Results for Bikeshare and Subway with Other Modal Demand

<table>
<thead>
<tr>
<th></th>
<th>Model 2. Bikeshare</th>
<th></th>
<th>Model 2. Subway</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>estimate</td>
<td>SE_citi</td>
<td>t</td>
<td>estimate</td>
</tr>
<tr>
<td>Phi (( \varphi ))</td>
<td>4.773E-01</td>
<td>2.139E-02</td>
<td>22.32</td>
<td>-4.657E-02</td>
</tr>
<tr>
<td>( \sigma_w1 )</td>
<td>-5.695E-03</td>
<td>1.249E-04</td>
<td>-45.58</td>
<td>-1.271E-05</td>
</tr>
<tr>
<td>( \sigma_w2 )</td>
<td>2.074E-05</td>
<td>1.412E-04</td>
<td>0.15</td>
<td>1.259E-01</td>
</tr>
<tr>
<td>( \sigma_v )</td>
<td>-4.845E-04</td>
<td>4.584E-04</td>
<td>-1.06</td>
<td>-3.228E-01</td>
</tr>
<tr>
<td>Precipitation (in.)</td>
<td>-1.473E-02</td>
<td>6.446E-04</td>
<td>-22.85</td>
<td>3.082E-02</td>
</tr>
<tr>
<td>Holiday Indicator</td>
<td>-2.750E-03</td>
<td>1.268E-03</td>
<td>-2.17</td>
<td>-1.313E+00</td>
</tr>
<tr>
<td>NYC Peak Travel (Sep-Dec)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8.133E-02</td>
</tr>
<tr>
<td>Peak Bike (May-Oct)</td>
<td>2.508E-03</td>
<td>6.486E-04</td>
<td>3.87</td>
<td>-</td>
</tr>
<tr>
<td>Number of Event Permits</td>
<td>-1.111E-05</td>
<td>2.956E-06</td>
<td>-3.76</td>
<td>-3.842E-04</td>
</tr>
<tr>
<td>Max Daily Temp (F)</td>
<td>2.813E-04</td>
<td>1.842E-05</td>
<td>15.27</td>
<td>-</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-------</td>
<td>---</td>
</tr>
<tr>
<td>TNC Demand (Mil. Trips)</td>
<td>1.605E-02</td>
<td>2.456E-03</td>
<td>6.53</td>
<td>3.068E+00</td>
</tr>
<tr>
<td>Taxi Demand (Mil. Trips)</td>
<td>-3.502E-02</td>
<td>3.521E-03</td>
<td>-9.95</td>
<td>9.799E+00</td>
</tr>
<tr>
<td>Bikeshare Demand (Mil. Trips)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.429E+01</td>
</tr>
<tr>
<td>Subway Demand (Mil. Trips)</td>
<td>2.569E-03</td>
<td>3.624E-04</td>
<td>7.09</td>
<td>-</td>
</tr>
<tr>
<td>MAPE</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Difference in MAPE</td>
<td>-15.46%</td>
<td>-7.55%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bringing the focus to the TNC demand variable (highlighted in red) in the 3 models in Tables 7 and 8, it is apparent that for the Taxi and Bikeshare models, the direction of influence swaps from negative to positive. Because the demand for all modes appears to be linked, the presence of multicollinearity is likely in the models including demand for other modes as exogenous predictors. To confirm this, the models were re-estimated with only the demand from one other mode as an additional predictor. In doing this, some substantial variations are noted in the modeled coefficients, which lend credence to the presence of issues. Future endeavors to overcome these issues are presented below.

As part of future work to confirm and improve the findings of Chapter 2, a comprehensive spatial analysis should be done to further explore and characterize the spatial influence on Uber demand. Also, to supplement and confirm the findings presented here, we anticipate further exploration through a primary survey which will collect disaggregate level information on users and their preferences for TNC providers and other modes.

To address the limitations in the second study, a few different methods are slated for future work. The first of these being a full multivariate DLM specification that allows for accurate covariate relationships to be specified. The second method to be explored in future work is vector based auto regression (VAR) using compositional time series similar to work done by Serhiienko et al (58). In this approach, the total shared demand is assumed to be captured by the
available modes and the proportion of each mode is what is investigated. This approach would allow for comments on any potential cannibalistic effects present across modes. Finally, spatio-temporal models investigating this data at a smaller spatial resolution will be explored to identify how location may impact shared demand.
REFERENCES


APPENDIX

Figure A1. Normality of Initial Modeled Residuals (Innovations) for models without demand for other modes
Figure A2. Normality of Initial Modeled Residuals (Innovations) for models with demand for other modes