University of California, Irvine

ABOUT TIME: A PEDAL IN THE RIGHT DIRECTION
A travel-time based estimation framework for modeling bikeshare demand

THESIS

submitted in partial satisfaction of the requirements
for the degree of

Master of Science

in

Transportation Science

by

Deep Suryakant Shah

Thesis Committee:
Professor Jean-Daniel Saphores, Chair
Assistant Professor Michael Hyland
Professor Michael G. McNally

2020

DEDICATION

Shraddha, Sagar, Rekha and Suresh

For your advice, your patience, your faith and your freedom.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

Dr. Jean-Daniel Saphores

Thank you for inspiring me to do a thesis. Performing research under your guidance has taught me a key life-lesson: keep it simple. Learning has been an adventure with you. Thank you for the five minutes of music before your classes, and for the five minutes of personal talks before our meetings. You will continue to be a role model for staying a lifelong learner. Your guidance has been an integral support in the successful completion of this thesis.

Dr. Michael Hyland

I was a completely different thinker in the transportation domain before taking your class. It has prompted me to spend some time on raising questions, and even more on answering them. Your expertise in the domain of bikeshare has been critical to making my thesis technically sound. Thank you.

Dr. Michael G. McNally

Attending your lectures has been equal parts technical and philosophical. Thank you for helping me realize that common beliefs need not be true. Your review has made my thesis, complete and me, confident.

ABSTRACT

ABOUT TIME: A PEDAL IN THE RIGHT DIRECTION
A travel-time based estimation framework for modeling bikeshare demand

Deep Suryakant Shah

Master of Science in Transportation Science

University of California, Irvine, 2020

Professor Jean-Daniel Saphores, Chair

The recent growth in bikeshare systems has received an enthusiastic response from the research community interested in understanding the factors that influence bikeshare demand. Many research efforts have modeled spatial interactions using a distance-based weight matrix. However, when biking for utilitarian reasons, biking time may be equally, if not more, relevant. To this end, this thesis explores the demand for bikesharing in downtown Los Angeles by contrasting two spatial SARAR models with the same explanatory variables but different weight matrices: one has a distance-based weight matrix, and the other a time-based weight matrix. To the best of my knowledge, this work is the first to contrast these models and to analyze the demand for Los Angeles' bike sharing program. Explanatory variables in my spatial models include socio-demographic, land use and transport characteristics in the proximity of bikeshare stations. The results show that (i) incorporating spatial interactions (spatial lag and spatial errors) is an important feature of bikesharing demand; and, (ii) that models with time-based weight matrices perform better than similar models with distance-based weight matrices for the models I

considered. Finally, my models show that bikesharing is seldom used as a mode for the ‘last mile’ travel to access transit in downtown Los Angeles.

1. INTRODUCTION

Bike sharing is a sustainable mode of transportation that generates no greenhouse gases, promotes health, saves money, could decrease congestion if used broadly, and could help address the ‘last mile’ problem (Shi et al. 2018). It was first introduced in Amsterdam in 1965, in what is now often referred to as a first generation bikeshare program. Since then, bike sharing as a transportation mode has gained worldwide popularity supported by technological advancements in the domains of mobile payment, GPS tracking, and big data. Campbell et al. (2016) identified the third generation bikeshare as one of the fastest growing modes of transport, characterized by docking stations, automated credit card payments, radio frequency ID tags and GPS technology. Globally, more than 800 cities currently offer one or more bikeshare programs (Fishman, 2015). In the US, there are close to 120 city bikeshare programs with a variety of functioning mechanisms including private, public and private-public partnerships.

This global rise in bikeshare systems has received an enthusiastic response from the bikeshare research community over the last decade. Si et al. (2019) organizes the bikeshare literature into four clusters: (a) factors and barriers, (b) system optimization, (c) behavior and impact, and (d) safety and health. The ‘factors and barriers’ cluster focusses on the effect of external variables on bikeshare demand; including built environment, infrastructure, weather and station location. Studies in this cluster analyze either trip, transaction or survey data at different scales, ranging from city-level to trip-level, and for different periods ranging from daily to monthly.

My thesis falls in the 'factors and barriers' cluster and is aimed at addressing the tactical challenges associated with the planning of a bikeshare system. The planning process includes siting and sizing of bikeshare stations, which is largely informed by demand models. The purpose of this thesis is to improve the existing demand modeling framework and subsequently improve the planning and modification of bikeshare systems. To capture the effects of external variables that influence the network design, this thesis adopts a station-level bikeshare demand modeling framework for the Los Angeles Metro Bikeshare Program using monthly data. In this context, my thesis makes two contributions to the existing literature.

First, the combination of short-duration trips (Fishman, 2015) and the requirement to start and end bike trips at designated bike stations makes it natural to expect spatial interactions between nearby bike stations. Earlier studies have attempted to incorporate these interactions using spatial models with network distance-based weight matrices. In this context, my first hypothesis is that travel behavior is influenced by actual biking distance, but even more so by biking time. To investigate this conjecture, I estimated three models: 1) a multiple linear regression model as a baseline; 2) a spatial econometric model with a weight matrix based on biking distance; and 3) a similar spatial econometric model with a weight matrix based on biking time. Results show that spatial interactions (spatial lag and spatial error) are important, and support my hypothesis that the use of travel time over travel distance offers a better estimation framework for modelling bikeshare demand. To the best of my knowledge, this study is the first one to explore time-based matrices for the spatial modeling of a bike share system.

My second contribution is to study the demand for bike sharing in Los Angeles, which is to the best of my knowledge, the largest system in the U.S. that has not yet received attention from academics. Even though the literature on bike sharing is growing, it is noteworthy that the same set of explanatory variables have been observed to cause contrasting effects in different bikeshare studies. For instance, Maurer (2011) identified the number of jobs to negatively impact bikeshare demand in Minnesota, while Rixey (2013) found it to positively impact bikeshare demand in Washington DC. Fishman et al. (2014) found more bikeshare demand at stations with relatively less accessible public transit opportunities in Melbourne and Brisbane, while Shaheen et al. (2014) identified bikeshare usage to be significantly higher near rail stations in Paris and Washington DC.

The remainder of this thesis is organized as follows. Chapter 2 summarizes selected papers from the bike-sharing literature and reviews some key papers describing the spatial econometric models used in this research. It also introduces the Los Angeles Metro Bikeshare Program. Chapter 3 presents my data, and Chapter 4 details the models estimated in this study. Chapter 5 then discusses my results. Finally, Chapter 6 summarizes key results, mentions some limitations of my analyses, and provides suggestions for future research.

2. LITERATURE REVIEW AND BACKGROUND

Bikeshare systems have been receiving increasing attention from researchers over the past decade as they have become more popular. For an excellent review of older papers, see Shaheen et al. (2010), who present a comprehensive overview on the history, present, and future of bikeshare systems across Europe, America, and Asia. Shaheen et al. (2010) also analyzed bikeshare systems, their business models, their environmental and social effects, and the lessons learned from their planning and implementation. The literature review in this chapter focuses on studies directly relevant to the modeling of station-level bikeshare usage and spatial econometric models. For a thorough review of bikeshare planning and implementation, please refer to Fishman (2016) and Si et al. (2019). Following Vogel et al.'s (2009) classification, I focus here on the strategic challenges associated with the sizing and siting of bikeshare stations in the Los Angeles' bikeshare program.

Section 2.1 discusses variables that were found to matter in selected station-level bikeshare demand studies. Section 2.2 then focuses on approaches used to model station-level bikeshare demand, with a particular interest for spatial econometric modeling, which is the methodology adopted in this thesis. Finally, Section 2.3 gives a brief overview of the Los Angeles bikeshare program.

2.1 Explanatory Variables in Station-Level Bikeshare Demand Modeling

This section reviews prior research studies that examine the relationships between station-level bikeshare usage and various explanatory factors. Table 1 presents a summary of previous studies that model station-level bikeshare demand; it includes the explanatory variables considered in each study.

The following socio-demographic variables are found significant in the literature: population (Buck and Buehler, 2012; Faghih-Imani et al., 2017), age (Fuller et al., 2011; Wang et al., 2015; Hyland et al., 2017), race (Maurer, 2011; Daddio 2012), education (Fuller et al., 2011; Rixey, 2013; Hyland et al., 2017), income (Rixey, 2013; Guidon et al., 2019), family structure (Faghih-Imani et al., 2017; Hyland et al., 2017), car ownership (Maurer, 2011; Daddio, 2012; Buck and Buehler, 2012), jobs (Maurer, 2011; Rixey, 2013; Faghih-Imani and Eluru (2015), El-Assi et al., 2018) and commute patterns (Maurer, 2011; Rixey, 2013, Hyland et al., 2017). However, the impacts of these variable in different studies are not necessarily similar. Faghih-Imani et al. (2017) find one-person households to have a negative influence on bikeshare demand, while Hyland et al. (2017) find family households to have a negative impact on bikeshare demand. While Maurer (2011) and Faghih-Imani and Eluru (2015) report that the number of jobs has a negative impact on bikeshare demand, Rixey (2013) and El-Assi et al. (2018) report instead that it has a positive impact. The impact of low-car ownership households on bikeshare demand is found to be both positive (Buck and Buehler, 2012) and negative (Maurer, 2011; Daddio, 2012).

The following land use variables are also found significant: residential (Kim et al., 2012; Sun and Chen, 2017), commercial (Kim et al., 2012, Sun and Chen, 2017), schools and universities (Wang et al., 2015; El-Assi et al., 2018), and parks (Kim et al., 2012; Faghih-Imani and Eluru, 2015). Moreover, the distance of a bikeshare station from various points of interest like restaurants and bars (Faghih-Imani et al., 2014; Hyland et al., 2017), grocery stores (Buck and Buehler, 2012), central business district (Wang et al., 2015; Faghih-Imani and Eluru, 2015), and hotels (Daddio 2012; Faghih-Imani et al., 2017) have also been found significant. Reported impacts of these variables again show contrasting impacts on bikeshare demand. Kim et al. (2012) find that residential and commercial land use has a positive impact on bikeshare demand, while Sun and Chen (2017) find the reverse. Similarly, the impact of hotels in the proximity on bikeshare demand is reported to be both positive (Faghih-Imani et al., 2017) and negative (Daddio 2012).

Another commonly employed variable to explain bikeshare demand is the proximity of bike share stations to transit stations (Fishman et al., 2014; Faghih-Imani and Eluru, 2015; Hyland et al., 2017; Guidon et al., 2019). While Fishman et al. (2014) find areas with lower transit accessibility to experience higher bikeshare demand, the remaining studies find proximity of transit to positively impact bikeshare demand. The location of bike stations with respect to other bike stations has also been a frequently studied aspect of bikeshare demand. Rixey (2013) concludes that adding more stations within 4800m of a bikeshare station increases its demand. Hyland et al. (2017) also find that the number of bikeshare stations within 1-5km of a given station has a positive impact on its usage; however, the number of stations within 0.8 km of the station has a negative impact on its usage. Faghih-Imani et al. (2014) reach a different conclusion: for them adding more stations within 250m of a bikeshare station increases its demand. The length of bike lane network in the proximity of a bikeshare station is another variable that positively influences its demand (Buck and Buehler, 2012; Rixey, 2013; Wang et al., 2015; Faghih-Imani and Eluru, 2016; Guidon et al., 2019). Finally, adverse weather conditions like precipitation and low temperatures have a negative impact the bikeshare demand (Rixey, 2013; Faghih-Imani et al., 2014; El-Assi et al. 2019).

Table 1. Summary of studies on station-level bikeshare demand

Author	Choice of variables (bikeshare system)
Maurer (2011)	Population, race, income, car ownership, commute patterns, proximity to jobs and transit, land use, bike lanes, station capacity (Minneapolis)
Fuller et al. (2011)	Age, education, employment, gender, commute mode, station proximity (Montreal)
Daddio (2012)	Age, race, income, land use, proximity to transit, distance from bike station cluster center, bike lanes, car ownership (Washington DC)
Kim et al. (2012)	Land use, weather, day of the week (Goyang, Korea)
Buck and Buehler (2012)	Population, mode-share, income, mode share, car ownership, bike infrastructure, proximity to transit and grocery (Washington DC)
Rixey (2013)	Population, income, race, education, built environment, car ownership, proximity to transit, weather (Washington DC, Denver, Minnesota)
Faghih-Imani et al. (2014)	Weather, biking infrastructure, station capacity, land use, proximity to transit, day of the week, population density, job density (Montreal)
Gebhart and Noland (2014)	Time of day, day of the week, month, weather (Washington DC)
Faghih-Imani and Eluru (2015)	Proximity to transit, station proximity, station capacity, land use, distance to CBD, population density, job density (Chicago)
Wang et al. (2015)	Race, age, business density, proximity to other stations, station age (Minneapolis - St. Paul)
Faghih-Imani and Eluru (2016)	Weather, biking facility, land use, station proximity, station capacity, population density, job density, proximity to transit (New York)
Hyland et al. (2017)	Age, education, family structure, jobs, homicides, car ownership, commute, proximity to transit, weather, bike lanes, land use (Chicago)
El-Assi et al. (2017)	Day of the week, weather, proximity to transit, population density, job density, land use, station capacity (Toronto)
Faghih-Imani et al. (2017)	Station elevation, land use, family structure, population density, time of the day, biking infrastructure, station density (Barcelona, Seville)
Sun and Chen (2017)	Weather, time of the day, day of the week, land use, household density, roadway design, transit usage (Seattle)
Guidon et al. (2019)	Weather, time of the day, day of the week, bike infrastructure, proximity to transit, population, income (Zurich)

The selection of variables and model specification in my thesis benefit from the existing research on station-level bikeshare demand modelling displayed in Table 1. Further, many of these variables have shown contrasting impacts in different studies. The justifications for these impacts by the individual papers will support the interpretation of model results for the current study. Moreover, since this is the first study using the LA bikeshare system, results from this study may differ from published results for other areas given the unique attributes of Downtown LA, which include the absence of single-family residences, a pleasant weather, and a dense network of bus stations.

2.2 Models used to Model Demand in Station-Level Bikeshare Systems

Spatial panel models have been extensively used for public governance, taxation, fiscal policy and economic growth (Elhorst, 2014). Spatial models have also been used to measure housing accessibility (Chalermpong, 2007; Mitra and Saphores, 2016), measure job accessibility (Wang and Chen, 2015), track land use change (Wang and Kockelman, 2006; Shen et al. 2014), analyze labor markets (Smith et al., 1981; Longhi and Nijkamp, 2007), forecast housing prices (Pace et al., 2000; Dubé et al. 2014) and analyze air-fare prices (Daraban and Fournier, 2008; Zhang and Wang, 2015).

However, spatial models are not commonly employed to study bikeshare demand. Most of the early literature uses linear regression to model station-level bikeshare demand (Maurer, 2011; Daddio, 2012; Kim et al., 2012; Buck and Buehler, 2012; Rixey, 2013). Both Faghih-Imani et al. (2014) and El-Assi et al. (2017) use a linear mixed model, while Hyland et al. (2017) employ a hybrid k-means clustering and multi-level mixed regression modeling approach. Faghih-Imani et al.

(2016), Sun and Chen (2017) and Guidon et al. (2019) appear to be the only studies so far that account for spatial and/or temporal dependencies between the bikeshare stations.

Faghih-Imani et al. (2016) compare results between (i) simple models without considering spatio-temporal effects; (ii) spatial error models with and without observed spatio-temporal effects; (iii) spatial lag models with and without observed spatio-temporal effects. They find that the inclusion of spatial and temporal effects leads to more accurate predictions and the proximity to gardens, subway stations, and areas with high population and job densities positively impact bikeshare demand.

Sun and Chen (2017) use a generalized mixed linear model that addresses potential temporal and spatial autocorrelations for the Pronto bikeshare system in Seattle. The study finds (i) proximity to transit stations have a substitution effect on bikeshare, (ii) workers from office buildings as the major source of public bikeshare customers.

Guidon et al. (2019) estimate a spatial regression model to analyze the electric bike trip transaction data from Smide bikeshare system in Zurich, Switzerland. They report that economic and social activity are key drivers of bikeshare demand, which is also positively impacted by public transportation availability.

As the first study to use a travel time-based weight matrix for spatial modeling, the current study benefits from the methods used for calculating weight matrices in the earlier studies. Moreover, the calculation and interpretation of impact measures in the earlier studies supports the interpretation of results for the two spatial models used in the current study.

2.3 Los Angeles' Metro Bikeshare Program

The Metro Bikeshare Program in LA is a public bicycle-sharing system operated by a partnership between LA Metro, the City of Los Angeles, and the Port of Los Angeles. This program started in August 2016. Its goal was 'to develop an affordable, user-friendly bikeshare program that increases ridership by integrating with the County's regional transit services' (Metro Bikeshare Business Plan, 2019-20). It is a docked bikeshare network, so users are required to start and end all their trips at designated bikeshare station locations. After starting with 61 stations and 700 bicycles, the program has progressively expanded and, as of the end of 2018, it included 82 stations and over 1,400 bicycles.

Registered users can subscribe for daily/monthly/annual passes, that can be purchased either at a station kiosk or online using a credit/debit card. Non-registered users can pay directly at the station kiosk using a credit/debit card. The Transit Access Pass (TAP) card made available by LA Metro can also be used to access bikeshare after a single-time online enrollment. Currently, the pricing is set at $150/year, $17/month, $5/day and $1.75/trip, with additional charges for trips longer than 30 minutes. The program is functional 24 hours/day and can be used by all individuals above the age of 16 years. However, riders under the age of 18 years need to have a parent/guardian purchase a pass for them.

3. DATA

This chapter describes the dependent and explanatory variables used in my models. Section 3.1 details my dependent variable, while Section 3.2 discusses my explanatory variables. Finally, Section 3.3 presents descriptive statistics of both dependent and explanatory variables.

3.1 Dependent Variable

This study focuses on supporting the strategic planning of bikeshare networks, rather than their operations. Since the purpose is to understand macro-level factors that influence bikeshare demand without accounting for temporal fluctuations, I selected the monthly count of trips originating at each station as my dependent variable. The same approach could be used to model trip arrivals (i.e. station attraction). In the interest of time and space, I only model and present trip generation.

I obtained the trip count data from Los Angeles' Metro Bikeshare program website (https://bikeshare.metro.net/about/data/). The data contain a record for each trip, along with its duration, timestamp, origin and destination station details, membership status, and a record of the bike used. For this analysis, I selected trips in the month of September 2018 for two reasons. First, according to the data, September was the month with the highest number of bike trips in 2018 (30,194 trips with a daily average of 1,006). Second, no new bike stations were opened during that month. Previous analyses suggest the first month a bikeshare station is in operation is unlikely to represent typical usage (Gebhart and Noland, 2014, Hyland et al., 2017). This was confirmed by empirical results for the related data. For the stations that opened in 2018, the average rise in the monthly trip count from the first to second month was 81%, followed by 13%, 7% and 5% for the respective months. After examining the data, I removed trips with the same origin and destination station with duration under 3 minutes (1,775 trips) because it is unlikely that such trips met the user's demand for travel. I also used the stations' coordinates to perform a series of data compilation exercises to create my exogenous variables.

In September 2018, there were a total of 93 bike-stations. They were clustered in three distinct regions –Downtown LA (67 stations - 20,885 trips), Venice Beach (14 stations - 805 trips) and Long Beach (12 stations - 6,729 trips). The individual clusters attracted different trip purposes, and thus could not be modeled together. Since the Downtown LA cluster experienced most (73.5%) of the monthly trips, I focused on this cluster. The final sample consists of 20,885 bike trips originating from 67 stations. Descriptive statistics are presented in Table 2. Figure 1 shows the location of the three clusters, the distribution of stations, and the demand for the Downtown LA cluster.

3.2 Explanatory Variables

The exogenous variables used in this study can be broadly divided into three categories: (i) socio-demographic variables; (ii) land use variables; and (iii) transport variables. Descriptive statistics of these variables are presented in Table 2. While existing studies incorporate some weather variables (Faghih-Imani et al., 2014; Hyland et al., 2018), I did not include any here because Los Angeles experiences pleasant year-around weather, with very few rainy days. In Downtown LA during September 2018, the temperature ranged between 61 °F and 91 °F, with an average of 71

°F. Moreover, the humidity during that time ranged from 28% to 94%, with an average of 71% (https://timeanddate.com). Additionally, since the bike stations were closely located, any impact of weather was equally experienced by all stations.

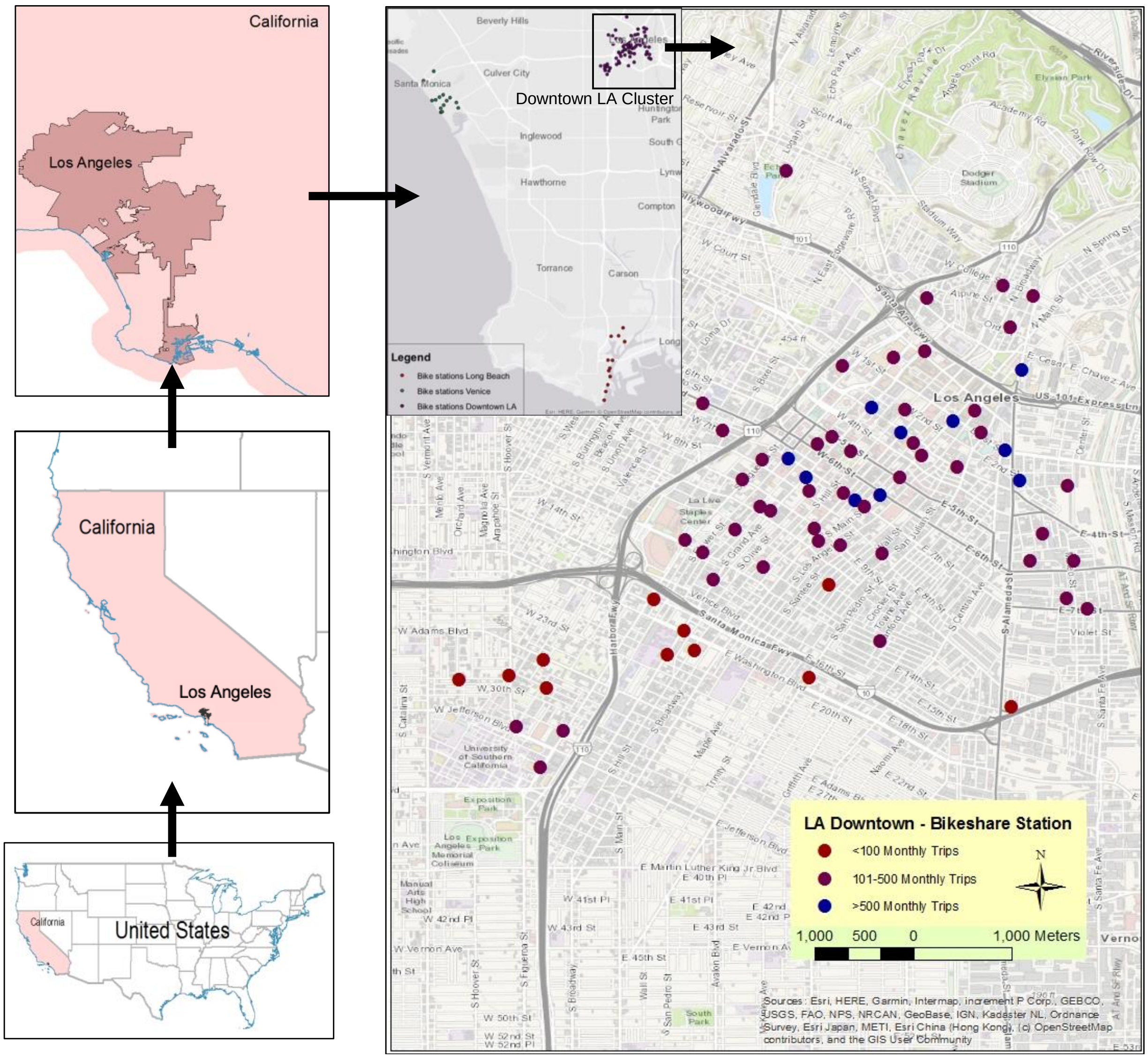

Figure 1. Location and Map of LA Downtown bikeshare cluster

3.2.1 Socio-demographic variables

For my socio-demographic variables, I relied on the 2017 American Community Survey five-year estimates (ACS, U.S. Census Bureau 2017), which provides socio-demographic information at the census block group level. The census block group is the lowest available level of aggregation for the required data. Based on the literature review (see Chapter 2), I collected population density, age, education, income, car ownership, and family structure data from ACS. I used ArcGIS to attach each bikeshare station with the attributes of the census block group it was located in.

Earlier studies have attempted to explain the effect of the younger population on bikeshare demand using different age brackets - Daddio (2012) used the 20-39 age bracket while Hyland (2017) used the 25-44 age bracket. Anyone above the age of 18 has unrestricted access to LA's Metro Bikeshare Program. Therefore, I started by using two separate age bracket variables – 18-29 years and 30-45 years. The two variables caused multicollinearity when used simultaneously. Thus, I selected a single variable with the combined 18-45 age bracket for my final model. Following Shaheen et al. (2010) and Buck and Buehler (2012), I used the number of carless households to represent car ownership, since these households are more likely to use a bicycle. I calculated the family structure variable as the number of households with children, since parents are less likely to use bike as a transport mode (Hosford et al., 2018). I also tried a variable for the population living in family households, but it increased the spatial error value over one in my spatial model with travel time-based weight matrix. Finally, I used the 2016 Business Patterns dataset to collect the employee data at the zip code level, which is the lowest level of aggregation available.

The simultaneous use of income (median household income), education (number of college graduates) and the number of car commuters as explanatory variables also resulted in multicollinearity. Trying multiple combinations in the model showed that two of these three variables need to be removed. Since car ownership can partially account for the number of car commuters, I removed it first from my models. To choose between the income and education variable, I ran individual models with each variable. The coefficient values and significance levels of all the other variables remained similar. Thus, I kept the model with the education variable on the account of its lower AIC and BIC values (difference of 11 each).

3.2.2 Land use variables

Kim et al. (2012), Faghih-Imani et al. (2014), as well as Sun and Chen (2017) have all used the shares of different types of land use near a bikeshare station to explain its demand. The unit of measurement used for these studies is either a census tract or a radial buffer ranging from 250m to 500m. For this study, I used SCAG's (Southern California Association of Governments) 2016 land use dataset, updated as of November 2018. This dataset includes 17,315 land parcels in the city of Los Angeles. The average parcel size is 1480 sq. m., excluding all roads and highways, as shown in Figure 2. To capture land use in my models, I calculated the share of areas for residential, commercial, open/recreational, public facilities and university (University of Southern California) land in the 500m radial buffer around every bike station using ArcGIS.

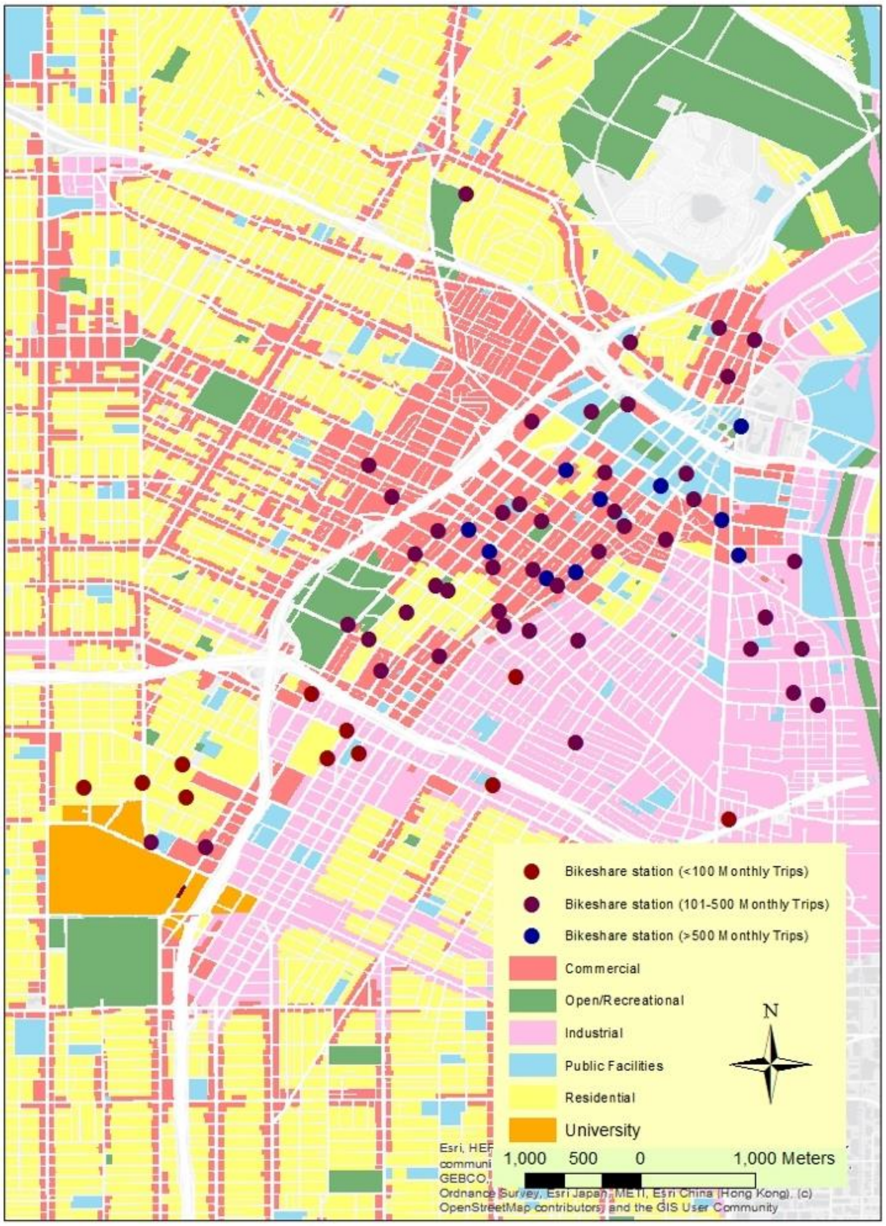

Figure 2. Land use near downtown LA bikeshare stations

Restaurant and bar density around a bikeshare station has also been shown to impact bikeshare demand (Hyland et al., 2017; Wang et al., 2018, Guidon et al., 2019). Considering the heavy share of residential and commercial land uses in Downtown LA, I expected restaurant locations to be frequently visited and thus, might influence bikeshare demand. To check for this influence, I incorporated data from the 'Restaurants in LA' dataset made available by the Office of Finance with the City of Los Angeles. Following Hyland et al. (2017), I created a variable that counts the number of restaurants within a 300m radial buffer of every bikeshare station.

3.2.3 Transport variables

One way to evaluate how bike sharing addresses the last-mile connectivity problem is by examining the relationship between transit arrivals and bikeshare departures in proximity. To that end, I gathered variables describing average daily passenger arrivals by bus/rail in the 100m buffer surrounding each bike station. I chose this radius because the average block size in LA is 100m x 100m, and also because bike stations are relatively close to each other. The transit data were obtained upon request from the Los Angeles County Metropolitan Transportation Authority. Figure 3 shows the location of transit stations with respect to the location of bikeshare stations. Based on the literature (Rixey, 2013; Faghih-Imani et al., 2014, Fishman et al., 2016), and the objective of LA's Metro Bikeshare Program of supporting last mile travel to increase transit usage, I expected the proximity to transit stations with high ridership to have a positive impact on the demand at a bikeshare station.

According to Shaheen et al. (2012), commuting (travel to/from work or school) is the most common trip purpose based on surveys of four of North America's largest bikeshare programs. Since one objective of LA's bikeshare program is to serve last-mile transportation needs and increase transit usage, measuring the extent of multimodal commute trips with transit could explain part of the demand for bikeshare. However, data that can connect intermediate trips in a

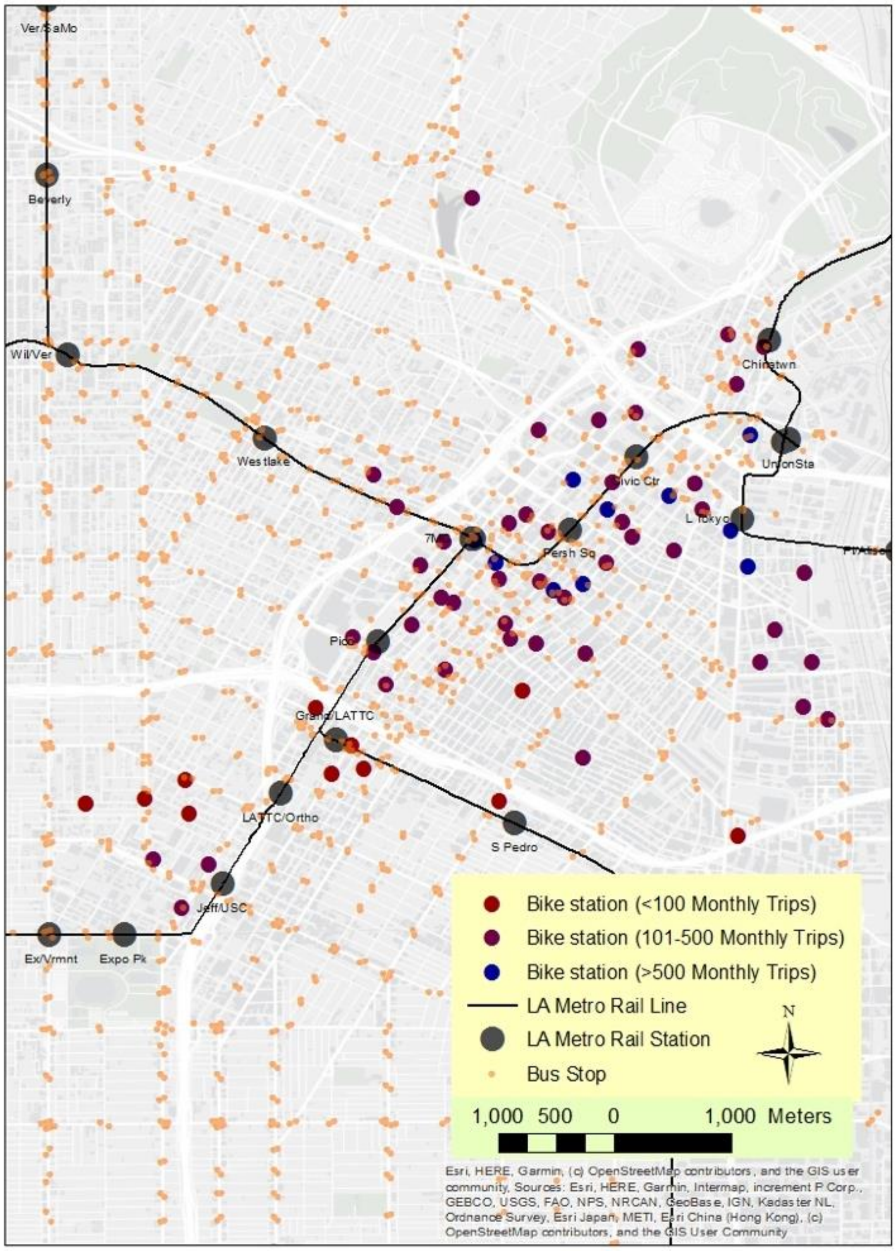

Figure 3. Transit stations near downtown LA bikeshare stations

bike-linked-transit trip are hard to gather. Since transfer penalties make multi-modal trips longer (Schakenbos, 2014; Garcia-Martinez, 2018), a possible proxy could be the number of transit passengers with longer commute times. Chakrabarti (2017) found that the average commute time using transit in Los Angeles is 69 minutes. Therefore, I used in my models the number of long-duration (>45 minutes) transit commuters at the census block group level from the 2017 ACS (U.S. Census Bureau, 2017). A positive coefficient would support the hypothesis that bikeshare is being used as a first/last mile mode to access transit.

Another variable likely to promote the demand for bikeshare is biking infrastructure (Wang et al., 2015, Wang and Chen, 2020). The provision of additional bike lanes not only improves the safety of riders but is also likely to reduce their travel time. To measure the relevant infrastructure, I used the length of the bike lane network within a 1.6-km radial buffer around each bikeshare station. Data were obtained from the LA Department of Transportation. I analyzed them using ArcGIS to create obtain station-specific values.

3.3 Descriptive Statistics

To improve model linearity based on a graphical exploration, I took the logarithmic transform of the model variables. Since the households with children, land use and transport variables include zero values, I added one to these variables before taking their logarithmic. Summary statistics are presented in Table 2.

Table 2. Summary of descriptive statistics for model variables

Variable (Source)	Unit (Spatial aggregation)	Min	Max	Mean	Std. Dev
Dependent variable					
Monthly trip count[a]	Count	20	1162	311.72	221.45
Socio-demographic variables					
Population density[b]	1000 persons/sq.km. (CBG)	0.16	24.54	4.62	5
College graduates[b]	1000 persons (CBG)	0.01	2.46	1.02	0.87
Age 18-45[b]	1000 persons (CBG)	0.45	6.36	1.60	1.07
Carless households[b]	Count (CBG)	4	1096	336.87	281.33
Households with children[b]	Count (CBG)	0	359	91.19	82.69
Employees in zip code[c]	1000 persons	3.75	42.79	27.87	10.71
Land use variables					
Residential area[d]	% (500-m buffer)	0	0.85	0.18	0.20
Commercial area[d]	% (500-m buffer)	0	0.90	0.40	0.27
Open space/Recreational[d]	% (500-m buffer)	0	0.17	0.03	0.04
Public Facilities[d]	% (500-m buffer)	0	0.60	0.09	0.14
University[d]	% (500-m buffer)	0	0.62	0.03	0.11
Restaurants[e]	Count (300-m buffer)	0	58	20.31	17.32
Transport variables					
Daily bus station arrivals[f]	1000 persons (0.1km buffer)	0	4.82	0.46	0.87
Daily rail station arrivals[f]	1000 persons (0.1km buffer)	0	55.02	0.95	6.77
Long-duration transit commuters[b]	Count (CBG)	0	340	75.97	69.99
Bike lane network length[g]	Kilometers (1.6-km buffer)	0	84.69	43.39	23.65

CBG: Census Block Group

Data sources:

a) LA Metro Bikeshare Program (https://bikeshare.metro.net/about/data/)
b) American Community Survey dataset by US Census Bureau (https://factfinder.census.gov)
c) Business Patterns dataset by US Census Bureau (https://factfinder.census.gov)
d) Land Use dataset by SCAG (http://gisdata-scag.opendata.arcgis.com/)
e) Restaurants in LA dataset by City of Los Angeles (https://data.lacity.org/)
f) LA Metro (made available upon request)
g) City of LA Bikeways dataset by LA Department of Transportation (http://geohub.lacity.org/)

4. MODELING APPROACH

This chapter presents an overview of the approaches used to model station-specific bikeshare demand. Section 4.1 provides an overview of the base modeling approach using regression. Section 4.2 describes the spatial model. Section 4.3 introduces the method for calculating the weight matrices. Finally, Section 4.4 details the method of interpretation for my models. For all the models, N designates the sample size (N=67) and Q is the number of explanatory variables including a constant (Q=17).

4.1 Overview and Base Modeling Approach

The base model approach is a multiple linear regression model without spatial interactions:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where,

Y is a *N x 1* vector of log transformed bike trip counts for each station;

X is a *N x Q* matrix of exogenous explanatory variables with log transformed continuous values;

β is a *Q x 1* vector of unknown coefficients; and

ε is a *N x 1* vector of independent and identically distributed errors

The plots of dependent variable and each independent log transformed variable show relatively linear trends. To check for multicollinearity, I relied on Variation Inflation Factors (VIF). Twelve of my sixteen variables have a VIF<5, and the remaining four have a VIF<7.

4.2 Spatial Dependence and Model

Owing to (i) the relatively short duration of most bikeshare trips, and (ii) the need for existence of a bikeshare station at the origin and destination in docked systems, it is natural to expect spatial interactions between nearby bike stations. Since the destination station is often proximate to the origin station, demand at the latter will influence demand at the former. It is well known (Anselin, 1988) that in the presence of spatial effects, OLS estimates may be biased and inconsistent. These spatial associations could operate via the demand (here, the number of trip departures), the structural characteristics of nearby stations, or be captured by the error terms. Using Moran's I statistic as the investigative tool (Cliff and Ord, 1981), I confirmed the presence of spatial autocorrelation among the log-transformed bikeshare trip counts in my dataset (p-value<0.01).

Earlier studies have captured spatial dependence using network distance to calculate the weight matrix (Wang et al., 2015; Becker et al., 2017), including studies to analyze bikeshare demand (Faghih-Imani et al., 2016; Guidon et al., 2019). Given that actual travel distance and time information is now readily available, I hypothesized that travel behavior is influenced by biking distance, but even more so by biking time. One purpose of this study is to investigate this conjecture. To test the first hypothesis that a model that accounts for spatial interactions yields improved results, I created a weight matrix of the actual biking distance between the bike stations and used it to estimate a spatial model. To test the second hypothesis that travel time is a better determinant of spatial interactions compared to travel distance, I created another weight matrix of the actual biking time between the bike stations and used it to estimate another spatial model.

The data for both matrices were collected from Google Places API (Google, 2019). Each API call response provided the biking distance (km) and biking time (minutes) between bikeshare stations *i* and *j*. Each data value was stored in their respective matrices. To decide the extent of spatial dependence, I plotted the Moran's I correlogram, which shows Moran's I versus potential bandwidth from the weight matrix to assess the extent of spatial dependence. For the biking distance matrix, the correlogram showed 0.75 km as an appropriate distance band, and for the biking time matrix, the correlogram showed 5 minutes as an appropriate time band. I used these bandwidths to define neighbors for the respective matrix.

I then performed Robust Lagrange Multiplier (RLM) tests for spatial lags and spatial errors. Lagrange Multiplier (LM) tests for different specifications while conducting spatial econometric analysis (Anselin, 1988):

1. Spatial residual autocorrelation in the presence of a spatially lagged dependent variable;
2. Spatial residual autocorrelation in the presence of heteroskedasticity.

An RLM test is a modified version of the LM test, which is robust to the error distributions and spatial layouts (Baltagi and Yang, 2012). Results are shown in Table 3.

Table 3. Results from Robust Lagrange Multiplier (RLM) Tests

Impedance Measure	**Distance-based weight matrix**		**Time-based weight matrix**	
Test	Magnitude	p-value	Magnitude	p-value
RLM Lag	7.00	0.009	8.78	0.003
RLM Error	4.48	0.034	6.08	0.014

Since both tests yield significant statistics, I estimated a combined spatial-autoregressive model with spatial autoregressive disturbances (SARAR; see Drukker et al., 2013). This SARAR model can be written:

$$
\begin{aligned}
\mathbf{Y} &= \lambda \mathbf{W}\mathbf{Y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\
\mathbf{u} &= \rho \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}
\end{aligned} \tag{2}
$$

where,

$\mathbf{Y}$ is an *N x 1* vector of log transformed bike trip counts for each station;

λ and ρ are unknown spatial lag and spatial error parameters respectively;

$\mathbf{W}$ is an *N x N* spatial weight matrix;

$\mathbf{X}$ is an *N x Q* matrix of exogenous explanatory variables with log transformed values;

$\boldsymbol{\beta}$ is a *Q x 1* vector of unknown coefficients;

$\mathbf{u}$ is an *N x 1* vector of correlated residuals; and

$\boldsymbol{\varepsilon}$ is an *N x 1* vector of independent and identically distributed errors.

In the first equation of (2), the term $\lambda\mathbf{WY}$ reflects the impact of trip counts of neighboring stations and it accounts for locally constant omitted variables (Drukker et al., 2013; Mitra and Saphores, 2016). The second equation of (2) captures the residual spatial autocorrelation. When $\rho = 0$, Equation (2) reduces to a spatial lag model; when $\lambda = 0$, it reduces to a spatial error model; and when $\rho = \lambda = 0$, we are back with a simple linear regression model. The SARAR model can be estimated using maximum likelihood (see Elhorst, 2014 for details on likelihood functions) or the method of moments (see Kelejian and Prucha, 1998 and 2009). In this thesis, to estimate my models I relied on the *sacsarlm* routine in R (Bivand et al., 2015), which performs a maximum likelihood estimation of the SARAR model.

4.3 Spatial Weight Matrix

While the weight matrix W reflects spatial interactions between neighbors, its specification is not theoretically determined. For example, W could be adjacent neighbors, k-nearest neighbors or distance-based neighbors (Bivand and Wong, 2018). Adjacency cannot be used here due to the lack of grid structures. For more details on weight matrix, see Elhorst (2014). Here, my weight matrix adopted the mathematical form of the inverse of the square of the value. Thus, if the station is less than 0.75km (distance bandwidth) away, its interaction coefficient in the distance-based weight matrix is the inverse of the squared distance value; else it is zero. Similarly, if the station is less than 5 minutes (time bandwidth) away, its interaction coefficient in the time-based weight matrix is the inverse of the squared time value; else it is zero. Since the weight matrix captures spatial interactions with nearby properties, its diagonal terms are 0. I normalized its rows to sum to 1 to facilitate the interpretation of my results. For robustness, I also tried using only inversed and exponential decay weight matrices. Since the results were similar, I discuss only the results for the inverse of a squared weight matrix.

4.4 Interpretation of spatial model results

Interpreting SARAR models is more involved than interpreting OLS results because of the presence of the spatial lag term λWY. The trip count of station *i* depends on the trip count of station *j*.

Similarly, the trip count of station *j* is dependent on the trip count of station *i*. This leads to an infinite feedback loop between neighboring stations. Assuming that $|\lambda| < 1$, we have:

$$\mathbf{V} \equiv (\mathbf{I} - \lambda\mathbf{W})^{-1} = \mathrm{I} + \lambda\mathbf{W} + \lambda^2\mathbf{W}^2 + \lambda^3\mathbf{W}^3 + \cdots \tag{3}$$

It is convenient to introduce $\boldsymbol{\omega}$, defined by:

$$\boldsymbol{\omega} \equiv (\mathbf{I} - \lambda\mathbf{W})^{-1}(\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon} \tag{4}$$

Using equations (3) and (4), the first equation of (2) can be rewritten:

$$\mathbf{Y} = \mathbf{VX\beta} + \boldsymbol{\omega} = \mathbf{X\beta} + \lambda\mathbf{WX\beta} + \lambda^2\mathbf{W}^2\mathbf{X\beta} + \lambda^3\mathbf{W}^3\mathbf{X\beta} + \cdots + \boldsymbol{\omega} \tag{5}$$

From Equation (5), the expected value of dependent variable depends on a mean value (term $\mathbf{X\beta}$) plus a linear combination of this mean value scaled by powers of the spatial lag parameter λ. Therefore, it is straightforward to derive the partial derivative of the dependent variable datapoint with respect to each of its corresponding explanatory variable datapoint. This derivative value resembles the percentage change in trip count for 1% change in the corresponding explanatory variable. Since a large number of partial derivatives could be non-zero, I followed LeSage and Pace (2009, pp. 36-37) and calculated a scalar summary measure for each explanatory variable $q \in \{1, 2, ..., Q - 1\}$.

The q^{th} Average Direct Impact (ADI_q) represents the average impact on each observation of changing its own q^{th} explanatory variable, including the feedback passing through neighbors back to each observation. Its value is obtained by averaging the main diagonal terms of $\beta_q\mathbf{V}$ as shown below:

$$\mathrm{ADI_q} = \beta_q \mathrm{N}^{-1} \sum_{i=1}^{N} \mathbf{V}_{ii} \tag{6}$$

The q^{th} Average Indirect Impact ($\mathrm{AII_q}$) represents the spillover impacts of changing its q^{th} explanatory variable on other observations only. Its value is obtained by averaging only the off-diagonal terms of $\beta_q\mathbf{V}$:

$$\mathrm{AII_q} = \beta_q \mathrm{N}^{-1} \sum_{i \neq j} \mathbf{V}_{ij} \tag{7}$$

The q^{th} Average Total Impact ($\mathrm{ATI_q}$) represents the resulting combined impact of both direct and indirect impacts of changing the q^{th} explanatory variable. Its value is obtained by summing the ($\mathrm{ADI_q}$) and ($\mathrm{AII_q}$), by averaging all row sums of $\beta_q\mathbf{V}$ matrix. Since I row-normalized the weight matrix, the result simplifies to:

$$\mathrm{ATI_q} = \frac{\beta_q}{(1 - \lambda)} \tag{8}$$

I further followed LeSage and Pace (2009) to calculate the statistical significance of these impacts. By assuming β, λ, ρ and σ^2 normally distributed and obtaining means and covariance matrix from (1), I calculated $\mathrm{ADI_q}$, $\mathrm{AII_q}$, and $\mathrm{ATI_q}$. I repeated these calculations for 10,000 draws and estimated their statistical significance based on the empirical distributions of these 10,000 draws. The same procedure was followed for both distance and time matrices.

5. RESULTS

To assess the influence of spatial interactions on modeling bikeshare demand, I estimated three models: 1) a multiple linear regression model as a baseline; 2) a SARAR model with a weight matrix based on biking distance; and 3) a similar SARAR model with a weight matrix based on biking time. For each model, I calculated the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), the Akaike Information Criteria (AIC), and the Bayesian Information Criteria (BIC). The results of these evaluation tools are summarized in Table 4. A comparison of the three models shows that the spatial model with a time-based weight matrix performs the best. For that spatial model, I calculated direct, indirect, and total impacts, and discussed results.

The remainder of this chapter is structured as follows. Section 5.1 presents an overview of my competing models and Section 5.2 discusses results for the preferred model. Section 5.3 discusses the robustness of my models.

5.1 Selected Diagnostics and Goodness of Fit

Table 4 presents a summary of results and goodness of fit measures for each of the three models estimated. A model with lower MAE, RMSE, AIC and BIC values is preferable in terms of goodness of fit.

From Table 4, we see that there is an improvement in the model's goodness of fit from the OLS model to both the SARAR models. The RMSE value reduces from 0.45 to 0.38 and the MAE value reduces from 0.38 to 0.31. Second, there is improvement in the model's goodness of fit from the SARAR model with distance-based weight matrix to the SARAR model with time-based weight matrix, observed in their AIC and BIC values, both of which reduce by five. This comparison provides evidence in support of my hypotheses that (i) incorporating spatial interaction (spatial lag and spatial error) and, (ii) the use of travel time over travel distance, offers a better estimation framework for modeling bikeshare demand.

Table 4. Summary of estimated models

Variable	OLS Regression	SARAR (Distance-based matrix)	SARAR (Time-based matrix)
Goodness of fit measures			
RMSE	0.45	0.38	0.37
MAE	0.38	0.31	0.30
AIC	120.30	110.08	105.62
BIC	159.99	154.18	149.72

Figure 4 shows a quantile-quantile plot for the residuals in all the three models. Their alignment along the straight-line curve suggests only a relatively mild departure from normality. Estimating SARAR models via maximum likelihood run the risk of biased and inconsistent estimators when errors are heteroskedastic (Arraiz et al., 2010). However, both the SARAR models used here (distance-based weight matrix and time-based weight matrix) do not show any presence of heteroskedasticity when tested graphically as seen in Figure 5. The plot of residuals vs trip count does not indicate a particular trend for either of the three models.

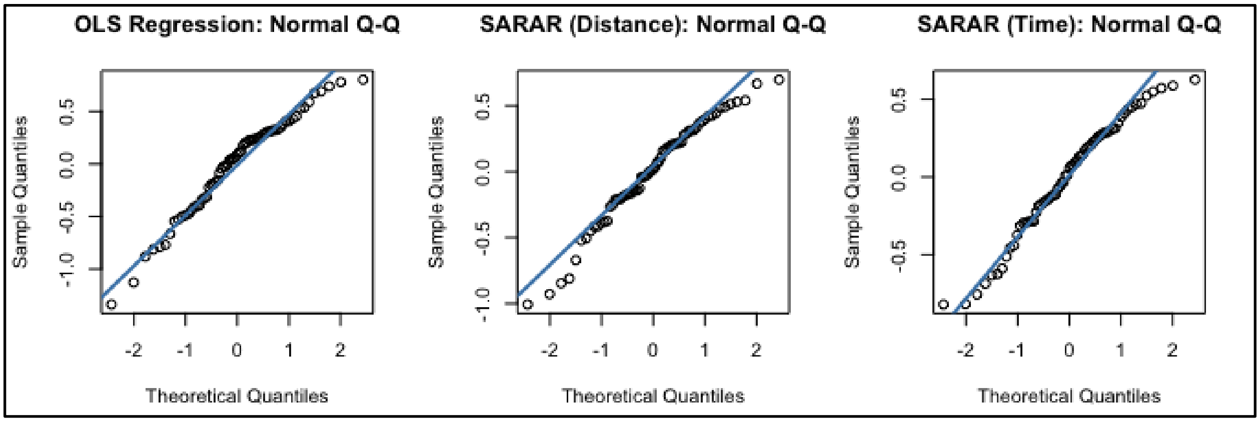

Figure 4. Results from tests for normality of residuals

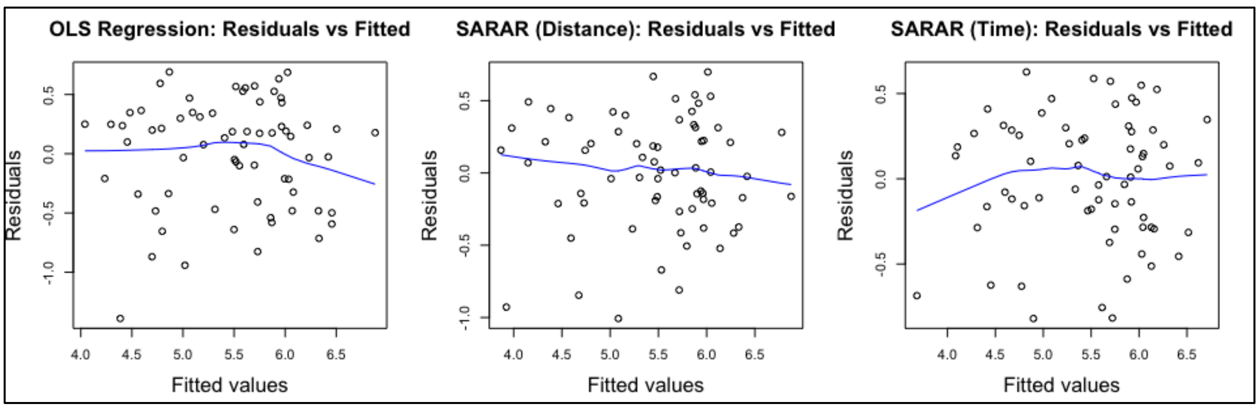

Figure 5. Results from tests for heteroskedasticity

5.2 Results

This section discusses my spatial model results. Section 5.2.1 discusses the spatial coefficients and Section 5.2.2 discusses the impact measures from my preferred model.

Table 5. Summary of spatial coefficients from estimated models

Spatial Coefficient	SARAR (Distance-based matrix)	SARAR (Time-based matrix)
λ (spatial lag)	0.153 ***	0.182 ***
ρ (spatial error)	-0.610 ***	-0.874 ***

Note: ***0.1%, **1%, *5%

5.2.1 Spatial coefficients

λ is the spatial autoregressive (spatial lag) coefficient and ρ is the spatial correlation (spatial error) coefficient. Their statistically significant values (Table 5) confirm the presence of spatial autocorrelation in my dataset, which was detected by the RLM tests (Table 3). Since I row-normalize the spatial weight matrices, the λ and ρ values need to be between between -1 and 1 (Elhorst, 2014). As expected, the positive λ here implies that the trip count at a bike station is positively influenced by the trip count of neighboring bike stations (similar to findings of Faghih-Imani et al., 2016). This may be reflective of a travel demand at neighboring bike stations that is influenced by the same set of attractions (home, job, transit stop). It can also be reflective of the bike station experiencing higher demand, thus reducing its availability of bikes and forcing riders to start trips from the neighboring bike stations. Another possible explanation could be return trips; a trip starting from the origin station to the nearby destination station would potentially generate a return trip from the destination station.

5.2.2 Impacts

Since the SARAR Model with a biking time-based weight matrix is better, I calculated the impacts of explanatory variables on the bikeshare demand for this model. The results are shown in Table 6. I discuss the direct impacts in detail as they give a measure of the effect of q^{th} explanatory variable on the demand for q^{th} bikeshare station. It is noticeable that the signs of all the three impacts are the same for all variables, meaning the nature of the influence of the q^{th} explanatory variable on the demand at the q^{th} station is the same as the nature of its influence on the demand at other neighboring stations. The magnitudes of indirect impacts are much smaller as compared to the magnitude of direct impacts, meaning the influence of the q^th^ explanatory variable on demand at the q^th^ bikeshare station is much stronger as compared to its influence on the demand at other neighboring stations. The total impacts are a sum of the direct and indirect impact.

Table 6. Impact measures for SARAR model with time-based weight matrix

Variable	Coefficient	ADI	AII	ATI
Socio-demographic variables				
Population density (1000 persons/sq.km.)	0.10	0.10	0.02	0.12
College graduates (1000 persons)	0.37 ***	0.38 ***	0.08 *	0.46 ***
Age 18-45 (1000 persons)	0.06	0.06	0.01	0.07
Carless households	-0.20 **	-0.20 **	-0.04 *	-0.24 **
Households with children	0.08	0.08	0.02	0.10
Employees in zip code (1000 persons)	-0.50 ***	-0.51 ***	-0.11 *	-0.61 ***
Land use variables				
Residential (%)	-0.04	-0.04	-0.01	-0.05
Commercial (% in 500m buffer)	0.00	0.00	0.00	0.00
Open/Recreational (%)	-0.03	-0.03	-0.01	-0.04
Public Facilities (%)	-0.01	-0.01	0.00	-0.01
University (%)	-0.02	-0.02	0.00	-0.02
Restaurants	0.12 *	0.12 *	0.03	0.15 *
Transport variables				
Daily bus station arrivals	0.03	0.03	0.01	0.03
Daily rail station arrivals	0.09 **	0.09 **	0.02 *	0.10 **
Long-duration transit commuters	-0.19 ***	-0.19 ***	-0.04 *	-0.23 ***
Length of bike lane network (km)	0.13 *	0.13 *	0.03	0.16 *

Note: ***0.1%, **1%, *5%

The impacts can be interpreted as scalar elasticity measures for a 1% change in the explanatory variable. For example, shared bike trips at a station increase with the number of college graduates (ADI=0.38***); meaning a 1% increase in the number of college graduates (in the occupying census block group) would cause the number of bike trips at the corresponding bike station to increase by 0.38%.

Three of the six socio-demographic variables are statistically significant – college graduates, carless households and employees in zip codes. The positive impact of college graduates on bikeshare demand is expected and consistent with results from earlier studies (Rixey, 2013; Fishman et al., 2014; Faghih-Imani et al., 2017; Hyland et al., 2017). A review of the bikeshare literature by Fishman (2016) also found that bikeshare users tend to be more educated than the general population. Based on direct impacts, an increase in the number of carless households in a census block group by 1% causes the bikeshare demand at the corresponding station to reduce by 0.2%. This seems counter-intuitive since the absence of car would result in the need for alternative modes for transport including bikeshare. A possible explanation is that the requirement for a debit/credit card to register may act as a barrier for low-income households. Earlier papers also found that bikeshare members exhibit a higher car ownership as compared to non-members (Shaheen et al., 2011).

The negative impact of employees may appear counter-intuitive result given the concentration of jobs in the Downtown LA area. Based on direct impacts, a 1% increase in the number of employees in a zip code causes the bikeshare demand at the corresponding station to reduce by 0.51%. Earlier papers report mixed impacts for the impact of the number of employees on the number of bike-sharing trips; Rixey (2013) and Fagihih-Imani and Eluru (2016) report that proximity to jobs has a positive impact on bikeshare demand while Maurer (2011) find it to have a negative impact. This result suggests that bikesharing in Los Angeles is often used for non-work-based trips. Additionally, zip code is a geographically large area. The 67 stations are distributed in 11 different zip codes, meaning the data may not reflect the actual business patterns in the close proximity of the station. Another possible explanation could be the access to free parking offered by the employers.

The coefficients of population density and age variables are not significant, even though earlier papers have found both to have a positive impact on bikeshare demand (Daddio, 2012; Rixey, 2013; Ricci, 2015; Faghih-Imani and Eluru, 2016). Similarly, the households with children variable is not significant even though Hosford et al. (2018) found households with children to be less-likely users of bikeshare in Vancouver.

With the exception of the number of restaurants within 300 meters, land use variables are not significant. Daddio (2012) also found the effect of park and university areas in the proximity of bikeshare stations to be not significant for Washington DC. However, other studies report one or more land use types to have significant impacts (Rixey 2013; Faghih-Imani and Eluru, 2014; Sun and Chen 2017). As in previous other studies (Faghih-Imani et al., 2014; Wang et al., 2015; Faghih-Imani and Eluru, 2016; Hyland et al., 2017), the proximity to restaurants variable (ADI = 0.12*) is statistically significant. A 1% increase in the number of restaurants within 300m of a bikeshare station would increase its bikeshare demand by 0.12%. This result further supports the use of bikeshare for non-work-based trips.

Three of the four transportation variables are statistically significant. The exception is the bus transit arrivals variable. Results show that rail transit arrivals within 100m of a bikeshare station have a positive impact on its demand. Earlier studies have found results (Rixey 2013; Hyland et al., 2017; Sun and Chen, 2017; El-Assi et al., 2017; Guidon et al., 2019). However, the transit variables used in these papers are different from the one used in this thesis. Guidon et al. (2019) used the number of passengers alighting transit daily as a variable, while all other studies have used the number of transit stations in proximity or distance to the nearest transit station as the explanatory variable. These variables are reflective of the general trends between the bikeshare and transit but do not to explain the demand for bikeshare as a last mile option. Unlike the variables stated above, I explained the bikeshare demand (departures) at a station using the transit demand (arrivals) within 100m of the bikeshare station. Connecting the transit arrivals and bikeshare departures in the vicinity provides more reliable evidence towards measuring the use bikeshare as a last mile mode. Since the LA bikeshare program is designed to integrate with the transit system and increase its ridership, I hypothesized the transit activity variables to have a strong positive impact on bikeshare demand. While the correlation for rail arrivals is positive, the magnitude of the impact is small. A 1% increase in the number of daily rail transit arrivals cause the monthly bike trip count to increase by a mere 0.09%, indicating that the LA bikeshare seldom serves as a last-mile connectivity option for using transit.

The above conclusions are further reinforced by the significant negative impact of the volume of longer commute trips using transit (>45 minutes, ADI = -0.19***) on bikeshare demand. A 1% increase in the number of long-duration transit commuters in the occupying census block group of a bikeshare station reduces its demand by 0.19%. Considering the count of long-duration commuters using transit as a proxy for the count of multimodal trip makers, the negative impact proves that areas with higher number of long-duration transit commuters experience lower bikeshare demand, and vice versa.

Finally, the positive impact of the length of bike lane network (ADI = 0.13*) on bikeshare demand is both intuitive and consistent with earlier papers (Buck and Buehler, 2012; Rixey, 2013; Wang et al., 2015; Faghih-Imani and Eluru, 2016; Guidon et al., 2019). A 1% increase in the length of bike lane network within the 1.6km radial buffer around a bikeshare station would increase its demand by 0.13%.

5.3 Model Robustness

I compare the goodness of fit measures for spatial models with three different weight matrix formulations; (i) inverse of square function, as discussed above (x^{-2}), (ii) inverse function (x^{-1}), and (iii) exponential decay function with power one ($\exp(x^{-1})$) to test the robustness of my model. Table 7 summarizes the results from this comparison. The values for all four goodness of fit measures – RMSE, MAE, AIC and BIC reduce from the SARAR model with distance-based matrix to SARAR model with time-based matrix across all three spatial weights formulations. This provides further evidence towards the superiority of a spatial model with time-based travel matrix over one with distance-based travel matrix. The impact measures of the explanatory variables across the three SARAR models with time-based weight matrix remained similar, while their significance was the same.

Table 7. Summary of estimated models with different weight matrix formulations

Model	OLS	SARAR (Distance)	SARAR (Time)	SARAR (Distance)	SARAR (Time)	SARAR (Distance)	SARAR (Time)
Weight Matrix	NA	Inverse Squared		Inverse		Exponential Decay	
RMSE	0.45	0.38	0.37	0.39	0.37	0.40	0.37
MAE	0.38	0.31	0.30	0.32	0.30	0.33	0.31
AIC	120.30	110.08	105.62	110.94	106.02	112.03	107.31
BIC	159.99	154.18	149.72	155.03	150.11	156.12	151.40

Moreover, I performed multiple permutations with the inclusion of variables during my modeling. However, the impact measure values of significant variables in my final model did not experience significant changes when new variables were added, or existing variables were removed. The changes include trying different radial buffers for the transit activity variables, alternative family structure variables and changing the number of land use types amongst others.

6. CONCLUSIONS

An increasing number of cities across the world have implemented bikeshare systems to provide a green, sustainable and economical mode of transport since 2010. Their popularity as solutions for the 'last mile' travel, urban traffic congestion, and carbon emissions have further prompted many cities to expand their systems. Identifying the factors influencing the demand for bikeshare systems is pivotal to planning for new systems and expanding the existing ones. Earlier research efforts have analyzed such factors, including few studies that have used spatial models with a network distance-based weight matrix to incorporate the spatial interactions of the bikeshare stations' demand. Given that actual travel distance and time information is now readily available, I hypothesized that travel behavior is influenced by biking distance, but even more so by biking time. To this extent, I estimated three models: 1) a multiple linear regression model as a baseline; 2) a SARAR model with a weight matrix based on biking distance; and 3) a similar SARAR model with a weight matrix based on biking time. I used socio-economic, land use and transport variables to explain the station-level bike trip departures for all the models.

The results clearly indicate that (i) incorporating spatial interactions (spatial lag and spatial errors) is an important feature of bikesharing demand; and, (ii) that there is an improvement in the models' goodness of fit from the SARAR models with distance-based weight matrices to the SARAR models with time-based weight matrices. The lower relative mean square error (RMSE) and mean absolute error (MAE) for the SARAR models with time-based matrix show that they outperform the usual SARAR models with distance-based matrix. This comparison provides strong evidence for the verification of my hypotheses that (i) incorporating spatial interaction (spatial lag and spatial error) and, (ii) the use of travel time over travel distance offers a better estimation framework for modeling bikeshare demand.

One of the objectives of the LA Metro Bikeshare Program was 'to develop an affordable, user-friendly bikeshare program that increases ridership by integrating with the County's regional transit services' (Metro Bikeshare Business Plan, 2019-20). Therefore, I hypothesized that bikeshare would be a popular mode choice for the last mile travel to access transit. To test for this, I used the number of transit trip arrivals within 100m of a bikeshare station as one of my explanatory variables. While the results show that rail trip arrivals in the proximity of a bike station have a positive impact on its demand, the magnitude is very small. A 1% increase in the proximate daily rail trip arrivals increases the monthly bikeshare departures by 0.09%. The results show that after controlling for other factors, higher nearby rail transit demand has minimal impact on bikeshare demand. This is further reinforced by the finding that bikeshare is not popular in areas with more long-duration transit commuters. Thus, there is little evidence to support the fulfillment of LA Metro Bikeshare Program's objective.

This is the first known spatial modeling study to use biking distance and biking time as a measure for the spatial interactions of the bikeshare stations' demand. It provides quantitative estimates of how socio-demographic, land use and transport patterns influence station-level bikeshare demand. The proposed framework and the quantitative impacts of exogeneous variables provide very important policy implications for planning and modeling for new bikeshare systems and modifying existing bikeshare systems. From a planning perspective, placing more stations in the proximity of existing stations should increase the overall system usage in LA. Demand for shared bikes would also be positively influenced by placing more stations in the proximity to restaurants and by increasing the bike lane network around it. Placing the bike stations in the vicinity of rail stations is not sufficient to promote the use of bikeshare as a last mile travel mode. From a modeling perspective, a similar framework can be employed in an iterative manner for different network designs to finalize one with optimum expected usage. This thesis provides evidence that the use of spatial model with a travel time-based matrix will give better forecasts. For existing systems, planners can employ this improved modeling framework to forecast the bikeshare usage when stations are added or removed.

This work is not without limitations. In my modeling framework, I consider the monthly trip count which does not account for temporal fluctuations. Future studies can extend this thesis by incorporating them alongside the existing spatial modeling framework to model the hourly panel data. This will also improve the travel time matrix by allowing for the time of the day criteria when collecting data from Google Places API. A similar method can also be used to model trip arrivals. Additionally, the use of TAP card data that links bike to transit trips would be a promising research direction to measure the extent of bikeshare use as the last mile mode. The statistical insignificance of the land use variables in my model could be driven by my choice of buffer (500m radius) and/or the dataset (SCAG). An extension of this study by using a different buffer sizes or using other data at the tax lot level may lead to different results. Finally, the SARAR model with travel time-based matrix shows promise but it needs to be tested on other bikeshare systems.

REFERENCES

Anselin, L., 1988. Spatial Econometrics: Methods and Models, vol. 4, Springer (1988).

Arraiz, I., Drukker, D. M., Kelejian, H. H., and Prucha, I. R. 2010., A spatial Cliff-Ord-type model with heteroskedastic innovations: Small and large sample results. Journal of Regional Science 50: 592–614.

Baltagi, B. and Yang, Z., 2012. Standardized LM Test for Spatial Error Dependence in Linear or Panel Regressions" (2012). Center for Policy Research. 188. https://surface.syr.edu/cpr/188

Becker, H., Ciari, F., Axhausen, K. W., 2017. Modeling free-floating car-sharing use in Switzerland: A spatial regression and conditional logit approach. Transportation Research Part C, (pp. 286{299). https://doi.org/10.1016/j.trd.2018.08.003

Bivand R. et al., 2015 Package ‘spdep’. https://mran.microsoft.com/snapshot/2017-08-23/web/packages/spdep/spdep.pdf

Bivand, R., Wong, D., 2018. TEST 27: 716. https://doi.org/10.1007/s11749-018-0599-x

Buck, D., Buehler, R., 2012. Bike lanes and other determinants of capital bikeshare trips. In: 91st Transportation Research Board Annual Meeting. Washington D.C.

Chalermpong, S., 2007. Rail transit and residential land use in developing countries: hedonic study of residential property prices in Bangkok, Thailand. Transport. Res. Rec.: J. Transport. Res. Board 2038 (1), 111–119. https://doi.org/10.3141/2038-15

Campbell, A. A., Cherry, C. R., Ryerson, M. S., Yang, X. M., 2016. Factors influencing the choice of shared bicycles and shared electric bikes in Beijing. Transportation Research Part C: Emerging Technologies, 67, 399–414. https://doi.org/10.1016/j.trc.2016.03.004

Chakrabarti, S., 2017. How can public transit get people out of their cars? An analysis of transit mode choice for commute trips in Los Angeles. Transp. Policy 2017, 54, 80–89. https://doi.org/10.1016/j.tranpol.2016.11.005

Cliff, A D, Ord, J K, 1981. Spatial Processes. Models and Applications (Pion, London).

Daddio, D., 2012. Maximizing Bicycle Sharing: An Empirical Analysis of Capital Bikeshare Usage. University of North Carolina. https://doi.org/10.17615/qv32-b860

Daraban, B., Fournier, G.M., 2008. Incumbent responses to low-cost airline entry and exit: a spatial autoregressive panel data analysis. Res. Transp. Econ. 24, 15–24. https://doi.org/10.1016/j.retrec.2009.01.004

Drukker, D. M., Prucha, I. R., Raciborski, R., 2013. Maximum Likelihood and Generalized Spatial Two-Stage Least-Squares Estimators for a Spatial-Autoregressive Model with Spatial-Autoregressive Disturbances. The Stata Journal, 13(2), 221–241. https://doi.org/10.1177/1536867X1301300201

Dubé, J., Legros, D., Thériault, M., Des Rosiers, F., 2014. A spatial difference-in-differences prices. Transp. Res. B Methodol. 64, 24–40. https://doi.org/10.1016/j.trb.2014.02.007

El-Assi, W.; Mahmoud, M., Habib, K., 2017. Effects of built environment and weather on bike sharing demand: A station level analysis of commercial bike sharing in Toronto. Transportation 2017, 44, 589–613. http://dx.doi.org/10.1007/s11116-015-9669-z

Elhorst, J.P., 2014. Spatial Econometrics: From Cross-Sectional Data to Spatial Panels, estimator to evaluate the effect of change in public mass transit systems on house Pp. 37–93. Springer, New York. https://link.springer.com/book/10.1007%2F978-3-642-40340-8

Faghih-Imani, A., Eluru, N., El-Geneidy, A.M., Rabbat, M., Haq, U., 2014. How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal. J. Transp. Geogr. 41, 306–314. https://doi.org/10.1016/j.jtrangeo.2014.01.013

Faghih-Imani, A., Eluru, N., 2015. Analyzing bicycle sharing system user destination choice preferences: an investigation of Chicago's Divvy System. J. Transp. Geogr., 44 (2015), pp. 53-64. https://doi.org/10.1016/j.jtrangeo.2015.03.005

Faghih-Imani, A., Eluru, N., 2016. Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: a case study of New York CitiBike system. J. Transp. Geogr., 54 (2016), pp. 218-227. https://doi.org/10.1016/j.jtrangeo.2016.06.008

Faghih-Imani, A., Hampshire, R., Marla, L., Eluru, N., 2017. An empirical analysis of bike sharing usage and rebalancing: evidence from Barcelona and Seville. Transp. Res. Part A: Policy Pract., 97 (2017), pp. 177-191. https://doi.org/10.1016/j.tra.2016.12.007

Fishman, E., Washington, S., Haworth, N., Mazzei, A. 2014. Barriers to bikesharing: an analysis from Melbourne and Brisbane. J. Transp. Geogr., 41 (2014), pp. 325-337. https://doi.org/10.1016/j.jtrangeo.2014.08.005

Fishman, E., 2015. Bikeshare: a review of recent literature. Transport Rev., 1–22.

Froehlich, J., Neumann, J., Oliver, N., 2009. Sensing and predicting the pulse of the city through shared bicycling. IJCAI 9, 1420–1426. https://www.aaai.org/ocs/index.php/IJCAI/IJCAI-09/paper/viewFile/578/910

Fuller, D., Gauvin, L., Kestens, Y., Daniel, M., Fournier, M., Morency, P., Drouin, L., 2011. Use of a new public bicycle share program in Montreal, Canada. Am. J.Prev. Med. 41 (1), 80–83. https://doi.org/10.1016/j.amepre.2011.03.002

Garcia-Martinez, A., Cascajo, R., Jara-Diaz, S. R., Chowdhury, S., Monzon, A., 2018. Transfer penalties in multimodal public transport networks. Transportation Research Part A: Policy and Practice 114, 52–66. https://doi.org/10.1016/j.tra.2018.01.016

Gebhart, K., Noland, R., 2014. Transportation (2014) 41: 1205. https://doi.org/10.1007/s11116-014-9540-7

Google, 2019. Google Places API.

Guidon, S., Becker, H., Dediu, H., Axhousen, K., 2019. Electric Bicycle-Sharing: A New Competitor in the Urban Transportation Market? An Empirical Analysis of Transaction Data. Journal of the Transportation Research Board, Volume 2673, Issue 4, 2019, pp 15-26. https://doi.org/10.1177/0361198119836762

Hyland M., Hong, Z., Pinto, H., Chen, Y., 2018. Hybrid cluster-regression approach to model bikeshare station usage. Transportation Research Part A, 115, 71-89. https://doi.org/10.1016/j.tra.2017.11.009

Hosford, K., Lear, S.A., Fuller, D. et al., 2018. BMC Public Health (2018) 18: 1326. https://doi.org/10.1186/s12889-018-6246-3

Kelejian, H.H., Prucha, I.R., 1998. The Journal of Real Estate Finance and Economics (1998) 17: 99. https://doi.org/10.1023/A:1007707430416

Kelejian HH, Prucha I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics. forthcoming https://doi.org/10.1016/j.jeconom.2009.10.025

Kim, D., Shin, H., Im, H., Park, J., 2012. The Impact of Weather on Bicycle Usage - Focus on Usage of Bike-sharing System in Goyang, Journal of Transport Research, Vol.19, No.3, pp.77-88.

LeSage, J. and Pace, R., 2009. Introduction to Spatial Econometrics. CRC Press, Boca Raton, FL.

Longhi, S., Nijkamp, P., 2007. Forecasting regional market developments under spatial autocorrelation. International Regional Science Review 30: 100–51. https://doi.org/10.1177/0160017606298428

Maurer, L.K., 2011. Suitability Study for a Bicycle Sharing Program in Sacramento. University of North Carolina, NY. https://doi.org/10.17615/7j3w-yk71

Metro Bikeshare Business Plan, FY 2019-20.

Mitra, S. K., Saphores, J-D., 2016. The value of transportation accessibility in a least developed country city – the case of Rajshahi City, Bangladesh. Transportation Research Part A: Policy and Practice, 89, 184–200. https://doi.org/10.1016/j.tra.2016.05.002

Nair, R., Miller-Hooks, E., Hampshire, R.C., Busic, A., 2013. Large-scale vehicle sharing systems: analysis of Vélib'. Int. J. Sustain. Transp. 7(1), 85–106 (2013). https://doi.org/10.1080/15568318.2012.660115

Pace R., Ronald Barry, K., Gilley, O., Sirmans, C.F., 2000. A method for spatial-temporal forecasting with an application to real estate prices. International Journal of Forecasting, 16 (2000), pp. 229-246. https://doi.org/10.1016/S0169-2070(99)00047-3

Ricci, M., 2015. Bike sharing: a review of evidence on impacts and processes of implementation and operation. Res. Transp. Bus. Manag., 15 (2015), pp. 28-38. https://doi.org/10.1016/j.rtbm.2015.03.003

Rixey, R., 2013. Station-level forecasting of bikesharing ridership. Transp. Res. Rec. 2387, 46–55. https://doi.org/10.3141/2387-06

Schakenbos, R., 2014. Valuation of a transfer in a multimodal public transport trip. A stated preference research into the experienced disutility of a transfer between bus/tram/metro and train within the Netherlands. University of Twente, Netherlands.

Shaheen, S., Guzman, S., Zhang, H., 2010. Bikesharing in Europe, the Americas, and Asia, 159-167. https://doi.org/10.3141/2143-20

Shaheen, S., Martin, E., Cohen, A. P. and Finson, R. 2012. Public bikesharing in North America: Early operator and user understanding, San Jose, CA: Mineta Transportation Institute. https://rosap.ntl.bts.gov/view/dot/24566

Shaheen S., Martin E., Chan N., Cohen A., Pogodzinski M., 2014. Public Bikesharing in North America During a Period of Rapid Expansion: Understanding Business Models, Industry Trends and User Impacts. Mineta Transportation Institute, San Jose, CA.

Shen, Y., Silva, J.D.A., Martínez, L.M., 2014. Assessing high-speed rail's impacts on land cover change in large urban areas based on spatial mixed logit methods: a case study of Madrid Atocha railway station from 1990 to 2006. J. Transp. Geogr. 41, 184–196. https://doi.org/10.1016/j.jtrangeo.2014.09.007

Shi, J., Si, H., Wu, G., Su, Y., Lan, J., 2018. Critical factors to achieve dockless bike-sharing sustainability in China: A stakeholder-oriented network perspective. Sustainability 2018, 10, 2090. https://www.mdpi.com/2071-1050/10/6/2090

Si, H., Shi, J., Wu, G., Chen, J., Zhao, X., 2019. Mapping the bike sharing research published from 2010 to 2018: A scientometric review. Journal of Cleaner Production., Vol. 213, 415-427. https://doi.org/10.1016/j.jclepro.2018.12.157

Smith, T.R. and Slater, P., 1981. A family of spatial interaction models incorporating information flows and choice set constraints applied to U.S. interstate labor flows. International Regional Science Review 6, 15-31.

Sun, F., Chen, P., 2017. Impact of built environment and temporal factors on public bikesharing. Transportation Research Board 96th Annual Meeting. Washington, D.C (2017).

Vogel, P., Greiser, T., Mattfeld, D.C., 2011. Understanding bike-sharing systems using data mining: exploring activity patterns. Procedia - Soc. Behav. Sci. 20, 514–523. http://dx.doi.org/10.1016/j.sbspro.2011.08.058.

Wang, X., Kockelman, K., 2006. Tracking land cover change in mixed logit model: recognizing temporal and spatial effects. Transp. Res. Rec. 1977, 112–120. https://doi.org/10.1177/0361198106197700114

Wang, X., Lindsey, G., Schoner, J., Harrison, A., 2015. Modeling bike share station activity: effects of nearby businesses and jobs on trips to and from stations. J. Urban Plann. Vol. 142 Iss. 1. https://ascelibrary.org/doi/full/10.1061/%28ASCE%29UP.1943-5444.0000273

Wang, C-H., Chen, N., 2015. A GIS-based spatial statistical approach to modeling job accessibility by transportation mode: case study of Columbus, Ohio. J. Transp. Geogr., 45 (2015), pp. 1-11. https://doi.org/10.1016/j.jtrangeo.2015.03.015

Wang, K., Akar, G., Chen, Y. J., 2018. Bike sharing differences among Millennials, Gen Xers, and Baby Boomers: Lessons learnt from New York City's bike share. Transportation Research Part A: Policy and Practice, 116, 1–14. W. https://doi.org/10.1016/j.tra.2018.06.001

Zhang, D., Wang, X., 2015. Investigating the dynamic spillover effects of low-cost airlines on airport airfare through spatio-temporal regression models. Networks Spatial Econ. (2015), pp. 1-16. http://dx.doi.org/10.1007/s11067-015-9300-z