University Transportation Research Center - Region 2

Final Report

The Spatial Effect of Socio-Economic Demographics on Transit Ridership: a Case Study in New York

Performing Organization: Manhattan College

March 2018

Sponsor:
University Transportation Research Center - Region 2

University Transportation Research Center - Region 2

The Region 2 University Transportation Research Center (UTRC) is one of ten original University Transportation Centers established in 1987 by the U.S. Congress. These Centers were established with the recognition that transportation plays a key role in the nation's economy and the quality of life of its citizens. University faculty members provide a critical link in resolving our national and regional transportation problems while training the professionals who address our transportation systems and their customers on a daily basis.

The UTRC was established in order to support research, education and the transfer of technology in the field of transportation. The theme of the Center is "Planning and Managing Regional Transportation Systems in a Changing World." Presently, under the direction of Dr. Camille Kamga, the UTRC represents USDOT Region II, including New York, New Jersey, Puerto Rico and the U.S. Virgin Islands. Functioning as a consortium of twelve major Universities throughout the region, UTRC is located at the CUNY Institute for Transportation Systems at The City College of New York, the lead institution of the consortium. The Center, through its consortium, an Agency-Industry Council and its Director and Staff, supports research, education, and technology transfer under its theme. UTRC's three main goals are:

Research

The research program objectives are (1) to develop a theme based transportation research program that is responsive to the needs of regional transportation organizations and stakeholders, and (2) to conduct that program in cooperation with the partners. The program includes both studies that are identified with research partners of projects targeted to the theme, and targeted, short-term projects. The program develops competitive proposals, which are evaluated to insure the mostresponsive UTRC team conducts the work. The research program is responsive to the UTRC theme: "Planning and Managing Regional Transportation Systems in a Changing World." The complex transportation system of transit and infrastructure, and the rapidly changing environment impacts the nation's largest city and metropolitan area. The New York/New Jersey Metropolitan has over 19 million people, 600,000 businesses and 9 million workers. The Region's intermodal and multimodal systems must serve all customers and stakeholders within the region and globally.Under the current grant, the new research projects and the ongoing research projects concentrate the program efforts on the categories of Transportation Systems Performance and Information Infrastructure to provide needed services to the New Jersey Department of Transportation, New York City Department of Transportation, New York Metropolitan Transportation Council , New York State Department of Transportation, and the New York State Energy and Research Development Authorityand others, all while enhancing the center's theme.

Education and Workforce Development

The modern professional must combine the technical skills of engineering and planning with knowledge of economics, environmental science, management, finance, and law as well as negotiation skills, psychology and sociology. And, she/he must be computer literate, wired to the web, and knowledgeable about advances in information technology. UTRC's education and training efforts provide a multidisciplinary program of course work and experiential learning to train students and provide advanced training or retraining of practitioners to plan and manage regional transportation systems. UTRC must meet the need to educate the undergraduate and graduate student with a foundation of transportation fundamentals that allows for solving complex problems in a world much more dynamic than even a decade ago. Simultaneously, the demand for continuing education is growing – either because of professional license requirements or because the workplace demands it – and provides the opportunity to combine State of Practice education with tailored ways of delivering content.

Technology Transfer

UTRC's Technology Transfer Program goes beyond what might be considered "traditional" technology transfer activities. Its main objectives are (1) to increase the awareness and level of information concerning transportation issues facing Region 2; (2) to improve the knowledge base and approach to problem solving of the region's transportation workforce, from those operating the systems to those at the most senior level of managing the system; and by doing so, to improve the overall professional capability of the transportation workforce; (3) to stimulate discussion and debate concerning the integration of new technologies into our culture, our work and our transportation systems; (4) to provide the more traditional but extremely important job of disseminating research and project reports, studies, analysis and use of tools to the education, research and practicing community both nationally and internationally; and (5) to provide unbiased information and testimony to decision-makers concerning regional transportation issues consistent with the UTRC theme.

Project No(s):
UTRC/RF Grant No: 49198-26-28

Project Date: March 2018

Project Title: The Spatial Effect of Socio-Economic Demographics on Transit Ridership: a Case Study in New York

Project's Website:
http://www.utrc2.org/research/projects/spatial-effect-socio-economic-demographics

Principal Investigator(s):
Matthew Volovski, Ph.D.
Assistant Professor
Department of Civil & Environmental Engineering
Manhattan College
Riverdale, NY 10471
Tel: (718) 862-7184
Email: mvolovski01@manhattan.edu

Performing Organization(s):
Manhattan College

Sponsor(s):
University Transportation Research Center (UTRC)

To request a hard copy of our final reports, please send us an email at utrc@utrc2.org

Mailing Address:

University Transportation Reserch Center
The City College of New York
Marshak Hall, Suite 910
160 Convent Avenue
New York, NY 10031
Tel: 212-650-8051
Fax: 212-650-8374
Web: www.utrc2.org

Board of Directors

The UTRC Board of Directors consists of one or two members from each Consortium school (each school receives two votes regardless of the number of representatives on the board). The Center Director is an ex-officio member of the Board and The Center management team serves as staff to the Board.

City University of New York
Dr. Robert E. Paaswell - Director Emeritus of NY
Dr. Hongmian Gong - Geography/Hunter College

Clarkson University
Dr. Kerop D. Janoyan - Civil Engineering

Columbia University
Dr. Raimondo Betti - Civil Engineering
Dr. Elliott Sclar - Urban and Regional Planning

Cornell University
Dr. Huaizhu (Oliver) Gao - Civil Engineering
Dr. Richard Geddess - Cornell Program in Infrastructure Policy

Hofstra University
Dr. Jean-Paul Rodrigue - Global Studies and Geography

Manhattan College
Dr. Anirban De - Civil & Environmental Engineering
Dr. Matthew Volovski - Civil & Environmental Engineering

New Jersey Institute of Technology
Dr. Steven I-Jy Chien - Civil Engineering
Dr. Joyoung Lee - Civil & Environmental Engineering

New York Institute of Technology
Dr. Nada Marie Anid - Engineering & Computing Sciences
Dr. Marta Panero - Engineering & Computing Sciences

New York University
Dr. Mitchell L. Moss - Urban Policy and Planning
Dr. Rae Zimmerman - Planning and Public Administration

(NYU Tandon School of Engineering)
Dr. John C. Falcocchio - Civil Engineering
Dr. Kaan Ozbay - Civil Engineering
Dr. Elena Prassas - Civil Engineering

Rensselaer Polytechnic Institute
Dr. José Holguín-Veras - Civil Engineering
Dr. William "Al" Wallace - Systems Engineering

Rochester Institute of Technology
Dr. James Winebrake - Science, Technology and Society/Public Policy
Dr. J. Scott Hawker - Software Engineering

Rowan University
Dr. Yusuf Mehta - Civil Engineering
Dr. Beena Sukumaran - Civil Engineering

State University of New York
Michael M. Fancher - Nanoscience
Dr. Catherine T. Lawson - City & Regional Planning
Dr. Adel W. Sadek - Transportation Systems Engineering
Dr. Shmuel Yahalom - Economics

Stevens Institute of Technology
Dr. Sophia Hassiotis - Civil Engineering
Dr. Thomas H. Wakeman III - Civil Engineering

Syracuse University
Dr. Baris Salman - Civil Engineering
Dr. O. Sam Salem - Construction Engineering and Management

The College of New Jersey
Dr. Thomas M. Brennan Jr - Civil Engineering

University of Puerto Rico - Mayagüez
Dr. Ismael Pagán-Trinidad - Civil Engineering
Dr. Didier M. Valdés-Díaz - Civil Engineering

UTRC Consortium Universities

The following universities/colleges are members of the UTRC consortium under MAP-21 ACT.

City University of New York (CUNY)
Clarkson University (Clarkson)
Columbia University (Columbia)
Cornell University (Cornell)
Hofstra University (Hofstra)
Manhattan College (MC)
New Jersey Institute of Technology (NJIT)
New York Institute of Technology (NYIT)
New York University (NYU)
Rensselaer Polytechnic Institute (RPI)
Rochester Institute of Technology (RIT)
Rowan University (Rowan)
State University of New York (SUNY)
Stevens Institute of Technology (Stevens)
Syracuse University (SU)
The College of New Jersey (TCNJ)
University of Puerto Rico - Mayagüez (UPRM)

UTRC Key Staff

Dr. Camille Kamga: *Director, Associate Professor of Civil Engineering*

Dr. Robert E. Paaswell: *Director Emeritus of UTRC and Distin*guished Professor of Civil Engineering, The City College of New York

Dr. Ellen Thorson: *Senior Research Fellow*

Penny Eickemeyer: *Associate Director for Research, UTRC*

Dr. Alison Conway: *Associate Director for Education/Associate Professor of Civil Enginering*

Nadia Aslam: *Assistant Director for Technology Transfer*

Nathalie Martinez: *Research Associate/Budget Analyst*

Andriy Blagay: *Graphic Intern*

Tierra Fisher*: Office Manager*

Dr. Sandeep Mudigonda*, Research Associate*

Dr. Rodrigue Tchamna*, Research Associate*

Dr. Dan Wan*, Research Assistant*

Bahman Moghimi*: Research Assistant;*
Ph.D. Student, Transportation Program

Sabiheh Faghih*: Research Assistant;*
Ph.D. Student, Transportation Program

Patricio Vicuna*: Research Assistant*
Ph.D. Candidate, Transportation Program

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2.Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle The Spatial Effect of Socio-Economic Demographics on Transit Ridership: a Case Study in New York		5. Report Date March, 2018
		6. Performing Organization Code
7. Author(s) Matthew Volovski, PhD		8. Performing Organization Report No.
9. Performing Organization Name and Address Manhattan College 4513 Manhattan College Parkway Riverdale, NY 10471		10. Work Unit No.
		11. Contract or Grant No. 49198-26-28
12. Sponsoring Agency Name and Address University Transportation Research Center Marshak Hall - Science Building, Suite 910 The City College of New York 138th Street & Convent Avenue ,New York, NY 10031		13. Type of Report and Period Covered
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Demand for vehicle and public transportation systems continues to increase in and around major urban centers. This increase is especially pronounced during the morning and evening commutes and is further complicated by the complex spatial interactions that influence the variation in system demand. In an effort to help agencies better understand this variability and develop better demand forecasts this research investigated the underlying factors impacting public transportation ridership regardless of transit mode, then uses this insight to estimate specific models to help forecast changes in subway ridership. The spatial database for the case study consisted of social, economic, and land use characteristics for the 2166 census tracts in the five boroughs of New York City, NY. The spatial models were found to have a better overall model fit compared to their non-spatial counterparts. Moreover, spatial dependence was found to be statistically significant in both models. Failure to account for spatial dependence in estimating public transportation use at the census tract or station level could lead to biased, inefficient or inconsistent parameter estimates. The completed research can help public agencies better address resource allocation by identifying locations that are over or underperforming in terms of expected ridership or identifying locations for network expansion.

17. Key Words Public Transportation, Spatial Econometrics, Census Tract		18. Distribution Statement	
19. Security Classif (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No of Pages 47	22. Price

Form DOT F 1700.7 (8-69)

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. The contents do not necessarily reflect the official views or policies of the UTRC or the Federal Highway Administration. This report does not constitute a standard, specification or regulation. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

Demand for vehicle and public transportation systems continues to increase in and around major urban centers. This increase is especially pronounced during the morning and evening commutes and is further complicated by the complex spatial interactions that influence the variation in system demand. In an effort to help agencies better understand this variability and develop better demand forecasts this research investigated the underlying factors impacting public transportation ridership regardless of transit mode, then uses this insight to estimate specific models to help forecast changes in subway ridership. The spatial database for the case study consisted of social, economic, and land use characteristics for all 2166 census tracts in New York City, NY's five boroughs. The data were used to estimate spatial econometric models for the percentage of commuters using public transportation at the census-tract level and the change in subway ridership between 2011 and 2016 at the subway station-level.

Analysis of the commuters indicate that census tracts with a higher average commute time, greater employed population, higher per capita income and lower median household income were found to have a higher percentage of commuters using public transportation. Additionally, this percentage increased if neighboring census tracts had a greater commercial space area or fewer buildings. Lastly, the percentage increased in a tract if it increased in neighboring tracts and vice-versa. The may be reflecting social norms or stigma related to public transportation versus personal vehicle ownership.

Results of the change in subway ridership between 2011 and 2016 indicate that subway stations that serve more train lines or are in areas comprised of census tracts with a greater number of tax units (residential, commercial, etc.) or lower mean household incomes

experienced a greater increase in ridership. Furthermore, subway stations located in areas surrounded by census tracts with more commercial property or higher median family income are also expected to have a greater increase in ridership. Lastly, ridership at a given station decreases due to an increase in ridership at neighboring stations. This may indicate that a change in ridership at a station is due, in part, to riders in a region changing which station they use instead of riders shifting from alternative modes of transportation.

The spatial models were found to have a higher overall model fit compared to their non-spatial counterparts. Moreover, spatial dependence was found to be statistically significant in both models. Failure to account for spatial dependence in estimating public transportation use at the census tract or station level could lead to biased, inefficient or inconsistent parameter estimates. The completed research can help public agencies better address resource allocation by identifying locations for network expansion or locations that are over or underperforming in terms of expected ridership.

1 INTRODUCTION

1.1 Background

Municipalities invest in public transportation systems, in part, to combat increasing road congestion. Investments in subway and bus rapid transit systems directly remove drivers from the road network onto alternative transportation systems. Investments into on-road public transportation, such as local and city public bus systems, help to alleviate congestion by increasing vehicle occupancy rates and thereby reducing the average space in the network consumer by each individual user. Investments in rail networks, such as region's commuter rail or city's subways, can result in users shifting off of the road network for a portion of their trips. The biggest impact can be expected during the busiest travel times, specifically during the morning and evening commute. Analysis of public transit use and commuters is complicated by the fact that public transit use in general and specifically for the purpose of commuting is not constant over space. There are census tracts in New York City that have less than 10% of commuters reporting using public transportation while other census tracts in the city report over 90%. In a city with easy access to public transportation (bus, subway, etc.), there is a need to explore the influential factors that contribute to this variation in use. This information could help public agencies better address resource allocation by identifying locations that are over or underperforming in terms of expected ridership or identifying locations for network expansion.

The impact that land use and socioeconomic demographics have on public transportation use for the purpose of commuting in not constant over space. Previous research on this topic have implicitly ignored the direct and underlying spatial processes (Cervero, 1993; Kitamura et al., 1997; Kyte et al; Kain and Liu 1999; Boarnet and Greenwald, 2000). In reality, it is observed that passengers embarking or disembarking at a given transit stop live, work, and recreate in both the immediate neighborhood as well as the adjacent neighborhoods. If public transportation use is high in a given census tract it could indicate that access to public transportation is especially convenient in the census tract which would could increase the percentage of commuters using public transportation in neighboring tracts. There could also be a negative stigma associated with public transportation in a given tract that could impact the decision making of those living in the census tract as well as those living in neighboring census tracts. Vehicle ownership is seen by some as a symbol of social status with public transportation perceived as a less desirable mode of travel. These social norms can propagate out of a region and effect the decision making of neighboring areas. These complex spatial interactions are crucial to understanding the observed variation in public transportation use.

1.2 Scope and Motivation

This research combined big-data visualization with spatial econometric modeling. Data collection included demographic and economic census data analyzed at the census tract-level paired with comprehensive land use and valuation based on tax lot records. This enabled the visualization of the complex interrelationships in the data. Spatial econometric models were used to capture the complex spatial trends that characterize the relationship between the influential factors and public transit use. The five boroughs of New York City (NYC) are used as a case study (Figure 1). The research investigates two important aspects of public transportation use. The first is public transportation used specifically for commuting regardless of transit mode. These insights were then used to develop models to estimate the five-year change in subway ridership. The underlying causes for variability in public transportation use and ridership are not constant over space which, if left unaccounted for in statistical and econometric models, will yield biased, inefficient, and inconsistent results (Anselin, 1988a; 2006; Anselin and Rey, 2014). The impact of spatial dependence can be investigated using lagged independent variables, known as cross-regressive terms, lagged dependent variables, and/or by applying a spatial process to the error term. In the context of the current research, cross-regressive terms quantify the change in public transportation use or subway ridership in a given census tract or at given subway station due to the land use and socioeconomic characteristics of neighboring census tracts. The lagged dependent variable captures the propensity for the public transportation use/subway ridership at one location to be impacted by the use/ridership in neighboring locations. In doing so, the lagged dependent variable accounts for global spillovers in that the use/ridership at one location is a function of the use/ridership of its neighbors, which is a function of their neighbors' use/ridership, and so on. The overall goal of the research is to develop efficient and unbiased models for commuters' use of public transportation and the five year change in subway ridership. The resulting models can be used by public agencies and decision makers to identify locations that are over or underperforming in terms of expected use/ridership for the purposes of resource allocation while also providing the framework to identify the best geographic areas for network expansion.

Figure 1. Study Area: New York City Boroughs

1.3 Review of Current Literature

Understanding the link between the socioeconomic conditions of a region and the propensity for those who reside in the region to use public transportation informs transportation policy-making at the state, regional, and local levels. Transportation agencies and policy makers have long grappled with shrinking budgets precipitating the need to optimize investment while ensuring an equitable distribution of the benefits of public transportation across user groups. In response to this need, research has continuously sought to better quantify underlying causes for variance in public transportation use in geographic regions. The body of literature on this topic has implicitly assumed that the causes of variance are constant over space. However, if this assumption is not reflected in the actual data then the resulting econometric models would have the potential to lead to biased, inefficient, and inconsistent results (Anselin, 1988a; 1988b; 2006; Anselin and Rey, 2014). People are not limited to the public transportation options provided in the census tract in which they live, but are more likely to use public transportation options closer to their homes. For this reason the propensity for the people living in a given area (census tract) to use public transportation in general or a specific mode is informed by the characteristics of their home and neighboring tracts.

Previous research has investigated the variation in transit use over time as a function of economic and land-use characteristics for a limited number of locations (Boarnet and Greenwald, 2000; Kain and Liu 1999). Research focusing on larger cross-sectional studies of public transportation for commuting and non-commuting trips (Cervero, 1993; Boarnet and Greenwald, 2000; Kitamura et al., 1997) have not accounted for spatial effects. Limited research has been conducted that directly investigates the underlying spatial processes evident in the data. The research that has addressed the issue focused on unobserved spatial process (spatial error) at the state-level without accounting for local spillovers (Chakrabortya and Mishrab, 2013). In additional to the limited research on public transportation, past research has used spatial econometric analysis to determine the relationship between socioeconomic factors and vehicle use and vehicle ownership (Badoe and Miller, 2000; Volovski, 2015). Some of the research estimated the average individual or household vehicle-miles-traveled (for a zip-code or census tract) as a function of local and lagged socioeconomic variables (Frank et al., 2000; Cook et al., 2012). In addition to the limited past research on spatial modeling of vehicle use and ownership data, there have been spatial econometric applications in other areas of transportation research, most notably in transportation safety data analysis and modeling. Spatial autocorrelation regression estimation techniques have been used to model crashes involving vehicles and pedestrians (LaScala et al., 2000; Schneider et al., 2000), and vehicles only (Boarnet and Greenwald, 2000; Li et al., 2007; Aguero-Valverde and Jovanis, 2008; Erdogan 2009). Furthermore, research has shown the influence of socioeconomic characteristics on vehicle crash rates across regions (Kirk et al., 2005; Stamatiadi and Puccini 1999).

2 DATA

Social, economic, and land use data were collected and analyzed. The data was used to investigate two measures public transportation use in large metropolitan centers. The first measure, commuter transit use, is defined as the percentage of commuters in a census tract that that use public transportation as their primary means of travel to and from work. The second measure, change in subway ridership, is defined as the five year change in subway ridership by station based on annual ridership data from 2011 to 2016.

Social, economic, and land use data was aggregated at the census tract-level due to the relative consistency across tracts in terms of key characteristics such as population size (between 2,000 and 8,000). High quality data for the 2166 census tracts (Figure 2) across the five boroughs of New York City is available from regional, state, and national sources. New York City was chosen as the case study location due to availability and accessibility of its public transportation system. Its extensive network of public transportation modes includes the nation's largest subway system in terms of ridership, length, and number of stops and largest bus system in terms of total ridership (APTA, 2017).

Figure 2. New York City Census Blocks

2.1 Data Collection

Data were obtained from multiple sources. Social and economic data, including commuter mode choice, were obtained from the 2015 American Community Survey conducted by the United States Census Bureau (U.S. Census, 2015). Land use data were obtained from the 2017 PLUTO database administered by the New York City Department of City Planning (PLUTO, 2017). Subway ridership data were obtained from the Metropolitan Transportation Authority. Table 1 provides a brief summary of the descriptive statistics for the factors that have been shown to influence transit use.

Table 1 Descriptive statistics for select census tract variables

	Average	Standard Deviation	1st Quartile	Median	3rd Quartile	Max
Percent commute with public transportation (0.000 – 1.000)	0.547	0.166	0.434	0.572	0.678	1.000
Mean travel time to work (minutes)	40.45	7.17	36.80	41.10	44.70	100.00
Mean household income (in 2014 inflation adjusted dollars)	78,555	44,948	52,538	69,434	89,512	435,803
Median household income (in 2014 inflation adjusted dollars)	58,543	28,919	38,608	53,868	73,044	250,000
Per capita income (in 2014 inflation adjusted dollars)	31,488	24,732	18,364	24,746	34,403	247,852
Percent with health insurance (0.000 – 1.000)	0.868	0.073	0.828	0.874	0.920	1.000
Building area (gross area for all buildings in 1,000,000ft^2)	2.527	2.955	1.174	1.788	2.690	54.138
Commercial area (gross commercial area in 1,000,000ft^2)	0.841	2.117	0.136	0.300	0.623	26.210
Land area (land area in ft^2)	2.976	8.755	1.098	1.313	2.134	214.976
Number of residential units (total)	1,625	1,290	841	1,338	2,008	14,338
Total assessed value (total of all tax lots in $1,000,000)	160.21	418.27	31.09	52.92	99.06	6,800.97

New York City subway ridership and station location data were obtained from two open-sourced databases compiled by the New York City Metropolitan Transportation Authority (MTA 2015; 2017). Ridership data included the annual ridership by borough and average weekday and weekend ridership for each subway station. The total subway ridership per borough is broken down in Figure 3.

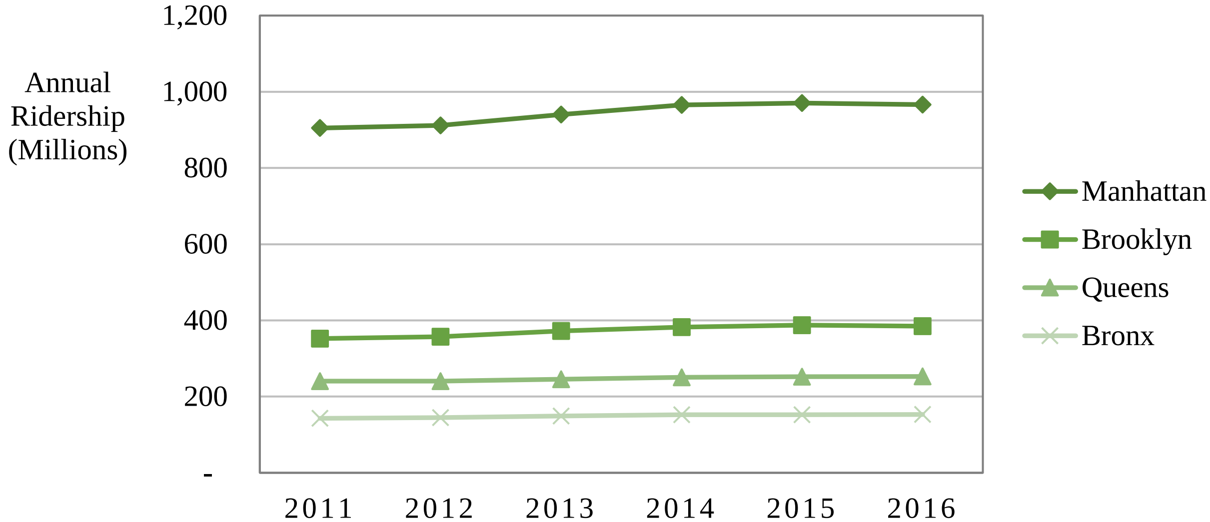


Figure 3. Annual New York City Subway Ridership by Borough

It is important to note that the ridership statistics reflect the number boarding at each station not the total volume of riders at the station (entering, exiting, and pass-through). Figure 4 shows the location of the subway stations. The stations are spread across all the boroughs except Staten Island. Station locations were obtained from the MTA station entrance database. This dataset included; the coordinates (latitude and longitude) for all entrances to each station, the coordinates of each station, the subway lines accessible at each station, and the ADA accessibility of each entrance. Since ridership data was station specific, the location and accessibility data obtained from the station entrance database had to be grouped by station before being merged with the ridership data.

Figure 4. New York City Subway Stations

It can be difficult to quantify the factors that influence subway ridership in a region where the variation in ridership is immense. For instance, in 2016 the average ridership at the ten busiest subway stations was over 32 million annual riders, whereas the average at the ten stations with the lowest ridership was just over 250 thousand annual riders. Investigating the change in ridership, in terms of magnitude or percent change, can reduce the variability in the data, thereby allowing for a more nuanced analysis of the underlying social, economic, and land use factors that influence subway ridership. The change in annual ridership and the percent change in annual rider between 2011 and 2016 is illustrated in Figure 5 and Figure 6, respectively.

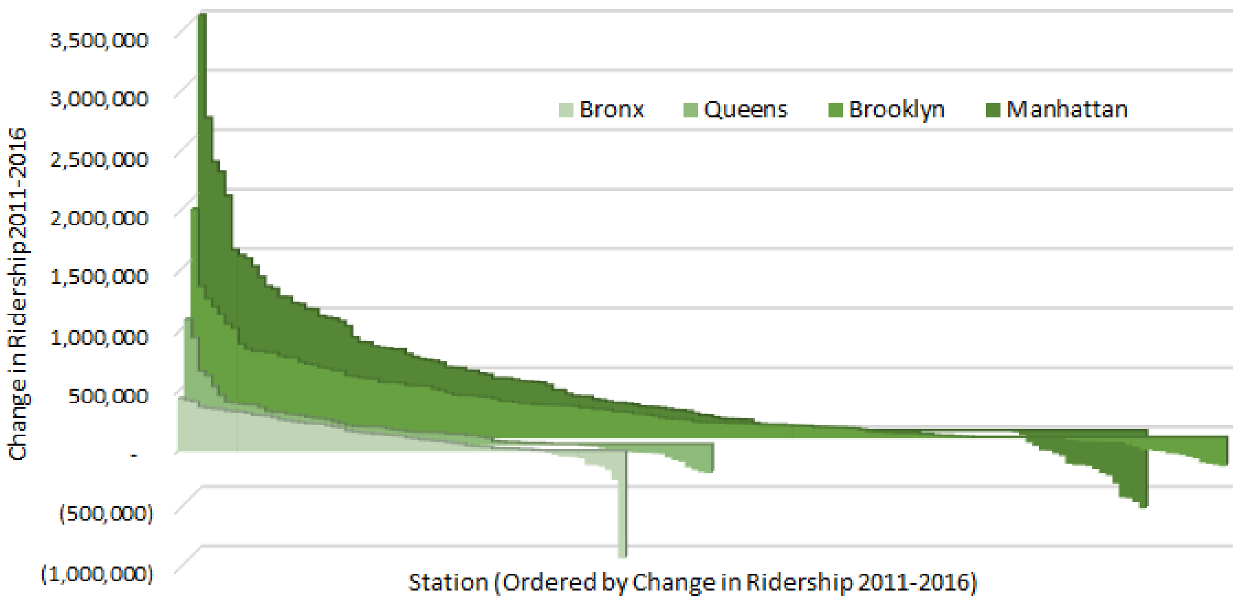


Figure 5. Change in Annual Subway Ridership 2011-2016 by Station

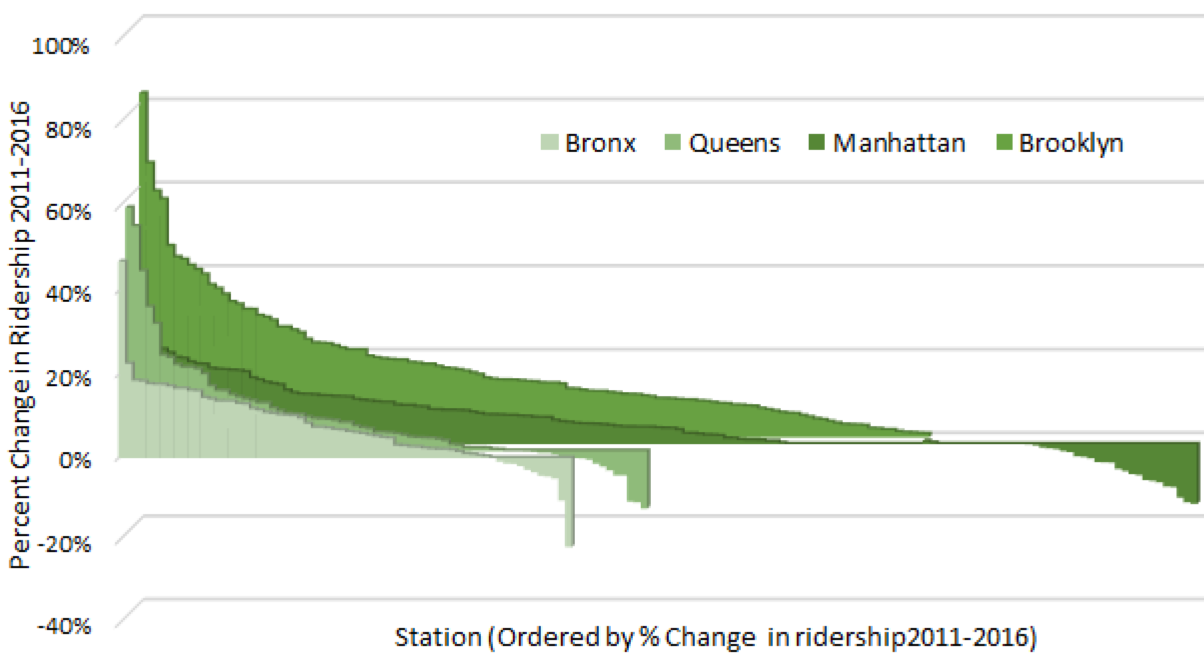


Figure 6. Percent Change in Annual Subway Ridership 2011-2016 by Station

3 METHODOLOGY

3.1 Overview

The social, economic, and land use data detailed in the previous section were used to investigate two measures of public transportation use in large metropolitan centers. The first measure, commuter transit use, is defined as the percentage of commuters in a census tract that use public transportation as their primary means of travel to and from work. The second measure, change in subway ridership, is defined as the change in subway ridership from 2011 to 2016. Its extensive network of public transportation modes includes the nation's largest subway system in terms of ridership, length, and number of stops and largest bus system in terms of total ridership (APTA, 2017*0*). Spatial econometric modeling techniques were used to investigate the factors effecting transit use while accounting for spatial processes. All spatial econometric modeling was completed using the spatial software GeoDa and GeodaSpace (Anselin et al., 2006a).

3.2 Spatial Weights Matrix

The spatial weights matrix is used to define the connectivity between a location and its neighbors. Connectivity can be defined by the form (rook, queen/king, k nearest neighbors, distance) and the extent (order or number). Connectivity in 1st order rook matrix are all regions that share an edge, a 1st order queen/king matrix is a rook matrix that includes regions that only share a single vertex, elements in a k-nearest neighbor matrix all have the same number of neighbors (k), and connectivity in a distance matrix is defined by the distance between the regions, typically measure between the regions' centroids (Anselin and Rey, 2014; Cliff and Ord, 1981).

3.2.1 Moran's I

The Moran's I is a measure of spatial autocorrelation of a given variable and is defined in Equation 1. The null hypothesis is that the variable exhibits spatial randomness, the alternative is spatial dependence. Moran's I takes the form:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} \left(X_i - \hat{X}\right)\left(X_j - \hat{X}\right)}{\sum_i \left(X_i - \hat{X}\right)^2} \qquad 1$$

Where $(x_i - x)$ is the rate of region i centered on the mean for $i \neq j$, N is the number of regions, and w_{ij} is the weight between region i and j. The statistical significance of Moran's I can be determined using random permutations where the variable of interest is randomly reassigned across the geographic units.

3.3 Models for Spatial Dependence and Spatial Heterogeneity

Spatial process models can take a variety of forms depending on which functional components (dependent variable, independent variables, and/or error) have a spatial process applied (Anselin, 1988a; 1988b; Anselin and Rey, 2014). Spatial error models can be estimated to ensure that regression estimation is efficient in instances where spatially correlated error terms are observed in the dataset (Anselin, 1988a). The spatial error model takes the form Anselin, 1988b; Anselin and Rey, 2014):

$$y = \beta x + \varepsilon \qquad 2$$

$$\varepsilon = \lambda W \varepsilon + \mu$$

where the dependent variable, y, is a function of a vector of independent variables, x, and a spatial error term ε. Spatial autocorrelation is accounted for in the error term by introducing the weights matrix, W. μ is a non-spatial error term. The estimated parameter λ can be tested to determine if the spatial error is statistical significance.

Spatial dependence can be accounted for by using a spatial lag or a cross regressive model. Failure to account for spatial dependence will result in biased and inconsistent parameter estimates (Anselin, 1988b; Anselin, 2006). The dependent variable in a spatial lag model is estimated as a function of the observation's independent variables and its neighbor's dependent variable. In the context of the commuter transit use model it means that the dependent variable, percentage of commuters using transit in a given census tract, is a function of the attributes of the census tract and the percentage of commuters that use transit in neighboring census tracts. Cross-regressive estimation allows for the dependent variable to be estimated as a function of the attributes of the given census tract and the attributes of neighboring census tracts.

The Spatial Durbin Model incorporates both spatial lag and cross-regressive terms. It takes the form (Anselin, 1988b; Anselin and Rey, 2014):

$$y = \rho W y + \beta x + \gamma W Z + \mu \tag{3}$$

where Wy is the spatial lag term, WZ is a vector of cross-regressive terms, and ρ and γ are coefficients for the lagged dependent variable and lagged independent variables (cross-regressive terms), respectively. It is important to note that regardless of the type of weights matrix chosen, the lagged independent variable and spatial error term will account for global spillovers. The endogeneity of the lagged dependent variable can be overcome with two-stage least squares estimation (2SLS), a special case of instrumental variables (IV) (Anselin and Rey, 2014). The General Spatial Durbin is a special case of the Spatial Durbin where spatial lag, spatial error, and cross regression are all statistically significant.

3.3.1 Tests for Spatial Lag and Spatial Error

The Lagrange Multiplier (LM) and the robust LM test for spatial lag are used to determine if spatial error, spatial lag, or spatial error and lag are statistically significant in the data (Anselin, L.,1988c; Anselin et al., 1996; Anselin and Rey, 2014,). The LM test for spatial error determines if the spatial error coefficient is statically different from zero. Likewise, the LM test for spatial lag determines the statistical significance of the spatial lag coefficient (Anselin, 2006). The null hypothesis for the LM test for error is that spatial error coefficient (λ) is equal to zero. Equation 4 details the LM test for error.

$$\mathrm{H}_0\colon \lambda = 0 \tag{4}$$

$$\mathrm{H}_A\colon \lambda \neq 0$$

$$\text{for } y = \beta x + \lambda \mathrm{W} + \varepsilon$$

$$LM\lambda = \left[\frac{\mathrm{e}'\mathrm{We}}{\mathrm{s}^2}\right]^2 / T \sim \chi_1^2$$

$$T = tr[(W' + W)]$$

$$s^2 = e'e/n$$

Likewise, the null hypothesis for the LM test for lag is that the spatial lag coefficient (ρ) is equal to zero. Equation 4 details the LM test for lag.

$$\mathrm{H}_0: \lambda = 0 \tag{5}$$

$$\mathrm{H}_A: \lambda \neq 0$$

for $y = \rho \mathrm{Wy} + \beta x + \mu$

$$LM\rho = \left[\frac{\mathrm{e'Wy}}{\mathrm{s}^2}\right]^2 \Big/ nJ_{\rho\beta} \sim \chi_1^2$$

$$J_{\rho\beta} = [(WX\beta)'M(WX\beta) + Ts^2]/n\,\mathrm{s}^2$$

$$M = I - X(X'X)^{-1}X'$$

The LM test for spatial lag and the LM test for spatial error are unidirectional in that they only test for the presence of one spatial process. The robust LM tests can be used to test for either spatial lag or spatial error while accounting for the presence of the other. The robust LM test for lag and error are provided in Equations 6 and 7, respectively

$$LM\lambda^* = \left[\frac{\mathrm{e'We}}{\mathrm{s}^2} - T\left(nJ_{\rho\beta}\right)^{-1}\frac{\mathrm{e'Wy}}{\mathrm{s}^2}\right]^2 \Big/ T[1 - T\left(nJ_{\rho\beta}\right)^{-1} \tag{6}$$

$$\left[\frac{\mathrm{e'Wy}}{\mathrm{s}^2} - \frac{\mathrm{e'We}}{\mathrm{s}^2}\right]^2 /[\, nJ_{\rho\beta} - T] \tag{7}$$

The Anselin-Kelejian (AK) test is a variant of the Moran's I statistic applied to the residuals of the estimation. If the AK test results are statistically significant it indicates that there is spatial autocorrelation left unaccounted for in the developed model (Anselin and Kelejian, 1997).

4 RESULTS: COMMUTERS USING PUBLIC TRANSPORTATION

New York City is home to one of the Nation's most robust public transportation networks. Even though there is nearly uninhibited access to the network, the variability in its use as a primary mode choice for commuters ranges from 10% to 90% across geographical units. This section presents the results of spatial econometric modeling to examine the social, economic, and land use characteristics that influence public transportation ridership for work trips. A non-spatial model, estimation without spatial error or lag, was estimated to provide a baseline for comparison. The non-spatial model results are presented in Table 2. The results indicate that the percentage of commuters using public transit living in a census tract is a function of the employed population, per capita income, mean travel time to work, and median family income. The R-squared and adjusted R-squared values indicate that the non-spatial model is explaining 32% of the variance exhibited in the census tract commuter transit data.

The non-spatial model has a corresponding multicollinearity condition number of 21.91. This value is used as an indication of the degree to which explanatory variables show a linear relationship. In statistics, it is generally agreed upon that multicollinearity should be addressed if the condition number is greater than 30 (Anselin and Rey, 2014). The Jarque-Bera test for non-normality of the error terms was significant at a 99% level of confidence (Jarque and Bera, 1980). Therefore, in order to test the residuals for homoscedasticity (consistency of the error variance) a Koenker–Basset test, a variant of the Breusch-Pagan test (Breusch and Pagan, 1979), was used because, unlike the Braush-Pagan test, it does not assume normality of the error terms. The Koenker–Basset test value was significant at a 99% level of confidence. To account for heterogeneity, White-adjusted standard errors that are robust to heteroskedasticity were used (White, 1980). A General Spatial Durbin specification was estimated to obtain the coefficient for the spatial error term, λ. It had an estimated coefficient of -0.042 and a corresponding 61% level of confidence, well below any acceptable threshold to reject the null hypothesis of no spatial error. Therefore, final model specification is the Spatial Durbin.

4.1 Spatial Autocorrelation

Spatial autocorrelation is a measure of the correlation a variable has with itself in space (Anselin, 1988a; 1988b; Anselin and Rey, 2014). Preliminary analysis of spatial autocorrelation can be completed by investigating a plot of the dependent variable over space to discern if there appears to be spatial clustering of relatively higher or lower values (positive autocorrelation). Figure 7 shows that there are clusters of census tracts with a higher percentage of commuters using public transit in northern Manhattan, selected areas of the Bronx and Brooklyn, whereas there are clusters of census tracts in Queens and Staten Island with relatively lower transit use.

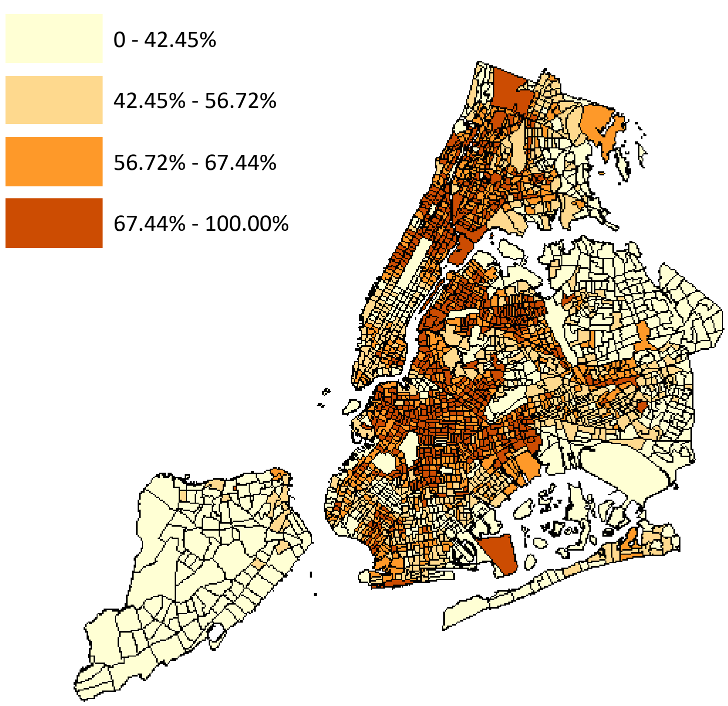


Figure 7. Quantile Plot for Percentage of Commuters Using Public Transportation

When spatial autocorrelation is positive it indicates that a census tract with a higher percentage of commuters using public transit correlates with neighboring tracts that also have a higher percentage (or smaller values correlate to smaller neighbor values). Negative autocorrelation occurs when greater values are correlated with smaller neighbor values (and vice-versa). Spatial clustering is visible in the local indicator of spatial association (LISA) cluster map shown in Figure 8. The LISA map is best defined by Anselin as: "The LISA for each observation [say, a small region among a set of regions] gives an indication of significant spatial clustering of similar values around that observation. The sum of LISAs for all observations is proportional to a global indicator of spatial association" (Anselin, 1995; Anselin et al., 1996).

Autocorrelation is found to exist in nearly half of the census tracts (960 out of 2106). Positive correlation was found in 917 tracts of which 503 were high-high spatial clustering and 414 were low-low spatial clustering. High-high clustering is primarily found in tracts spread across northern Manhattan, southern Bronx, and northern Brooklyn. In these areas, a census tract with a higher percentage of transit use is more likely to have neighboring census tracts with higher transit use. There were only 43 instances of negative autocorrelation which means a census tract with high transit use is more likely to have a neighbor with low transit use. There are three neighbor-less census tracts which are located on islands with a small enough population to require a single census tract (unlike an island such as Roosevelt Island that is comprised of multiple census tracts).

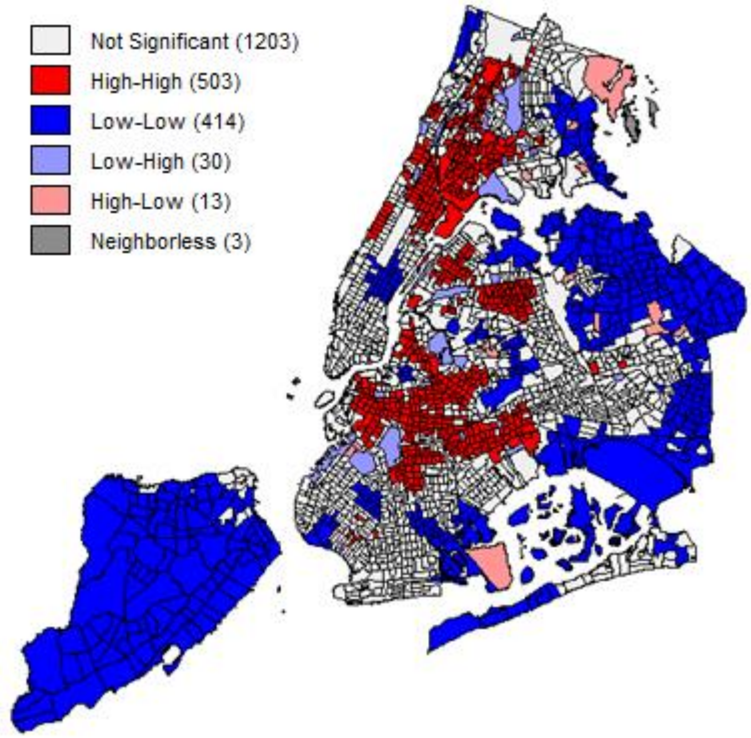


Figure 8. LISA Cluster Map for Percentage of Commuters Using Public Transportation

The statistical measure, Moran's I, is used to measure the spatial autocorrelation in the dataset and can be compared to a randomized spatial set to test the hypothesis of the presence of spatial autocorrelation (Cliff and Ord, 1981). A plot of the Moran's I is presented in Figure 9. The Moran's I was calculated to be 0.58047 with a corresponding z-score of 45.36 and p-value of 0.002 based on 999 random permutations. This indicated that there is spatial heterogeneity at a 99.8% level of confidence.

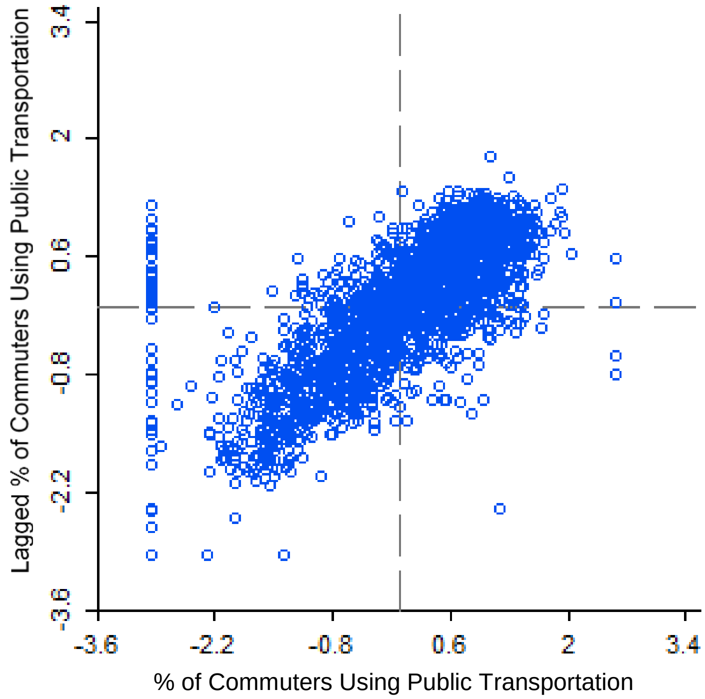


Figure 9. Moran's I for 1st Order Queen Matrix (Percentage of Commuters Using Public Transportation)

4.2 Model for Spatial Dependence

It was determined that the commuter transit use data exhibits spatial processes demonstrated by a lagged dependent variable and two cross-regressive terms. The spatial processes are best captured using a 1st order queen matrix. Figure 10 provides a histogram of the 1st-order queen connectivity of census tracts in NYC in terms of the number of neighbors each tract has.

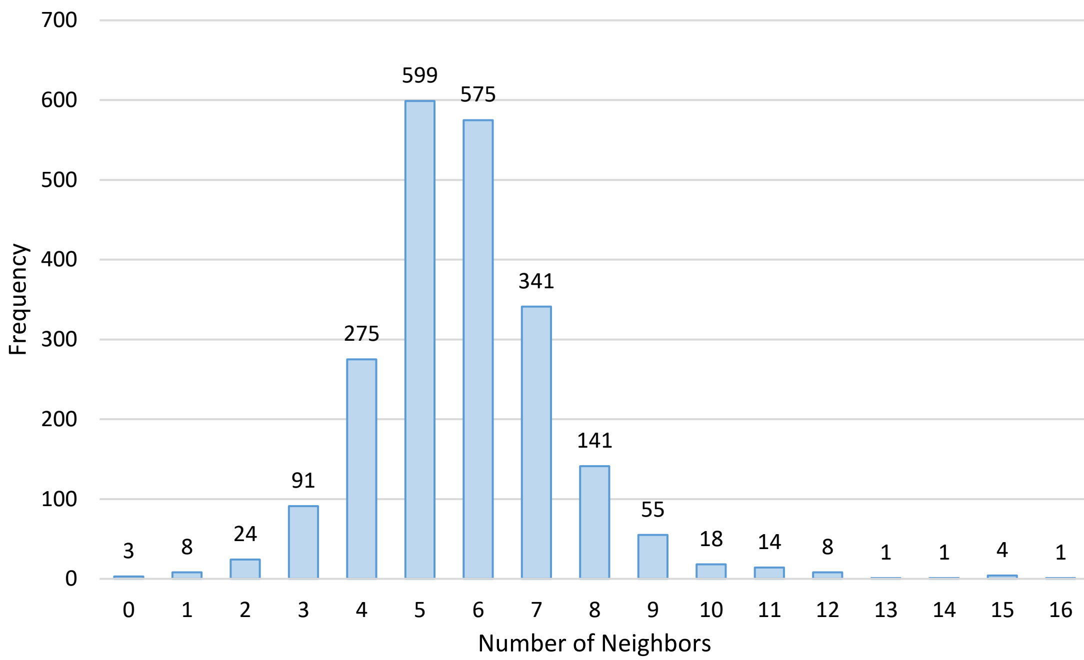


Figure 10. Connectivity Frequency Distribution for 1st Order Queen Matrix (Percentage of Commuters Using Public Transportation)

The resulting Spatial Durbin model is presented in Table 2. The results were used to investigate the influence each variable had on the expected percentage of commuters using public transportation in each census tract.

Table 2. Commuters Using Public Transportation Modeling Specification Results

Response Variable: Percent of Commuters using Public Transit in the Census Tract (0.000 – 1.000)	Non-spatial	Spatial Durbin w/ White Adj. Std. Errors
Independent Variables	**Coefficient**	**Coefficient**
Constant	0.2007***	-0.1243***
Mean travel time to work (minutes)	0.0086***	0.0068***
Employed population (age 16 and older in 1,000s)	0.0295***	0.0162***
Per capita income (in $10,000s)	0.0410***	0.0127***
Median family income (in 2014 inflation adjusted dollars)	-0.0409***	-0.0120***
Total assessed value of all tax lots (in $100,000,000)	-0.0293***	-0.0206***
Percent with health insurance (0.000 – 1.000)	0.0592**	0.1298***
Cross-Regressive Variables (1st Order Queen Weights Matrix)		
Number of buildings (total for all tax lots)		-0.0001***
Percent commercial floor area (com/total; 0.000 – 1.000)		0.1110***
Spatial Lag Variable (1st Order Queen Weights Matrix)		
Percent of commuters using public transit (0.000 – 1.000)		0.5789***
Model Statistics		
Number of Observations	2166	2166
R-squared	0.3248	
Adjusted R-squared	0.3232	
Pseudo R-squared		0.6993
Spatial Pseudo R-squared		0.5167

***99% Level of Confidence, **95% Level of Confidence, *90% Level of Confidence

The estimated coefficients were determined to be statistically significant at the 99% level of confidence. The resulting model experienced an increased statistical fit compared to the non-spatial model exhibited by a spatial pseudo R-squared value 0.5167.

The results show that census tracts with a higher average commute time have a higher percentage of commuters using public transportation. This may indicate that more secluded census tracts are better served by public transportation compared to alternative travel modes such as driving your own vehicle. It was shown that census tracts with a greater employed population have a higher percentage of commuters using public transit. Two variables were included in the model to capture the relative wealth of the individuals in the census tract. An increase in per capita income increases the expected percentage commuters relying on public transportation whereas the median household income had an inverse relationship with the dependent variable. Census tracts with a higher household income would be expected to have a greater number of households that can afford at least one vehicle, which is the main alternative to public transportation use for commuting. This relationship is supported by the total assessed value variable. Census blocks with more valuable real estate would be expected to house a greater percentage of commuters that could more easily afford the costs associated with car ownership in the city. It would also be expected that these employed people would be more likely to have jobs with flexible work hours allowing them to avoid rush hour traffic. An increase in health insurance increases the percentage of commuter trips using public transportation. This may indicate that there is a stigma or distrust of public agencies among certain subsets of the population regardless of whether the agency provides health insurance or access to transportation.

The framework identified two statistically significant cross-regressive terms. An increase in the number of buildings in neighboring census tracts reduces the percentage of commuters using public transportation. This may indicate that people feel safer walking to work if they are in more built-up areas and therefore wouldn’t rely as heavily on the bus or subway. Also, the number of buildings might translate to an increased number employment opportunities allowing a larger percentage of commuters to walk to work as opposed to using public transportation. It was determined that as the ratio of commercial floor space relative to total floor space of all buildings increases in neighboring tracts, the percentage of commuters using public transportation increases. Deliveries and customers arriving to these commercial spaces may make driving through these census tracts difficult thereby increasing the percentage of commuters using public transportation options to avoid driving to work.

Lastly, the lagged dependent variable was determined to be statistically significant. This means that an increase in the percentage of commuters who use public transportation increases in neighboring tracts increases the percentage of commuters who use public transportation in the given tract. This impact propagates back from the neighbors' neighboring tracts, and so on, providing for global spillovers. These impacts were shown in the raw data quantile and Lisa plots, Figure 7 and Figure 8, respectively. This means that use of public transportation is increased in a region if its use is high in neighboring regions, perhaps reflecting social norms. This also means that social pressures may reduce public transportation use if users feel that their neighbors aren't using public transportation. The hypothesis that these global effects were instead a function of an unobserved underlying spatial process was rejected because the spatial error coefficient was not statistically significant.

5 RESULTS: SUBWAY RIDERSHIP

A metropolis's subway system is typically the most expensive public transportation network. As such it is especially importation to understand the underlying causes of ridership fluctuations so that demand can be accurately forecasted. The current research focused on the five year change in ridership at subway stations between 2011 and 2016. The land use characteristics in the immediate area of the subway station and the socioeconomic characteristics of those individuals living in the area was used to investigate ridership change. The area associated with each station was defined according to the shortest arc distance between a location and subway station. Figure 11 illustrates this process. In the figure, all locations whose closest subway station (black dots) are considered to be the station's attributable area (blue polygon). The socioeconomic and land use characteristics of the census tracts (gray polygons) were joined with the subway data using GIS applications in the software R (R Core Team, 2017). These census-tract level characteristics were weighted according the percentage of the attributable area comprised by each tract. Some manual data cleaning was required to remove census tracts attributed to a given subway station that did not have a way for the individuals of that census tract to access the subway station (this typically occurred in the limited instances where a station's attributed area crossed a natural boarder such as water).

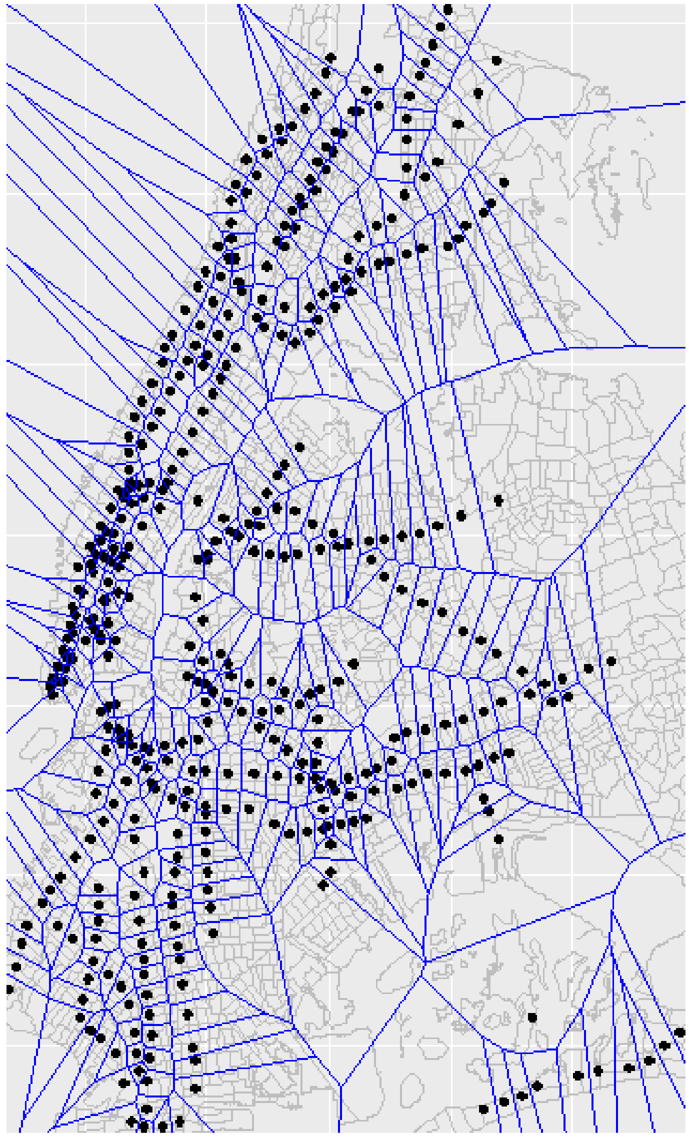


Figure 11. Subway Station Attributable Area

As is evident in Figure 11, the geographic area attributed to each subway station is much greater as one moves east from the Bronx and Manhattan to Queens and Brooklyn due to a decrease in station density. For this reason the subsequent subway ridership analysis was limited to the boroughs of Manhattan and the Bronx. In these regions individuals have greater access to the subway network with an increased likelihood of using a station that isn't the closest to their residence if it serves a different rail-line thereby reducing an individual's overall trip duration. It is therefore expected that spatial processes will be especially pronounced in these boroughs.

5.1 Spatial Autocorrelation

Figure 12 illustrates spatial distribution of the change in subway ridership between 2011 and 2016. A preliminary indication of spatial autocorrelation is evident in the appearance of spatial clustering of higher values in mid and lower Manhattan and northwestern sections of the Bronx.

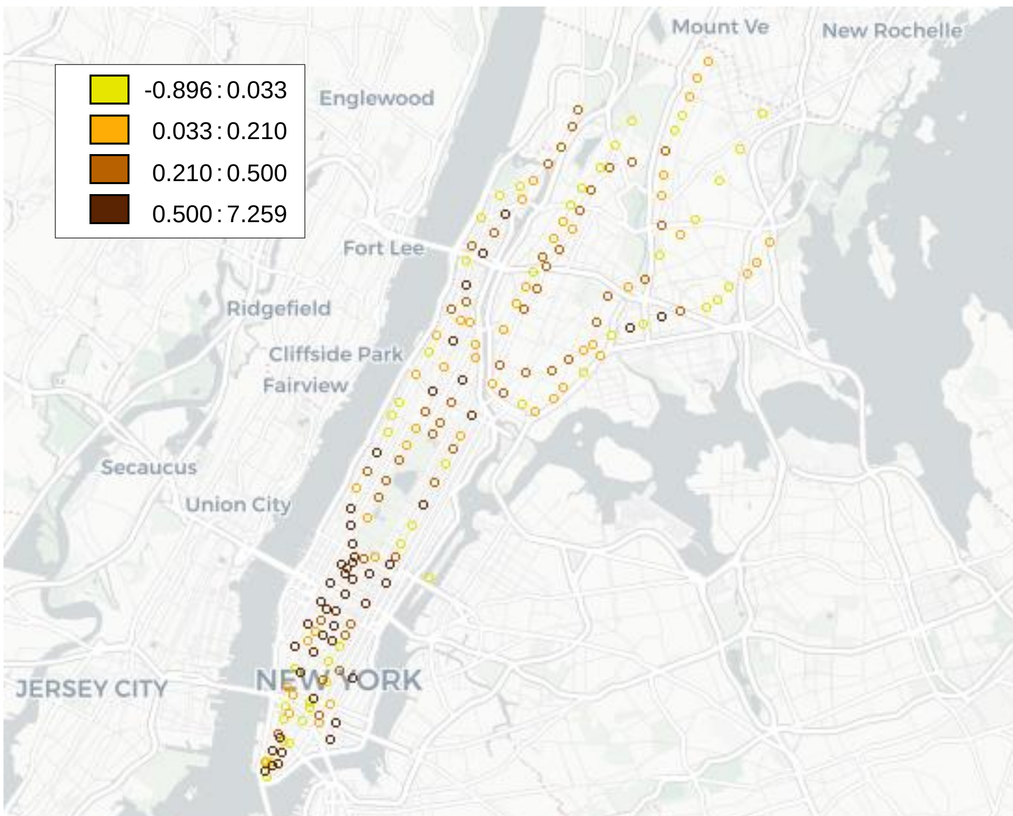


Figure 12. Quantile Plot for Change in Subway Ridership 2011-2016 (in Millions)

Distance-based weights matrices were used to investigate at what distance are the lagged independent and dependent variables significant. The first defined neighbors as subway stations separated by less than 1.0 miles the second defined neighbors as stations separated by less than 1.5 miles. An example of 1.0 mile neighbor is provided in Figure 13 where the 8 stations indicated by gray dots are defined as the neighbors of the station indicated with a black dot.

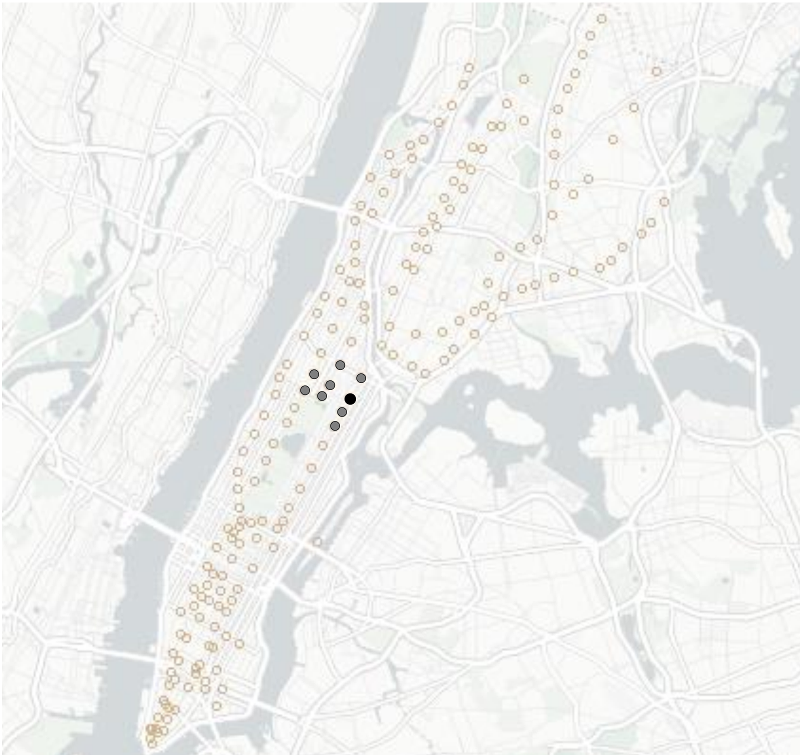


Figure 13. Example Neighbor Diagram (1.0 Mile Distance Matrix)

Figure 14 provides the corresponding histogram of the 1.0 mile connectivity of subway stations in NYC in terms of the number of neighbors for each station.

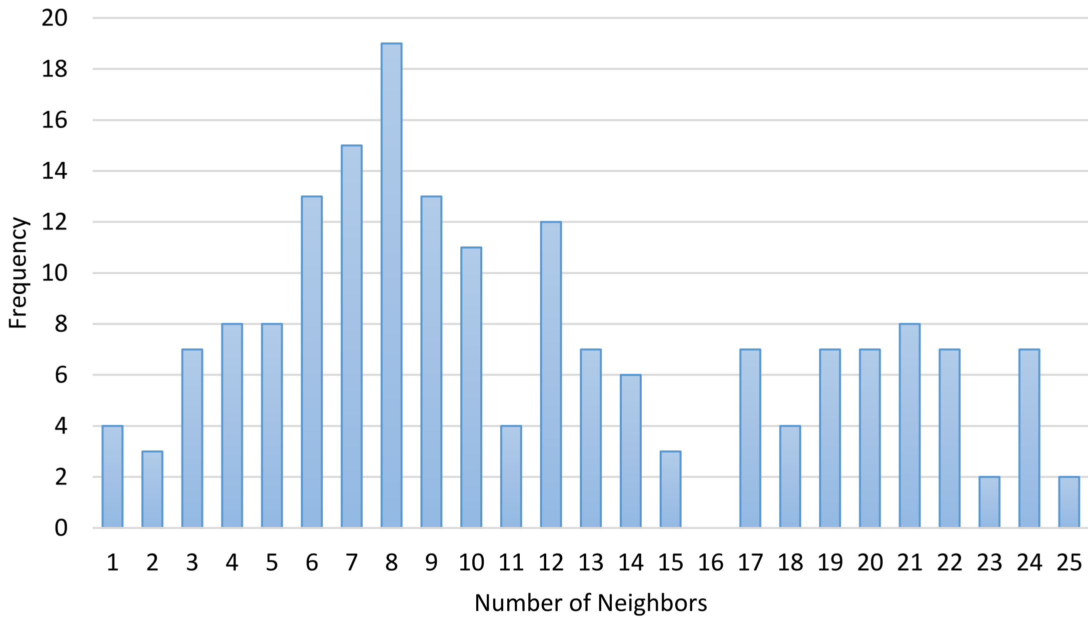


Figure 14. Connectivity Frequency Distribution for 1.0 Mile Weights Matrix

A plot of the Moran's I value for the change in subway ridership for 1.0 mile neighbor's weight matrix is presented in Figure 15. The Moran's I was calculated to be 0.096 with a corresponding z-score of 3.48 and p-value of 0.002 determined using 999 random permutations. This indicated spatial heterogeneity at a 99.8% level of confidence. This process was repeated for the 1.5 mile neighbor weights matrix (Figure 16) with a corresponding Moran's I of 0.069 which is statistically significant at a 99.1% level of confidence.

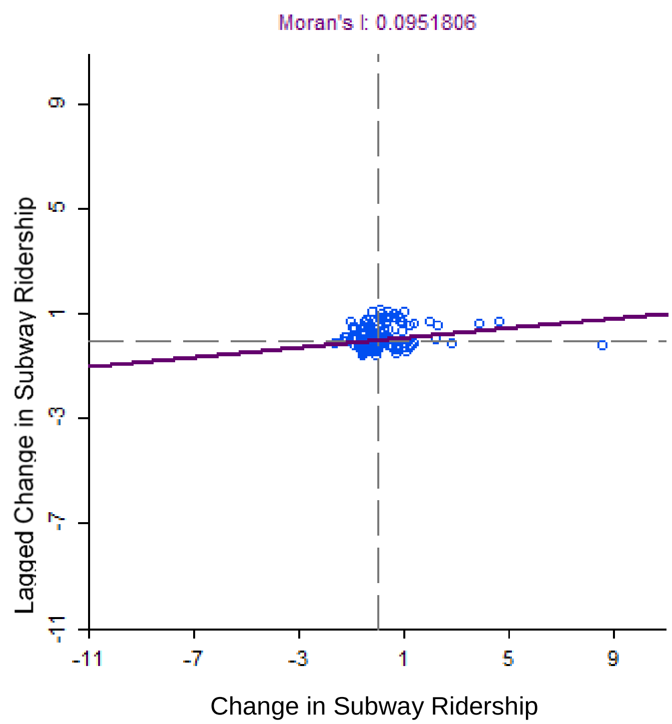


Figure 15. Moran's I for Change in Ridership 2011-2016 (1.0 Mile Distance Matrix)

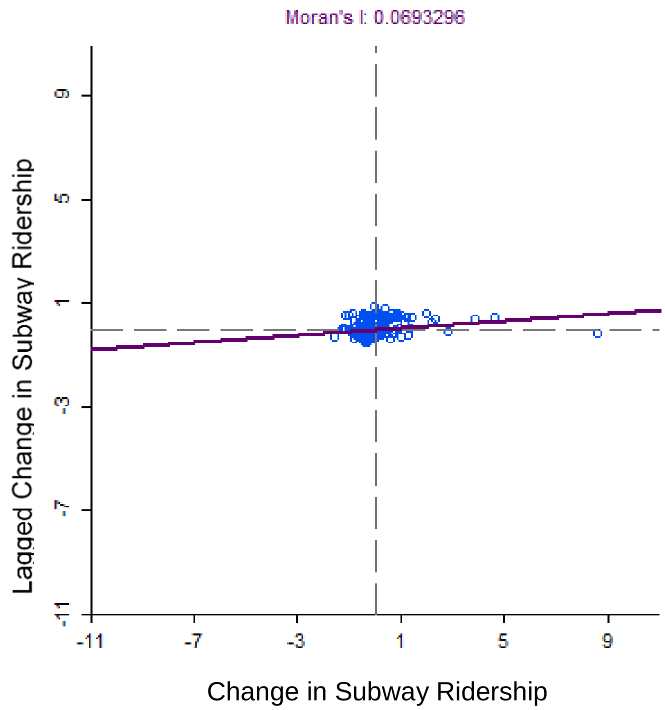


Figure 16. Moran's I for Change in Ridership 2011-2016 (1.5 Mile Distance Matrix)

5.2 Model for Spatial Dependence

The change in subway ridership is best modeled using lagged dependent and independent variables. It was determined that the spatial lag of the dependent variables was best captured using a 1.0 mile distance-based weights matrix, whereas the lagged independent variable should be weighted using a 1.5 mile distance based matrix. The LM test for spatial lag was significant at a 90% level of confidence whereas the LM test for spatial error was not statistically significant. The robust LM test for spatial lag and the robust LM for spatial error were both statically significant at a 95% level of confidence. This indicates that the best model for the data is one that incorporates spatial lag and may need to incorporate spatial error. A General Spatial Durbin model (incorporating both spatial lag and error) was estimated using generalized methods of moments. The spatial error coefficient was determined to be statically insignificant. Therefore the final model specification was a Spatial Durbin with three independent variables, two cross-regressive variables and a lagged dependent variable. Results of the Anselin-Kelejian Test do not indicate the presence of un-accounted for spatial processes at a statistically significant level. Table 3 provides the results of the non-spatial and Spatial Durbin models. The non-spatial model was calculated to illustrate the importance of the underlying spatial processes. The non-spatial model had an adjusted r-squared value of 0.414. The inclusion of the lagged variables increased the model fit as evidenced by the pseudo and spatial pseudo R-squared values of 0.471 and 0.461, respectively. What is important to note is the change in the estimated parameters in the Spatial Durbin model compared to the non-spatial base case.

Table 3. Commuters Using Public Transit Modeling Specification Results

Response Variable: Change in Ridership from 2011-2016 (in Millions)		
Independent Variables	**Non-spatial Coefficient**	**Spatial Durbin Coefficient**
Constant	-0.4111***	-0.3154***
Number of train lines	0.3191***	0.2969***
Total units (in 1,000s)	0.1403 ***	0.0988**
Mean household income (in $10,000s)	-0.0200 ***	-0.0390***
Cross-Regressive Variables (1.0 Mile Weights Matrix)		
Gross commercial area (in 1,000,000ft^2)		0.1000**
Median family income (in $10,000s)		0.0370***
Spatial Lag Variable (1.5 Mile Weights Matrix)		
Change in Ridership from 2011-2016 (in Millions)		-0.7384*
Model Statistics		
Number of Observations	185	185
R-squared	0.423	
Adjusted R-squared	0.414	
Pseudo R-squared		0.471
Spatial Pseudo R-squared		0.461

***99% Level of Confidence, **95% Level of Confidence, *90% Level of Confidence

Results of the Spatial Durbin model indicate that subway stations that serve more train lines experienced a greater increase in ridership. This result is especially interesting because variables that quantified the total ridership in 2011 and average ridership per line in 2016 were both statistically insignificant. This means that it is not necessary the volume of initial passengers that is best indication of future demand change, rather it is the extent of services, in this case the number of available train lines (routes). The results also indicate that stations located in areas comprised of census tracts with a greater number of tax units (residential, commercial, etc.) and lower mean household incomes experienced a greater increase in ridership. The result of the income is consistent with previous research and captures the greater reliance on subway for lower income households. Two statistically significant cross-regressive variables were identified. Subway stations located in areas surrounded by census tracts with more commercial property or higher median family income are expected to have a greater increase in ridership. Lastly the estimated parameter for the spatial lag variable was negative and statistically significant. This indicates that the ridership at a given station decrease in response to an increase in ridership at neighboring stations. This may indicate that a portion of the change in ridership at a station is due to riders in a region changing which station they use instead of riders shifting from alternative modes of transportation.

6 SUMMARY AND CONCLUSIONS

In an effort to help metropolitan planning agencies better understand current and future demand of their public transportation networks this research investigates the underlying social, economic, and land use factors that impact transit ridership regardless of transit mode, then uses this insight to estimate specific models to help forecast changes in subway ridership. A spatial dataset that covered all of the nearly 2,200 census tracts in the city was used as the basis of the research. The dataset consisted of US Census social and economic data paired with New York City Department of City Planning land use data and subway ridership data. The results showed that the underlying causes for variability in public transportation ridership was not constant over space. Moreover, the spatial processes were determined to be statistically significant in both the commuter transit and subway ridership model. To account for these spatial processes (spatial autocorrelation, heterogeneity, and dependence) the models were estimated using the Spatial Durbin specification.

Census tracts with a higher average commute time, greater employed population, higher per capita income and lower median household income were determined to have a greater percentage of commuters using public transportation. Local spatial effects also played a significant role in public transportation use for work trips. Census tracts were found to have a greater percentage of commuters using public transportation if neighboring census tracts had more commercial space or fewer buildings. The lagged dependent variable accounted for global spatial effects. The significance of this variable paired with the insignificance of the spatial error term shows that social norms and expectations can influence people’s mode choice.

Examination of subway ridership between 2011 and 2016 shows that subway stations serving more train lines or are in areas comprised of census tracts with a greater number of tax units (residential, commercial, etc.) or lower mean household incomes experienced a greater increase in ridership over the study period. Furthermore, subway stations located in areas surrounded by census tracts with more commercial property or higher median family income are also expected to have a greater increase in ridership. Lastly, ridership at a given station decreases due to an increase in ridership at neighboring stations. This may indicate that a change in ridership at a station is due, in part, to riders in a region changing which station they use instead of riders shifting from alternative modes of transportation. The completed research can help public agencies better address resource allocation by identifying locations that are over or underperforming in terms of expected ridership or identifying locations for network expansion.

7 ACKNOWLEDGEMENTS

The author would like to thank the Region 2 University Transportation Research Center for supporting this research endeavor. The author would also like to thank Manhattan College undergraduate research assistants Nicola Grillo and Conor Varga for their hard work and contributions. The contents of this paper reflect the views of the author, who is responsible for the facts and the accuracy of the data presented herein, and do not necessarily reflect the official views or policies of any of the agencies included in the paper, nor do the contents constitute a standard, specification, or regulation.

8 REFERENCES

Aguero-Valverde, J., & Jovanis, P. (2008). Analysis of Road Crash Frequency with Spatial Models. *Transportation Research Record*: Journal of the Transportation Research Board, 55–63.

Anselin, L. (1988a). Spatial Econometrics: Methods and Models (Kluwer, Dordrecht).

Anselin, L. (1988b). Model Validation in Spatial Econometrics: A Review and Evaluation of Alternative approaches, *International Regional Science Review 11*, 279-316.

Anselin, L. (1988c). Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity, *Geographical Analysis,* No. 20, 1-17.

Anselin, L. (1995). Local Indicators of Spatial Association—LISA, *Geographical Analysis*, No. 27, 93-115

Anselin, L. (2006). Palgrave Handbook of Econometrics, Volume 1: Econometric Theory, Chapter 26: Spatial Econometrics, Edited by Terence C. Mills, Kerry Patterson

Anselin, L., & Kelejian, H. (1997). Testing for Spatial Error Autocorrelation in the Presence of Endogenous Regressors. *International Regional Science*, 435-444.

Anselin, L., and Rey, S. (2014). Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace

Anselin, L., Anil, Bera., Florax, R., & Yoon, M. (1996). Simple Diagnostic Tests for Spatial Dependence, *Regional Science and Urban Economics*, 26, 77-104

Anselin, L., Syabri I., & Kho, Y., (2006). GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis* 38 (1), 5-22

APTA. (2017). Transit Ridership Report Fourth Quarter 2016. American Public Transportation Association. http://www.apta.com/resources/statistics/Pages/ridershipreport.aspx. Accessed 07-01-217

Badoe, D., & Miller, E., (2000).Transportation Land Use Interaction: Empirical Findings in North America, and Their Implications for Modeling. *Transportation Research Part D: Transport and Environment*.

Boarnet, M and Greenwald, M. (2000). Land use, urban design, and nonwork travel: reproducing other urban areas' empirical test results in Portland, Oregon. *Transportation Research Record: Journal of the Transportation Research Board*, 1722, pp. 27-37

Breusch, T., & Pagan, A. (1979). Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 1287–1294

Cervero, R. (1993). Mixed Land-Use and Commuting: Evident from the American Housing Survey, *Transportation Research Part* A: *Policy and Practice*. Vol. 30, No. 5, pp 361-377.

Chakrabortya, A. and Mishrab, S. (2013). Land use and transit ridership connections: Implications for state-level planning agencies. *Land Use Policy*. Volume 30, Issue 1, pp. 458-469

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

Cook, J., Alon, D., Sanchirico, J., & Williams, J., (2012). Driving Intensity in California: Exploring Spatial Variation in VMT and its Relationship to the Built Environment. Western Energy Policy Research Conference, Boise, ID

Erdogan, S. (2009). Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. *Journal of Safety Research*, 341–351.

Frank, L., Stone, B., & Bachman, W., (2000). Linking Land Use with Household Vehicle Emissions in the Central Puget Sound: Methodological Framework and Findings. *Transportation Research Part D: Transport and Environment*, pp.173–196.

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and Serial Independence of Regression Residuals. *Economics Letters*.

Kain, J. and Liu, Z. (1999). Secrets of success: assessing the large increases in transit ridership achieved by Houston and San Diego transit providers. *Transportation Research Part A: Policy and Practice*, vol. 33 pp. 601-624.

Kirk, A. J., Pigman, J. G., & Agent, K. R. (2005). Socioeconomic Analysis of Fatal Crashes. Lexington: Kentucky Transportation Center.

Kitamura, R., Mokhtarian, M., and Laidet, L. (1997). A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. *Transportation*, vol. 24 (2), pp. 125-158.

Kyte, M., Stoner, J., and Cryer, J. (1988). A time-series analysis of public transit ridership in Portland, Oregon, 1971–1982 *Transportation Research Part Al*, 22 (5), pp. 345-359

LaScala, E. A., Gerber, D., & Gruenewald, P. J. (2000). Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. *Accident Analysis & Prevention*, 32(5), 651-658.

Li, L., Zhu, L., & Sui, D. (2007). A GIS Based Bayesion approach for analyzing spatial-temporal patterns of intra-city vehicle crashes. *Journal of Transport Geography*, 274-285.

MTA. (2015). NYC Transit Subway Entrance and Exit Data. Metropolitan Transportation Authority, https://data.ny.gov/Transportation/NYC-Transit-Subway-Entrance-And-Exit-Data/i9wp-a4ja/data, Accessed: 10/01/2017

MTA. (2017). Introduction to Subway Ridership. Metropolitan Transportation Authority, web.mta.info/nyct/facts/ridership/, Accessed: 10/01/2017

PLUTO (2017). PLUTO and MapPLUTO. New York City Department of City Planning. http://www1.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page, Accessed: 07/01/2017

R Core Team. (2017). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

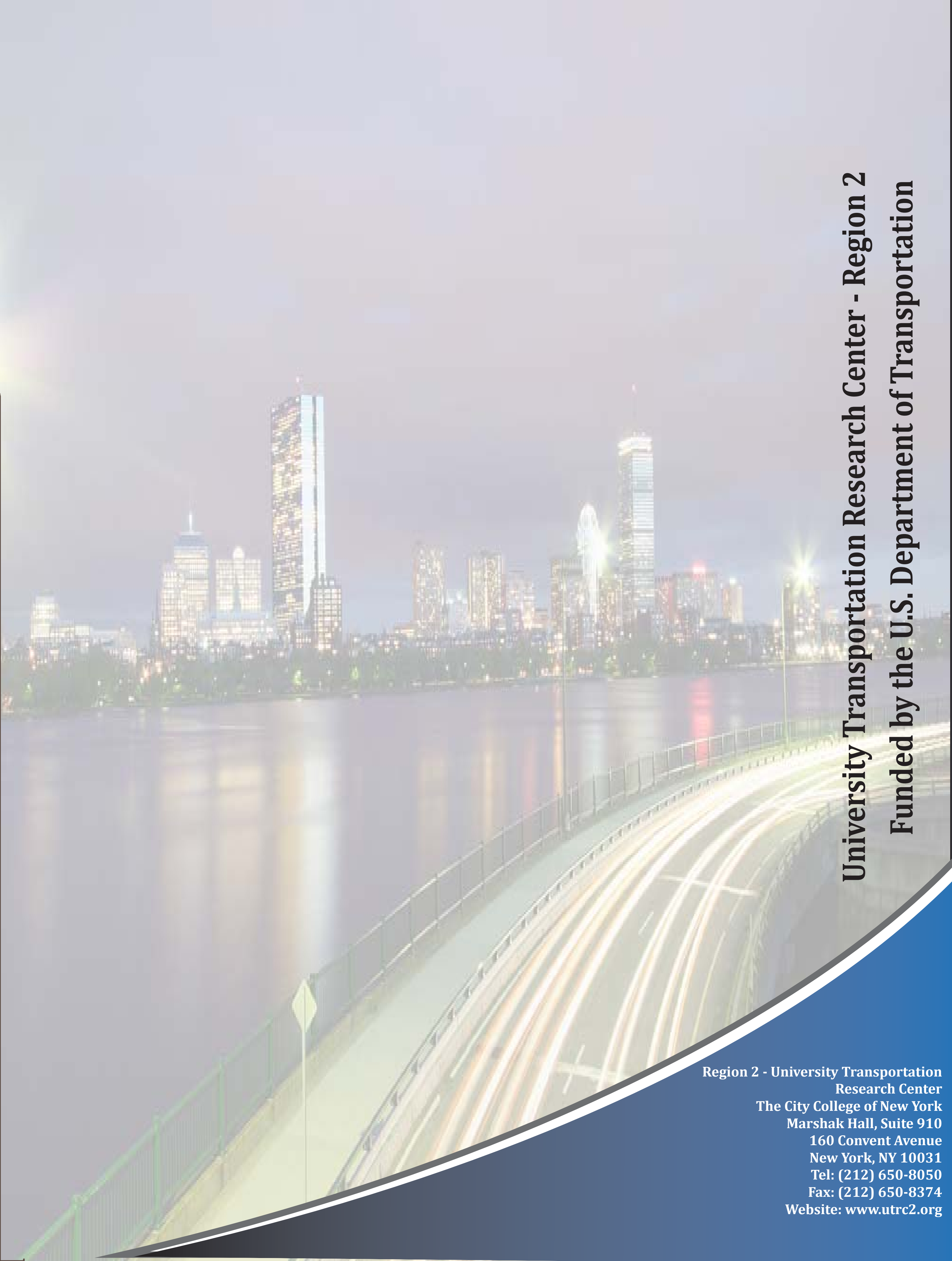
Schneider, R. J., Ryznar, R. M., & Khattak, A. J. (2000). An accident waiting to happen: a spatial approach to. *Accident Analysis and Prevention*, 193–211.

Stamatiadis, N. & Puccini, G. (1999). Fatal Crash Rates in Southeastern United States: Why are They Higher? *Transportation Research Record*: Journal of the Transportation Research Board, 118-124.

U.S. Census. (2015). American Fact Finder. United States Census Bureau, U.S. Department of Commerce, Washington D.C., www. factfinder.census.gov, Accessed: 07/1/2016

Volovski, M. (2015). Spatial analysis of passenger vehicle use and ownership and its impact on the sustainability of highway infrastructure funding. Dissertation, Purdue University, West Lafayette, Indiana 47906

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48 (4): 817–838.



University Transportation Research Center - Region 2

Funded by the U.S. Department of Transportation

Region 2 - University Transportation Research Center
The City College of New York
Marshak Hall, Suite 910
160 Convent Avenue
New York, NY 10031
Tel: (212) 650-8050
Fax: (212) 650-8374
Website: www.utrc2.org