2018

Econometric Modeling Analysis of Public Transit Ridership: Application for Orlando Region

Moshiur Rahman
University of Central Florida

ECONOMETRIC MODELING ANALYSIS OF PUBLIC TRANSIT RIDERSHIP: APPLICATION FOR ORLANDO REGION

by

MOSHIUR RAHMAN
B.Sc. Bangladesh University of Engineering and Technology, 2012
M.Sc. University of Central Florida, 2018

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Civil, Environmental and Construction Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Summer Term
2018

Major Professor: Naveen Eluru

ABSTRACT

Policy makers are considering several alternatives to counter the negative externalities of personal vehicle dependence. Towards this end, public transit investments are critical in growing urban regions such as Orlando, Florida. Transit system managers and planners mostly rely on statistical models to identify the factors that affect ridership as well as quantifying the magnitude of the impact on the society. These models provide vital feedback to agencies on the benefits of public transit investments which in turn act as lessons to improve the investment process. We contribute to public transit literature by addressing several methodological challenges for transit ridership modeling. Frist, we examine the impact of new transit investments (such as an addition of commuter rail to an urban region) on existing transit infrastructure (such as the traditional bus service already present in the urban region). The process of evaluating the impact of new investments on existing public transit requires a comprehensive analysis of the before and after measures of public transit usage in the region. Second, we accommodate for the presence of common unobserved factors associated with spatial factors by developing a spatial panel model using stop level public transit boarding and alighting data. Third, we contribute to literature on transit ridership by considering daily boarding and alighting data from a recently launched commuter rail system (SunRail). The model system developed will allow us to predict ridership for existing stations in the future as well as potential ridership for future expansion sites. Fourth, we accommodate for potential endogeneity between bus headway and ridership by proposing a simultaneous model system of headway and ridership. Finally, a cost benefit analysis exercise is conducted for examining the impact of Sunrail on the region.

This thesis work is dedicated to my mother, Anwara Begum, who has been inspiring me since my childhood for higher educations, my father Md. Zillullah, my wife Evana Ahmed, a constant source of support and encouragement during the challenges of graduate school and life. I am truly thankful for having you in my life. This work is also dedicated to my brother, sisters and all member of my family, who have always loved me unconditionally and whose good examples have taught me to work hard for the things that I aspire to achieve.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to God for His guidance and protection throughout my doctoral study.

There are a number of people without whom this thesis might not have been written, and to whom I am greatly indebted.

I would like to thank my advisor, Professor Naveen Eluru (Department of Civil, Environmental, and Construction Engineering) at the University of Central Florida (UCF), for his advice, and support on this dissertation works. His financial assistance, words of encouragements and knowledge positively impact this Ph.D. research work. I would like to thank Dr. Shamsunnahar Yasmin, Post-Doctoral Research Fellow (Department of Civil, Environmental, and Construction Engineering) at the University of Central Florida (UCF) for her continuous support throughout my dissertation works. Thanks go to Dr. Mohamed Abdel-Aty and Dr. Nizam Uddin for serving my dissertation committee. I recognize and acknowledge those two Professors for their encouragements and helpful reviews of my Dissertation.

I would like to thank all of the member from our "Transportation Econometric Modeling Group (TEMG)" research group for their support. Dr. Sabreena Anowar, for her encouragements and supports; Salah, Nowreen, Tanmoy, Bibhas, and Naveen for helping me during the course works and research works throughout my Ph.D. life. Besides that, I would thank all the members and staff of my department and all the members of UCF Bangladesh Student Community.

I especially thank my family for their love, support, and encouragement throughout the journey of my dissertation. To my mother, Anwara Begum, who gave me my first step towards learning, growing, and development and who has been a source of encouragement and inspiration to me throughout my life. To my dear wife, Evana, who remains willing to engage in the struggle, sacrifice her carrier and ensuing discomfort, of having a partner like you is a blessing from GOD. A very special thank you for your practical and emotional support throughout my Ph.D. journey. I am also very grateful to all of the members of my family, especially my elder brother, Mizanur Rahman, who guided me and encouraged me after my father died, who support our family, my in-law's family for their support. I would like to take this opportunity to say warm thanks to all my beloved friends, who have been so supportive along the way of doing my thesis. Without my family's unending trust, confidence and support, I would not be where I am today.

Last but not least, deepest thanks go to all people who took part in making this thesis real.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

1.1 Background

The economic development and the associated growth in household incomes in the United States during the post-Second World War resulted in an increased household and vehicle ownership, population and employment decentralization and urban sprawl. While population has increased nearly 72% between 1950 and 1990, the aggregate population in central cities declined by 17% (Baum-Snow, 2007). Population and employment changes resulted in a drastic reduction in public transit ridership. In terms of commute to central cities, only 38% of commute trips in 2000 were destined to central cities; a 66% reduction from 1960 (Baum-Snow, 2010). In fact, in fifty years since 1940, transit ridership in the US reduced by 31% - a drop of about 4 billion trips (Baum-Snow and Kahn, 2000). The ridership reduction occurred while a near doubling of the population happened in the same time frame (O'Sullivan, 1996). Not surprisingly, the rapid decline in public transit ridership is associated with nearly 44% growth in personal vehicle miles traveled.

The consequences of the drastic transformation of the transportation system include negative externalities such as traffic congestion and crashes, air pollution associated environmental and health concerns, and dependence on foreign fuel (Schrank, et. al, 2012). For instance, in 2014, traffic congestion has resulted in a loss of about 6.9 billion hours and 3.1 billion gallons of fuel amounting to a cumulative cost of nearly 160 billion dollars (Schrank et al., 2015). Furthermore, the increased private vehicular travel contributes to increasing air pollution and greenhouse gas (GHG) emissions - a matter receiving substantial attention given the significant impact on health and safety of future generations (Woodcock et al., 2009). In an endeavor to counter the negative externalities of personal vehicle dependence, policy makers have often found the development of an efficient multi-modal public transportation system to be the most suitable solution. Many urban regions, across different parts of North America, are considering investments in public transportation alternatives such as bus, light rail, express bus service, metro and bicycle sharing systems (see TP, 2016 for public transportation projects under construction or consideration). While non-motorized modes of transportation are beneficial in the urban core, public transit with its reach to serve populations residing throughout the urban region can enhance mobility for a large share of urban residents.

1.2 Motivation

In recent years, transportation professionals and policymakers have recognized the potential of public transit in enhancing mobility for urban residents as well as reversing (or at least reducing) the negative externalities of car dependence. Several major investments in public transit projects are under consideration in cities including New York, San Francisco, Los Angeles, Detroit, Charlotte and Orlando (Barber, 2017). These investments include bus and subway system expansions, streetcar additions, light rail and commuter rail system addition (and expansion). The public transit investments are particularly critical in growing urban regions such as Orlando, Florida. In recent years, Greater Orlando region has experienced rapid growth. In fact, according to the US Census Bureau, among the country’s thirty large urban regions, Orlando is the fastest growing one (Brinkmann, 2016). It is reported that the majority (about 74%) of the population growth in this region is driven by domestic and international migration. The rapid growth in population increases the stress on the existing transportation system. Thus, it is not surprising that several transportations and public transit investments are underway in the region to alleviate traffic congestion and improve mobility for Greater Orlando residents.

Recent construction for I-4 highway expansion causes excessive traffic congestion near downtown Orlando thus increasing the travel time and safety risk factors. SunRail system provides viable transit options for Central Florida residents who live along the I-4 corridor. The service is expected to alleviate congestion along I-4 corridor that is currently under multi-year construction associated with its expansion. Further, the system has the potential for improving overall livability, property values, and reducing overall carbon footprint. An important tool to evaluate the influence of these public transit investments on transit ridership is the application of statistical models. Transit system managers and planners mostly rely on statistical models to identify the factors that affect ridership as well as quantifying the magnitude of the impact on the society (see Chakour and Eluru, 2016 and Pulugurtha and Agurla, 2012 for example). These models provide vital feedback to agencies on the benefits of public transit investments which in turn act as lessons to improve the investment process.

While earlier research has explored the benefits of public transit ridership, the approach to quantifying the benefits from public transit investments is a field in its infancy. This is particularly so in the context of disaggregate level public transit analysis (such as ridership at a stop or route level). The growing emphasis of sustainability and livability improvements from transportation systems require us to undertake a rigorous analysis to quantify benefits form public transit investments. The greater Orlando region, serves as an ideal test bed to contribute research approaches to evaluate the impact of transit investments on public transit system usage.

1.3 Objectives of the Research

The specific objectives for the dissertation are described here:

Objective 1. *Evaluating the Impact of a Newly Added Commuter Rail System on Bus Ridership: A Grouped Ordered Logit Model Approach.*

The dissertation examines the impact of new transit investments (such as an addition of commuter rail to an urban region) on existing transit infrastructure (such as the traditional bus service already present in the urban region). The process of evaluating the impact of new investments on existing public transit requires a comprehensive analysis of the before and after measure of public transit usage in the region. The main emphasis of the research is to develop a comprehensive and statistically valid framework to study the impact of new public transportation infrastructure (such as commuter rail) on existing public transit infrastructure (such as bus). Specifically, the current research effort contributes to transit literature by evaluating the influence of a recently inaugurated commuter rail system on traditional bus service. We examine the before and after impact of "SunRail" commuter rail system in the Orlando metropolitan region on the "Lynx" bus system. Given the relatively long-time span required for the influence of large scale public transportation system changes, any analysis of the value of new investments should consider adequate data before the system installation and after the system installation. The current research effort is focused on addressing two important data techniques. First, by employing data on stop level ridership (weekday boarding and alighting) for three 4-month time periods before and after commuter rail installation in a large metropolitan area, the current research effort makes a unique empirical contribution identifying the commuter rail impact while controlling for all other factors affecting ridership. Second, the study contributes methodologically, by developing a panel joint grouped response ordered modeling framework. The proposed model accommodates for common unobserved factors affecting boarding and alighting as well as repeated measures for each stop. Furthermore, the grouped response structure allows for flexible specification of the dependent variable while also not being restricted by additional threshold parameters to be estimated (see Chakour and Eluru, 2016). Additionally, the influence of SunRail on ridership has a positive temporal trend indicating the strengthening of the impact with the time of operation, a healthy metric for potential future expansion.

Objective 2. *Incorporating the Impact of Spatio-Temporal Interactions on Bus Ridership.*

The dissertation accommodates for the presence of common unobserved factors associated with spatial factors by developing a spatial panel model by using stop level public transit boarding and alighting data, Specifically, two spatial models: 1) Spatial Error Model (SEM) and 2) Spatial Lag Model (SAR) are estimated for boarding and alighting separately by employing several exogenous variables including stop level attributes, transportation and transit infrastructure variables, built environment and land use attributes, sociodemographic and socioeconomic variables in the vicinity of the stop and spatial and spatio-temporal lagged variables. The repeated observation data at a stop-level offers multiple dimensions of unobserved factors including stop-level, spatial and temporal factors. In our analysis, we apply a framework proposed by Elhorst (Elhorst, J.P., 2014) to accommodate for the aforementioned observed and unobserved factors. The results from the spatial error and lag models are compared with the results from traditional linear regression models to identify the improvement in model fit with accommodation of spatial unobserved effects and panel repeated measures. In the earlier literature on bus transit ridership has not accommodated for observed and unobserved spatial effects on ridership. Toward addressing these limitations, we formulate and estimate a spatial panel model structure that accommodates for repeated ridership data for the same stop as well as the impact of spatial and temporal observed and unobserved factors.

Objective 3. *Examining Determinants of Commuter Rail ridership: A Case Study of the Orlando SunRail System.*

The main objective is to identify the factors that affect the SunRail ridership in Orlando region. The current study contributes to literature on transit ridership by considering daily boarding and alighting data from a recently launched commuter rail system. With the rich panel of repeated observations for every station, the potential impact of observed and unobserved factors affecting ridership variables are considered. Specifically, an estimation framework that accounts for these unobserved effects at multiple levels – station, station-week and station day are proposed and estimated. In addition, the study examines the impact of various observed exogenous factors such as station level attributes, transportation infrastructure variables, transit infrastructure variables, land use and built environment attributes, sociodemographic and weather variables on ridership. Separate models are developed for boarding and alighting. The model system developed will allow us to predict ridership for existing stations in the future as well as potential ridership for future expansion sites.

Objective 4. *Controlling for endogeneity between bus headway and bus ridership: A Case Study of the Orlando region.*

In transit ridership analysis, headway is considered an important determinant of ridership. However, the choice of headway at a bus stop is not made in isolation. Rather it is in response to expected demand. Thus, as headway reduces between buses it is likely to result in increased ridership. In traditional ridership studies, this is often neglected and headway is considered as a pure exogenous variable. The assumption violates the requirement that the dependent variable does not affect the independent variable. In this dissertation, we address this limitation by developing a headway prediction model and using its residual as an exogenous variable in the ridership model.

Objective 5. *Benefit cost analysis of Sunrail.*

Given the limited financial resources for urban transportation planning organizations it is important to quantitatively analyze the impacts of transportation investments in an effort to maximize the resource allocation efficiency across different transport needs. Cost-benefit analysis (CBA) is considered to be one of the most appropriate tools in evaluating transportation policies and projects (Litman, 2001). A comprehensive CBA would allow analysts to predict several direct and/or indirect impacts of improvements in existing system or proposed new infrastructures. In terms of investments for transport infrastructure; spending money for transit infrastructures are often a low priority compared with investments on roads, improvements to traffic flow and other government expenditure. However, more recently investments in transit infrastructures have gained traction from transport authorities as a measure of reducing negative externalities of increasing private auto mode usage. A comprehensive CBA of public transit mode investments would assist the planners and policy makers to evaluate the "real" benefit of these investments and provide evidence to justify allocation of more funding for improving/building public transit infrastructures. The current research report focuses on CBA for Sunrail in Orlando region.

1.4 Dissertation Structure

This dissertation is divided by several chapters. A details overview of each chapter is given below.

In Chapter 2, a detailed literature review is conducted on public transit ridership research efforts. Traditional travel demand modeling research has focused on automobile travel. In recent years, an increased number of studies are undertaking detailed analysis of transit systems and associated ridership. These studies examine transit ridership to identify the impact of socioeconomic characteristics, built environment, and transit attributes on ridership across different contexts. In this chapter, we focus on different dimensions of transit mode such as bus transit (including bus rapid transit), light rail, subway and commuter rail. Besides the literature review on transit ridership, we will discuss some previous study on the cost benefit analysis

Chapter 3 describes the data source and data preparation for analysis. The ridership data was obtained from Lynx transit authority and SunRail authority. The exogenous variable information was generated based on multiple data sources including 2010 US census data, American Community Survey (ACS), Florida Geographic Data Library (FDGL), and Florida Department of Transportation (FDOT) databases. Details on data source and data preparation process is described in chapter 3.

Chapter 4 examines the impact of new public transportation infrastructure (SunRail) on existing public transit infrastructure (Lynx) in the Orlando metropolitan region. This research formulates and estimates an innovative grouped ordered response model structure for the ridership analysis. The proposed model accommodates for common unobserved factors affecting boarding and alighting as well as repeated measures for each stop. To measure the impact of commuter rail on stop level bus ridership (defined as boarding and alighting), the model system controls for a host of exogenous variables including stop level attributes, transportation infrastructure variables, transit infrastructure variables, land use, built environment attributes, sociodemographic and socioeconomic variables. The results while highlighting the impact of the exogenous variables provide strong evidence of the positive impact of SunRail system on the ridership. Furthermore, the influence of SunRail on ridership has a positive temporal trend indicating the strengthening of the impact with the time of operation.

Chapter 5 presents details on the development of a spatial panel model that accommodates for impact of spatial and temporal observed and unobserved factors on bus ridership. Two spatial models: Spatial Error Model (SEM) and Spatial Lag Model (SAR) are estimated for boarding and alighting separately by employing several exogenous variables including stop level attributes, transportation and transit infrastructure variables, built environment and land use attributes, sociodemographic and socioeconomic variables in the vicinity of the stop and spatial and spatio-temporal lagged variables. These models are expected to provide feedback to agencies on the benefits of public transit investments while also providing lessons to improve the investment process.

Chapter 6 describes the study that contributes to literature on transit ridership by considering daily boarding and alighting data from a recently launched commuter rail system – SunRail in Orlando region. The analysis is conducted based on daily boarding and alighting data for ten months for the year 2015. With the rich panel of repeated observations for every station, the potential impact of common unobserved factors affecting ridership variables are considered. The research develops an estimation framework that accounts for these unobserved effects at multiple levels – station, station-week and station day. In addition, the study examines the impact of various observed exogenous factors such as station level attributes, transportation infrastructure variables, transit infrastructure variables, land use and built environment attributes, sociodemographic and weather variables on ridership. Separate models are developed for boarding and alighting. The model system developed will allow us to predict ridership for existing stations in the future as well as potential ridership for future expansion sites. Finally, a policy analysis is performed to demonstrate the implications of the developed models.

Chapter 7 discusses the impact of bus frequency on bus ridership. Earlier research in public transportation has identified headway as one of the primary determinants affecting ridership. The stops with higher headway between buses are likely to have lower ridership. While this is a perfectly acceptable conclusion, most (if not all) studies in public transit literature ignore that the stop level headway was determined (by choice) in response to expected ridership i.e. stops with lower headway were expected to have higher ridership numbers. This potential endogeneity is often neglected and headway is considered as an independent variable. The approach violates the requirement that the unobserved factors that affect the dependent variable do not affect the independent variable. In this study, we address this limitation by proposing to model headway itself as a choice dimension and then using the residuals from headway model as an independent variable in modeling ridership.

Chapter 8 discusses the cost benefit analysis of SunRail transit system in Orlando region. Transit systems are an integral part of the development of a community. But comprehensive benefits of these systems often are not estimated or remain unmeasured. Though the capital cost of developing a transit system is significantly higher, total benefits accrued from a transit system operation in the long run is likely to surpass the higher investment cost. With the focus of encouraging more people to use sustainable transportation alternatives, FDOT is constructing a new, 17.2-mile extension to the existing 31-mile SunRail commuter rail. A comprehensive CBA of the existing operational SunRail system would assist planners and policy makers to evaluate the "real" benefit of these investments and provide evidence to justify allocation of more funding for improving/building transit infrastructures.

Finally, chapter 9 discusses the summary of the study and benefits from my study to society. The chapter also identifies future directions of research and concludes the dissertation.

CHAPTER TWO: LITERATURE REVIEW

2.1 Literature Review on Bus Ridership

Traditional travel demand modeling research has focused on automobile travel. Only recently studies have begun to undertake detailed analysis of transit systems and associated ridership. Examining the performance and/or the impact of public transportation systems is a burgeoning area of research. Of particular relevance to our research is earlier work examining transit ridership. While there have been few studies that explore transit ridership from a nation or regional perspective (see for example Taylor et al., 2009), a large number of studies examine transit ridership focusing on a specific urban region. These studies examine transit ridership to identify the impact of socioeconomic characteristics, built environment, and transit attributes on ridership across different contexts (Chakour & Eluru, 2016). These studies broadly examine macro-level ridership (Chakraborty & Mishra, 2013 and Taylor et. al., 2009), study impact of financial attributes such as fares, fuel price and parking cost (Chen et. al., 2011, Currie & Phung, 2007, Hickey, R., 2005, Lane, B.W., 2010, Lane, B.W., 2012 and Mattson, J. W, 2008), and effect of transit attributes and built environment on transit ridership. The research on ridership can be broadly classified based on the public transit mode under consideration along two streams: (1) rail and metro ridership and (2) bus ridership. As the focus of our current work is bus transit ridership, we limit our review to bus ridership studies. For bus ridership studies, at the bus-stop level, the most common dependent variables of interest include daily level or time-period specific boarding and alighting variables or a sum of boarding and alighting variables. A brief review of most relevant literature follows.

The *first stream of studies* on rail and metro ridership examined the influence of station characteristics, transit service attributes, and urban sociodemographic patterns and built environment. A number of studies that examined station choice dimension observed that station attributes including parking space availability and bicycle standing areas, amenities and train frequency, vehicle ownership patterns affect station choice (see Debrezion et al., 2007, 2009; Fan et al., 1993; Wardman & Whelan, 1999; Chakour and Eluru, 2014). In a study evaluating rail ridership in Atlanta, Brown and Thompson (2008) observed that employment decentralization was responsible for drop in ridership. Transit Oriented Development (TOD) that comprises of dense commercial developments is expected to affect ridership positively (Shoup, 2008; Sung and Oh, 2011). Population and job density variables are likely to positively influence ridership (Guerra and Cervero, 2011). Studies exploring ridership at metro stations found that retail, service and government land use, accessibility by bus, presence of transfer terminals, walkability in the vicinity of stations are positively correlated with ridership (Chan & Miranda-Moreno, 2013; Gutiérrez, 2001; Gutiérrez et al., 2011; Lin & Shin, 2008).

The *second stream of studies*, closely related to the effort of current study, examine the impact of built environment and urban form at the stop level on bus ridership. The transit ridership variables considered include daily ridership computed as sum of boardings and alightings at a stop level (Ryan and Frank, 2009), daily boardings (Johnson, 2003; Chu, 2004; Banarjee et al., 2005; Estupiñán and Rodríguez, 2008; Pulugurtha and Agurla, 2012), time period specific boarding's and alighting's (Chakour and Eluru, 2016). The methodologies employed for the analysis range from simple linear or log-linear regression models, geographically weighted negative binomial count models, composite likelihood based ordered regression models. Major exogenous variables identified to affect transit ridership include land use and urban form and sociodemographic characteristics in the vicinity of the stop, walkability measures, real-time bus schedules transportation system attributes, transit system operational attributes and unobserved factors that simultaneously affect boardings and alightings (Johnson, 2003; Chu, 2004; Banarjee et al., 2005; Estupiñán and Rodríguez, 2008; Pulugurtha and Agurla, 2012; Dill et al, 2013; Tang and Thakuriah, 2012; Chakour and Eluru, 2016). Tang and Thakuriah (Tang and Thakuriah, 2012) highlight the value of real-time bus information is slightly increasing the bus ridership in Chicago.

2.2.1 Literature Review on endogeneity on bus ridership

Transit ridership has been widely explored in transportation literature. Broadly, the earlier literature can be categorized into two groups. The first group of studies focus on the factors that affect transit adoption at a disaggregate level by exploring individual perceptions and behavioral responses (see Acker, et al, 2010; Handy, S. 1996; Handy, et al, 2005; Balcombe, 2004; Eavns 2004; McCollom and Pratt, 2004; Pratt and Evans, 2004, Debrezion et al., 2007, 2009; Fan et al., 1993; Wardman & Whelan, 1999; Chakour and Eluru, 2014). The second group of studies examine the impact of various factors on system level (or route level) ridership measures (Seskin and Cervero, 1996; Johnson, 2003; Babalik-Sutcliffe, 2002; Mackett and Babalik-Sutchliffe, 2003; FitzRoy and Smith, 1998; Kain and Liu, 1999; Ma et al., 2018). The proposed research effort falls into the second group of studies. A detailed review of all these studies is beyond the scope of the paper. The reader is referred to a recent study Rahman et al., 2017 that provides a detailed summary of literature across these two groups. In this section, we focus on literature particularly relevant to our research effort. We begin with an overview of studies in transportation that attempt to accommodate for endogeneity. Subsequently, we examine studies that consider endogeneity within transit literature.

Addressing endogeneity in transportation

The travel behavior field has extensively examined the influence of endogeneity across various decision processes. Specifically, these studies have explored the potential impact of residential location choice – labelled as residential self-selection - on various travel behavior choices (see Bhat and Guo, 2007 Mokhtarian and Cao, 2008; Pinajri et al., 2009; Bhat and Eluru, 2009; Cao, et al, 2010; Walker et al., 2011; Aditjandra, T., 2012; Vij and Walker, 2014; Ding, et. Al, 2017; Ettema & Nieuwenhuis, 2017). There are examples from other fields including seat belt choice in driver injury severity models (Eluru and Bhat, 2007; Abay et al., 2013); emergency medical response time affecting fatality timeline (see Yasmin et al., 2015) and bicycle sharing system station capacity decision influencing bicycle sharing demand (Faghih-Imani and Eluru, 2016). The most commonly employed modeling approaches in these studies include developing a choice model for the endogenous variable to reduce/eliminate the bias associated with the endogenous variable. The endogenous variables and the choice variables could be examined as continuous or discrete indicators. Based on the nature of the variables involved, several approaches such as instrument variables regression, two-stage residual inclusion approach and Roy's (1951) endogenous system or the treatment effects model (see Maddala, 1983; Chapter 9; Heckman and Vytlacil, 2005) and joint econometric modeling approaches (see Eluru and Bhat, 2007) are employed.

Research in transit field accommodating endogeneity

Given the prevalence of modeling approaches for addressing endogeneity bias in transportation field, it is not surprising that multiple studies have either alluded to the presence of endogeneity or specifically employed approaches to control for it in the context of public transit analysis. Earlier research in transit ridership analysis have discussed potential endogeneity of transit ridership and transit price, service and automobile ownership dimensions (Crutzig, 2014). Holmgren, (2007) conducted a meta-analysis of elasticity estimates of bus demand in transit literature and recommended that service variable (headway) should be treated as endogenous while other variables such as car ownership, fuel price and ticket price be considered as exogenous variables. The studies that considered endogeneity have controlled for different dimensions governed by the author's judgement. Voith (1991) develop community transit demand models while accommodating for the interaction between transit fare prices and service decisions on ridership. The authors estimate a dynamic fixed effects panel model with Instrumental Variables (IV) using data from Southeastern Pennsylvania Transportation Authority (SEPTA). Voith (1997) extends the model developed in Voith (1991) with a larger data sample with IV approach developing separate equations for price and service.

Fitzroy and Smith (1999) developed a framework to examine the impact of season tickets on transit ridership across four Swiss cities. To account for the potential impact of investments on road and transit infrastructure on overall ridership the authors employed an IV approach. Further, the authors control for potential contemporaneous unobserved correlation by developing seemingly unrelated regression approach. Deka, 2002 examined the potential endogeneity of automobile ownership and transit availability in the Los Angeles region. Specifically, the author estimated a model for transit availability and employed its predicted value as an independent variable in modeling automobile ownership. Novak and Savage, (2013) studied the cross-elasticity between fuel price and transit usage for the Chicago region for various rail and bus services. The authors indicate that adopting a two stage least squares approach leads to counter-intuitive results in their data analysis. The reader would note that a majority of these studies develop models at a system level i.e. employ aggregate measures of ridership. Table 1 shows the studies done by the researcher where endogeneity was considered.

Table 1. Summary of Literatures on Bus Ridership Analysis for endogenous variables

Paper	Study Region/Data Source	Methodological Approach	Dependent Variables	Endogenous Variables
Endogeneity in Transportation				
Bhat & Guo, 2007	Alameda County in the San Francisco Bay Area (2000)	Unified mixed multinomial logit-ordered response structure	Travel behavior	Residential choice and car ownership decisions
Mokhtarian and Cao, 2008	Review Paper		Travel behavior	Residential self-selection
Pinajri et al., 2009	Alameda County in the San Francisco Bay Area (2000)	Joint mixed Multinomial Logit–Multiple Discrete-Continuous Extreme Value (MNL–MDCEV) structure	Residential choice and individual activity time-use behavior	Residential self-selection
Bhat and Eluru, 2009	Alameda County in the San Francisco Bay Area (2000)	Copula Methods	Travel choice	Residential self-selection
Cao, et al, 2010	Raleigh, NC (2006)	Propensity score matching (PSM) technique	Vehicle miles driven (VMD) per day	Residential self-selection
Walker et al., 2011	Amsterdam and Amstelveen, Netherlands (1992-1997)	Berry, Levinsohn, and Pakes (BLP) method and linear regression model	Mode choice	Residential choice
Aditjandra, T., 2012	The metropolitan area of Tyne and Wear, North East of England, UK	Structural Equations Modelling (SEM) approach	Travel choice	Residential self-selection
Vij and Walker, 2014	Nine county San Francisco Bay Area of California (2000)	Latent Class Choice Models (LCCMs)	Travel/Mode choice	Several factors (travel time, Residential location etc)
Ding, et. Al, 2017	Baltimore metropolitan area (NHTS data)	Structural Equation Model (SEM) and Discrete Choice Model (DCM)	Travel/Mode choice	Residential choice and car ownership decisions
Ettema & Nieuwenhuis, 2017	Hague, Netherlands (2014)	Statistical control approach	Travel/Mode choice	Residential self-selection
Eluru and Bhat, 2007	2003 General Estimates System (GES) data for 60 areas across the U.S	Ordered-Response models	Injury Severity	Seat belts use
Abay et al., 2013	Denmark (2002-2008)	Multivariate ordered-response probit model	Injury Severity of drivers	Seat belts use
Yasmin et al., 2015	Fatality Analysis Reporting System (FARS) (2010)	Mixed Generalized Ordered Logit (MGOL) model	Fatality Timeline	Emergency Medical Service (EMS) response time
Faghih-Imani and Eluru, 2016	New York city	Spatial Panel Model	Bicycle Sharing System Station Capacity Decision	Bicycle Sharing Demand

Paper	Study Region/Data Source	Methodological Approach	Dependent Variables	Endogenous Variables
Endogeneity in transit field				
Crutzig, 2014	---	Alonso– Mills–Muth model of a monocentric city	Public Transit fare	Fuel price and Urban form
Holmgren, 2007	---	Two Stage Least Squares (2SLS) /Regression model	Bus Demand	Headway
Voith, 1991 and Voith 1997	Southeastern Pennsylvania Transportation Authority (SEPTA)	Dynamic fixed effects panel model with Instrumental Variables (IV)	Transit demand model	Transit Fare Prices and Service Decisions
Fitzroy and Smith, 1999	Basel, Bern, Geneva & Zurich, Switzerland	Instrumental Variables (IV) approach	Transit ridership	Season Tickets
Deka, 2002	Los Angeles region	Logit Model/Regression Model	Transit Availability	Automobile Ownership
Novak and Savage, 2013	Chicago region	Two Stage Least Squares (2SLS) Approach	Transit ridership	Price of gasoline

2.2 Literature Review on Rail Ridership

In recent years, an increased number of studies are undertaking detailed analysis of transit systems and associated ridership. These studies examine how various exogenous variables influence system level ridership. Literature has focused on different dimensions of transit mode such as bus transit (including bus rapid transit), light rail, subway and commuter rail. A comprehensive review of literature along all these dimensions is beyond the scope of the paper (See Chakour & Eluru, 2016 for a review). In our review, we focus our attention only on the rail alternative. Table 2 provides a summary of the literature on rail ridership with information on study region, the level of analyses (macro or micro), modeling methodology, consideration for repeated observations, and attributes considered in ridership analysis. Based on the review of the literature, it is clear that rail ridership is typically analyzed along two streams – macro level and micro level.

The macro level studies examine ridership for multiple urban regions or at the national level. In this stream, ridership is modeled as a function of population and employment, gasoline prices and transit fares, and transit service facilities. The preferred modeling approach employed is the multivariate linear regression and its variants such as time series models, generalized least squares and auto-regressive models. The studies have spanned various countries including U.S., Canada, Greece, and Great Britain. It is interesting to note that across macro level studies a reasonable proportion of studies accounted for the presence of common unobserved factors in panel data (or data with repeated observations).

The second stream of research is conducted at the micro-level (or station level) with the objective of identifying the determinants of ridership. In these studies, the emphasis is on station level infrastructure, transportation infrastructure in the vicinity of the station, urban form and built environment and socio-demographics. Multiple linear regression approach has been widely used in micro level rail ridership estimation at the station level. Advanced approaches considered include fixed effects linear regression models, distance-decay weighted regression models, network kriging regression. Within micro studies, accommodating for presence of repeated observation is not as common as the application of these methods is in macro level studies. It is possible that data availability at multiple time points is not as readily available. In micro level ridership analysis, most of the studies find significant effect of gasoline prices, transit fares, accessibility and reliability and land use patterns surroundings the rail station. In table 2, summary of the literature review of rail ridership is given.

Table 2. Summary of Literatures on Rail Ridership Analysis

Paper	**Study Region**	**Methodological Approach**	**Level of Analysis**	**Panel data/ Time series**	**Accessibility**	**Transportation & stop level Infrastructures**	**Sociodemographic characteristics**	**Socioeconomics characteristics**	**Road network characteristics**	**Fuel Price Travel Cost**	**Built environment**
Baum-Snow and Kahn	Boston, Atlanta, Chicago, Portland, and Washington DC	Multivariate regression	Macro	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Baum-Snow and Kahn	16 cities of US	Regression analysis	Macro	Yes	No	No	Yes	Yes	No	No	Yes
Robert	Montgomery County, Maryland	Multinomial mode choice model	Macro	No	Yes	No	Yes	Yes	No	Yes	Yes
Kohn	Canada	Multiple regression analysis	Macro	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Chen et al.	New Jersey to New York	ARFIMA (auto-regressive fractionally integrated moving average) model	Macro	Yes	Yes	No	No	No	No	Yes	Yes
Kain and Liu	Houston	Cross-section and time series model	Macro	Yes	Yes	Yes	Yes	Yes	No	Yes	No
Kim et al.	St. Louis Metro Link	Multinomial logit (MNL) model	Macro	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Lane	35 cities of USA	Multiple regression analysis	Macro	No	Yes	Yes	No	No	No	No	Yes
Taylor	265 urbanized areas of USA	Multiple regression analysis and single-stage OLS model	Macro	No	Yes	Yes	Yes	Yes	Yes	No	No
Chiang et al.	Metropolitan Tulsa	Regression analysis (with autoregressive error correction), neural networks, and ARIMA models	Macro	Yes	No	No	Yes	Yes	No	Yes	No
Gkritza et al.	Athens, Greece	Generalized least squares method	Macro	Yes	No	No	Yes	Yes	No	Yes	No

Paper	Study Region	Methodological Approach	Level of Analysis	Panel data/ Time series	Accessibility	Transportation & stop level Infrastructures	Sociodemographic characteristics	Socioeconomics characteristics	Road network characteristics	Fuel Price Travel Cost	Built environment
Paulley et al.	Great Britain	Comparison	Macro	No	Yes	No	No	Yes	No	Yes	No
Kuby et al.	Nine cities in USA	Cross-sectional/Linear regression analysis	Micro, Station level	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Voith	Southeastern Pennsylvania	Fixed-effects ridership level model	Micro, Station level	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Lee et al.	Korea	Sketch level ridership models Linear Regression	Micro, Block level	No		No	Yes	Yes	No	No	No
Gutiérrez et al.	Madrid, Spain	Distance-decay weighted regression model	Micro, Station level	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Huang et al.	Wuhan, China	Accessibility-weighted ridership model	Micro, Station level	Yes	Yes	No	No	Yes	No	No	Yes
Liu et al.	Maryland	Direct ridership models (DRM)	Micro, station level	No	Yes	Yes	Yes	Yes	No	No	Yes
Beko	Slovenia	Multivariate Regression	Micro, Station level	No	No	No	Yes	Yes	No	Yes	No
Saur et al.	California	Multivariate Regression	Micro, Station level	No	No	Yes	Yes	Yes	No	No	No
Lane et al.	17 U.S. regions	Multivariate Regression	Micro, Station level	No	No	Yes	Yes	Yes	No	No	Yes
Choi et al.	Seoul, Korea	Multiplicative model and the Poisson regression model	Micro, Station level	No	Yes	Yes	Yes	Yes	No	No	Yes
Parks et al.	U.S regions	Linear Regression	Micro, station level	No	Yes	No	Yes	Yes	No	No	Yes
Zhao et al.	Nanjing, China	Linear, Multiplicative Regression	Micro, station level	No	Yes	No	Yes	Yes	No	No	Yes

Paper	Study Region	Methodological Approach	Level of Analysis	Panel data/ Time series	Accessibility	Transportation & stop level Infrastructures	Sociodemographic characteristics	Socioeconomics characteristics	Road network characteristics	Fuel Price Travel Cost	Built environment
Zhang and Wang	New York	Network Kriging regression	Micro, station level	No	Yes	No	Yes	Yes	No	No	Yes
Sun et al.	Beijing, China	Direct ridership models (DRM)/Multiple Regression Analysis	Micro, station level	No	No	No	No	No	No	No	Yes

2.3 Literature Review of Cost-Benefit Analysis Studies

Given the limited financial resources for urban transportation planning organizations it is important to quantitatively analyze the impacts of transportation investments in an effort to maximize the resource allocation efficiency across different transport needs. Cost-benefit analysis (CBA) is considered to be one of the most appropriate tools in evaluating transportation policies and projects (Litman, 2001). A comprehensive CBA would allow analysts to predict several direct and/or indirect impacts of improvements in existing system or proposed new infrastructures. A comprehensive CBA of public transit mode investments would assist the planners and policy makers to evaluate the "real" benefit of these investments and provide evidence to justify allocation of more funding for improving/building public transit infrastructures. The current research report focuses on reviewing existing literature of CBA for transit infrastructure investments. The literature review will enable the research team to identify several factors that are generally considered in different components of CBA and thus aid in developing a template for CBA for the Central Florida region.

Several studies have evaluated CBA in terms of transit infrastructure investments. Weisbrod et al. (2014) performed an economic impact analysis of public transportation investments. From the long-term impact analysis, the study concluded that increased transit investments have potential for significant economic gain as well as societal benefits. They showed that a programme of enhanced public transit investment over twenty years will lead to an increase in income that is equivalent to approximately 50,000 additional jobs per $1 billion invested. Litman (2004) provided a framework for evaluating CBA of a particular transit service or improvements. The author pointed out that the conventional transport evaluation model is usually developed based on financial cost to government, vehicle operating cost, travel speed, crash risk and project construction environmental impacts. These studies overlook many benefits factors; such as downstream congestion impact, parking cost, environmental impacts, strategic land use impact, equity impact, public health and transportation diversity value.

Godavarthy et al. (2014) have documented and quantified benefits of small urban and rural transit systems in the US by employing CBA. The authors categorized transit benefits in three components: transit cost savings benefits (vehicle ownership and operation expenses, chauffeuring cost savings, taxi trip cost savings, travel time cost savings, crash cost savings and emission cost savings), low-cost mobility benefits and economic impact benefits. Cost component included capital, operation and maintenance costs. From the extensive analysis results, the authors concluded that the benefits (benefit-cost ratio greater than 1) provided by transit services in rural and small urban areas are greater than the costs of these services. With respect to rail transit system, Gordon and Kolesar (2011), in an effort to perform CBA for rail transit system in modern American cities, also considered non-user benefit in the benefits component other than conventional benefit measures. The non-user benefits included was number of auto trips avoided by any new-to-transit passengers. Based on the analysis, the authors found that rail transit system into modern American cities cannot be justified on economic ground even after accounting for non-user benefits in the assessments.

Bus Rapid Transit (BRT) has emerged as an attractive public transit system to enhance level of accessibility, mobility and system capacity. Some of the studies have conducted CBA for BRT system as well. Ang-Olson and Mahendra (2011) discussed a methodology of CBA for evaluating the potential benefits of converting a mixed traffic lane to an exclusive BRT lane at a corridor, local and regional level. The costs quantified in the analysis were capital cost, operation and maintenance costs. The benefits component included change in crash cost, travel time change cost, travel cost savings, emission and noise reduction costs and indirect social benefits (land development impacts, savings in parking costs, accessibility impacts and system reliability impacts). From the analysis of a hypothetical project, the authors showed that converting an arterial traffic lane for BRT can result in positive net benefits if the arterial has high person throughput and relatively high pre-project transit mode share. Blonn et al. (2006) analyzed costs and benefits of implementing a BRT system in the greater Madison metropolitan area. The analysis was conducted by considering several costs (raising local revenue, capital cost, operations and maintenance costs) and benefits (reduced travel time, reduced vehicle user cost, reduced emission and reduced crash cost). Based on the CBA, the authors concluded that implementing a BRT system in the greater Madison metropolitan area would return negative net benefits and hence would not be justified to implement on efficiency grounds.

CHAPTER THREE: DATA SOURCE AND DATA PREPARATION

3.1 Study Area

Orlando metropolitan region is the 24th largest metropolitan area in the United States. Greater Orlando region has experienced rapid growth. In fact, according to the US Census Bureau, Orlando is the fastest growing urban region among the country's thirty large urban regions (Brinkmann, 2016). The rapid growth in population increases the stress on the existing transportation system. Thus, it is not surprising that several transportation and public transit investments are underway in the region to alleviate traffic congestion and improve mobility for Greater Orlando residents. The Greater Orlando region with a population of around 3.2 million in 2016 is a typical American city in the south with an automobile oriented transportation system with the following mode share: automobile (85.7%), Public transit (1.0%), walk (9.2%) and bike (1.2%). The main public transit service in the region is the Lynx system that serves an area of approximately 2,500 square miles within Orange, Seminole, Osceola and Polk County in central Florida. The bus system operates 77 daily routes with average weekday ridership of around 105,000. SunRail, a commuter rail system has been introduced in the city on May 1, 2014. SunRail system is 31 miles long with 12 stations that connect Volusia county and Orange county. The system served an average of 3,800 passengers on weekdays in 2015. Figure 1 represents the study area along with Lynx bus route, bus stop, SunRail line and SunRail station locations.

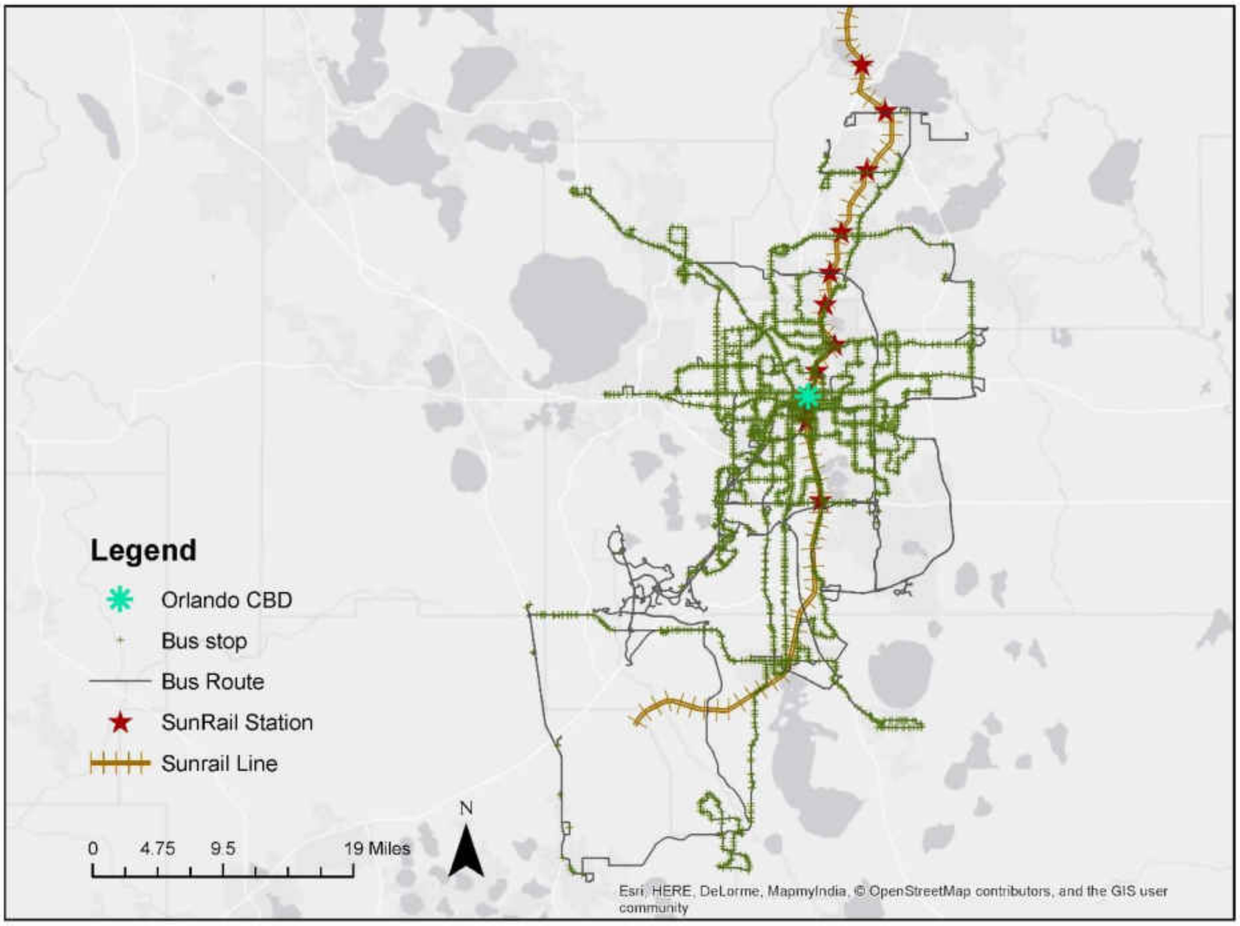


Figure 1. Public Transit System (LYNX and SUNRAIL) of Orlando

3.2 Data Source and Preparation for Bus Ridership

3.2.1 Data Source

The bus ridership data was obtained from Lynx transit authority. GIS shape files from Lynx were used to identify the number of bus stops, bus route length. For creating the exogenous variables, we considered various buffer distances (800m, 600m, 400m, and 200m) from each bus stop. The exogenous variable information was generated based on multiple data sources including 2010 US census data, American Community Survey (ACS), Florida Geographic Data Library (FDGL), and Florida Department of Transportation (FDOT) databases.

3.2.2 Data Preparation

For the purpose of our analysis, stop level average weekday boarding and alighting ridership data for 6-time periods of 4-month each are considered. These include the following 6-time period: May through August 2013, September through December 2013, January through April 2014, May through August 2014, September through December 2014, January through April 2015. The ridership information was processed for all the 6-time periods and analyzed to ensure data availability and accuracy. The resulting data provided ridership information for 3,745 stops across the 6-time periods. The ridership data was augmented with stop level headway, route length as well as route to stop correspondence for Lynx across the 6-time periods. A summary of the system level ridership (boarding and alighting) are provided in Table 3. The average weekday boarding (alighting) across the 6-time periods range from 71,006 (71,029) to 77,940 (76,725).

Table 3. Summary Statistics of Lynx Bus Ridership (August 2013 to April 2015)

Time-period	Quarter Name	Number of Observations	Boarding		Alighting	
			Mean	Standard Deviation	Mean	Standard Deviation
1	August-13	20970	22.30	160.51	21.95	152.86
2	December-13		20.88	151.85	20.61	143.49
3	April-14		20.54	157.83	20.32	151.89
4	August-14		21.51	162.01	21.38	154.30
5	December-14		20.32	151.18	20.39	146.65
6	April-15		20.65	156.02	20.52	149.57

We consider thirteen categories/bins for analysis ridership as per the frequency of ridership and these categories/bins are: Bin 1 = 0~5; Bin 2 = >5~10; Bin 3 = >10~20, Bin 4 = >20~30, Bin 5 = >30~40, Bin 6 = >40~50, Bin 7 = >50~60, Bin 8 = >60~70, Bin 9 = >70~80, Bin 10 = >80~90, Bin 11 = >90~100, Bin 12 = >100~120 and Bin 13= >120 ridership. Figure 2 and table 4 shows the frequency distribution for both boarding and alighting categories/bins.

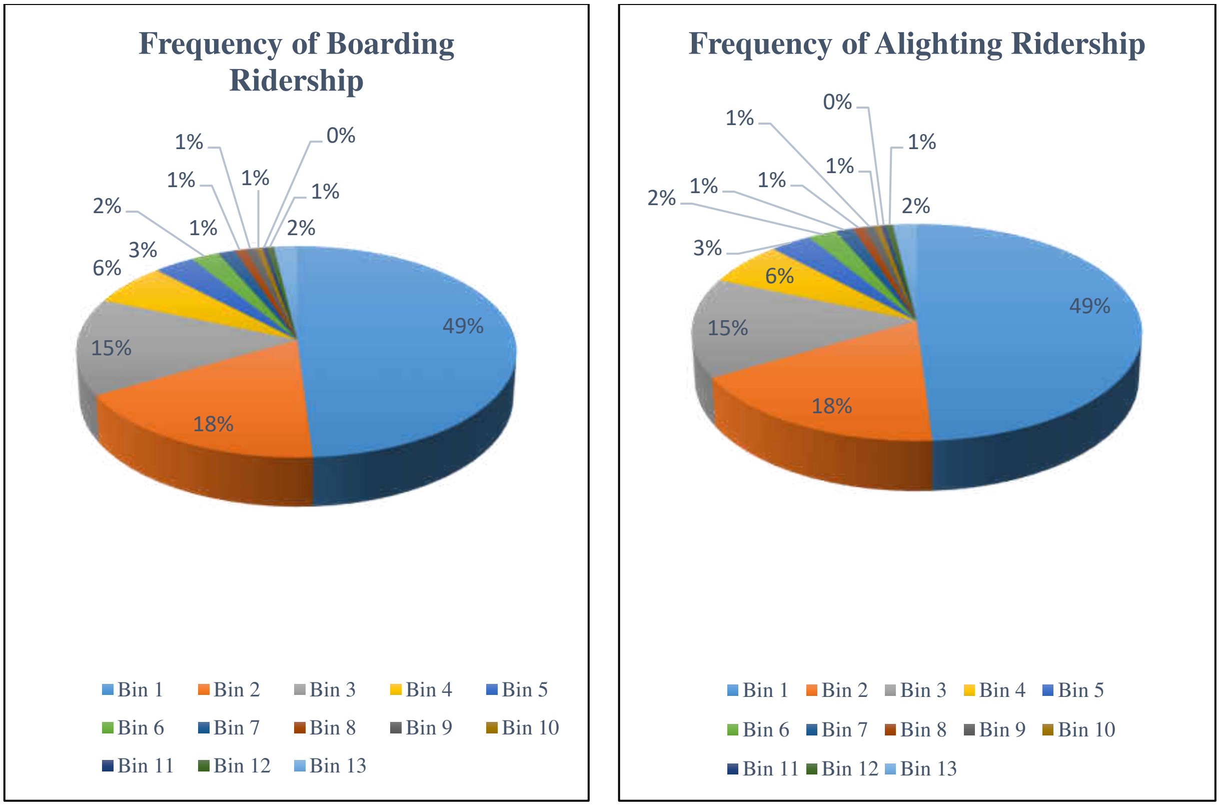


Figure 2. Frequency Distribution for boarding and alighting

Table 4. Frequency distribution of each ridership category for boarding and alighting

Ridership Category	**Frequency**		**Percent**		**Cumulative Percent**	
	Boarding	**Alighting**	**Boarding**	**Alighting**	**Boarding**	**Alighting**
1	16182	15544	52.5	50.5	52.5	50.5
2	5315	5306	17.3	17.2	69.8	67.7
3	4224	4433	13.7	14.4	83.5	82.1
4	1594	1906	5.2	6.2	88.7	88.3
5	888	982	2.9	3.2	91.6	91.5
6	581	683	1.9	2.2	93.5	93.7
7	468	383	1.5	1.2	95.0	94.9
8	302	298	1	1.0	96.0	95.9
9	218	231	0.7	0.8	96.7	96.6
10	157	158	0.5	0.5	97.2	97.2
11	113	108	0.4	0.4	97.5	97.5
12	182	190	0.6	0.6	98.1	98.1
13	576	578	1.9	1.9	100.0	100.0

We identified specific bus routes that intersect or pass through the SunRail system. Of the 77 bus routes operated by Lynx, we found that 60 routes are within the SunRail influence zone (i.e. pass through SunRail). These routes account for 3,321 out of the 3,745 stops considered in our analysis. To allow stops in the proximity of different SunRail stations, we identify influence stops separately for different stations. To capture the realization that the effects of SunRail on bus ridership would be only after the SunRail came into operation, interaction terms representing influence of SunRail and quarters representing SunRail operational period (May through August 2014, September through December 2014, January through April 2015) are generated. Further, these interactions terms (SunRail synced stops*SunRail operation period) are employed as exogenous variables in the current study context.

The exogenous variables considered for the empirical analysis can broadly be categorized as stop level attributes, transportation infrastructure characteristics, built environment attributes, demographic and socioeconomic characteristics, temporal effects and SunRail effects. Stop level attributes include headway, number of bus stops in a buffer around stops. Transportation infrastructure characteristics include bus route, side walk and rail road lengths in a buffer around stops. Built environment attributes include land use mix[1] in a buffer around stops and distance of stop from central business district (CBD). Demographic and socioeconomic characteristics include number of population aged 17 and less, number of population with education at some college level, number of population with education at bachelor level, number of households with low income level and number of owned households by residents. The demographic and socioeconomic characteristics are generated at the census tract level. In terms of Temporal effect, we introduced

[1] Land use mix = $\left[\frac{-\sum_k (p_k(lnp_k))}{lnN}\right]$, where k is the category of land-use, p is the proportion of the developed land area devoted to a specific land-use, N is the number of land-use categories within 1mile buffer of the roadway segment.

a variable called "time elapsed" which is the time difference between the most recent quarters from the base quarter (May through August 2013) considered in the current study context. In our case, for the 6-time periods, the variable takes the following values: 0, 1, 2, 3, 4 and 5. Finally the SunRail effect includes variables representing the interaction of SunRail synced stops and SunRail operation period. Temporal lagged variables were calculated for each bus stop by computing the boarding (alighting) variables from previous time period. Temporal and spatio-temporal lagged variables (such as stop boarding (alighting) in the last time period) is also considered. Spatio-temporal lagged variables were created based on stops within the buffer. The boarding (alighting) from previous time period for stops within the buffer were generated for spatio-temporal lag variables.

Several buffer sizes - 800m, 600m, 400m, and 200m - around the bus stop were employed for variable generation. A summary of the exogenous variables generated is provided in Table 5.

Table 5. Descriptive Statistics of Exogenous Variables for bus ridership

Variable Name	Variable Description	No of obs., n	Minimum	Maximum	Mean
Stop Level Attributes					
Headway	Headway in minutes	20970	1.11	60.00	37.63
Headway	Ln of headway		0.11	4.09	3.49
No of Bus stop in a					
800 m buffer	Scale: (Number of bus stops in 800m buffer)/10		0.10	9.30	1.79
Transportation Infrastructure around the stop					
Bus route Length in a	Bus route length in kilometers	20970			
600 m buffer	(Bus route length in 600 m buffer)/10		0.11	6.06	0.51
400 m buffer	(Bus route length in 400 m buffer)/10		0.05	4.17	0.27
Side walk length in a	Side walk length in kilometers				
800 m buffer			0.00	13.27	3.16
Secondary highway length in a	Secondary highway length in kilometers				
800 m buffer	Secondary highway length in 800 m buffer / Total road length in 800 m buffer		0.00	1.00	0.34
Rail road length in a	Rail road length in kilometers				
800 m buffer			0.00	6.04	0.31
Local road length in a	Local road length in kilometers				
800 m buffer	Local road length in 800 m buffer / Total road length in 800 m buffer		0.00	1.00	0.65
Presence of shelter in bus stop	(1 = Yes/0 = No)		0.00	1.00	0.23
Built environment around the stop					
Residential area in a	Residential area in square kilometers	20970			
800 m buffer	Residential area in 800 m buffer / Total area in 800m buffer		0.00	1.00	0.32
600 m buffer	Residential area in 600 m buffer / Total area in 600m buffer		0.00	1.00	0.31
Land use mix area in an 800 m buffer	Land use mix = $\left[\frac{-\sum_k \left(p_k (ln p_k)\right)}{ln N}\right]$, where $\boldsymbol{k}$ is the category of land-use, $\boldsymbol{p}$ is the proportion of the developed land area devoted to a specific land-use, $\boldsymbol{N}$ is the number of land-use categories within 1mile buffer of the roadway segment.		0.001	0.810	0.501

Variable Name	Variable Description	No of obs., n	Minimum	Maximum	Mean
Household density	HH Density = HH size / Census area/1000		0.005	3.718	0.476
Employment density	Employment Density = Employment / Census area/1000		0.000	37.339	1.096
Central Business area distance (km)	(Central Business area distance)/10		0.00	5.06	1.18
Sociodemographic and socioeconomic variables in census tract					
Age 0 to 17 years	Ln of (People age 0 to 17 years)/Census Area		-6.584	3.682	-0.282
Age 65 and up	Ln of (People age 65 and up)/Census Area		-6.36	3.23	-1.07
Education level - 9 to 12 grade	Ln of (Education level 9 to 12 grade / Census Area)		-8.04	2.41	-1.50
Low Income Category (<30k)	Ln of (Low income People (<30k)/Census Area)	20970	-8.55	2.85	-0.77
High Income Category (>80k)	Ln of (High income People (>80k)/Census Area)		-8.526	2.740	-1.827
Vehicle Ownership - No vehicle	Ln of (Vehicle Ownership - No Vehicle / Census Area)		-8.55	1.58	-2.11
Household ownership	Ln of (Household Ownership / Census Area)		-6.87	3.36	-0.53
Spatial and Spatio-Temporal Effect					
Temporal lagged variables 1 for boarding	Ln of temporal lagged variables 1 for boarding		0.00	8.857	1.459
Temporal lagged variables 1 for alighting	Ln of temporal lagged variables 1 for alighting		0.00	8.820	1.490
Spatio-Temporal lagged variables 1 for boarding in a 800 m buffer	Ln of spatio-temporal lagged variables 1 for boarding in a 800 m buffer	20970	0.00	9.623	3.811
Spatio-Temporal lagged variables 1 for alighting in a 800 m buffer	Ln of spatio-temporal lagged variables 1 for alighting in a 800 m buffer		0.00	9.584	3.815

3.3 Data Source and Preparation for Rail Ridership

3.3.1 Data Source

The main data source of SunRail daily ridership is the SunRail authority. In our study, the rail ridership analysis is focused on the 12 active stations shown in Figure 3.

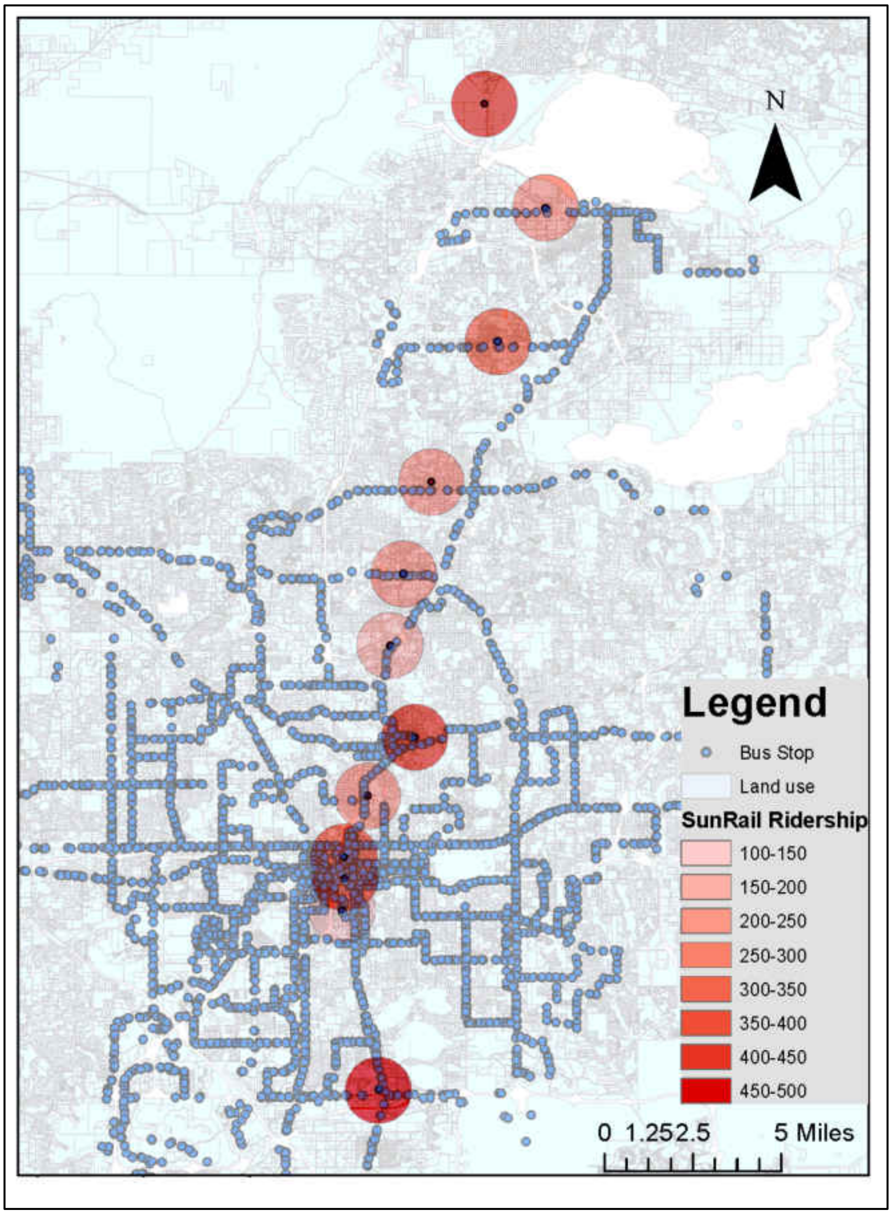


Figure 3. SunRail line and station locations.

In addition to the rail ridership, we assembled variables from multiple sources including 2010 US census data, American Community Survey (ACS), Florida Geographic Data Library (FDGL), Florida Department of Transportation (FDOT) and Florida Automated Weather Network (FAWN) databases. For the empirical analysis, the explanatory variables can be grouped into three broad categories: temporal and seasonal variables, transportation infrastructure, land use variables, sociodemographic variables, and weather variables.

3.3.2 Data Preparation

We have compiled stop level daily boarding and alighting ridership data for ten months from January 2015 to October 2015. The daily ridership data includes weekdays only as SunRail did not operate during weekends during the data collection period. This ridership data is processed and analyzed to ensure data availability and accuracy. A summary of the system level ridership (boarding and alighting) is provided in Table 6. The average daily boarding (alighting) across the 10-month periods range from 124.26 (134.09) to 451.17 (512.18). It is interesting to observe that the two end stations (Sand Lake and Debary Stations) have the highest difference in daily boarding and alighting values relative to other stations. The 10-month, 12 station data provided us 2,496 observations. Out of 2,496 observations, 2,124 observations were randomly selected for model estimation and remaining 372 observations were set aside for model validation.

Table 6. Summary Statistics for SunRail Average Daily Ridership (January 2015 to October 2015)

Station Name	**No of Observations, n**	**Boarding**		**Alighting**	
		Mean	**Standard Deviation**	**Mean**	**Standard Deviation**
Sand Lake Station (SLR)	2124	451.168	82.127	512.178	111.112
Amtrak Station (ARTRAK)		124.260	20.507	134.091	16.969
Church Street Station (CSS)		393.135	79.184	400.962	96.775
Lynx Central Station (LCS)		403.769	35.282	377.813	34.610
Florida Hospital (FLHS)		201.976	26.562	224.168	29.862
Winter Park Station (WPS)		411.707	205.107	443.433	203.524
Maitland Station (MLS)		180.962	27.084	183.697	23.986
Altamonte Springs station (ATSS)		244.163	40.788	251.135	35.830
Longwood Station (LWS)		240.909	36.959	227.024	29.418
Lake Mary Station (LMS)		337.005	55.139	312.221	51.052
Sanford Station (SFS)		258.952	45.735	235.202	38.199
Debary Station (DBS)		445.178	90.608	391.260	93.938

For the empirical analysis, the explanatory variables can be grouped into three broad categories: temporal and seasonal variables, transportation infrastructure, land use variables, sociodemographic variables, and weather variables. The data at the station level was generated by creating a buffer around the rail station using ArcGIS. However, the influence buffer size area may vary across different variables (see Chakour & Eluru, 2016). To accommodate for such an effect on transit ridership, we have computed attributes of different variables by using 1500m, 1250m, 1000m, 750m, and 500m buffer sizes. Temporal and seasonal variables considered include day of week and month of the year. Transportation infrastructure variables considered include local roadway length, number of bus stops, and presence of free parking facilities at stations. Land use variables considered include number of commercial centers, number of educational centers, number of financial centers and land use mix. Sociodemographic variables considered include number of households with zero vehicle ownership level. Finally, weather variables considered include temperature, average wind speed and rainfall. Table 6 offers a summary of the sample characteristics of the exogenous factors used in the estimation data set. Table 7 represents the definition of variables considered for final model estimation along with the minimum, maximum and average values of the exogenous variables.

Table 7. Descriptive Statistics of Exogenous Variables for rail ridership

Variable Name	Variable Description	No of obs. n	Minimum	Maximum	Mean
Temporal and Seasonal Variables					
Day of week					
Monday	Rail ridership on Monday (Proportion)	2124	0.000	1.000	0.190
Friday	Rail ridership on Friday (Proportion)		0.000	1.000	0.206
Month of the Year 2015					
January	Rail ridership on January 2015 (Proportion)	2124	0.000	1.000	0.094
February	Rail ridership on February 2015 (Proportion)		0.000	1.000	0.095
March	Rail ridership on March 2015 (Proportion)		0.000	1.000	0.109
April	Rail ridership on April 2015 (Proportion)		0.000	1.000	0.105
May	Rail ridership on May 2015 (Proportion)		0.000	1.000	0.095
June	Rail ridership on June 2015 (Proportion)		0.000	1.000	0.106
July	Rail ridership on July 2015 (Proportion)		0.000	1.000	0.111
August	Rail ridership on August 2015 (Proportion)		0.000	1.000	0.103
Transportation Infrastructures					
Local roadway length in a 1500 m buffer	Local roadway length in kilometers	2124	16.113	141.443	77.956
Number of bus stops in a 1500 m buffer	Number of Lynx bus stop in 1500 m buffer from SunRail station		0.000	205.000	55.667
Free Parking Facility	Free Parking Facility (Yes and No)		0.000	1.000	0.667
Land Use Patterns					
Number of Commercial centers in a 1500 m buffer		2124	0.000	6.000	2.750
Number of Educational centers in a 1500 m buffer			0.000	11.000	4.250

Variable Name	Variable Description	No of obs. n	Minimum	Maximum	Mean
Number of Financial centers in a 1500 m buffer			0.000	55.000	17.833
Land Use mix in a 1500 m buffer			0.263	0.811	0.638
Sociodemographic Variables					
Vehicle Ownership – No vehicle 1500 m buffer	Vehicle Ownership – No Vehicle	2124	52.000	4532.000	1326.250
Weather Variables					
Average Temperature in air	Average Temperature in air at 2 m height in degree Celsius	2124	4.889	30.204	23.222
Average Wind speed in air	Average wind speed in air at 10 m height in miles per hour		2.892	12.040	5.566
Rainfall	Sum of rainfall at 2 m in inches		0.000	1.577	0.132

CHAPTER FOUR: BUS RIDERSHIP ANALYSIS

4.1 Introduction

The major focus of the proposed research effort is to evaluate the influence of recently inaugurated commuter rail system "SunRail" in Orlando on bus ridership while controlling for host of other exogenous variables including stop level attributes, transportation infrastructure variables, transit infrastructure variables, land use and built environment attributes and sociodemographic and socioeconomic variables. Given the relatively long-time span required for the influence of large scale public transportation system changes, any analysis of the value of new investments should consider adequate data before the system installation and after the system installation. The data for the study is drawn from bus ridership information for six 4-month time periods - 3 prior to installation of SunRail and 3 after installation of SunRail - allowing us to study time varying effects of SunRail system on ridership.

4.2 Current Study in Context

While several research efforts have explored the influence of a host of exogenous variables on transit ridership, it is evident from the literature review (presented in section 2.1), that no earlier research effort has examined the impact of new transit investment on existing transit infrastructure. Of course, data availability was a major impediment for the analysis. Further, the earlier research studies on ridership have heavily focused on linear or log-linear regression approaches (with some exceptions). These approaches impose an implicit structure on the impact of exogenous variables. Chakour and Eluru (2016), in their recent research relaxed this assumption by estimating a flexible non-linear specification in the form of an ordered regression model. While the approach is definitely less restrictive relative to linear or log-linear models, it adds an additional burden for model estimation with the need to estimate threshold parameters. The number of threshold parameters are associated with the number of ordered alternatives considered. Chakour and Eluru (2016) considered only 5 categories thus minimizing the additional burden. However, in cases where the range of ridership varies substantially, it might necessitate a large number of threshold parameters thus increasing the burden required for parameter estimation.

The current research effort is focused on addressing these two aforementioned limitations. First, by employing data on stop level ridership (weekday boarding and alighting) for three 4-month time periods before and after commuter rail installation in a large metropolitan area, the current research effort makes a unique empirical contribution identifying the commuter rail impact while controlling for all other factors affecting ridership. Second, the study contributes methodologically, by developing a panel joint grouped response ordered modeling framework. The proposed model accommodates for common unobserved factors affecting boarding and alighting as well as repeated measures for each stop. Furthermore, the grouped response structure allows for flexible specification of the dependent variable while also not being restricted by additional threshold parameters to be estimated (see Chakour and Eluru, 2016). Through our grouped response model structure, we avoid the estimation of thresholds by recognizing that the thresholds of bus ridership are observed and the propensity can be tied to the observed thresholds while relaxing the standard normal or logistic assumption for the variance. Thus, irrespective of the number of ridership categories generated there is no additional parameter burden. In fact, the approach allows us to estimate exactly the same number of parameters as in the linear or log-linear regression approaches. To be sure, the proposed application of the simple grouped response model is not the first of its kind in literature. Eluru et al. (2009) have employed the grouped response structure in a different empirical context (for examining residential mobility). However, the study does not explicitly provide details of the advantages of the framework. The reader would also note that the panel joint grouped response structure proposed in our paper is different from the approach employed in Eluru et al. (2009), and is the first application of this methodology in transportation literature as well as econometric literature in general.

4.3 Methodology for Bus Ridership

The focus of this study is to examine stop-level boarding and alighting ridership simultaneously. Let q (q = 1, 2,…, Q) be an index to represent bus stops, let t (t = 1, 2, 3,…, T) represent the different time periods and j (j = 1, 2, 3,…, J = *13*) be an index to represent the number of boardings or alightings. We consider thirteen categories for ridership analysis and these categories are: Bin 1 = ≤5; Bin 2 = 5-10; Bin 3 = 10-20, Bin 4 = 20-30, Bin 5 = 30-40, Bin 6 = 40-50, Bin 7 = 50-60, Bin 8 = 60-70, Bin 9 = 70-80, Bin 10 = 80-90, Bin 11 = 90-100, Bin 12 = 100-120 and Bin 13= >120. Then, the equation system for modeling boarding's and alighting's jointly may be written as follows:

$$B_{qt}^{*} = (\alpha' + \gamma_q')x''_{qt} + (\theta' + \mu'_q)h_{qt} \pm (\eta'_q)y_{qt} + \varepsilon_{qt}, B_{qt} = j \ \ if \ \psi_{j-1} < B_{qt}^{*} \leq \psi_j \quad (1)$$

$$A_{qt}^{*} = (\beta' + \delta_q')x''_{qt} + (\theta'' + \mu''_q)h_{qt} \pm (\eta'_q)y_{qt} + \xi_{qt}, A_{qt} = j \ \ if \ \psi_{j-1} < A_{qt}^{*} \leq \psi_j \quad (2)$$

In equations 1 and 2, B_{qt}^{*} (A_{qt}^{*}) is the latent propensity for stop level boardings (alightings) of stop q for the t^{th} time period. This latent propensity B_{qt}^{*} (A_{qt}^{*}) is mapped to the actual grouped ridership category j by the ψ thresholds, in the usual ordered-response modeling framework. In our case, we consider J = 13 and thus the ψ values are as follows: -∞, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, and +∞.x''_{qt} is a matrix of attributes that influences stop level boarding and alighting. ; α (β)is the corresponding vector of mean coefficients and $\gamma_q(\delta_q)$ is a vector of coefficients representing the impact of unobserved factors moderating the influence of corresponding element of x'_{qt} (x''_{qt}) for boardings (alightings), h_{qt} represents the headway variables generated from H_{qt} for consideration in boarding and alighting. θ' (θ'') represents the corresponding vector of mean coefficients and μ'_q (μ''_q) is a vector of coefficients representing the impact of unobserved factors moderating the influence of corresponding element h_{qt} for boardings (alightings). ε_{qt} (ξ_{qt}) is an idiosyncratic random error term assumed independently logistic distributed across choice stops and choice occasions for boardings (alightings) with variance λ_B^2 (λ_A^2). The variance vectors for boarding's and alighting's are parameterized as a function of independent variables as follows: $\lambda_B = \exp(\theta' z_{qt})$ and: $\lambda_A = \exp(\vartheta' z_{qt})$. The parameterization allows for the variance to be different across the bus stops accommodating for heteroscedasticity.

η_q present in all three equations represents the vector of coefficients that accommodates for the impact of stop level common unobserved factors that jointly influence boardings, alightings and headway. The $'\pm'$ sign indicates the potential impact could be either positive or negative. A positive sign implies that unobserved factors that increase the headway for a given reason will also increase the propensity for boarding/alighting, while a negative sign suggests that unobserved individual factors that increase the propensity for headway will decrease the propensity for boarding/alighting. In our empirical context, we expect the relationship to be positive.

To complete the model structure of the Equations (1) and (2), it is necessary to define the structure for the unobserved vectors γ_q, δ_q, σ_q, μ_q (combined vector of μ'_q and μ''_q and η_q. In this paper, we assume that the two vectors are independent realizations from normal distributions as follows: $\gamma_{qn} \sim N(0, \kappa_n^2)$ $\delta_{qn} \sim N(0, \nu_n^2)$, $\mu_{qn} \sim N(0, o_n^2)$ and $\eta_{qn} \sim N(0, \varrho_n^2)$.

With these assumptions, the probability expressions for the ridership category may be derived. Conditional on γ_{qm}, δ_{qm} and η_{qm}, the probability for stop q to have boarding and alighting in category j in the t^{th} time period is given by:

$$P(B_{jt})|\gamma,\eta = \Lambda\left[\frac{\psi_j-\left((\alpha'+\gamma_q')x''_{qt}+(\rho_j')z_{qjt}+(\theta'+\mu'_q)h_{qt}\pm(\eta'_q)y_{qt}\right)}{\lambda_B}\right]- \Lambda\left[\frac{\psi_{j-1}-\left((\alpha'+\gamma_q')x''_{qt}+(\rho_j')z_{qjt}+(\theta'+\mu'_q)h_{qt}\pm(\eta'_q)y_{qt}\right)}{\lambda_B}\right] \quad (3)$$

$$P(A_{jt})|\delta,\eta = \Lambda\left[\frac{\psi_j-\left((\beta'+\delta_q')x''_{qt}+(\tau_j')z_{qjt}+(\theta''+\mu''_q)h_{qt}\pm(\eta'_q)y_{qt}\right)}{\lambda_A}\right] - \Lambda\left[\frac{\psi_{j-1}-\left((\beta'+\delta_q')x''_{qt}+(\tau_j')z_{qjt}+(\theta''+\mu''_q)h_{qt}\pm(\eta'_q)y_{qt}\right)}{\lambda_A}\right] \quad (4)$$

where Λ (.) is the cumulative standard logistic distribution. z_{qjt} is a vector of attributes specific to stop q and ridership category alternative j , while ρ_j and τ_j is the vector of corresponding Ridership category-specific coefficients for boarding and alighting components, respectively.

The complete set of parameters to be estimated in the joint model system of Equations (3) and (4) are $\alpha, \beta, \rho, \tau, \theta$ and ϑ vectors and the following standard error terms: σ_m , ν_m and ϱ_m. Let Ω represent a vector that includes all the standard error parameters to be estimated. Given these assumptions the joint likelihood for stop level boarding and alighting is provided as follows

$$Lq\ |\Omega = \prod_{t=1}^{T}\prod_{j=1}^{J}\left[\left(P(B_{jt}|\gamma,\eta)\right)\right]^{d_{bjt}}\left[\left(P(A_{jt}|\delta,\eta)\right)\right]^{d_{ajt}} \quad (5)$$

where d_{bjt} and d_{ajt} are dummy variables taking a value of 1 if stop q has ridership within the j^{th} category for the t^{th} time period and 0 otherwise. Finally, the unconditional likelihood function may be computed for stop q as:

$$L_q = \int_{\Omega} (L_q|\Omega) f(\Omega) d\Omega \tag{6}$$

The log-likelihood function is given by

$$\mathrm{Ln(L)} = \sum_{q=1}^{Q} \ln L_q \tag{7}$$

The likelihood function in Equation (7) involves the evaluation of a multi-dimensional integral of size equal to the number of rows in Ω. We apply Quasi-Monte Carlo simulation techniques based on the scrambled Halton sequence to approximate this integral in the likelihood function and maximize the logarithm of the resulting simulated likelihood function (See Bhat, 2001; Yasmin and Eluru, 2013 for more details). The likelihood functions are programmed in Gauss (Aptech 2016).

4.4 Model Specification and Overall Measures of Fit

The empirical analysis involves estimation of different models: 1) independent grouped ordered logit (IGOL) models for boarding and alighting, 2) joint panel mixed grouped ordered logit (JPMGOL) model for boarding and alighting without correlation parameterization, and 3) joint panel mixed grouped ordered logit (JPMGOL) model for boarding and alighting with correlation parameterization. The independent models were estimated to establish a benchmark for comparison. Prior to discussing the estimation results, we compare the performance of these models in this section. We employ the Bayesian Information Criterion (BIC) to determine the best model between independent and joint models. The BIC for a given empirical model is equal to:

$$BIC = -2LL + K\,ln(Q) \quad (8)$$

where LL is the log likelihood value at convergence, K is the number of parameters, and Q is the number of observations. The model with the lower BIC is the preferred model. The log-likelihood values at convergence for the models estimated are as follows: (1) IGOL (with 30 parameters) is -65,230.750, (2) JPMGOL (with 37 parameters) without parameterization is -44,234.747 and (3) JPMGOL (with 38 parameters) with parameterization is -44,232.650. The BIC values for the final specifications of IGOL, JPMGOL without parameterization and JPMGOL with parameterization are 130,760.025, 88,837.675 and 88,843.432, respectively. The comparison exercise clearly highlights the superiority of the joint model with the correlation parameterization in terms of data fit compared to independent model.

4.5 Variable Effects

The final specification of the model development was based on removing the statistically insignificant variables in a systematic process based on statistical significance (95% significance level). The specification process was also guided by prior research and parsimony considerations. In estimating the models, several functional forms and variable specifications were explored. The functional form that provided the best result was used for the final model specifications. For variables in various buffer sizes, each variable for a buffer size was systematically introduced (starting from 800m to 200m buffer size) and the buffer variable that offered the best fit was considered in the final specification. In presenting the effects of exogenous variables, we will restrict ourselves to the discussion of the JPMGOL model with parameterization. For simplicity, we will refer JPMGOL with parameterization as JPMGOL in the following sections. The model estimates for boarding, alighting and joint effects are presented in Table 8. The variable results across different exogenous variable categories are presented below.

Table 8. Group Ordered Logit Model Results for bus ridership

Variable Name	Boarding		Alighting	
	Estimates	**t-stat**	**Estimates**	**t-stat**
Constant	-8.062	-4.634	-6.779	-4.828
Stop Level Attributes				
Headway	-1.015	-48.520	-0.710	-40.330
No of Bus stop in a				
800 m buffer	-9.051	-21.032	-7.810	-19.086
Transportation Infrastructure around the stop				
Bus route Length in a				
800 m buffer	-	-	9.91	26.995
600 m buffer	16.479	26.689	-	-
Side walk length in a				
800 m buffer	4.645	23.496	3.518	19.328
Rail road length in a				
600 m buffer	-	-	-7.044	-11.654
400 m buffer	-17.429	-14.379	-	-
Built environment around the stop				
Land Use mix area in a				
800 m buffer	-	-	22.357	11.985
400 m buffer	14.110	7.969	-	-
Central Business area distance (km)	-13.849	-27.009	-9.696	-21.332
Sociodemographic and socioeconomic variables in census tract				
Age up to 17	10.816	17.363	8.256	14.462
Education at some college level	-4.771	-12.647	-	-
Education bachelor	-7.822	-18.026	-6.722	-17.780
Low income (<30K)	7.720	12.399	4.717	8.141
HH Ownership	-5.733	-10.349	-6.160	-12.325
SunRail Effect				
Temporal ID (0,1,2,3,4,5)	-	-	-0.466	-6.005

Variable Name	Boarding		Alighting	
	Estimates	**t-stat**	**Estimates**	**t-stat**
Bus stop sync with Church streets station and before after of SunRail	-4.098	-4.543	0.963	2.301
Bus stop sync with AMTRAK station and before after of SunRail	3.605	3.391	-	-
Alternative Specific effect				
Constant – Alternative 1 (0-5 ridership)	50.755	106.590	28.919	74.165
Constant – Alternative 2 (5-10 ridership)	24.148	67.405	13.248	42.599
Scale parameter				
Constant	3.211	565.330	1.672	218.060
Correlation Parameter				
Variable Name	Estimates		t-stat	
Constant	55.137		133.697	
Temporal ID (0,1,2,3,4,5)	1.945		28.823	
Headway	0.400		40.647	

4.5.1 Stop Level Attributes

As is expected, headway at the stop level has a significant influence on ridership. We observe that with increasing headway, boarding and alighting are likely to reduce. The result highlights how transit frequency directly affects ridership. The results for number of Lynx bus stops in the 800m buffer indicates that the presence of more number of bus stops in an 800m buffer contributes to reduced ridership. The result is in contradiction to earlier work (see Chakour and Eluru, 2016). The result is perhaps indicating competition across the stops for the same ridership population.

4.5.2 Transportation Infrastructure Characteristics

Transportation infrastructure offered quite complex effects on total ridership. Bus route length in the buffer has a positive impact on ridership for both boarding and alighting. Interestingly, the influence of buffer size is slightly different for boarding and alighting. The bus route length in the 600m buffer offered the best fit for boarding whereas the corresponding buffer for alighting was 800m. The result clearly demonstrates that increasing route length (an indication of higher transit accessibility) is correlated with higher ridership. A similar positive impact is observed for side walk length variables. On the other hand, increasing rail length in the different buffer size around a stop is related to lower boarding and alighting bus ridership. The rail length in the 600m buffer best fitted the results for alighting and corresponding buffer size for boarding is 400m. The presence of higher rail road length is a surrogate for the land use in the vicinity.

4.5.3 Built Environment Attributes

Built environment variable estimates indicate significant influence on bus ridership at the stop level. Land use mix variables in different buffer size near bus stop significantly increased the boarding and alighting ridership in Orlando. The impact of land use mix is observed for the 400 m buffer for boarding and the 800 m buffer for alighting. The distance from the central business district (CBD) variable highlights how in Orlando region, ridership reduces as the distance from CBD increases.

4.5.4 Demographic and Socioeconomic Characteristics

The demographic and socioeconomic variables based on census tract of the bus stop significantly affects the bus ridership in Orlando. The presence of larger share of young population (age 17 and below) indicates increased level of boarding and alighting. The presence of higher proportion of education level at bachelor level reduces ridership. After their bachelor degree, most of the people are capable to buy their own automobiles and thus reduces ridership. The increased presence of low income population is likely to be positively associated with bus ridership, as is expected. On the other hand, increased share of household ownership has a negative influence on public transit ridership, presumably is reflecting higher economic wealth and more private auto inclination of this group of population.

4.5.5 Temporal effects and SunRail Effect

The major objective of the paper was to study the influence of SunRail system while controlling for all other attributes. The variable for SunRail impact is present only for the last three time-periods. As described earlier, we consider two variables related to SunRail: (1) Bus stop synchronized with SunRail stop and (2) time elapsed since SunRail inception in time periods. The two variables have a significant influence on the ridership components. Specifically, Bus stop synchronized with SunRail stop indicates a significant influence of bus ridership. The Church Streets SunRail station is synchronized with lynx bus stop and the interaction term between these variables along with SunRail before after variables positively affected the alighting ridership but opposite for boarding ridership. This is therefore, people are using SunRail to go downtown Orlando (as church streets station is at downtown) mostly but they are not using SunRail to return home. The AMTRAK SunRail station is synchronized with bus stop and the interaction term of this variables and before after of SunRail variables also significantly increased the boarding ridership but does not have any impact on alighting ridership. With time elapsed, we observe that the negative influence of SunRail increases over time i.e. alighting ridership is likely to less with longer time elapsed but do not have any impact on boarding ridership. While, we recognize that the coefficient is estimated on only 3 time periods, it is still an encouraging finding. The result will provide further impetus to the SunRail expansion projects.

4.5.6 Alternative Specific Effects

In the grouped ordered specification of the joint model, we also estimate alternative specific constants for categories considered across different ridership components. It is worthwhile to mention here that it is possible to estimate group-specific effects for each group considered across different components. However, in our joint model specifications, we estimate group-specific effects if it improves data fit. The results of these group specific effects are presented in second row panel of Table 7. With respect to boarding and alighting, group-specific components are estimated for one (ridership ≤5) and two (ridership 6-10) categories, respectively. Adding more group-specific components did not improve the data fit further in the current study context and hence are not included in our final joint model specifications. These parameters are similar to constants in discrete choice models and do not really have a substantive interpretation.

4.5.7 Scale Parameter

As indicated earlier, in the JPMGOL model specification, we introduce scale parameters both in the boarding and alighting components to reflect the variance of the unobserved portion for each group. From Table 3, in the second to last row panel, we can see that the scale parameters are significant for both the dimensions. The result confirms the presence of heteroscedasticity across stops highlighting the appropriateness of the proposed model structure.

4.5.8 Correlation Effects

The estimation results of the correlation effects are presented in last row panel of Table 7. We can see that the dependence effects are significant. Further, from the estimated results we can see that the dependencies are characterized by additional exogenous variables. This provides support to our hypothesis that the dependency structure is not the same across the observations. The various exogenous variables that contribute to the dependency include temporal effect and headway. The parameters represent common correlation between boarding and alighting. As shown in Equation 2, the correlation between the two components could be either positive or negative. In our analysis, we found the positive sign to offer better fit for common correlation. Overall, the results clearly support our hypothesis that common unobserved factors influence the two components.

4.6 Model Validation

We also performed a validation exercise to evaluate the performance of the estimated models. To examine the fit of the model we used aggregate measures on the validation sample with 250 stops for 6 time periods (1,500 records). The most common approach of performing validation exercise for aggregate level model is to evaluate the in-sample predictive measures. To evaluate the in-sample goodness-of-fit measures, we employ different fit measures that are widely used in statistical analysis. For this models, we compute root mean square error (RMSE) and mean absolute deviation (MAD). These fit measures quantify the error associated with model predictions and the model with lower fit measures provides better predictions of the observed data. These measures are computed as:

$$RMSE = \sqrt{\left[\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}\right]} \tag{9}$$

$$MAD = \frac{\sum_{i=1}^{n}|\hat{y}_i - y_i|}{n} \tag{10}$$

where, $\hat{y}_i$ and y_i are the predicted and observed values for event i (i be the index for event ($i = 1,2,3, \dots, N$)) and n is the number of events. Table 9 presents the values for these measures for this model. Overall, the validation exercise indicates satisfactory performance of the proposed model.

Table 9. Predictive performance evaluation

Bin	Boarding						Alighting					
	Observed (y)	Predicted ($\hat{y}$)	$(\hat{y}_i - y_i)$	RMSE	$\lvert\hat{y}_i - y_i\rvert$	MAD	Observed (y)	Predicted ($\hat{y}$)	$(\hat{y}_i - y_i)$	RMSE	$\lvert\hat{y}_i - y_i\rvert$	MAD
1	848.000	804.81	-43.19		43.19		851.000	811.45	-39.55		39.55	
2	254.000	216.82	-37.18		37.18		255.000	159.45	-95.55		95.55	
3	204.000	194.12	-9.88		9.88		187.000	165.94	-21.06		21.06	
4	76.000	46.56	-29.44		29.44		74.000	62.68	-11.32		11.32	
5	45.000	41.24	-3.76		3.76		31.000	61.00	30.00		30.00	
6	23.000	35.76	12.76		12.76		16.000	60.76	44.76		44.76	
7	12.000	30.37	18.37	22.07	18.37	18.99	18.000	56.41	38.41	35.74	38.41	25.86
8	4.000	25.31	21.31		21.31		15.000	38.36	23.36		23.36	
9	6.000	20.79	14.79		14.79		4.000	18.04	14.04		14.04	
10	5.000	16.92	11.92		11.92		10.000	9.36	-0.64		0.64	
11	8.000	13.74	5.74		5.74		4.000	6.75	2.75		2.75	
12	4.000	20.32	16.32		16.32		15.000	15.39	0.39		0.39	
13	11.000	33.23	22.23		22.23		20.000	34.40	14.40		14.40	
		Sum	-0.000007		246.90			Sum	-0.000002		336.23	

4.7 Policy Analysis

In order to highlight the effect of various attributes over time on boarding and alighting ridership, an elasticity analysis is also conducted (see Eluru and Bhat (2007) for a discussion on the methodology for computing elasticities). We investigate the change in ridership, due to the change in selected exogenous variables. Specifically, we compute the change in ridership (both boarding and alighting) for change in headway, sidewalk length, route length, and low income population percentage, CBD distance from bus stop, Young population percentage and Temporal ID for the thirteen ridership categories/bins considered. The total boardings and alightings are calculated for all the above categories/bins for the percentage changes of those exogenous variables considered. The results for the elasticity analysis are presented in Table 10.

Several observations can be made from the results presented in Table 10. First, headways, sidewalk length, CBD distance from bus stop and route length are the most important variables in terms of high ridership categories. These results indicate that ridership is more sensitive to transit attributes which endorse the need to invest in improving transit infrastructure and service in order to encourage transit usage. Second, the effect of higher percentage of low income population in HH further indicates that reduced accessibility to private automobile increases more transit usage. Thirdly, the increases of young population (aged between 0 to 17 years old), reduces the ridership over time. Finally, and most importantly, with time the SunRail temporal effect results in increased ridership – an encouraging result for SunRail expansion project under consideration. From the above policy analysis, it is clear that in the Orlando region addition of commuter rail has contributed to increased ridership in stops influenced by SunRail. Further, to increase the ridership, services related to public transit (improvement of headway and route length increasing) should be considered.

Table 10. Elasticity Analysis

Categories	**Bin = 1**	**Bin = 2**	**Bin = 3**	**Bin = 4**	**Bin = 5**	**Bin = 6**	**Bin = 7**	**Bin = 8**	**Bin = 9**	**Bin = 10**	**Bin = 11**	**Bin = 12**	**Bin = 13**
Boarding													
Headway													
10% Decrease	-4.21%	1.42%	3.10%	4.06%	4.44%	4.80%	5.14%	5.46%	5.75%	6.03%	6.29%	6.62%	7.30%
25% Decrease	-9.59%	3.19%	8.19%	11.40%	12.74%	14.05%	15.33%	16.57%	17.76%	18.92%	20.02%	21.49%	24.82%
Sidewalk at 800 m buffer													
10% Increase	-1.52%	0.07%	0.98%	1.62%	1.90%	2.18%	2.46%	2.74%	3.03%	3.33%	3.62%	4.01%	5.15%
25% Increase	-3.77%	3.98%	4.72%	5.46%	6.21%	6.99%	7.80%	8.64%	9.49%	10.68%	14.30%	-3.77%	3.98%
Route Length at 600m buffer													
10% increase	-0.84%	0.00%	0.51%	0.89%	1.06%	1.23%	1.40%	1.59%	1.79%	2.00%	2.21%	2.49%	3.66%
25% increase	-2.08%	-0.03%	1.24%	2.21%	2.65%	3.08%	3.53%	4.01%	4.52%	5.07%	5.64%	6.46%	9.89%
Low Income population													
10% increase	-0.61%	0.21%	0.49%	0.69%	0.78%	0.88%	0.98%	1.07%	1.15%	1.23%	1.28%	1.33%	1.35%
25% increase	-1.52%	0.47%	1.20%	1.73%	1.98%	2.25%	2.51%	2.76%	3.00%	3.20%	3.37%	3.52%	3.60%
CBD from bus stop													
10% Decrease	-1.69%	0.60%	1.37%	1.82%	2.01%	2.18%	2.36%	2.54%	2.71%	2.88%	3.04%	3.21%	3.56%
25% Decrease	-4.09%	1.41%	3.48%	4.78%	5.31%	5.83%	6.34%	6.86%	7.38%	7.90%	8.38%	8.97%	10.11%
Young population (Age 0 to 17 years old)													
10% increase	0.32%	-0.11%	-0.26%	-0.36%	-0.41%	-0.48%	-0.54%	-0.62%	-0.69%	-0.75%	-0.78%	-0.78%	-0.63%
25% increase	0.81%	-0.38%	-0.68%	-0.88%	-0.98%	-1.10%	-1.22%	-1.36%	-1.49%	-1.59%	-1.64%	-1.57%	-1.12%
Alighting													

Categories	Bin = 1	Bin = 2	Bin = 3	Bin = 4	Bin = 5	Bin = 6	Bin = 7	Bin = 8	Bin = 9	Bin = 10	Bin = 11	Bin = 12	Bin = 13
Headway													
10% Decrease	-3.59%	0.88%	2.64%	3.04%	3.35%	3.98%	3.84%	5.03%	6.20%	6.00%	5.53%	6.63%	7.26%
25% Decrease	-8.25%	-8.25%	-8.25%	-8.25%	-8.25%	-8.25%	-8.25%	-8.25%	-8.25%	-8.25%	-8.25%	-8.25%	-8.25%
Sidewalk at 800 m buffer													
10% Increase	-1.47%	0.08%	0.80%	0.98%	1.85%	1.90%	2.08%	2.37%	3.46%	3.88%	3.79%	4.18%	5.26%
25% Increase	-3.64%	-0.05%	2.09%	2.11%	4.69%	4.83%	5.28%	5.72%	8.82%	10.30%	10.77%	10.81%	15.06%
Route Length at 800m buffer													
10% increase	-1.11%	-0.04%	0.50%	0.81%	1.28%	1.48%	1.75%	1.68%	2.81%	3.93%	3.36%	3.20%	4.69%
25% increase	-2.70%	-0.29%	1.25%	2.06%	3.56%	3.18%	4.21%	4.94%	7.07%	8.87%	10.10%	9.54%	13.12%
Low Income population													
10% increase	-0.47%	0.21%	0.31%	0.40%	0.34%	0.43%	0.88%	0.81%	1.26%	1.42%	1.05%	0.90%	0.93%
25% increase	-1.17%	0.45%	0.77%	0.98%	1.02%	0.91%	2.15%	2.09%	3.20%	3.85%	2.54%	2.26%	2.48%
CBD from bus stop													
10% Decrease	-1.46%	0.35%	1.17%	1.35%	1.54%	2.01%	1.86%	2.23%	2.46%	2.56%	2.88%	3.83%	3.00%
25% Decrease	-3.53%	0.76%	2.89%	3.67%	3.87%	5.15%	5.20%	6.01%	6.91%	6.42%	7.72%	10.93%	8.56%
Young population (Age 0 to 17 years old)													
10% increase	0.30%	-0.11%	-0.20%	-0.35%	0.03%	-0.25%	-0.91%	-0.38%	-0.92%	-1.20%	-1.17%	-0.69%	-0.24%
25% increase	0.78%	-0.48%	-0.60%	-0.65%	0.33%	-1.00%	-2.26%	-0.91%	-1.43%	-2.26%	-2.80%	-1.94%	0.13%
Temp_ID													
2016 (6,7,8,9,10,11)	3.53%	-0.94%	-2.58%	-3.57%	-3.91%	-4.48%	-4.76%	-6.92%	-8.99%	-7.80%	-8.21%	-9.77%	-10.11%
2017 (9,10,11,12,13,14)	3.42%	-0.95%	-2.65%	-3.68%	-4.07%	-4.67%	-4.95%	-7.40%	-9.78%	-8.36%	-8.83%	-10.75%	-11.21%
Note: Bin 1 = 0~5; Bin 2 = 5~10; Bin 3 = 10~20, Bin 4 = 20~30, Bin 5 = 30~40, Bin 6 = 40~50, Bin 7 = 50~60, Bin 8 = 60~70, Bin 9 = 70~80, Bin 10 = 80~90, Bin 11 = 90~100, Bin 12 = 100~120 and Bin 13= 120+ ridership in each stop													

4.8 Summary

In this study, we examined the impact of new transit investments (such as an addition of commuter rail to an urban region) on an existing public transit system (such as the traditional bus service already present in the urban region). Specifically, the study developed a comprehensive and statistically valid framework in studying the impact of new public transportation infrastructure (such as commuter rail, "SunRail") on existing public transit infrastructure (such as bus, "Lynx) in the Orlando metropolitan region.

Two variables representing the impact of SunRail on bus ridership –and time elapsed since SunRail inception in time periods – were found to have significant impacts on bus ridership. In our research, in order to highlight the effect of various attributes over time on boarding and alighting ridership, an elasticity analysis was also presented. We investigated the change in ridership due to the change in selected exogenous variables. From the above policy analysis, it is clear that in the Orlando region adding of commuter rail has contributed to increased ridership in stops influenced by SunRail. Further, to increase the ridership, services related to public transit (improvement of headway and route length increasing) should be considered.

CHAPTER FIVE: SPATIO-TEMPORAL FACTORS ON BUS RIDERSHIP ANALYSIS

5.1 Introduction

Orlando provides an ideal test bed to identify factors influencing public transit ridership due to its increasing popularity and tourism. Drawing on stop level public transit boarding and alighting data for 6 four-month periods from May 2013 to April 2015, the current study estimates stop-level ridership models. Specifically, we apply a spatial panel regression model that accommodates for the influence of observed exogenous factors as well as unobserved factors. The repeated observation data at a stop-level offers multiple dimensions of unobserved factors including stop-level, spatial and temporal factors. In our analysis, we apply a framework to identify the observed and unobserved factors.

5.2 Current Study in Context

The review of earlier research (presented in section 2.1), indicates the burgeoning research in the bus transit ridership field. However, the literature is not without limitations. First, earlier work is usually based on a cross-sectional – a single time snapshot - ridership data. Second, earlier literature on bus transit ridership has not accommodated for observed and unobserved spatial effects on ridership. Toward addressing these limitations, we formulate and estimate a spatial panel model structure that accommodates for repeated ridership data for the same stop as well as the impact of spatial and temporal observed and unobserved factors.

5.3 Econometric Methodology

Let $q = 1, 2, \ldots, Q$ (in our study Q=3,495) be an index to represent each station (spatial unit) and $t = 1, 2, \ldots, T$ (in our study T=6) be an index for each time period. A pooled linear regression model for panel data considering spatial specific effects without considering spatial dependency can be written as:

$$y_{qt} = \beta' x_{qt} + \mu_q + \epsilon_{qt} \tag{11}$$

Where y_{qt} is the log-normal of boarding and alighting, x_{qt} is a column vector of attributes at station q and time t, and β is the corresponding coefficient column vector of parameters to be estimated. The random error term, ϵ_{qt}, is assumed to be an independently and identically distributed normal error term for q and t with zero mean and variance σ^2 , and μ_q represent a spatial specific effect to account for all the station-specific time-invariant unobserved attributes. This spatial specific effect can be treated as fixed effects or random effects. In the fixed effects model, for every station a dummy variable is created while in the random effects model, μ_q is treated as random term that is independently and identically distributed with zero mean and variance ${\sigma_\mu}^2$. The spatial random effects and random error term are assumed to be independent. The fixed effects methodology is not appropriate in the presence of time-invariant independent variables. In addition, the fixed effects models estimate a large number of parameters (one parameter specific to each station) thus are computationally cumbersome for large systems as ours. Therefore, in the current study, we restrict ourselves to spatial random effects.

In traditional econometric literature, spatial dependency is incorporated in model in two main forms: 1) by a spatially lagged dependent variable known as spatial lag or spatial autoregressive model (SAR), or 2) by a spatial autocorrelation process in the error term known as spatial error model (SEM). The first model comprises endogenous interactions effects with dependent variable at other stops and in the second model the spatial interaction is capture through the error term.

A spatial lag model can be written as follows:

$$y_{qt} = \delta \sum_{j=1}^{Q} w_{qj} y_{jt} + \beta' x_{qt} + \mu_q + \epsilon_{qt} \tag{12}$$

Where δ is called the spatial autoregressive coefficient and w_{qj} is an element from a spatial weight matrix *W*. The diagonal elements of W matrix are zero and define the spatial arrangement of the stops. Again, in some literature, other types of spatial matrices are introduced. In our study, the spatial W matrix is a 3495×3495 matrix with elements equal to 1 for the stations that are within 800m buffer area of each other and zeros for the rest of the elements. It must be noted that diagonal of W matrix is set to be zero to prevent the use of y_{qt} to model itself. For stability in estimation, a row-normalized form of the W matrix is employed as our spatial weight matrix (see Elhorst, 2014 for more details on W matrix).

A spatial error model may be written as follows:

$$y_{qt} = \beta' x_{qt} + \mu_q + \varphi_{qt} \tag{13a}$$

$$\varphi_{qt} = \rho \sum_{j=1}^{Q} w_{qj} \varphi_{jt} + \epsilon_{qt} \tag{13b}$$

where φ_{qt} accounts for the spatial auto correlated error term and ρ reflects the spatial autocorrelation coefficient. Both spatial lag model and spatial error model can be estimated using maximum likelihood approach (see Elhorst, 2014 for details on likelihood functions). In this paper, we use Matlab routines provided by Elhorst (Elhorst, 2014 ; Elhorst, 2003), to estimate pooled spatial lag and error models with spatial specific random effects.

5.4 Model Specification and Overall Measures of Fit

The empirical analysis in our study is based on two different models: 1) Spatial Error Model (SEM) and 2) Spatial Lag Model (SAR) for boarding and alighting ridership. The log linear independent models were estimated to serve as bench mark for advanced models. In this section, we compare SEM and SAR model. For each model type, the log likelihood at convergence, R square value, the number of parameters estimated, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were calculated. The AIC and BIC for a given empirical model are equal to:

$$AIC = 2K - 2LL \tag{14}$$

$$BIC = -2LL + K\,ln(Q) \tag{15}$$

where LL is the log likelihood value at convergence, K is the number of parameters, and Q is the number of observations. The model with the lower AIC or BIC is the preferred model. The log-likelihood values at convergence for the models estimated are as follows: (1) simple linear regression model for boarding (with 18 parameters) is -22,957.537, (2) simple linear regression model for alighting (with 18 parameters) is -22,911.193, (3) SEM for boarding (with 16 parameters) is –13,029.935, (4) SEM for alighting (with 15 parameters) is –12,361.319, (5) SAR for boarding (with 13 parameters) is –12,801.731 and (6) SAR for alighting (with 11 parameters) is –12,022.572. The BIC (AIC) values for the six models are as follows: (1) simple linear regression for boarding – 46,094.188 (45,951.073), (2) simple linear regression for alighting – 46,001.501 (45,858.386), (3) SEM for boarding is – 24,752.690 (26,091.870), (4) SEM for alighting is – 24,871.903 (26,219.084), (5) SAR for boarding is – 24,067.144 (25,629.462) and 6) SAR for alighting is – 24,154.603 (25,732.823). Based on the information criteria, SAR model performs better for boarding and alighting. However, the number of explanatory variable are higher in SEM model. Hence, we consider both frameworks for our discussion. The results from the models for boarding and alighting are presented in Table 11.

Table 11. Spatial Error Model (SEM) and Spatial Lag Model (SAR) Results

Variable Name	Boarding				Alighting			
	SEM		SAR		SEM		SAR	
	Estimates	t-stat	Estimates	t-stat	Estimates	t-stat	Estimates	t-stat
Constant	2.423	19.260	1.723	172.504	3.084	27.137	2.090	182.354
Stop Level Attributes								
Headway (Ln of headway)	-0.526	-29.285	-0.403	-3.473	-0.510	-28.956	-0.346	-3.894
Transportation Infrastructure Around the Bus Stop								
Bus route length in a 600m buffer	0.307	7.222	0.208	5.502	0.303	7.623	0.208	5.555
Side walk length in a 800m buffer	0.044	5.360	-	-	0.058	7.383	-	-
Secondary highway length in a 600m buffer	0.769	7.047	0.677	36.325	-	-	-	-
Local road length in a 800m buffer	0.708	10.919	0.528	-16.331	-	-	-	-
Rail road length in a 800m buffer	-	-	-	-	-0.071	-3.006	-	-
Presence of shelter in a bus stop	0.775	19.904	0.739	39.254	0.553	14.185	0.518	27.966
Built environment around the stop								
Land use mix area in a 800m buffer	0.409	2.712	0.316	3.230	0.628	4.027	0.472	41.242
Household density	-	-	-	-	-0.114	-2.115	-	-
Employment density	-0.016	-2.242	-	-	-	-	-	-
Central Business area distance (km)	-0.110	-5.460	-0.064	-3.920	-0.148	-6.901	-0.055	-3.517
Sociodemographic and Socioeconomic Variables in Census Tract								
Age 0 to 17 years	0.116	4.685	0.102	1.725	0.100	4.165	-	-
Age 65 and up	-0.106	-5.086	-0.087	-4.737	-0.095	-4.591	-	-
High income (>80k)	-0.054	-4.122	-	-	-0.067	-5.178	-0.048	-3.941
Household rent	0.051	2.518	-	-	0.065	3.114	0.056	1.741

Variable Name	Boarding				Alighting			
	SEM		SAR		SEM		SAR	
	Estimates	t-stat	Estimates	t-stat	Estimates	t-stat	Estimates	t-stat
Spatial and Spatio-Temporal Effect								
Temporal lagged variables 1 (Ln of TL)	0.052	13.320	0.050	0.349	0.051	13.513	0.048	0.344
Spatio-temporal lagged variables 1 in a (Ln of STL) 800 m buffer	-0.032	-12.685	-0.025	-6.305	-0.027	-11.098	-0.023	-6.087
Spatial auto correlated term	1.617	39.268	-	-	1.710	104.83	-	-
Spatial autoregressive term	-	-	0.336	174.130	-	-	0.374	200.094

5.5 Variable Effects

The final specification of the model development was based on removing the statistically insignificant (90% significance level) variables from the model. We considered various buffer size (800m, 600m, 400m and 200m buffer size) and considered the buffer size that offered the best data fit. Columns 2 through 5 present results from SEM and SAR models for boarding while columns 6 through 9 present results from SEM and SAR models for alighting. The model results are described by variable categories below.

5.5.1 Stop level Variables

The headway between buses at a stop has a significant influence on ridership. The result from all models confirm this. An increase in headway is associated with significant drop in ridership. The findings are in accordance with the previous literature (Turnquist, 1981; Kuah & Perl, 1988; CHien, 2005; Ruan, 2009; Abkowitz & Tozzi, 1986; Ding & Chien, 2001).

5.5.2 Transportation Infrastructures Variables

Several transportation infrastructure variables significantly affect boarding and alighting. Bus route length in a 600m buffer is associated with increase in boarding and alighting across all models. Sidewalk length in an 800m buffer is observed to positively influence boarding and alighting in the SEM model. The corresponding coefficient was not significant in the SAR models. The secondary highway length in a 600m buffer and local road length in an 800m buffer is positively associated with boarding for SEM and SAR models. However, these variables are statistically insignificant in the alighting models. Rail road length in an 800m buffer is negatively associated with alighting in only the SEM model. Finally, the presence of bus shelter at the bus stop is likely to positively influence boarding and alighting in SEM and SAR models.

5.5.3 Built Environment Variables

Several built environment variables are found to influence boarding and alighting. Land use mix variable is associated positively for boarding and alighting in SEM and SAR models. The result is quite encouraging policies favoring mixed land use developments in urban regions. An increase in household density of census tract, where the bus stop is located, is negatively associated with alighting in SEM model. On the other hand, increasing employment density (of census tract) is negatively associated with boarding in SEM model. The distance of the stop from CBD variable impact follows an expected trend. Specifically, as the stop is away from CBD, the ridership is likely to reduce.

5.5.4 Sociodemographic and Socioeconomic Variables

Several sociodemographic and socioeconomic variables based on census tract, where the bus stops are located, were found to significantly influence boarding and alighting. The proportion of people aged between 0 to 17 years is observed to positively influence boarding in both SEM and SAR model. The result is intuitive as an increase in the proportion of young individuals' increases, population without access to car is also likely to increase. For alighting, the variable has a significant influence only in the SEM model. An increase in proportion of individuals 65 and higher is associated with a reduction in boarding and alighting (except for alighting in SAR model). The result while counter intuitive on first glance is representative of vehicle access among this age group. As the number of Households in the high-income category increase, the model results indicate a possible reduction in boarding and alighting (except for boarding SAR model). The result is expected in a city like Orlando where high income individuals are more likely to use their personal vehicle for travel. Finally, the number of households renting in a census tract is positively associated with boarding and alighting (except for boarding SAR model). The relationship between rent and ridership is along expected lines.

5.5.5 Spatial and Spatio-temporal Effects

The temporal lagged variables are positively associated with boarding and alighting ridership for SEM and SAR models. On the other hand, spatio-temporal lag variables present a reverse trend. To elaborate, the results indicate that stops with larger ridership in adjacent station for previous time period are likely to have a lower ridership. The result is indicative of competition from nearby stops. The result is indicative of how the same ridership in the urban region is being split across stops.

5.5.6 Spatial Error and Spatial Lag Effects

The study estimated SEM and SAR models to account for the presence of spatial effects. The model fit measures clearly confirmed our hypothesis. In the SEM model, the results indicate the presence of a significant spatial auto-correlated error term. In the SAR model, the spatial autoregressive coefficient indicates a significant impact of unobserved effects.

5.6 Model Validation

A hold-out sample of 250 stops (250*6=1500 observation) was set aside for validation purposes. We used both SEM and SAR model to compute predicted boarding and alighting at the station level. The predicted rates were compared with the observed boarding and alighting in the sample. We computed Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to compute the deviation from observed values. The MAE (RMSE) values for the four models are as follows: (1) boarding SEM – 0.815 (1.011), (2) boarding SAR – 0.837 (1.083), (3) alighting SEM – 0.809 (1.016), and (4) alighting SAR 0.897 (1.123). The results indicate a satisfactory performance for boarding and alighting models across the two systems. Overall, between the two model systems the SEM models perform slightly better.

5.7 Summary

Two spatial models: 1) Spatial Error Model (SEM) and 2) Spatial Lag Model (SAR) are estimated for boarding and alighting separately by employing several exogenous variables including stop level attributes, transportation and transit infrastructure variables, built environment and land use attributes, sociodemographic and socioeconomic variables in the vicinity of the stop and spatial and spatio-temporal lagged variables. The model fit measures clearly confirmed our hypothesis that spatial unobserved effects influence boarding and alighting through the presence of spatial auto-correlated error term in SEM model and the spatial autoregressive coefficient in SAR model. Further, the validation exercise results confirmed that the two-model performed adequately. In our model, we have considered both boarding and alighting model separately. The observed and unobserved factors for boarding and alighting ridership at the same stop can have an impact on ridership. Incorporating such station level dependency between boarding and alighting along with spatial unobserved factors requires the development of an advanced model and is a potential avenue for future research.

CHAPTER SIX: RAIL RIDERSHIP ANALYSIS

6.1 Introduction

With the increasing investments in public transit, federal transit administration and various agencies supporting these initiatives are interested in examining the influence of investments on transit ridership. A major analytical tool to analyze the impact of these investments is the development of statistical models that consider the impact of various exogenous factors on ridership. The current study contributes to literature on transit ridership evaluation by considering daily boarding and alighting data form a recently launched commuter rail system - SunRail that began operating in May 2014 in the greater Orlando region. The service has potential to alter travel patterns in the Orlando region. The current study develops an estimation framework that accounts for these unobserved effects at multiple levels – station, station-week and station day.

6.2 Current Study in Context

Based on the literature review (presented in section 2.2), it is evident that earlier research on transit ridership has provided significant insights. However, the literature is not without limitations. At the micro level, the application of methodologies that accommodate for repeated observations is considered in only two studies. Even in these studies the authors have only accommodated for unobserved factors at a single level (such as station). However, transit ridership could potentially be influenced by unobserved factors at multiple levels. For example, in an urban region, regular weekend concerts could potentially influence Friday ridership at downtown stations. Thus, Fridays from different weeks are likely to exhibit potential correlation. Similar dependency can be envisioned for weeks with festivals in the city core. Thus, to get an accurate estimation of various exogenous factors, accommodating for presence of unobserved effects at multiple configurations is beneficial. The current study contributes to transit ridership literature by developing a flexible panel linear regression model that accommodates for the presence of unobserved factors for various levels (such as station, station-week, station-day). The most appropriate model structure for the unobserved factors is guided by intuition and data fit metrics.

6.3 Methodology for Rail Ridership

The focus of our study is to model average daily boarding and alighting by employing panel linear regression (PLR) modeling approach. The econometric framework for the PLR model is presented in this section.

Let i $(i = 1,2,3, \dots, N)$ be an index to represent weekdays, q $(q = 1,2,3, \dots, Q)$ be the index to represent different level of repetition measures (station, station-day or station-week) and r $(r = 0,1,2, \dots, R)$ be an index to represent the number of boarding or alighting. Then, the equation system for modeling boarding/alighting may be written as follows:

$$y_{ir} = (\boldsymbol{\beta_r} + \boldsymbol{\delta_{ir}} + \boldsymbol{\gamma_{qr}})\boldsymbol{x}_{ir} + \varepsilon_q \tag{16}$$

where, $\boldsymbol{x_{ir}}$ is a vector of exogenous variables specific to weekday i and ridership component r, $\boldsymbol{\beta_r}$ is the associated vector of unknown parameters to be estimated (including a constant). $\boldsymbol{\delta_{ir}}$ is a vector of unobserved factors moderating the influence of attributes in $\boldsymbol{x_{ir}}$. $\boldsymbol{\gamma_{qr}}$ is another vector of unobserved effects specific to repetition level q and ridership component r. ε_q is normal distributed error term.

In estimating the PLR model, it is necessary to specify the structure for the unobserved vectors $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ represented by Ω. In this paper, it is assumed that these elements are drawn from independent realization from normal population: $\Omega{\sim}N(0, (\boldsymbol{\pi^2}, \boldsymbol{\sigma}_q^2))$. Thus, conditional on Ω, the likelihood function for the panel model can be expressed as:

$$L_{qr} = \int_{\Omega} \left(\prod_{q=1}^{Q} \prod_{i=1}^{N} (y_{ir})\right) d\Omega \tag{17}$$

Finally, the log-likelihood function is:

$$LL = \sum_{q} Ln(L_{qr}) \tag{18}$$

The parameters to be estimated in the PLR model are: $\boldsymbol{\beta_r}$, $\boldsymbol{\pi}$ and $\boldsymbol{\sigma_q}$. In the current study context, we estimate $\boldsymbol{\sigma_q}$ for different levels of repetition measures (q). Specifically, we evaluate unobserved effects at station, station-day and station-week levels. In accommodating unobserved effects at different levels, random numbers are assigned to the appropriate observations of the repetition measures. For example, at station level, we have 12 stations. Thus, in evaluating unobserved effect at the station level, 12 sets of different random numbers are generated specific to 12 stations and assigned to the data records based on their station ID. The station-day level repetition measure represents unobserved effects across different day of week (from Monday to Friday) at each station level. Thus, the station-day has a total 60 (12 stations*5days) records and in evaluating the unobserved effect at the station-day level, 60 sets of different random numbers are generated assigned to the data records based on their station-day combinations. Finally, the station-week level repetition measure represents unobserved effect across different weeks at a station level. In our data, we have total 43 weeks of ridership records for each station resulting in 516 (12 stations*43 weeks) records. Thus, in evaluating unobserved effect at the station-week level, 516 sets of different random numbers are generated and assigned to the data records based on their station-week combinations**.** All the parameters in the model are estimated by maximizing the logarithmic function LL presented in equation 18.

6.4 Model Specification and Overall Measures of Fit

The empirical analysis of SunRail ridership is estimated based on Panel Linear Regression model (PLR). A simple linear regression model was estimated to serve as a benchmark for the panel models. The log-likelihood values for simple linear regression (LR) model of boarding and alighting are -11815.132 (with 23 parameters) and -12090.381 (with 23 parameters), respectively.

The log-likelihood values at convergence for the boarding and alighting models estimated are as follows: PLR for boarding (with 25 parameters) is -11,781.170, and PLR for alighting (with 24 parameters) is -12,051.406. Prior to discussing the estimation results, we compare the performance of these models in this section. We employ log-likelihood ratio test for comparing these models. The log-likelihood test statistic is computed as $2[\mathrm{LL_U} - \mathrm{LL_R}]$, where $\mathrm{LL_U}$ and $\mathrm{LL_R}$ are the log-likelihood of the unrestricted and the restricted models, respectively. The computed value of the LR test is compared with the $\aleph^2$ value for the corresponding degrees of freedom (dof). The resulting LR test values for the comparison of LR/PNL for boarding and alighting models are 67.926 (2 dof) and 77.951 (1 dof), respectively. The log-likelihood ratio test values indicate that PLR models outperform the LR models at any level of statistical significance for boarding and alighting models.

6.5 Variable Effects

The estimated results for boarding and alighting are presented in Table 12. In PLR models, the positive (negative) coefficient corresponds to increased (decreased) ridership propensities. The constant does not have any substantive interpretation after adding exogenous variables. The variable results across different exogenous variable categories are discussed below.

Table 12. Station-Week Level Panel Linear Regression Model Results

Variable Name	**Boarding Ridership**		**Alighting Ridership**	
	Coefficient	**t-stat**	**Coefficient**	**t-stat**
Constant	410.053	20.191	228.535	8.818
Temporal and Seasonal Variables				
Day of week (Base: Tuesday, Wednesday, Thursday)				
Monday	-21.058	-3.978	-22.072	-3.492
Friday	48.155	11.852	48.004	10.604
Season/Month of the Year (Base: September, October)				
January	51.085	5.908	61.701	6.111
February	48.283	4.248	53.774	4.305

Variable Name	Boarding Ridership		Alighting Ridership	
	Coefficient	**t-stat**	**Coefficient**	**t-stat**
March	69.643	10.948	74.101	9.798
April	40.127	5.655	44.357	5.125
May	23.001	2.670	24.675	2.660
June	43.559	4.368	41.215	4.078
July	48.178	6.392	46.287	5.135
August	26.462	3.803	28.013	3.246
Transportation Infrastructures				
Local roadway length in a				
1500 m buffer	-7.189	-38.125	-6.948	-36.956
Number of bus stop in a				
1500 m buffer	9.587	22.573	10.096	23.146
Free Parking Facility	18.315	2.210	91.194	10.437
Land Use Patterns				
Number of Commercial centers in a				
1500 m buffer	50.317	13.918	68.541	16.568
Standard Deviation	1.869	25.513	2.068	31.388
Number of Educational centers in a				
1500 m buffer	-46.088	-10.034	-38.291	-14.896
Number of Financial centers in a				
1500 m buffer	5.442	5.924	-	-
Land Use mix in a				
1500 m buffer	347.969	20.089	538.002	29.858
Sociodemographic Variables				
Vehicle Ownership - No vehicle				
1500 m buffer	-0.307	-18.523	-0.326	-21.788
Weather Variables				
Average Temperature in air	1.753	2.813	1.844	2.257
Average Wind speed in air	-3.924	-3.603	-3.832	-3.036
Rainfall	-27.756	-4.028	-25.528	-2.962
Standard error of estimates	4.066	405.301	4.183	444.830
Panel Effects				
Standard deviation at Station level	2.545	9.689	2.844	14.972

6.5.1 Temporal and Seasonal Variables

The day of the week variables offer interesting results. Specifically, the result indicate that boarding and alighting are likely to be lower on Mondays while on Fridays an opposite trend is observed. The higher ridership value on Friday is possibly associated with transit being adopted for cultural, sports and social activities (such as Orlando Lions football games or restaurants) in downtown Orlando with limited parking. To accommodate for seasonal variation in ridership we also consider the month variable. Based on the estimates, month of March is associated with largest impact on boarding and alighting. Months of September and October have the lowest impact (as they are the base). It is also observed that the association of various months with boarding and alighting are very similar.

6.5.2 Transportation Infrastructures

Several transportation infrastructure variables for various buffer sizes were considered in the model. Local highway length for a 1500m buffer area around rail stations presents a significant negative impact on boarding and alighting. On the other hand, number of bus stops within 1500m buffer variable highlights the symbiotic influence of bus transit on rail ridership. For both boarding and alighting, increase in number of bus stops is associated with higher ridership. The result while encouraging is also possibly indicative of presence of higher number of bus stops near the rail station. Finally, the availability of free parking space at SunRail stations also significantly affect both boarding and alighting ridership. The parking facilities have significantly higher impact on alighting relative to boarding.

6.5.3 Land Use Variables

Land use variables including presence of commercial centers, educational centers and financial centers within 1500 m distance from SunRail station have significant influence on ridership. The presence of higher commercial centers in 1500m buffer surrounding the station positively influences boarding and alighting. The number of commercial centers variable impact varies substantially across the stations as evidenced by the significant standard deviation parameters for both boarding and alighting models. The presence of financial centers affects boarding positively while having no impact on alighting. SunRail stations are located near downtown Orlando and provide access to commercial and financial hubs of Orlando city. In these locations, availability of parking spaces, cost of parking, and traffic congestion encourage the adoption of SunRail. On the other hand, the presence of education centers around rail stations reduces rail ridership. The result is quite intriguing. It is possible that driving is the preferred option to educational centers; particularly for parents driving their children to the education center and then proceeding to another location.

6.5.4 Sociodemographic Variables

Several socioeconomic variables under several buffer sizes were tested in the boarding and alighting models. Of these variables only one variable offered a statistically significant impact. The number of households with access to no vehicles in the 1500m buffer around the station is negatively associated with boarding and alighting. While the result is counter intuitive on first glance, it is possible that the result is a surrogate for lower job participation in these neighborhoods. The result warrants more detailed analysis.

6.5.5 Weather Variables

We also account for the impact of weather variables on ridership. While we cannot control weather patterns, these variables are included in the model to ensure that the impact of other attributes is accurately determined. The average temperature variable indicates that with higher temperature, boarding and alighting are likely to be higher. On the other hand, higher average wind speed is associated with lower boarding and alighting. The wind speed might be an indicator for possible wind gusts from hurricanes in the Orlando region. Finally, rain occurrence discourages rail usage as indicated by the negative coefficient in boarding and alighting components. The result is expected for any public transit alternative.

6.5.6 Station Specific Unobserved Effects

In estimating SunRail daily average ridership models (for boarding and alighting), we estimated several station specific unobserved effects. Specifically, we estimated unobserved effects at station, station-day and station-week level. Among different considered levels, we found that the station level effects have significant influence on both boarding and alighting components of ridership. The estimation results of the station specific standard deviation is presented in last row panel of Table 11. The significant standard deviation parameters at station level provide evidence toward supporting our hypothesis that it is necessary to incorporate these unobserved effects in examining rail ridership. The station specific standard deviation variables for boarding and alighting indicate that the daily average ridership may vary for different stations based on the unobserved effects.

6.6 Model Validation

We also performed a validation exercise with the data set aside to evaluate model performance. To examine the fit of the model, we used (31*12 = 372) 372 records. We calculated the observed mean and predicted mean for panel regression model. The predictive mean for PLR models are calculated as 309.31 and 310.72 for boarding and alighting, respectively. The values are almost similar for observed mean ridership for the validation sample (309.42 and 308.13). The validation exercise shows that the predictive performance of the panel model is good.

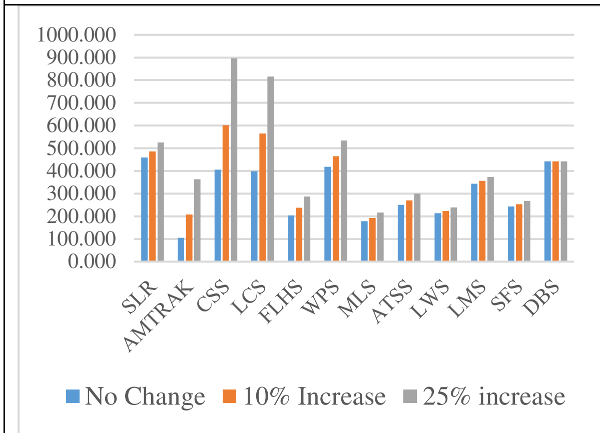
6.7 Policy Analysis

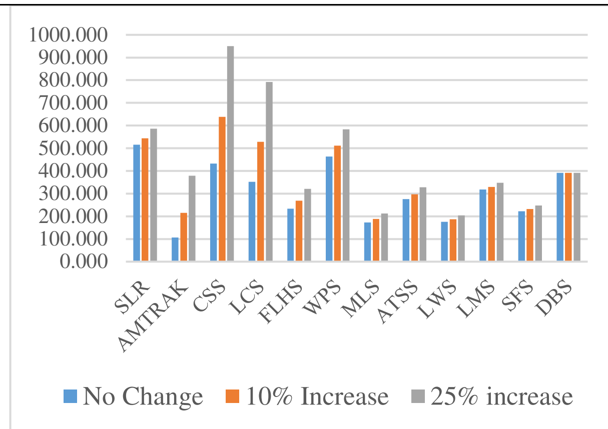
The parameter effects of exogenous variables in Table 11 do not directly provide the magnitude of the effects on exogenous variables on SunRail ridership. For this purpose, we compute aggregate level "elasticity effects" of exogenous variables. Specifically, we identified the average daily boarding and alighting ridership for changes in some selected exogenous variables. We consider the number of bus stops, land use mix and the number of commercial centers in 1500 m buffer around the SunRail stations for this purpose. In calculating the expected average predicted daily ridership, we increase the value of these variable by 10% and 25%. The computed ridership due to the change in these variables are shown in Figure 4 along with the observed daily ridership.

Boarding Ridership

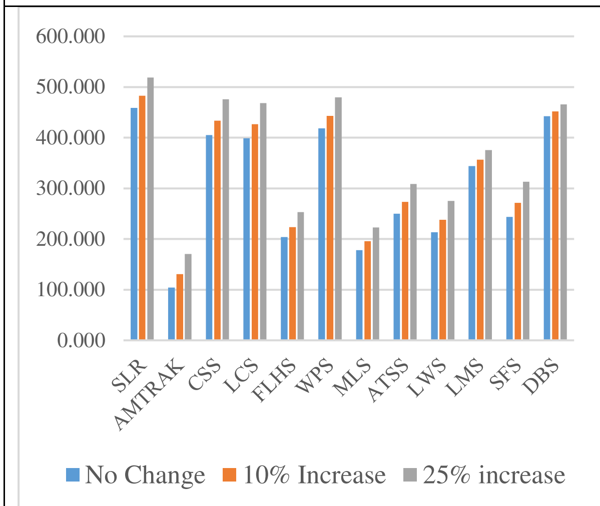
Alighting Ridership

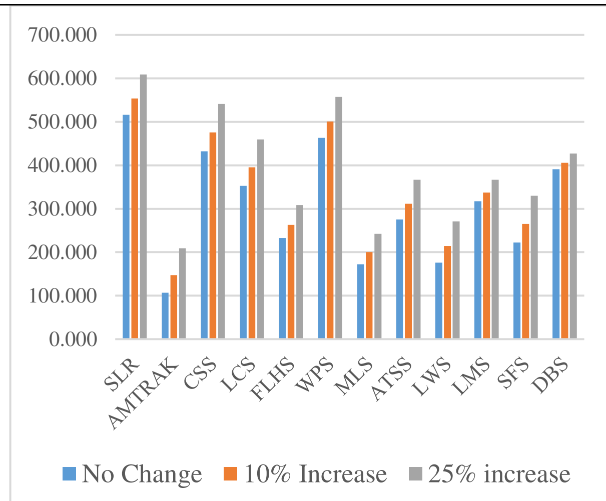
Number of bus stop increased in 1500 m buffer



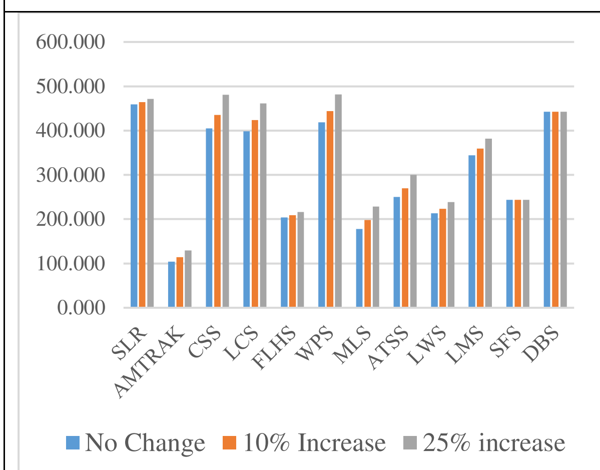


Land use mix increased in 1500 m buffer





Number of commercial center increased in 1500 m buffer



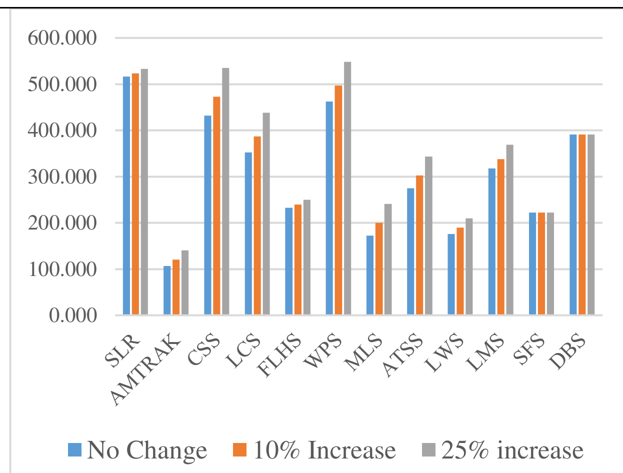


Figure 4. Policy analysis for rail ridership.

Several observations can be made from Figure 4. First, increased number of bus stops in 1500 m buffer have higher impacts in increasing the ridership on almost every SunRail station, with highest impact on AMTRAK, Church Street and Lynx Central stations. This results indicates that in the downtown area, the ridership is sensitive to bus stops around SunRail station; thus supporting =investments on transit infrastructure for encouraging an integrated transit system. Second, the effect of land use mix indicates that improving the mix of land use patterns has positive impact on ridership. The land-use mix variable has almost similar impact across all stations. Finally, increasing the number of the commercial centers also considerably increases the ridership. However, there was no impact on ridership for SFS and DBS stations. The elasticity analysis conducted provides an illustration on how the proposed model can be applied for policy evaluation for SunRail ridership.

6.8 Summary

The current study contributes to literature on transit ridership by considering daily boarding and alighting data from a recently launched commuter rail system - SunRail that began operating in May 2014 in the greater Orlando region. The analysis is conducted based on daily boarding and alighting data for ten months for the year 2015. With the rich panel of repeated observations for every station, the potential impact of common unobserved factors affecting ridership variables are considered. The current study developed an estimation framework that accounts for these unobserved effects at multiple levels – station, station-week and station day. In addition, the study examined the impact of various observed exogenous factors such as station level attributes, transportation infrastructure variables, transit infrastructure variables, land use and built environment attributes and sociodemographic and weather variables on ridership. Separate models were developed for boarding and alighting. The final specification of the model development was based on removing the statistically insignificant variables in a systematic process (at the 95% confidence level). For variables in various buffer sizes, each variable for a buffer size was systematically introduced (starting from 1500m to 500m buffer size) and the buffer variable that offered the best fit was considered in the final specification.

The day of the week variables offer interesting results. Specifically, the result indicate that boarding and alighting are likely to be lower on Mondays while on Fridays an opposite trend is observed. Based on the estimates, month of March is associated with largest impact on boarding and alighting. Local highway length and number of bus stop for a 1500m buffer area around rail stations presents a significant impact on boarding and alighting. The availability of free parking space at SunRail stations also significantly affect both boarding and alighting ridership. Land use variables including presence of commercial centers, educational centers and financial centers within 1500 m distance from SunRail station have significant influence on ridership. The number of households with access to no vehicles in the 1500m buffer around the station is negatively associated with boarding and alighting. The average temperature variable indicates that with higher temperature, boarding and alighting are likely to be higher. On the other hand, higher average wind speed is associated with lower boarding and alighting. Rain occurrence discourages rail usage as indicated by the negative coefficient in boarding and alighting components. In estimating SunRail daily average ridership models (for boarding and alighting), we estimated several station specific unobserved effects at station, station-day and station-week level. Among different considered levels, we found that the station level effects have significant influence on both boarding and alighting components of ridership. The station specific standard deviation variables for boarding and alighting indicate that the daily average ridership may vary for different stations based on the unobserved effects. The model system developed will allow us to predict ridership for existing stations in the future as well as potential ridership for future expansion sites. Finally, a policy analysis was performed to demonstrate the implications of the developed models.

CHAPTER SEVEN: CONTROLLING FOR ENDOGENEITY BETWEEN BUS HEADWAY AND BUS RIDERSHIP

7.1 Introduction

According to 2016 American Community Survey data, transit mode only accounts for about 5% of the commute trips in the United States (ACS, 2016). In fact, in recent years, several urban transit systems have experienced declines in ridership (Gomez-Ibanez, 1996; Garrett and Taylor, 1999; Siddiqui 2018; Bliss 2017; Schmitt 2017; Lewyn 2018). Ideally, in the presence of a well-designed public transit system, urban residents irrespective of their ethnicity, household income, and vehicle ownership should have similar access to activity participation opportunities or employment opportunities. Several researchers have found evidence to the contrary while examining the influence of transportation on employment opportunities (e.g., Shen, 2001; Wenglenski and Orfeuil, 2004; Kawabata and Shen, 2006, 2007; Grengs, 2010; Boarnet et al., 2017). These studies identified that access to employment by transit is substantially lower than access to employment by car mode. However, several public transit riders own no cars and are reliant on public transportation to arrive at work. Existing public transportation systems are either facing ridership declines and/or facing challenges with regards to providing equitable services to residents. In urban regions, public transportation systems ought to provide an equitable, safe and accessible transportation mode for residents. Thus, there is a need to examine public transportation system design and operations to enhance transit adoption and equity for urban residents.

Policy makers and urban agencies across different parts of North America, are considering investments in various public transportation alternatives including bus, light rail, commuter rail, and metro (see TP, 2016 for public transportation projects under construction or consideration). A critical component to evaluating the success of these investments is the development of appropriate statistical tools to examine the impact. Our proposed research contributes to public transit literature by developing econometric models that consider the potential endogeneity of stop level headway in modeling ridership. To elaborate, earlier research in public transportation has identified headway (alternatively bus frequency) as one of the primary determinants affecting ridership. The stops with higher headway (lower frequency) between buses are likely to have lower ridership. While this is a perfectly acceptable conclusion, most (if not all) studies in public transit literature ignore that the stop level headway was determined (by choice) in response to expected ridership i.e. stops with lower headway were expected to have higher ridership numbers. In traditional ridership studies, this potential endogeneity is often neglected and headway is considered as an independent variable. The approach violates the requirement that the unobserved factors that affect the dependent variable do not affect the independent variable. If this is the case, the estimated impact of headway on ridership would be biased (potentially over-estimated). More importantly, the estimated impact of all other variables (such as land use factors, bus infrastructure) will also be biased (possibly under-estimated). Traditional ridership models also consider transit ridership at a single time point for analysis using cross-sectional datasets. Ideally, it would be beneficial to consider data from multiple time points. The consideration of data from multiple time points is of particular value in accommodating for the impact of headway associated endogeneity.

In this study, we address these challenges by proposing a simultaneous equation system that considers headway and ridership in a joint framework while accounting for the influence of common unobserved factors affecting headway and ridership. The proposed model is developed employing ridership data from Orlando region for the Lynx bus transit system. The ridership data includes stop level average weekday boarding and alighting information for 11 four-month time periods from May 2013 to December 2016. The presence of multiple data points for each stop allows us to develop panel models for headway, boarding and alighting. The headway variable is modeled using a panel ordered logit model while the ridership variables are modeled using panel group ordered logit models. In addition to unobserved effects in the form panel random effects, several exogenous variables including stop level attributes (such as number of bus stop), transportation infrastructure variables (such as secondary highway length, rail road length and local road length, sidewalk length), transit infrastructure variables (bus route length, presence of shelter and distance of bus stop from central business district (CBD)), land use and built environment attributes (such as land use mix, residential area, recreational area, institutional area, office area, etc.) and sociodemographic and socioeconomic variables in the vicinity of the bus stop (income, vehicle ownership, age and gender distribution) were considered in the model estimation. The model estimation results identify that headway, number of the bus stops in the 800m buffer, presence of shelter at the bus stop, sidewalk length in a 400m buffer, bus stop distance from the central business district (CBD), distance between Sunrail station and bus stop, and automobile ownership are likely to impact bus ridership in Orlando. The bus route length in an 800m buffer is negatively affected the bus ridership in Orlando which is opposite of author's earlier work (Rahman, et. al. 2017) because, in the earlier paper, endogeneity of headway in bus ridership was not considered but in this study, we have considered the endogeneity. This is a clear indication of the impact of the endogenous variable on the dependent variable.

7.2 Current Study in Context

The literature review highlights how well recognized the issue of endogeneity is within the transit filed. However, the literature is not without limitations. First, while several studies have explicitly considered/controlled for endogeneity the study frameworks focus on aggregate transit ridership metrics such as monthly boardings at the system level. There is no study that has examined the endogeneity issue at a more disaggregate level such as bus route or stop level. The aggregate level models are adequate for planning at a system level. However, for any analysis of changes to the existing service for various bus routes, more detailed analysis at the bus route or stop level is warranted. Second, earlier analysis was explored using cross-sectional or panel data with very small data samples. This is expected because the analysis was conducted at a system level yielding smaller data samples. Third, while several studies developed IV and/or 2SLS approaches there is no effort in the discrete choice realm controlling for endogeneity. The current research effort addresses these limitations by undertaking a disaggregate stop level ridership analysis (for boarding and alighting) while controlling for endogeneity associated with stop-level headway. For the Orlando region, while headway is a continuous value in minutes, due to the nature of the service in the region, it is more accurate to consider headway as a discrete variable. In our study, we have considered three categories for headway model: (i) Category 1 (0-15 minutes), (ii) Category 2 (15-30 minutes) and (iii) Category 3 (>30 minutes). Hence, we have considered headway as an ordered discrete variable. Further, to model ridership, building on our earlier work (Rahman et al., 2017), we categorize the boardings and alightings as grouped ordered variables. Thus, the overall econometric methodology employed results in a panel multivariate ordered system with three separate equations (for headway, boarding and alighting). The proposed model system is estimated using data for eleven 4-month periods from May 2013 to December 2016. The proposed joint panel modeling approach is the first of its kind for transit ridership analysis to the best of the author's knowledge.

7.3 Methodology

The focus of this study is to examine stop-level boarding, alighting and headway simultaneously. Let q (q = 1, 2,…, Q) be an index to represent bus stops, let t (t = 1, 2, 3,…, T)

represent the different time periods j (j = 1, 2, 3,…, J = *13*) be an index to represent the number of boardings or alightings and m (m = 1,2,…M=3) be an index to represent headway categories. The thirteen categories for ridership analysis are: Bin 1 = ≤5; Bin 2 = 5-10; Bin 3 = 10-20, Bin 4 = 20-30, Bin 5 = 30-40, Bin 6 = 40-50, Bin 7 = 50-60, Bin 8 = 60-70, Bin 9 = 70-80, Bin 10 = 80-90, Bin 11 = 90-100, Bin 12 = 100-120 and Bin 13= >120. For headway component, we consider three categories: category 1 = 0 to 15 minutes; category 2= 15 to 30 minutes and category 3= > 30minutes. Then, the equation system for modeling headway, boarding and alighting jointly can written as:

$$H^*_{qt} = \left(\nu' + \sigma'_q\right)x'_{qt} + (\eta'_q)y_{qt} + \Delta_{qt}, H_{qt} = m \;\; if \;\; \varpi_{m-1} < H^*_{qt} \leq \varpi_m \tag{19}$$

$$B^*_{qt} = \left(\alpha' + \gamma'_q\right)x''_{qt} + \left(\theta' + \mu'_q\right)h_{qt} \pm (\eta'_q)y_{qt} + \varepsilon_{qt}, B_{qt} = j \;\; if \;\; \psi_{j-1} < B^*_{qt} \leq \psi_j \tag{20}$$

$$A^*_{qt} = \left(\beta' + \delta'_q\right)x''_{qt} + \left(\theta'' + \mu''_q\right)h_{qt} \pm (\eta'_q)y_{qt} + \xi_{qt}, A_{qt} = j \;\; if \;\; \psi_{j-1} < A^*_{qt} \leq \psi_j \tag{21}$$

In equation 19, H^*_{qt} is the latent (continuous) propensity for headway at stop q for the t^{th} time period. This latent propensity H^*_{qt} is mapped to the actual grouped headway category m by the ϖ thresholds, in the usual ordered-response modeling framework. x'_{qt} is a matrix of attributes that influences stop level headway, ν is the vector of mean coefficients and σ_q is a vector of coefficients representing the impact of unobserved factors moderating the influence of corresponding element of x'_{qt}.

In equations 20 and 21, B^*_{qt} (A^*_{qt}) is the latent propensity for stop level boardings (alightings) of stop q for the t^{th} time period. This latent propensity B^*_{qt} (A^*_{qt}) is mapped to the actual grouped ridership category j by the ψ thresholds, in the usual ordered-response modeling framework. In our case, we consider J = 13 and thus the ψ values are as follows: -∞, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, and +∞.x''_{qt} is a matrix of attributes that influences stop level boarding and alighting. ; α (β)is the corresponding vector of mean coefficients and $\gamma_q(\delta_q)$ is a vector of coefficients representing the impact of unobserved factors moderating the influence of corresponding element of x'_{qt} (x''_{qt}) for boardings (alightings), h_{qt} represents the headway variables generated from H_{qt} for consideration in boarding and alighting. θ' (θ'') represents the corresponding vector of mean coefficients and μ'_q (μ''_q) is a vector of coefficients representing the impact of unobserved factors moderating the influence of corresponding element h_{qt} for boardings (alightings). ε_{qt} (ξ_{qt}) is an idiosyncratic random error term assumed independently logistic distributed across choice stops and choice occasions for boardings (alightings) with variance λ_B^2 (λ_A^2).

η_q present in all three equations represents the vector of coefficients that accommodates for the impact of stop level common unobserved factors that jointly influence boardings, alightings and headway. The $' \pm '$ sign indicates the potential impact could be either positive or negative. A positive sign implies that unobserved factors that increase the headway for a given reason will also increase the propensity for boarding/alighting, while a negative sign suggests that unobserved individual factors that increase the propensity for headway will decrease the propensity for boarding/alighting. In our empirical context, we expect the relationship to be positive.

Further, to accommodate for ridership category specific effects z_{qjt} is a vector of attributes specific to stop q and ridership category alternative j and ρ_j is the vector of corresponding ridership category-specific coefficients.

To complete the model structure of the Equations (19), (20) and (21), it is necessary to define the structure for the unobserved vectors γ_q, δ_q, σ_q, μ_q (combined vector ofμ'_q and μ''_q and

η_q. In this paper, we assume that the three vectors are independent realizations from normal distributions as follows: $\gamma_{qn} \sim N(0,\kappa_n^2)$ $\delta_{qn} \sim N(0,\nu_n^2)$, $\sigma_{qn} \sim N(0,\varsigma_n^2)$ $\mu_{qn} \sim N(0,o_n^2)$ and $\eta_{qn} \sim N(0,\varrho_n^2)$.

With these assumptions, the probability expressions for the ridership category may be derived. Conditional on γ_{qn}, δ_{qn}, σ_{qn}, μ_{qn} and η_{qn}, the probability for stop q to have boarding, alighting and headway in the t^{th} time period is given by:

$$P(H_{mt})|\sigma,\eta = \Lambda\left[\varpi_m - \left((\nu' + \sigma_q')x'_{qt} + (\eta'_q)y_{qt}\right)\right] - \Lambda\left[\varpi_{m-1} - \left((\nu' + \sigma_q')x'_{qt} + (\eta'_q)y_{qt}\right)\right] \tag{22}$$

$$P(B_{jt})|\gamma,\eta = \Lambda\left[\frac{\psi_j - \left((\alpha'+\gamma_q')x''_{qt} + (\rho_j')z_{qjt} + (\theta'+\mu'_q)h_{qt} \pm (\eta'_q)y_{qt}\right)}{\lambda_B}\right] - \Lambda\left[\frac{\psi_{j-1} - \left((\alpha'+\gamma_q')x''_{qt} + (\rho_j')z_{qjt} + (\theta'+\mu'_q)h_{qt} \pm (\eta'_q)y_{qt}\right)}{\lambda_B}\right] \tag{23}$$

$$P(A_{jt})|\delta,\eta = \Lambda\left[\frac{\psi_j - \left((\beta'+\delta_q')x''_{qt} + (\tau_j')z_{qjt} + (\theta''+\mu''_q)h_{qt} \pm (\eta'_q)y_{qt}\right)}{\lambda_A}\right] - \Lambda\left[\frac{\psi_{j-1} - \left((\beta'+\delta_q')x''_{qt} + (\tau_j')z_{qjt} + (\theta''+\mu''_q)h_{qt} \pm (\eta'_q)y_{qt}\right)}{\lambda_A}\right] \tag{24}$$

where Λ (.) is the cumulative standard logistic distribution. z_{qjt} is a vector of attributes specific to stop q and ridership category alternative j , while ρ_j and τ_j is the vector of corresponding Ridership category-specific coefficients for boarding and alighting components, respectively.

Let Ω represent a vector that includes all the standard error parameters to be estimated. Given these assumptions the joint likelihood for stop level boarding and alighting is provided as follows:

$$L_q|\Omega = \prod_{t=1}^{T}\left[\prod_{m=1}^{M}[P(H_{mt})|\sigma,\eta]^{d_{hmt}} * \left\{\prod_{j=1}^{J}\left[\left(P(B_{jt})|\gamma,\eta\right)\right]^{d_{bjt}}\left[\left(P(A_{jt})|\delta,\eta\right)\right]^{d_{ajt}}\right\}\right] \tag{25}$$

where d_{hmt} is a dummy variable taking a value of 1 if stop q has headway within the m^{th} category for the t^{th} time period and 0 otherwise; d_{bjt}, and d_{ajt} are dummy variables taking a value of 1 if stop q has ridership within the j^{th} category for the t^{th} time period and 0 otherwise. Finally, the unconditional likelihood function may be computed for stop q as:

$$L_q = \int_{\Omega} \left(L_q|\Omega\right) f(\Omega) d\Omega \tag{26}$$

The log-likelihood function is given by

$$\text{Ln(L)} = \sum_{q=1}^{Q} \ln L_q \tag{27}$$

The likelihood function in Equation (27) involves the evaluation of a multi-dimensional integral of size equal to the number of rows in Ω. We apply Quasi-Monte Carlo simulation techniques based on the scrambled Halton sequence to approximate this integral in the likelihood function and maximize the logarithm of the resulting simulated likelihood function (See Bhat, 2001; Yasmin and Eluru, 2013 for more details). The likelihood functions are programmed in Gauss (Aptech 2016).

7.4 Model Specification and Overall Measures of Fit

The empirical analysis involves estimation of different models: 1) Independent ridership-headway (IRH) model that does not accommodate for headway endogeneity and 2) Joint ridership-headway (JRH) model that explicitly accommodates for headway endogeneity. Prior to discussing the estimation results, we compare the performance of these models in this section. We employ the Bayesian Information Criterion (BIC) to determine the best model between independent and joint model. The BIC for a given empirical model is equal to:

$$BIC = -2LL + K\, ln(Q) \quad (28)$$

where LL is the log likelihood value at convergence, K is the number of parameters, and Q is the number of observations. The model with the lower BIC is the preferred model. The log-likelihood values at convergence for the models estimated are as follows: (1) Independent ridership-headway (IRH) model (with 55 parameters) is -110,705.364 (2) Joint ridership-headway (JRH) model (with 49 parameters) is -105,059.724. The BIC values for the final specifications of IRH and JRH are 221,979.168 and 210,625.876 respectively. The comparison exercise clearly highlights the superiority of the Joint ridership headway (JRH) in terms of data fit compared to independent ridership-headway (IRH) model.

7.5 Variable Effects

The final specification of the model was based on by removing the statistically insignificant variables at 95% confidence level, which was determined by prior research and knowledge. In this research, various buffer sizes (800m, 600m, and 400m buffer size) were considered during analysis and best fitted buffer size was taken into consideration for the final model. In presenting the effects of the exogenous variables, we will restrict ourselves to the discussion of the joint model. Table 13 presents the estimation results of the joint model. Specifically, columns 2 and 3 provide the variable impacts of the headway component while columns 4 through 7 present the results of boarding and alighting components. The model results are discussed by model component.

Table 13. Group Ordered Logit Model Results

Variable Name	Headway Model		Alighting Model		Boarding Model	
	Estimates	t-stat	Estimates	Estimates	Estimates	t-stat
Constant	-	-	-8.439	-10.286	-20.193	-20.379
Threshold Value 1	-3.889	-73.979	-	-	-	-
Threshold Value 2	0.399	7.916	-	-	-	-

Variable Name	Headway Model		Alighting Model		Boarding Model	
	Estimates	t-stat	Estimates	Estimates	Estimates	t-stat
Stop Level Attributes						
Headway (Base: Category 1) *Dummy for headway category 2*	-	-	-49.429	-107.635	-54.287	-106.974
Dummy for headway category 3	-	-	-80.448	-153.226	-86.460	-147.837
No of Bus stop in a 800m buffer	-	-	-4.382	-28.617	-4.411	-25.989
Presence of shelter in bus stop	-	-	19.677	74.191	34.034	109.754
Bus route Length in an 800m buffer	-0.820	-71.485	-2.649	-17.144	-3.932	-23.983
Transportation Infrastructures						
Side walk length in an *400 m buffer*	-	-	2.698	14.783	2.642	13.108
Bike road length in an *800 m buffer*	-0.203	-26.537	-	-	-	-
Secondary road length in an *800 m buffer*	-0.517	-39.033	7.900	36.461	5.941	25.169
Local road length in an *800 m buffer*	0.298	20.398	5.082	24.659	5.150	21.397
Rail road length in an *800 m buffer*	-0.627	-52.824	-	-	-	-
Built environment and land use attributes						
Land use area type in an 800m buffer						
Institutional area	-1.810	-17.247	24.727	13.257	6.155	2.768
Residential area	1.821	32.010	-	-	17.128	20.615
Office area	-1.952	-24.983	39.989	42.699	42.241	31.761
Recreational area	-0.517	-2.380	-75.610	-25.432	-64.925	-19.209
Industrial Area	5.260	42.726	-	-	-	-
Central business district (CBD) distance	0.502	45.345	-2.884	-15.057	-2.993	-14.496
Sociodemographic and socioeconomic variables						
Zero vehicle in HH	-2.174	-14.200	75.952	28.658	72.740	24.276
High income population	-0.304	-4.244	-	-	-	-
Household rent	-	-	31.596	48.830	35.839	49.835
SunRail effects						
Distance Decay Function for SunRail*SunRail operation period	-	-	-5.367	-19.593	-5.188	-17.740
Variance						
Constant – Alternative 1 (0-5 ridership)	-	-	37.550	124.964	42.178	123.004
Constant – Alternative 2 (5-10 ridership)	-	-	17.905	82.714	20.074	82.247
Scale parameter						
Scale variables	-	-	3.270	752.608	3.343	707.846
Random Effect						
Constant	1.726			154.121		
Route Length in 800m buffer	0.777			102.920		

7.5.1 Headway Components:

The positive (negative) coefficient corresponds to increased (decreased) proportion for headway categories.

7.5.1.1 Transportation Infrastructure Characteristics

The bus route length of 800m buffer has a negative impact on headway. The variable impact is expected. Bus stops with larger bus route length are likely to have higher frequency of bus arrivals i.e. lower headway. A negative impact of the presence of bike length in 800m vicinity of the bus stop on headway is also along expected lines. The presence of bicycle infrastructure serves as a proxy for denser neighborhoods encouraging non-automobile alternatives. The presence of increased secondary highway length in the 800m buffer decreases the headway while a corresponding increase in local road length increases headway. The roadway length variable is possibly serving as an indicator of type of urban locations. The results also indicate that in the presence of a rail road headway is likely to be lower. The result warrants further investigation.

7.5.1.2 Built Environment Attributes

The built environment around a bus stop has a significant impact on bus frequency. The presence of industrial and residential areas within a 800m buffer of a bus stop is likely to increase the headway. On the other hand, in the presence of institutional, recreational and office area (800m buffer) the headway is likely to be lower. The results are intuitive. An increase in the stop distance from the central business district (CBD) is likely to increase the headway (as expected).

7.5.1.3 Demographic and Socioeconomic Characteristics

In terms of demographic and socioeconomic variables vehicle ownership variable has a significant impact. Specifically, locations with higher proportion of households with no vehicle are likely to have a lower headway value. The result is symptomatic of households with no vehicles being captive to transit mode.

7.5.2 Boarding and Alighting components:

7.5.2.1 Stop Level Attributes

Headway (here headway category headway) at the stop level has a significant impact on ridership (as expected). By increasing the headway, the boarding and alighting ridership are likely to decrease. This result indicates that if the frequency of the bus increases in stop level than the ridership of that stop leads to higher ridership. If there is higher demand of bus in a stop, it is likely to increases the bus frequency as well to accommodate the demand. The results for the number of the bus stop in the 800m buffer presented that if the number of bus stops increasing in the 800m buffer of a stop than the ridership will reduce at that stop which supports author earlier work (see Rahman et. Al., 2017). The main reason may be the bus spend more time for boarding/alighting and red lights and there might be some competition among the stop. A study (El-Geneidy, et. Al., 2005) found that by merging nearby stops is nearly increased 6 percent bus speeds and also increased the ridership. By prioritizing which bus stop should stay (considering high ridership, locations), Transit center can improve the ridership at that location. The presence of shelter at the bus stop also increases the ridership in Orlando. Waiting for the bus can be brutal as it tricks passenger about the actual time they are waiting for the bus. By having shelters in bus stop can do the opposites and thus people feel more satisfied when they have shelters at the bus stop (Jaffe, 2014).

7.5.2.2 Transportation Infrastructure Characteristics

The bus route length of 800m buffer has a negative effect on both boarding and alighting ridership which is expected but in auther earlier works this impact came positive because in the earlier works, we did not considered the endogeiety of the headway on bus ridership. The presence of headway variables directly at exogenous variables impact the effect of the bus route length of

800m variable effect. Bus stops with larger route length are likely to be lower headway value as well as Lynx does not have any stop along the interstate and also for increasing the unlinked trips. A positive impact on sidewalk length of the 400m buffer of the stop found for both boarding and alighting ridership in Orlando. By improving the pedestrian facility, walkability and safety, people are willing to ride on the bus and thus increasing the ridership. Along with the sidewalk, local road and secondary highway in 800m buffer are also increasing the ridership as a Lynx bus authority does not provide any stop along the major highway (Interstate and Expressway).

7.5.2.3 Built Environment Attributes

The built environment around a bus stop has a significant influence on bus ridership at the stop level. The presence of office area and the institutional area in 800m buffer within a stop significantly increase the bus ridership in Orlando. The presence of school/college and office helps people to take a bus rather than taking automobile as huge traffic congestion during School/college time and morning and an evening pick hour in Orlando. The proportion of residential area has positive effects on boarding ridership of 800m buffer, but no impact on alighting ridership. On the other hand, the presence of recreation area within a 800m buffer of a stop is decreasing the bus ridership as people usually take their bike/automobiles/family car to go to recreation center rather than taking a bus. The distance from the central business district (CBD) from bus stop negatively impacts the bus ridership as the distance from CBD increases, the bus ridership will reduce (expected outcome). The sum of squares distance inverse from Sunrail station to bus stop also negatively impacts the bus ridership as the distance increased then the multimodal facility and scope are decreasing thus that reduce the bus ridership as well. If the connecting between the Lynx bus and Sunrail improve then the bus ridership is more likely improve in this region.

7.5.2.4 Demographic and Socioeconomic Characteristics

The demographic and socioeconomic variables based on census tract of the bus stop significantly affect the bus ridership in Orlando. The increased share of the household renters in Orlando is likely to increase the bus ridership. The automobile ownership also positively impacts the bus ridership. People having no vehicle in their household merely increase the bus ridership as expected as the bus or public transit is the only feasible solutions for them to commute.

7.6 Summary

In urban regions, public transportation systems ought to provide an equitable, safe and accessible transportation mode for residents. Thus, there is a need to examine public transportation system design and operations to enhance transit adoption and equity for urban residents. Policy makers and urban agencies across different parts of North America, are considering investments in various public transportation alternatives including bus, light rail, commuter rail, and metro. A critical component to evaluating the success of these investments is the development of appropriate statistical tools to examine the impact. Our proposed research contributes to public transit literature by developing econometric models that consider the potential endogeneity of stop level headway in modeling ridership. Most (if not all) studies in public transit literature ignore that the stop level headway was determined (by choice) in response to expected ridership i.e. stops with lower headway were expected to have higher ridership numbers. In traditional ridership studies, this potential endogeneity is often neglected and headway is considered as an independent variable. The approach violates the requirement that the unobserved factors that affect the dependent variable do not affect the independent variable. If this is the case, the estimated impact of headway on ridership would be biased (potentially over-estimated). More importantly, the estimated impact of all other variables (such as land use factors, bus infrastructure) will also be biased (possible under-estimated).

In this study, we address these challenges by proposing a simultaneous equation system that considers headway and ridership in a joint framework that accounts for the influence of common unobserved factors that affect headway and ridership. The proposed model is developed employing ridership data from Orlando region from the Lynx bus transit system. The ridership data includes stop level average weekday boarding and alighting information for 11 four-month time periods from May 2013 to December 2016. The presence of multiple data points for each stop allows us to develop panel models for headway, boarding and alighting. The headway variable is modeled using a panel ordered logit model while the ridership variables are modeled using panel group ordered logit models. In addition to unobserved effects in the form panel random effects, several exogenous variables including stop level attributes (such as number of bus stop), transportation infrastructure variables (such as secondary highway length, rail road length and local road length, sidewalk length), transit infrastructure variables (bus route length, presence of shelter and distance of bus stop from central business district (CBD)), land use and built environment attributes (such as land use mix, residential area, recreational area, institutional area, office area, etc.) and sociodemographic and socioeconomic variables in the vicinity of the bus stop (income, vehicle ownership, age and gender distribution) were considered in the model estimation.

The model estimation results identify that headway, number of the bus stops in the 800m buffer, presence of shelter at the bus stop, sidewalk length in a 400m buffer, bus stop distance from the central business district (CBD), distance between Sunrail station and bus stop, and automobile ownership are likely to impact bus ridership in Orlando. The bus route length in an 800m buffer is negatively affected the bus ridership in Orlando which is opposite of author's earlier work because, in the earlier paper, endogeneity of headway in bus ridership was not considered but in this study, we have considered the endogeneity. This is a clear indication of the impact of the endogenous variable on the dependent variable. In our research, in order to highlight the effect of various attributes over time on boarding and alighting ridership, an elasticity analysis was also presented. We investigated the change in ridership due to the change in selected independent variables. The elasticity analysis highlights a worrisome trend of reducing transit ridership with time. Significant investments in transit infrastructure can arrest this trend.

To be sure, the research is not without the limitations. We examined the effect of headway variables and endogeneity of headway on bus ridership. However, we just compute the endogeneity of headway on bus ridership, it will be interesting to consider another variable that might be endogenous with bus ridership.

CHAPTER EIGHT: COST BENEFIT ANALYSIS OF SUNRAIL

8.1 Introduction

The objective of this chapter is to document and present the cost-benefit analysis (CBA) of the recently added SunRail transit system in Orlando. Transit systems are an integral part of the development of a community. But comprehensive benefits of these systems often are not estimated or remain unmeasured. Though the capital cost of developing a transit system is significantly higher, total benefits accrued from a transit system operation in the long run is likely to surpass the higher investment cost. CBA is considered to be one of the most appropriate tools in evaluating net benefits of a transportation system (Litman, 2001). With the focus of encouraging more people to use sustainable transportation alternatives, FDOT is constructing a new, 17.2-mile extension to the existing 31-mile SunRail commuter rail. A comprehensive CBA of the existing operational SunRail system would assist planners and policy makers to evaluate the "real" benefit of these investments and provide evidence to justify allocation of more funding for improving/building transit infrastructures. To that extent, in this research effort, we present and discuss CBA result for the existing 31-mile SunRail system.

8.2 Cost-Benefit Analysis for Sunrail

SunRail is in operation since May 2014 in greater Orlando. The existing operational SunRail system comprises of 31-mile rail length along with 12 active stations - Sand Lake Station, Amtrak Station, Church Street Station, Lynx Central Station, Florida Hospital Station, Winter Park Station, Maitland Station, Altamonte Springs station, Longwood Station, Lake Mary Station, Sanford Station and Debary Station. In this research effort, we focus on this existing SunRail system for the CBA. We projected cost and benefit for 30 years (from 2014 to 2044) considering 2014 as base year.

8.2.1 Factors Considered

The potential cost-benefit components of SunRail is identified based on literature review and the components identified in Task 1. With regards to cost component, the factors we consider included: (1) capital costs and (2) operation and maintenance costs. In terms of the benefit component, the factors we consider included: (1) personal automobile cost savings, (2) crash cost savings, (3) parking cost savings, (4) energy conservation savings, and (5) assessed property value increase. In the current study context, we assume that SunRail trips has an impact on personal automobile mode only. However, SunRail could have potential impact on individuals using other modes including bus, walk or bike. However, in computing benefits, we assume that SunRail trip would have negligible effect on other modes since we did not have information on actual modal shifts that may have induced by SunRail.

8.2.2 Demand Attributes

Transit demand attributes (such as ridership, passenger miles travelled, frequencies, headway etc.) determine the magnitude of benefits from any transit investments as these attributes represents the demand and efficiency of the system. Therefore, the first step of CBA is to identify these demand attributes. In this research effort, we compute the benefit factors as function of daily ridership, passenger miles travelled and train frequency. In this section, we describe the procedure for computing these attributes.

Daily Ridership

For the purpose of identifying average daily ridership of SunRail at a system-level, we have compiled stop level daily boarding and alighting ridership data for ten months from January 2015 to October 2015. The daily ridership data includes weekdays only as SunRail did not operate during weekends over the data collection period. The 10-month, 12 station data provided us 2,496 observations. A summary of the system level ridership (boarding and alighting) is provided in Table 14. From Table 14, we can see that the average daily system-level ridership is 3,693.163. Therefore, for the current study, we consider an average daily ridership of 3,700 at a system-level for computation of benefit factors.

Table 14. Summary Statistics for SunRail Average Daily Ridership (January 2015 to October 2015)

Station Name	Mean	
	Boarding	**Alighting**
Sand Lake Station	451.168	82.127
Amtrak Station	124.260	20.507
Church Street Station	393.135	79.184
Lynx Central Station	403.769	35.282
Florida Hospital	201.976	26.562
Winter Park Station	411.707	205.107
Maitland Station	180.962	27.084
Altamonte Springs station	244.163	40.788
Longwood Station	240.909	36.959
Lake Mary Station	337.005	55.139
Sanford Station	258.952	45.735
Debary Station	445.178	90.608
Total	**3,693.183**	**3,693.183**

Passenger Miles Travelled

For the purpose of identifying passenger miles travelers, we selected station level ridership for a random day. From the stop-level daily ridership information including boarding and alighting, we computed the train occupancy between stations. The occupancy and station to station distance was employed to generate person level mileage on the system. Table 15 represents the passenger miles travelled computation details. From Table 15, we can see that on an average a passenger travelled about 16.57 miles by using SunRail on a typical weekday. Therefore, we have considered 17 miles as average passenger miles travelled for computation of benefit factors.

Table 15. Passenger Miles Travelled Calculations for SunRail

SOUTHBOUND							
No.	**Stations**	**Distance from station to station (miles)**		**Number of passenger**			**Total passenger miles (Remained boarded*Distance from station to station)**
				Boarded	**Alighted**	**Remained boarded**	
1	DeBary Station	1-2	5	451	0	451	2255.00
2	Sanford Station	2-3	4.5	253	15	689	3100.50
3	Lake Mary Station	3-4	5.5	331	18	1002	5511.00
4	Longwood Station	4-5	3	207	39	1170	3510.00
5	Altamonte Springs Station	5-6	3	167	72	1265	3795.00
6	Maitland Station	6-7	3.5	129	42	1352	4732.00
7	Winter Park Station	7-8	2.5	152	266	1238	3095.00
8	Florida Hospital Station	8-9	2.3	70	157	1151	2647.30
9	Lynx Central Station	9-10	0.7	64	322	893	625.10
10	Church Street Station	10-11	1.2	46	299	640	768.00
11	AMTRAK Station	11-12	5.7	13	118	535	3049.50
12	Sand Lake Road Station	--	---	0	535	---	---
	Total Southbound			1883	1883		33088.40
NORTHBOUND							
No.	**Stations**	**Distance from station to station (miles)**		**Number of passenger**			**Total passenger miles (Remained boarded*Distance from station to station)**
				Boarded	**Alighted**	**Remained boarded**	
1	Sand Lake Station	1-2	5.7	395	0	395	2251.50
2	Amtrak Station	2-3	1.2	109	13	491	589.20
3	Church Street Station	3-4	0.7	326	41	776	543.20
4	Lynx Central Station	4-5	2.3	343	62	1057	2431.10
5	Florida Hospital	5-6	2.5	139	86	1110	2775.00

Table 15 (Continued): Passenger Miles Travelled Calculations for SunRail

NORTHBOUND								
No.	**Stations**	**Distance from station to station (miles)**		**Number of passenger**			**Total passenger miles (Remained boarded*Distance from station to station)**
				Boarded	**Alighted**	**Remained boarded**	
6	Winter Park Station	6-7	3.5	243	175	1178	4123.00
7	Maitland Station	7-8	3	48	153	1073	3219.00
8	Altamonte Springs station	8-9	3	92	177	988	2964.00
9	Longwood Station	9-10	5.5	41	203	826	4543.00
10	Lake Mary Station	10-11	4.5	17	314	529	2380.50
11	Sanford Station	11-12	5	10	235	304	1520.00
12	Debary Station	---	---	0	304	---	---
Total Northbound				1763	1763		27339.50
Total Passenger miles travelled		**33088.40 + 27339.50 = 60427.90**					
Average passenger miles travelled		**60427.90/(1883+1763) = 16.57**					

Train frequency

We identify train frequency based on SunRail train frequency operation. The frequency of SunRail is 18 in each direction, therefore, we consider train frequency as 36 per day (representing both direction run) for computation of benefit factors.

8.3 Cost Factors

In our current study, we consider two cost factors: (1) capital costs, and (2) operation and maintenance costs. Capital costs include costs for planning, design and constructing the infrastructure for SunRail operation along with costs for buying the trains. Operation and maintenance costs include compensation cost of train operators, operation and maintenance personnel, electricity bills, buying replacement parts, supplies from vendors and other regular operation cost. For the current research purposes, we consider SunRail capital costs as $615 million. In terms of operation and maintenance costs, we consider it as $34.4 million for the base year (sourced from FDOT, 2016; FDOT, 2017). For 30 year cost projection, we assume an increase rate of 2.8% per year in computing operation and maintenance cost.

8.4 Benefit Factors

8.4.1 Personal Automobile Cost Savings

Personal automobile cost (PAC) savings refers to the cost saving to riders due to the shift from personal automobile to transit mode. There are marginal costs associated with driving a personal vehicle in terms of fuel usage, depreciation, insurance, maintenance, parking cost and vehicle ownership cost. By shifting from driving to transit, travelers are likely to reduce their annual transportation costs related to owning and operating a personal vehicle. In fact, Litman (2004) computed the savings to be $1,300 per household in cities with established rail transit system. Thus, there is likely to be cost savings for train riders from reduced personal automobile usage. For our current research purpose, we assume PAC savings to be \$0.65 per vehicle-mile (AAA, 2013). The value is identified by assuming that a vehicle is driven approximately 15,000 miles per year and the cost includes operating (gas, maintenance, and tires) and ownership (insurance, depreciation, license, registration, taxes, and finance charge) components of driving personal automobile. Further, in identifying PAC savings per person, we assume that the average occupancy of a vehicle is 1.67 (NHTSA, 2011). Thus, the PAC cost savings is computed for a person as $\frac{\$0.65}{1.67\ person-mile}$. Table 16 provides our estimates of per year PAC savings of SunRail.

Table 16. Personal Automobile Cost Savings

Cost category	**Unit cost (\$/rider-miles)**	**Average train-miles travelled (miles/rider-day)**	**Personal automobile cost savings (\$/rider-day)**
Personal automobile cost savings	$\frac{0.65}{1.67}$	17	$\frac{0.65}{1.67} * 17$
Total personal automobile savings $\left(\frac{\$}{\text{year}}\right) = \frac{\mathbf{0.65}}{\mathbf{1.67}} * \mathbf{17} * \mathbf{3700} * (\mathbf{5} * \mathbf{52}) = \$\mathbf{6,365,329.34}$			

Note: $(5 * 52)$ *represents 5 days of the week and 52 weeks operation period of SunRail per year*

8.4.2 Crash Cost Savings

In general, public transportation has better safety record per unit of travel relative to passenger vehicle. As documented by Litman (2014), death rate of commuter rail from road traffic crashes is 0.43 per billion passenger mile, while the crash rate for passenger vehicle is 7.28. The value clearly signify the benefit of transit mode in terms of road safety. In our current research effort, we compute the crash cost savings of SunRail by subtracting SunRail crash cost from the automobile crash cost for trips to reflect the net benefit of replacing automobile trips with transit mode. For computing crash cost savings, we assume crash cost of automobile as \$0.10 per vehicle mile and crash cost of SunRail as (\$0.258 (external risk)+0.05*occupant(internal risk)) per vehicle mile. Table 17 provides our estimates of per year crash cost savings of SunRail (following Litman, 2012).

Table 17. Crash Cost Savings

<table>
<tr><th>Cost category</th><th>Unit cost
($/rider-miles)</th><th>Average train-miles travelled
(miles/rider-day)</th><th>Automobile crash cost
($/rider-day)</th></tr>
<tr><td>Automobile crash cost</td><td>$\frac{0.10}{1.67}$</td><td>17</td><td>$\frac{0.10}{1.67} * 17$</td></tr>
<tr><td colspan="4">$\mathbf{Total\ automobile\ crash\ cost} \left(\frac{\$}{\mathbf{year}}\right) = \frac{\mathbf{0.10}}{\mathbf{1.67}} * \mathbf{17} * \mathbf{3700} * \mathbf{5} * \mathbf{52} = \mathbf{\$954,910.18}$</td></tr>
<tr><td>Cost category</td><td>Train-miles (per day)</td><td>External cost
($/day)</td><td><b>Internal cost
($/day)</td></tr>
<tr><td>SunRail crash cost</td><td>$31 * 36$</td><td>$0.258 * 31 * 36$</td><td>$0.05 * 17 * 3700$</td></tr>
<tr><td colspan="4">$\mathbf{Total\ SunRail\ crash\ cost} \left(\frac{\$}{\mathbf{year}}\right) = (\mathbf{0.258} * \mathbf{31} * \mathbf{36} * \mathbf{20} + \mathbf{0.05} * \mathbf{17} * \mathbf{3700}) * \mathbf{5} * \mathbf{52} = \mathbf{\$866,452.72}$</td></tr>
<tr><td colspan="4">$\mathbf{Total\ crash\ cost\ savings} \left(\frac{\$}{\mathbf{year}}\right) = \mathbf{\$954,910.18} - \mathbf{\$866,452.72} = \mathbf{88457.46\$}$</td></tr>
</table>

8.4.3 Emission Cost Savings

One of the major benefits of transit over automobile is emission reduction benefits (Gallivan et al., 2015). Automobile and bus are likely to emit carbon monoxide, nitrogen dioxide, car dioxide and hydrocarbon in air. On the other hand, light rail is likely to produce 99% less hydrocarbons and carbon monoxide emissions per mile relative to that of automobile (Garrett, 2004). In our current study, we use air pollution cost as $0.08 per vehicle mile (Blonn et al., 2006), reflecting the fact that SunRail is located in urban area and the rail system also generates some air emissions. Thus, we compute emission cost savings as "change in automobile miles travelled*emission cost per automobile mile travelled". Table 18 provides our estimates of per year emission cost saving of SunRail.

Table 18. Emission Cost Savings

Cost category	Unit cost ($/rider-miles)	Average train-miles travelled (miles/rider-day)	Emission cost savings ($/rider-day)
Emission cost savings	$\frac{0.08}{1.67^{¥}}$	17	$\frac{0.08}{1.67} * 17$
$\textbf{Total emission cost savings} \left(\frac{\$}{\textbf{year}}\right) = \frac{\mathbf{0.08}}{\mathbf{1.67}} * \mathbf{17} * \mathbf{3700} * \mathbf{5} * \mathbf{52} = \$\mathbf{763928.14}$			

¥ average vehicle occupancy is considered as 1.67

8.4.4 Parking Cost Savings

Parking personal automobiles are often associated with cost of parking spaces and time spent to find the space. Unlike automobile mode, transit mode does not have parking cost associated with it (except park and ride option). In our current study, we compute parking cost savings for trip to reflect the net benefit of replacing automobile trips with transit mode. For computing cost savings, we assume parking cost of automobile as $0.36 per vehicle mile (following Litman, 2018). Table 19 provides estimates of per year parking cost savings of SunRail.

Table 19.Parking Cost Savings

Cost category	Unit cost ($/rider-miles)	Average train-miles travelled (miles/rider-day)	Parking cost savings ($/rider-day)
Parking cost savings	$\frac{0.36}{1.67^{¥}}$	17	$\frac{0.36}{1.67} * 17$
$\textbf{Total parking cost savings} \left(\frac{\$}{\textbf{year}}\right) = \frac{\mathbf{0.36}}{\mathbf{1.67}} * \mathbf{17} * \mathbf{3700} * \mathbf{5} * \mathbf{52} = \$\mathbf{3,467,376.65}$			

¥ average vehicle occupancy is considered as 1.67

8.4.5 Energy Conservation Savings

Transit mode can provide significant energy efficiency. Shapiro et al. (2002) found that an average automobile consumes about double the energy per passenger-mile travel relative to transit mode. In our current research effort, we use energy conservation savings as \$0.03 per vehicle miles (following Litman, 2018). Table 20 provides estimates of per year energy conservation cost savings of SunRail.

Table 20. Energy Conservation Savings

Cost category	Unit cost ($/rider-miles)	Average train-miles travelled (miles/rider-month)	Energy conservation savings ($/rider-month)
Energy conservation savings	$\frac{0.03}{1.67^{¥}}$	17	$\frac{0.03}{1.67} * 17$
$\textbf{Total energey conservatio savings} \left(\frac{\$}{\textbf{year}}\right) = \frac{\textbf{0.03}}{\textbf{1.67}} * \textbf{17} * \textbf{3700} * \textbf{5} * \textbf{52} = \textbf{286,473.05\$}$			

¥ average vehicle occupancy is considered as 1.67

8.4.6 Assessed Property Value Increase

Development of transit infrastructure increases overall accessibility which in turn is likely to increase land values around transit stops/stations. Moreover, higher accessibility attributable to transit development is likely to attract more economic development, higher active transportation friendly environment, more activities, higher density and mixed-use community development. Clearly, there are positive impacts of transit development on land use value. In our current study, we also consider the change in land use values surrounding the SunRail stations as one of the elements in benefit computation. In calculating the land use values, we consider assessed property value or just value as a surrogate measure of direct land use value. Just value (land just value, building value and special feature value) of a property includes: present cash value; use; location; quantity or size; cost; replacement value of improvements; condition; income from property; and net proceeds if the property is sold. The net proceeds equal the value of the property minus 15% of the true market value. This accounts for the cost of selling the property. In the following sections, we refer assessed property value as property value for simplicity.

To capture the change in property value, we collected and compiled parcel level data from Department of Revenue (DOR) for 2011 to 2016. The data has tax information of each parcel along with parcel boundaries from the Florida Department of Revenue's tax database. Each parcel polygon (Parcel ID) has information on property/feature value, land value, land area in square feet, owner name, owner address, physical address, physical zip code, building details and land use type. From the land use categories of parcel data, we have considered six major land use categories for identifying the impact of SunRail on property value change. The considered land use categories are: (1) Single family residential, (2) Multiple family residential, (3) Institutional, (4) Industrial, (5) Recreational and (6) Retail/Office area. For our current research, we assume that one mile buffer area around each SunRail station is the influence area of SunRail for property value impact computation. We labeled the parcels within the SunRail influence area as “Case Parcels”. For these case parcels, we computed property value by six land use types identified. To be sure, we have computed property value for case parcels from six years from 2011 to 2016. 2011 to 2013 period is considered to understand the change in property value before SunRail operation period, while 2014 to 2016 period shows the change in property value reflecting after SunRail operation period. Figure 5 and Figure 6 represent the spatial distribution of land use categories and property values for 2011 (before) and 2016 (after) within the SunRail influence area. From spatial representations, we can see that even though there are not much visible changes in land use categories from 2011 to 2016, the property values, on the contrary, have changed significantly after SunRail has become operational.

Year 2011	Year 2016

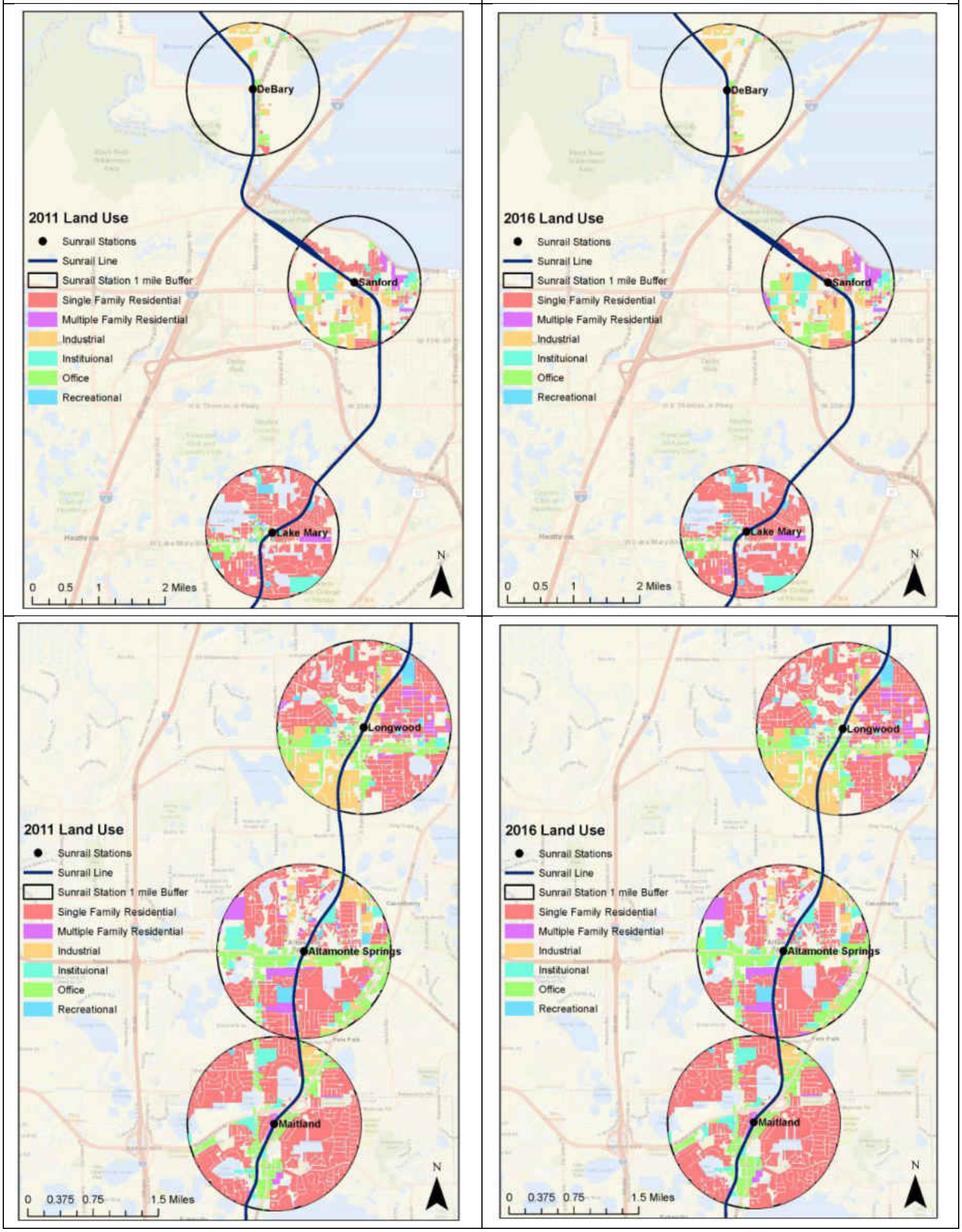
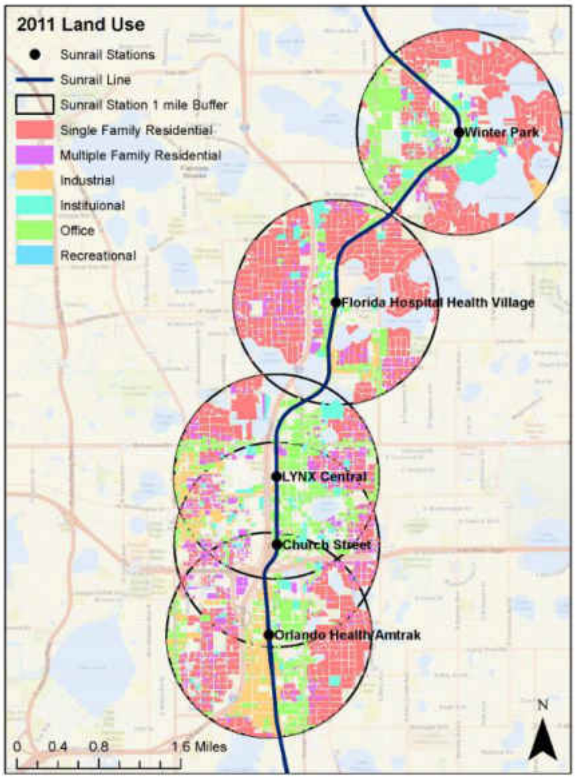
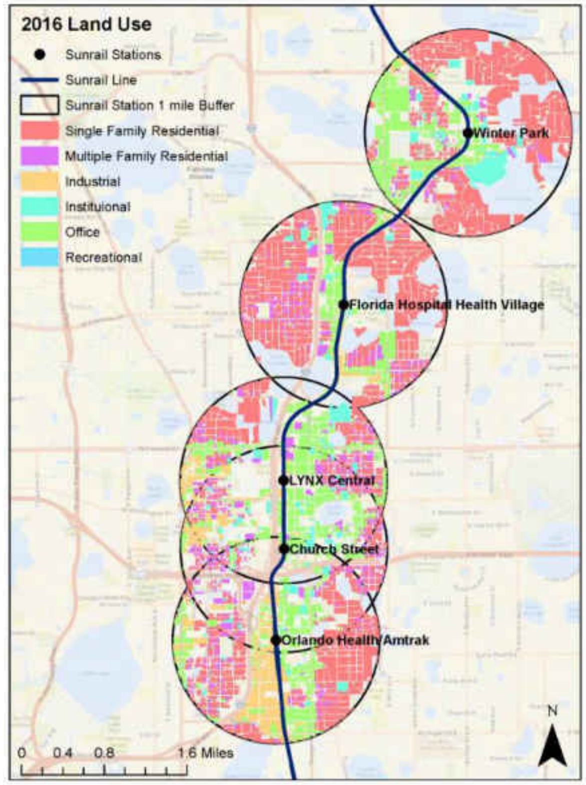
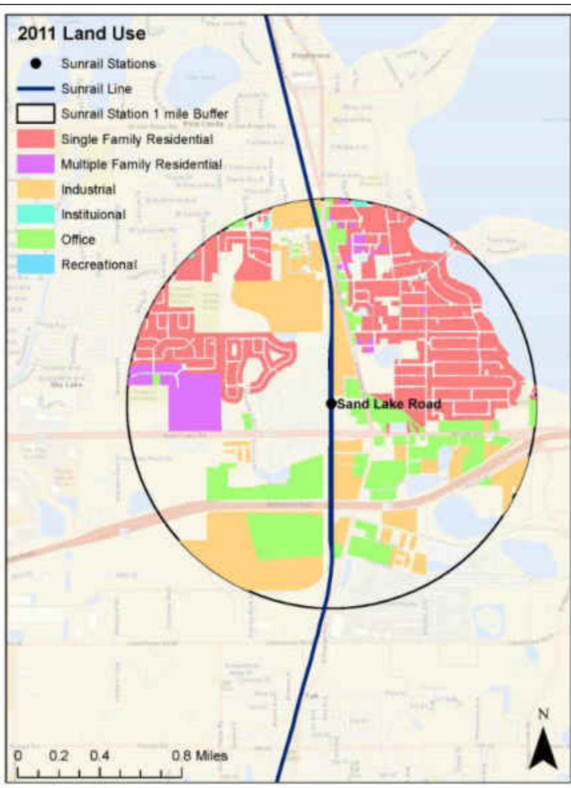


Figure 5. Land Use Types within SunRail Influence area for 2011 and 2016

Year 2011	Year 2016







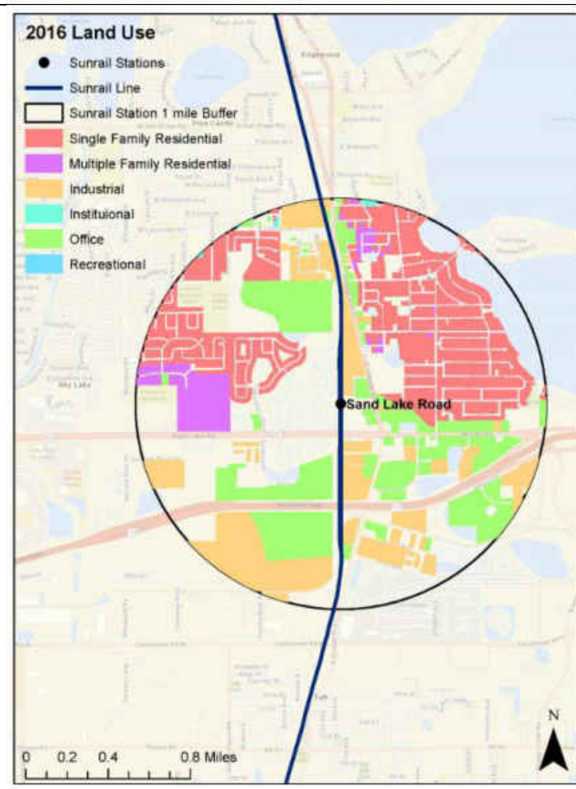


Figure 5 (Continued): Land Use Types within SunRail Influence area for 2011 and 2016

Year 2011	Year 2016

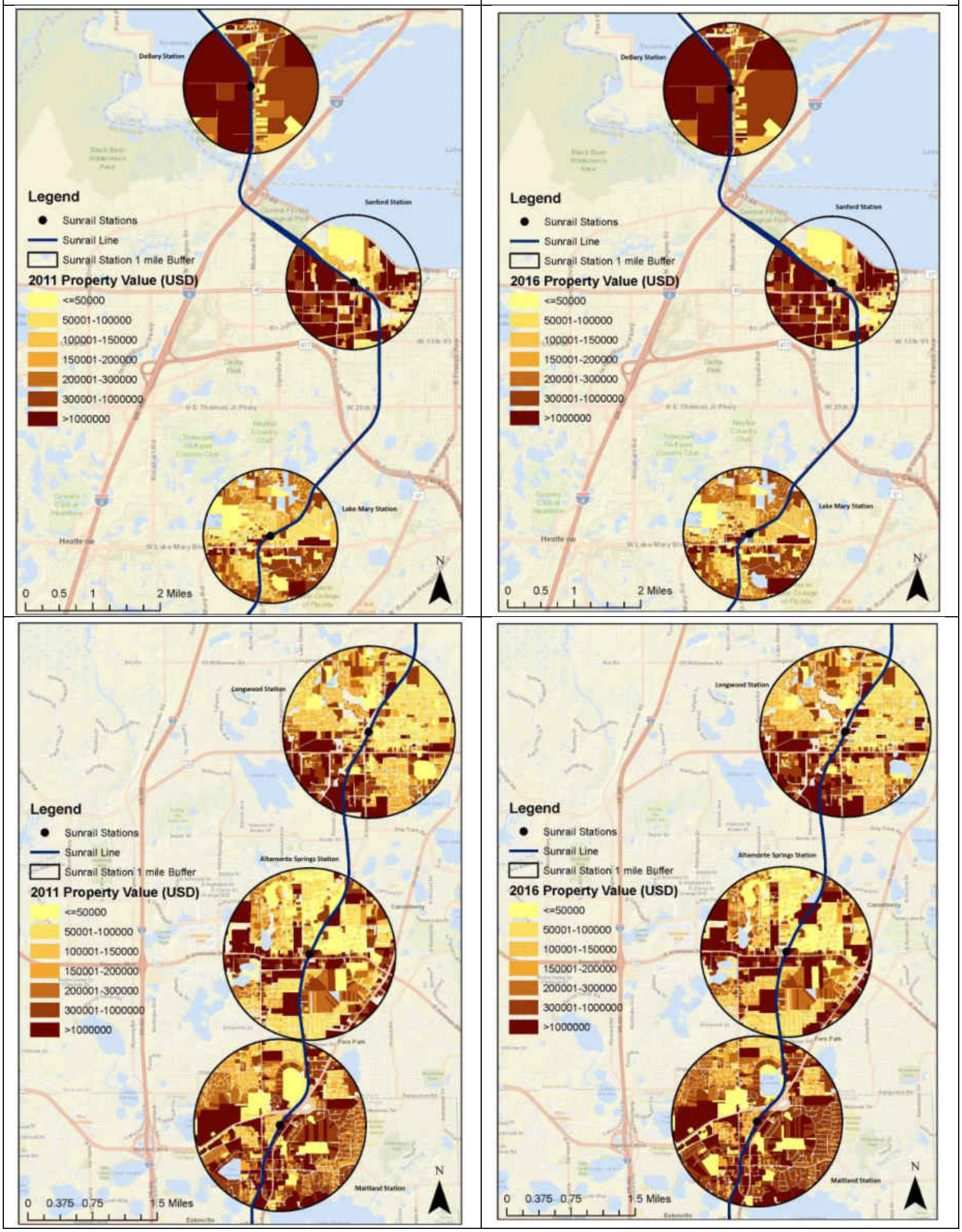


Figure 6. Property Values within SunRail Influence area for 2011 and 2016

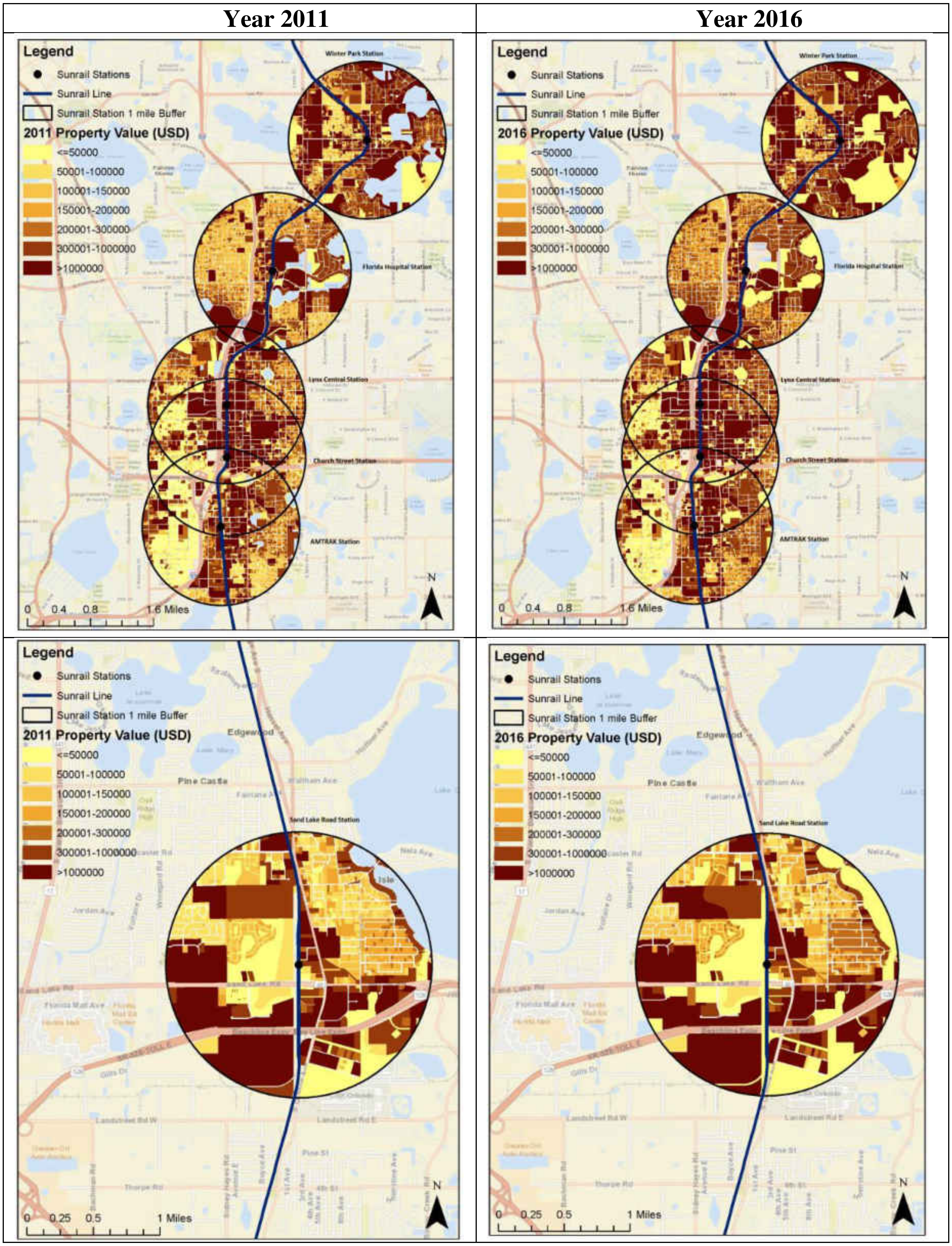


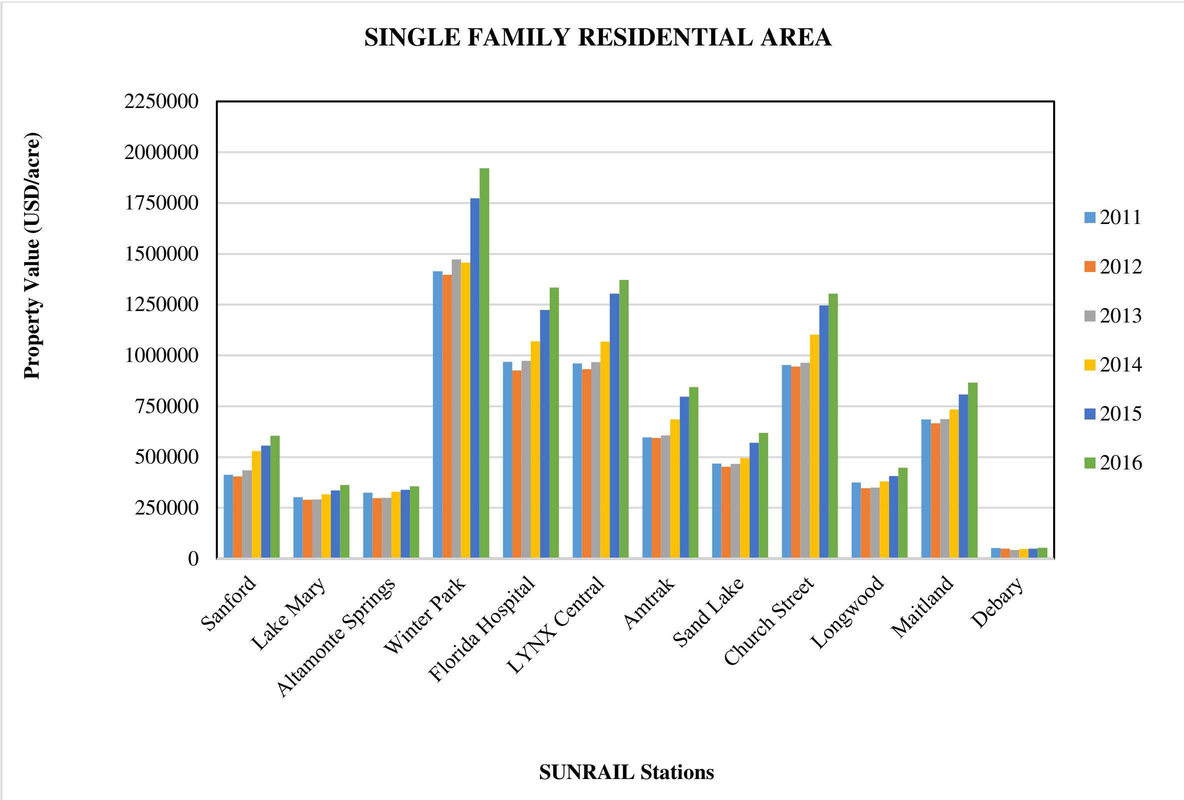
Figure 6. (Continued): Property Values within SunRail Influence area for 2011 and 2016

For CBA, we are interested in the overall system-level impact of SunRail on property value. However, for future investment and improvement proposals, it is also important for us to understand the station-level impacts. Therefore, in this study effort, we also compute the property values of the influence area across different stations. However, as is evident from Figure 5 and 6, certain portion of the influence areas for some stations are not exclusive. For some stations, buffer areas within 1-mile radius overlap with each other. We allocate the parcels within the overlapping area to a particular station by using nearest distance or proximity to or from station (Hess and Almeida, 2007). For example, Lynx Central station and Church Street station are the closest stations in the downtown area. For taking care of the overlapping problem, we draw a straight line from the parcel to each station by using ArcGIS tool and then we assign the parcel to the nearest station in computing station-level property values. Figure 6 represents the property value per acre of different land use categories across twelve stations.

From Figure 6, we can observe that, compared to other stations, the property value is very high around Church Street station for multi-family residential, retail/office and institutional area categories while in case of single family residential and industrial area, Winter park station is found to be the expensive one. As expected, property value per unit area by land use category had increased over the years for almost every station. One interesting trend that can be observed from Figure 7 is that across all the land use categories, property price declined a little bit from 2011 to 2012 for all land use types except for multifamily residential. On the other hand, there is a huge increase in property price from 2014 to 2015 (after SunRail period) for industrial, single family residential, multi-family residential and office area around the Winter Park, Lynx Central, Florida Hospital and Church Street station. On the other hand, for recreational areas, property price did not change much over the years for almost all stations except for Maitland station which shows a

25% increase in this category. For multifamily residential area, the property price has almost doubled from 2014 to 2016 for the Lynx Central, Florida Hospital and Winter Park stations.

In the current research effort, our main objective is to identify the effect of SunRail on property value. However, based on the property value change within the vicinity of station areas, it is not accurate to attribute all of these changes to the introduction of SunRail. It is possible that the Greater Orlando region experienced a boom in property price. To address this, we identify parcels outside the influence area to estimate changes in property values. In other words, we need to identify some controls in order to compute the SunRail specific effect of property value. In our study, we identify "Control Parcels" from the area which are outside 2-mile buffer boundary of SunRail stations but from within 8-mile buffer area. We randomly selected control parcels based on their land use category and the property value. If the parcel values of control parcels are within 25% range of case parcels, we selected those as control parcels and we repeated this procedure for all land use categories.



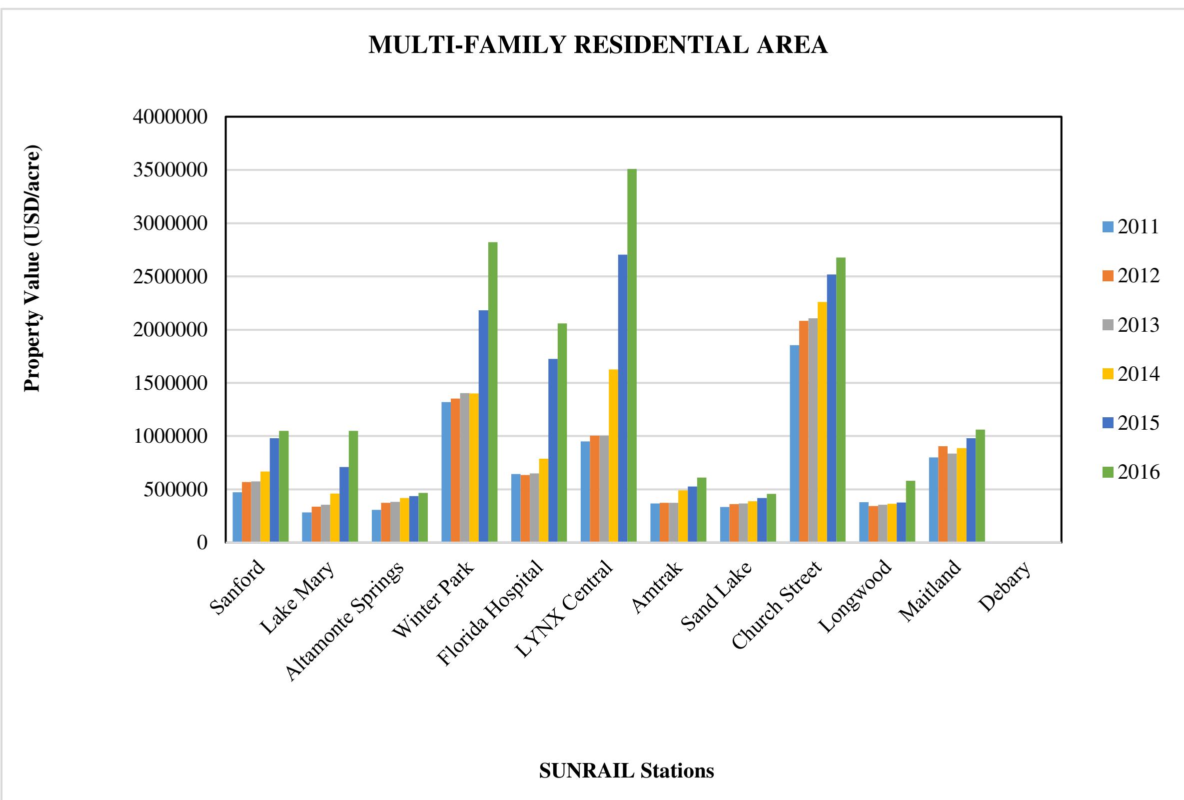
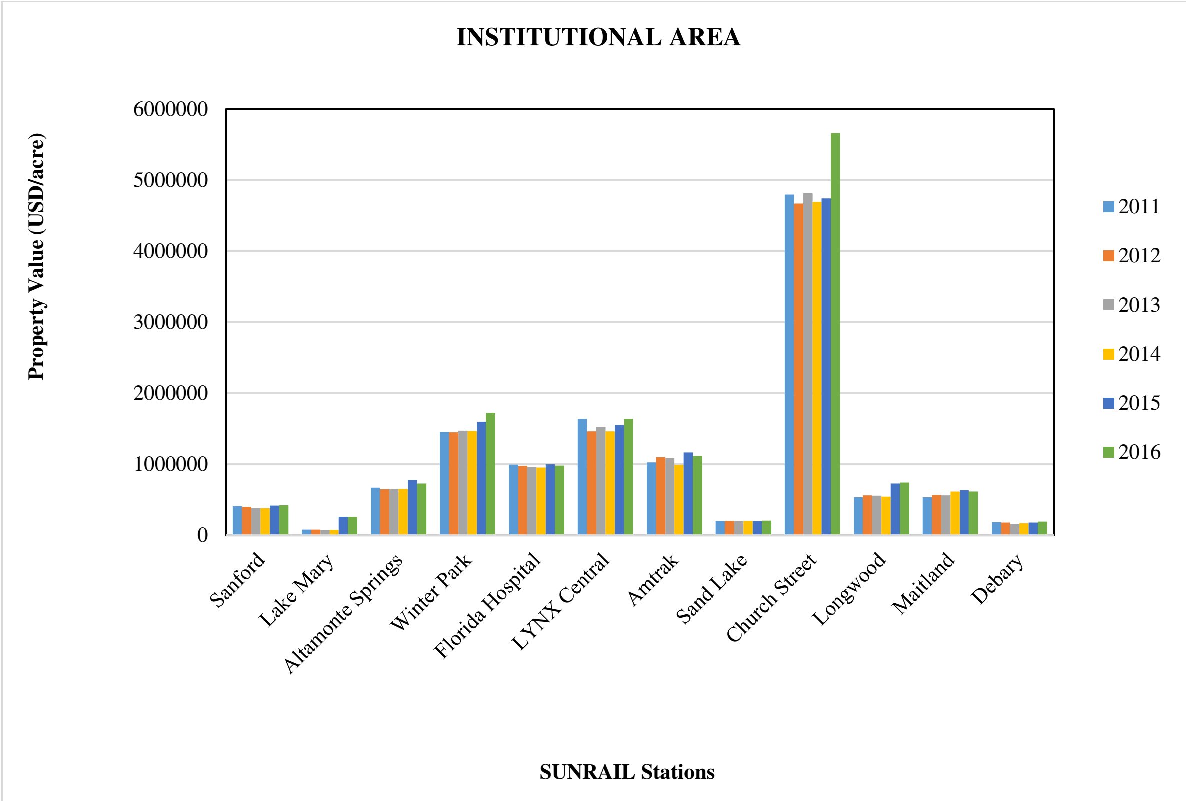


Figure 7. Station-level Property Value per Acre for Different Land use Types



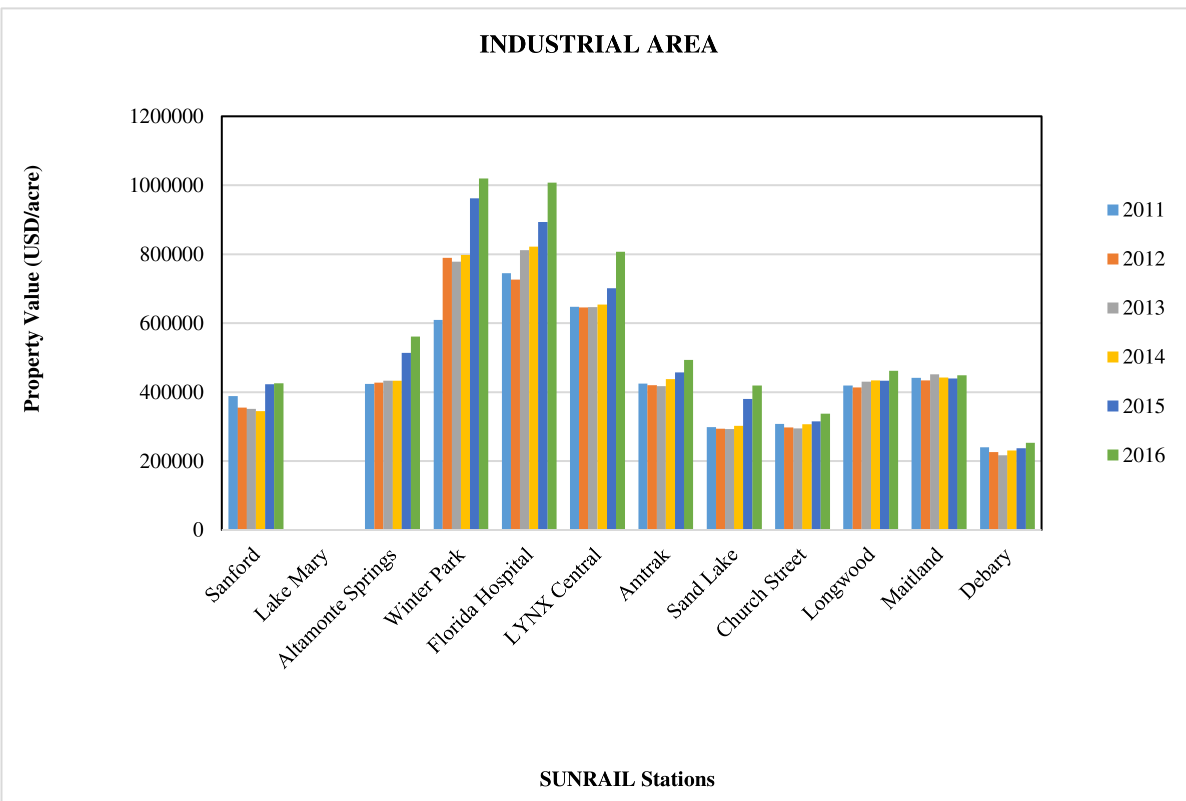
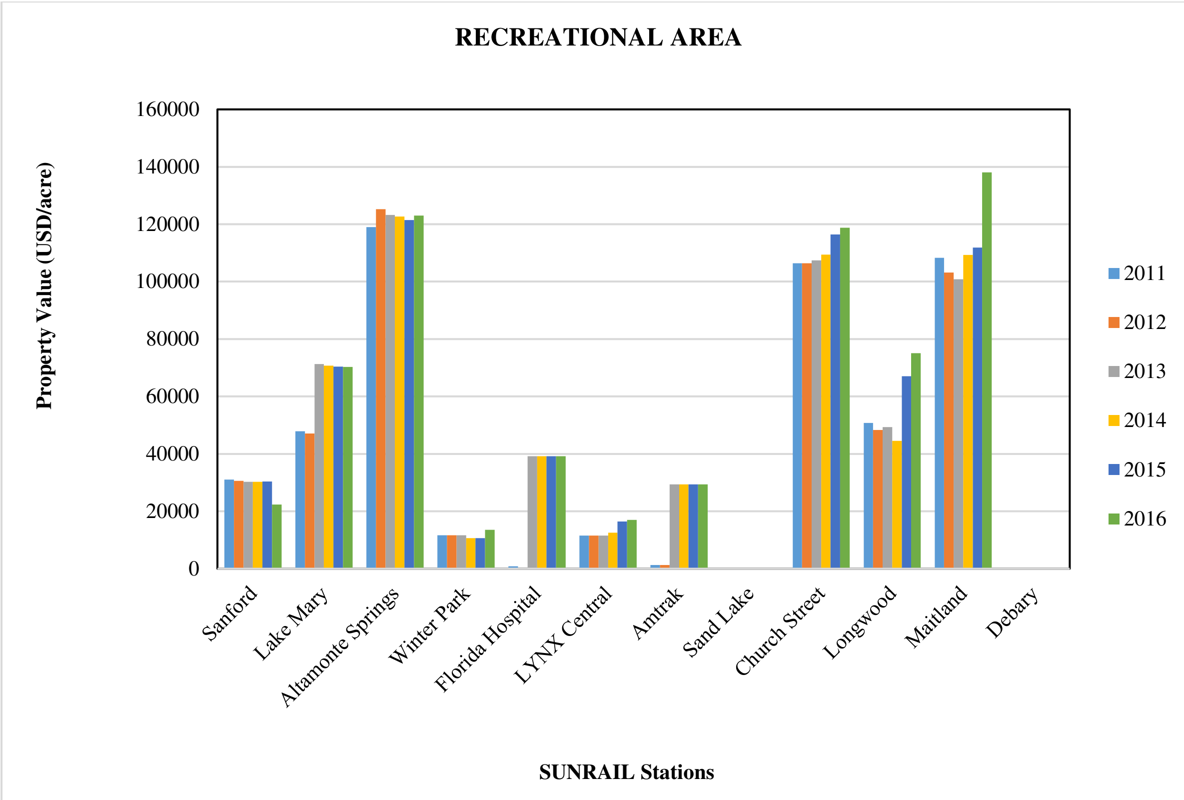


Figure 7. (Continued): Station-level Property Value per Acre for Different Land use Types



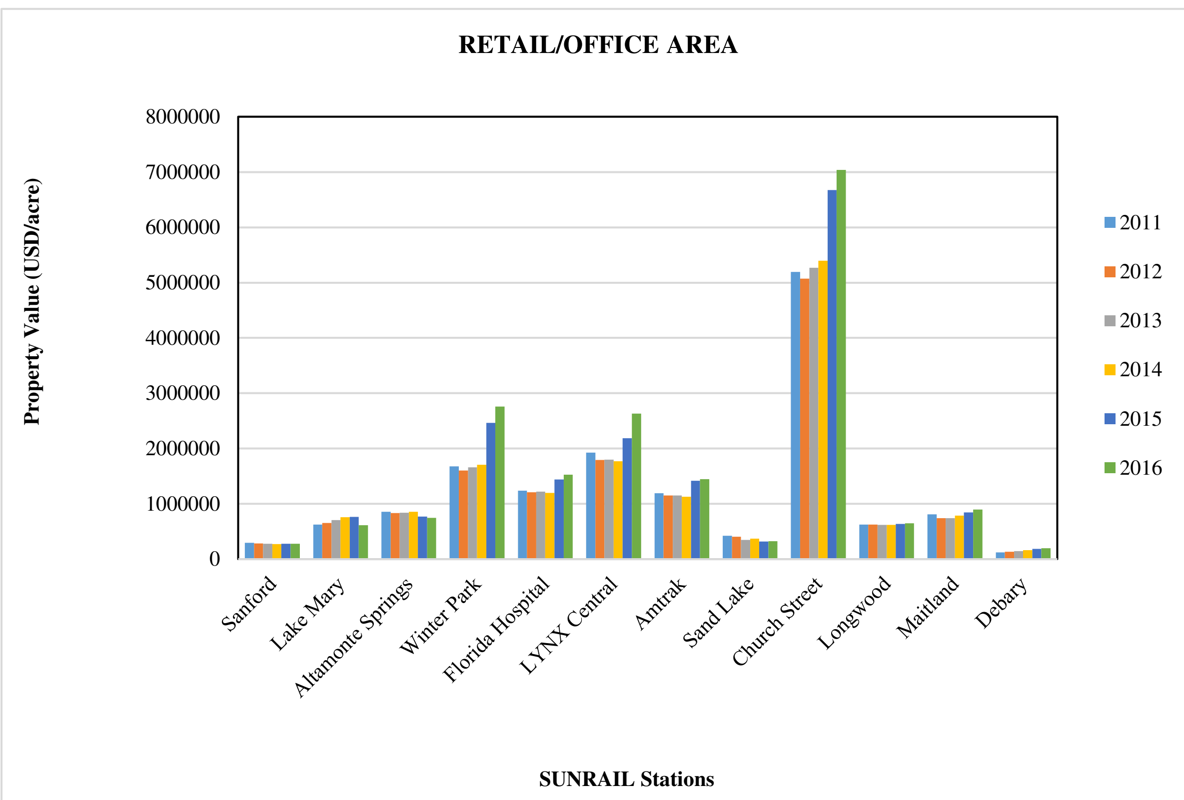


Figure 7. (Continued): Station-level Property Value per Acre for Different Land use Types

It is also important for us to recognize that the parcels within downtown area have different impact than those outside the downtown area since downtown area was already mostly developed before SunRail introduction. To reflect this, we have identified control parcels for downtown and outside downtown area separately. We have considered three stations as downtown stations (Lynx Central, Church Street, and AMTRAK station) and the rest 9 stations as outside downtown stations (DeBary station, Sanford Station, Lake Mary, Longwood Station, Altamonte Station, Maitland station, Winter Park station, Florida Hospital and Sand Lake road). By following this procedure, we finally consider as many control parcels as we have as case parcels. Finally, we compute the assessed base year property value increase of areas within the vicinity of SunRail stations as:

$$BYPVI = 0.85 * BP * [P_A^{cases} - P_B^{cases} - P^{control}] \tag{29}$$

Where,

$BYPVI$ = Base year Property value increase for SunRail influence area

BP = Base year Property value for case parcels

P_A^{cases}= Annual percentage change in property value for case parcels from 2014-2016

P_B^{cases} = Annual percentage change in property value for case parcels from 2011-2013

$P^{control}$ = Annual percentage change in property value of control parcels

The factor 0.85 is employed to allow for a safety margin on the impact of SunRail. In addition to accounting for growth in the control parcels we attribute only 85% of the increase in property values to SunRail. This can be viewed as a conservative estimate of SunRail associated property increase. For the base year, the computed property value increase across different land use types are presented in Table 21.

Table 21. Computed Property Value Increase for Base Year

Land use types	Property value increase	
	Downtown	Outside downtown
Single family residential	800,244,624.92	4,250,778,859.61
Multiple family residential	464,788,552.54	424,960,294.01
Industrial	136,904,784.32	392,667,602.42
Institutional	307,379,096.55	441,908,986.35
Recreational	29,485.69	9,515,762.34
Retail/Office	2,123,586,528.71	1,686,474,314.84

8.5 Result of Cost-Benefit Analysis

In performing the CBA, we assume that the useful life of the existing SunRail project will be 30 years with the beginning year as 2014. Therefore, we projected the costs and benefit values for 30 years, from 2014 to 2044, and computed the net benefit and benefit-cost ratio. In the current study context, we perform CBA for different scenarios as presented in Table 22. In evaluating net benefits of SunRail, we perform scenario analysis by assuming change in annual ridership and change in annual property value increase. Specifically, with respect to ridership change, we consider three scenarios:

Scenario 1: No change in SunRail ridership over 30 years (Monthly ridership is 3700).

Scenario 2: SunRail ridership increases by 2% each year over 30 years (Monthly ridership is 3700 for the base year 2014).

Scenario 3: SunRail ridership increases by 10% each year over 30 years (Monthly ridership is 3700 for the base year 2014).

In terms of property value, we have considered seven different property value increase conditions for each ridership scenario. The scenarios consider projected growth rate as a function of previous year growth rate. We evaluate the impact of property price increase under various reducing growth rate scenarios with and without a threshold level. The rationale for these scenarios is to evaluate how the property value impacts change under various growth rate scenarios.

Overall, the total numbers of scenarios considered are twenty-one (3*7). We consider change in ridership to reflect the possible ridership addition from Phase II and Phase III operations of SunRail in the future. To be sure, in computing the benefit components for scenario 2 and 3, we have updated the values of all the benefit components considered for cost-benefit analysis, since those factors are assumed to be a function of ridership. The computed net benefits and benefit-cost ratio for all the considered scenarios described are presented in Table 22. Positive net benefit and benefit-cost ration greater than 1 reflect the overall surplus over investment and operation costs of SunRail operation.

Table 22. Scenarios of Cost-Benefit Analysis

Scenarios	Description
Scenario 1: No change in SunRail ridership over 30 years (Monthly ridership is 3700)	
Scenario 1.1	➢ $Property\ value\ growth\ rate\ (PVGR) = \left(\frac{PVGR computed\ in\ Section\ 4.3}{3}\right)^{for\ first\ 15\ years} \sim \left(\frac{PVGR computed\ in\ Section\ 4.3}{6}\right)^{for\ last\ 15\ years}$ ➢ Everything else remain same
Scenario 1.2	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = \left(\frac{PVGR\ for\ the\ year\ \tau-1}{2}\right)$ ➢ Everything else remain same
Scenario 1.3	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = Maximum\left(\frac{PVGR\ for\ the\ year\ \tau-1}{2},\ 3.00\%\right)$ ➢ Everything else remain same
Scenario 1.4	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = Maximum\left(\frac{PVGR\ for\ the\ year\ \tau-1}{2},\ 2.00\%\right)$ ➢ Everything else remain same
Scenario 1.5	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = \left(\frac{PVGR\ for\ the\ year\ \tau-1}{5}\right)$ ➢ Everything else remain same
Scenario 1.6	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = Maximum\left(\frac{PVGR\ for\ the\ year\ \tau-1}{5},\ 3.00\%\right)$ ➢ Everything else remain same
Scenario 1.7	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = Maximum\left(\frac{PVGR\ for\ the\ year\ \tau-1}{5},\ 2.00\%\right)$ ➢ Everything else remain same
Scenario 2: SunRail ridership increases by 2% each year over 30 years (Monthly ridership is 3700 for the base year 2014)	
Scenario 2.1	➢ $Property\ value\ growth\ rate\ (PVGR) = \left(\frac{PVGR computed\ in\ Section\ 4.3}{3}\right)^{for\ first\ 15\ years} \sim \left(\frac{PVGR computed\ in\ Section\ 4.3}{6}\right)^{for\ last\ 15\ years}$ ➢ Adjusted benefit components due to the change in ridership
Scenario 2.2	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = \left(\frac{PVGR\ for\ the\ year\ \tau-1}{2}\right)$ ➢ Adjusted benefit components due to the change in ridership
Scenario 2.3	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = Maximum\left(\frac{PVGR\ for\ the\ year\ \tau-1}{2},\ 3.00\%\right)$ ➢ Adjusted benefit components due to the change in ridership
Scenario 2.4	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = Maximum\left(\frac{PVGR\ for\ the\ year\ \tau-1}{2},\ 2.00\%\right)$ ➢ Adjusted benefit components due to the change in ridership
Scenario 2.5	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = \left(\frac{PVGR\ for\ the\ year\ \tau-1}{5}\right)$ ➢ Adjusted benefit components due to the change in ridership

Table 22. (Continued): Scenarios of Cost-Benefit Analysis

Scenarios	Description
Scenario 2.6	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = Maximum\left(\frac{PVGR\ for\ the\ year\ \tau-1}{5},\ 3.00\%\right)$ ➢ Adjusted benefit components due to the change in ridership
Scenario 2.7	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = Maximum\left(\frac{PVGR\ for\ the\ year\ \tau-1}{5},\ 2.00\%\right)$ ➢ Adjusted benefit components due to the change in ridership
Scenario 3: SunRail ridership increases by 10% each year over 30 years (Monthly ridership is 3700 for the base year 2014)	
Scenario 3.1	➢ $Property\ value\ growth\ rate\ (PVGR) = \left(\frac{PVGR computed\ in\ Section\ 4.3}{3}\right)^{for\ first\ 15\ years} \sim \left(\frac{PVGR computed\ in\ Section\ 4.3}{6}\right)^{for\ last\ 15\ years}$ ➢ Adjusted benefit components due to the change in ridership
Scenario 3.2	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = \left(\frac{PVGR\ for\ the\ year\ \tau-1}{2}\right)$ ➢ Adjusted benefit components due to the change in ridership
Scenario 3.3	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = Maximum\left(\frac{PVGR\ for\ the\ year\ \tau-1}{2},\ 3.00\%\right)$ ➢ Adjusted benefit components due to the change in ridership
Scenario 3.4	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = Maximum\left(\frac{PVGR\ for\ the\ year\ \tau-1}{2},\ 2.00\%\right)$ ➢ Adjusted benefit components due to the change in ridership
Scenario 3.5	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = \left(\frac{PVGR\ for\ the\ year\ \tau-1}{5}\right)$ ➢ Adjusted benefit components due to the change in ridership
Scenario 3.6	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = Maximum\left(\frac{PVGR\ for\ the\ year\ \tau-1}{5},\ 3.00\%\right)$ ➢ Adjusted benefit components due to the change in ridership
Scenario 3.7	➢ $Property\ value\ growth\ rate\ (PVGR)\ for\ year\ \tau = Maximum\left(\frac{PVGR\ for\ the\ year\ \tau-1}{5},\ 2.00\%\right)$ ➢ Adjusted benefit components due to the change in ridership

Table 23. Cost-benefits analysis of SunRail over 30 Years

Scenarios	Property Value increase	Other benefits	Total benefits (Property value increase + Other benefits)	Total Costs	Net benefit (Total benefits - Total costs)	Benefit-Cost ratio (Total benefits/Total Costs)
Scenario 1: No change in SunRail ridership over 30 years (Monthly ridership is 3700)						
Scenario 1.1	4,868,083,957.13	323,503,544.15	5,191,587,501.28	1,674,985,000.00	3,516,602,501.28	3.10
Scenario 1.2	569,084,731.14	323,503,544.15	892,588,275.29	1,674,985,000.00	-782,396,724.71	0.53
Scenario 1.3	9,791,139,652.85	323,503,544.15	10,114,643,197.00	1,674,985,000.00	8,439,658,197.00	6.04
Scenario 1.4	5,802,933,688.77	323,503,544.15	6,126,437,232.92	1,674,985,000.00	4,451,452,232.92	3.66
Scenario 1.5	238,746,982.13	323,503,544.15	562,250,526.28	1,674,985,000.00	-1,112,734,473.72	0.34
Scenario 1.6	9,733,889,988.41	323,503,544.15	10,057,393,532.56	1,674,985,000.00	8,382,408,532.56	6.00
Scenario 1.7	5,656,322,022.42	323,503,544.15	5,979,825,566.57	1,674,985,000.00	4,304,840,566.57	3.57
Scenario 2: SunRail ridership increases by 2% each year over 30 years (Monthly ridership is 3700 for the base year 2015)						
Scenario 2.1	4,868,083,957.13	438,194,196.42	5,306,278,153.56	1,674,985,000.00	3,631,293,153.56	3.17
Scenario 2.2	569,084,731.14	438,194,196.42	1,007,278,927.57	1,674,985,000.00	-667,706,072.43	0.60
Scenario 2.3	9,791,139,652.85	438,194,196.42	10,229,333,849.27	1,674,985,000.00	8,554,348,849.27	6.11
Scenario 2.4	5,802,933,688.77	438,194,196.42	6,241,127,885.20	1,674,985,000.00	4,566,142,885.20	3.73
Scenario 2.5	238,746,982.13	438,194,196.42	676,941,178.56	1,674,985,000.00	-998,043,821.44	0.40
Scenario 2.6	9,733,889,988.41	438,194,196.42	10,172,084,184.83	1,674,985,000.00	8,497,099,184.83	6.07
Scenario 2.7	5,656,322,022.42	438,194,196.42	6,094,516,218.84	1,674,985,000.00	4,419,531,218.84	3.64
Scenario 3: SunRail ridership increases by 10% each year over 30 years (Monthly ridership is 3700 for the base year 2015)						
Scenario 3.1	4,868,083,957.13	1,783,400,526.10	6,651,484,483.24	1,674,985,000.00	4,976,499,483.24	3.97
Scenario 3.2	569,084,731.14	1,783,400,526.10	2,352,485,257.25	1,674,985,000.00	677,500,257.25	1.40
Scenario 3.3	9,791,139,652.85	1,783,400,526.10	11,574,540,178.95	1,674,985,000.00	9,899,555,178.95	6.91
Scenario 3.4	5,802,933,688.77	1,783,400,526.10	7,586,334,214.88	1,674,985,000.00	5,911,349,214.88	4.53
Scenario 3.5	238,746,982.13	1,783,400,526.10	2,022,147,508.24	1,674,985,000.00	347,162,508.24	1.21
Scenario 3.6	9,733,889,988.41	1,783,400,526.10	11,517,290,514.51	1,674,985,000.00	9,842,305,514.51	6.88
Scenario 3.7	5,656,322,022.42	1,783,400,526.10	7,439,722,548.52	1,674,985,000.00	5,764,737,548.52	4.44

From Table 23, we can observe that increased ridership is the most important factor in achieving an overall net benefit over long term for SunRail. The result has significant implication in terms of SunRail extension. With Phase II addition, it has the potential to increase ridership. It is also interesting to observe that property value increase plays an important role in accruing overall positive net benefit with a benefit-cost ratio over 1. The result is perhaps indicating benefits of transit oriented development for a personal automobile governed city like Orlando. Based on this result, we can argue that the SunRail commuter system has potential in promoting overall transit oriented development community concept in encouraging sustainable transportation alternatives.

8.6 Summary

The chapter summarized cost-benefit analysis for the existing operation SunRail system (Phase I). With regards to cost component, the factors we considered included: (1) capital costs and (2) operation and maintenance costs. In terms of the benefit component, the factors we considered included: (1) personal automobile cost savings, (2) crash cost savings, (3) parking cost savings, (4) energy conservation savings, and (5) assessed property value increase. For cost-benefit analysis, we considered total 21 hypothetical scenarios reflecting the change in ridership and property value increase rate over thirty years. Based on this result, we can conclude that the SunRail commuter system has potential in promoting overall transit oriented development community concept in encouraging sustainable transportation alternatives.

In promoting sustainable urban transportation, policy makers are more focused on encouraging travelers to walk, bike or take transit among Floridians like many other auto oriented states and cities in the US. In Orlando, other than SunRail, another such initiative is Juice Bike share system of Downtown Orlando. It might also be interesting and worth investigating the cost-benefit analysis for Juice bike share system. The cost-benefit analysis for Juice bike share system would allow the policy makers to take such other initiative in consideration. The research team did not have any detailed data and information available on the bike share investment project and hence the cost-benefit analysis was not evaluated. However, the same framework, as presented in this technical report for SunRail, is applicable for performing cost-benefit analysis of Juice bike share system, which might be considered as a future research avenue.

CHAPTER NINE: CONCLUSION

9.1 Summary of this study

The economic development and the associated growth in household incomes in the United States during the post-Second World War resulted in an increased household and vehicle ownership, population and employment decentralization and urban sprawl. Population and employment changes resulted in a drastic reduction in public transit ridership. The consequences of the drastic transformation of the transportation system include negative externalities such as traffic congestion and crashes, air pollution associated environmental and health concerns, and dependence on foreign fuel. Furthermore, the increased private vehicular travel contributes to increasing air pollution and greenhouse gas (GHG) emissions - a matter receiving substantial attention given the significant impact on health and safety of future generations. In an endeavor to counter the negative externalities of this personal vehicle dependence, many urban regions, across different parts of North America, are considering investments in public transportation alternatives such as bus, light rail, express bus service, metro and bicycle sharing systems.

The public transit investments are particularly critical in growing urban regions such as Orlando, Florida. The greater Orlando region, serves as an ideal test bed to contribute research approaches to evaluate the impact of transit investments on public transit system usage. Transit system managers and planners mostly rely on statistical models to identify the factors that affect ridership as well as quantifying the magnitude of the impact on the society. These models provide vital feedback to agencies on the benefits of public transit investments which in turn act as lessons to improve the investment process.

In our study, we examine the impact of new transit investments (such as an addition of commuter rail to an urban region) on existing transit infrastructure (such as the traditional bus service already present in the urban region). The process of evaluating the impact of new investments on existing public transit requires a comprehensive analysis of the before and after measures of public transit usage in the region. The current research effort contributes to transit literature by evaluating the influence of a recently inaugurated commuter rail system on traditional bus service. We examine the before and after impact of “SunRail” commuter rail system in the Orlando metropolitan region on the “Lynx” bus system. Given the relatively long-time span required for the influence of large scale public transportation system changes, any analysis of the value of new investments should consider adequate data before the system installation and after the system installation. A panel joint grouped response ordered modeling framework that accommodates for common unobserved factors affecting boarding and alighting as well as repeated measures for each stop. Additionally, the influence of SunRail on ridership has a positive temporal trend indicating the strengthening of the impact with the time of operation, a healthy metric for potential future expansion.

We also accommodate for the presence of common unobserved factors associated with spatial factors by developing a spatial panel model by using stop level public transit boarding and alighting data, Specifically, two spatial models: 1) Spatial Error Model (SEM) and 2) Spatial Lag Model (SAR) are estimated for boarding and alighting separately by employing several exogenous variables including stop level attributes, transportation and transit infrastructure variables, built environment and land use attributes, sociodemographic and socioeconomic variables in the vicinity of the stop and spatial and spatio-temporal lagged variables. The repeated observation data at a stop-level offers multiple dimensions of unobserved factors including stop-level, spatial and temporal factors. In our analysis, we apply a framework proposed to accommodate for the aforementioned observed and unobserved factors. The results from the spatial error and lag models are compared with the results from traditional linear regression models to identify the improvement in model fit with accommodation of spatial unobserved effects and panel repeated measures.

Another objective of this study is to identify the factors that affect the SunRail ridership in Orlando region. The current study contributes to literature on transit ridership by considering daily boarding and alighting data from a recently launched commuter rail system. With the rich panel of repeated observations for every station, the potential impact of observed and unobserved factors affecting ridership variables are considered. Specifically, an estimation framework that accounts for these unobserved effects at multiple levels – station, station-week and station day - is proposed and estimated. In addition, the study examines the impact of various observed exogenous factors such as station level attributes, transportation infrastructure variables, transit infrastructure variables, land use and built environment attributes, sociodemographic and weather variables on ridership. The model system developed will allow us to predict ridership for existing stations in the future as well as potential ridership for future expansion sites.

Our proposed research contributes to public transit literature by developing econometric models that consider the potential endogeneity of stop level headway in modeling ridership. Most (if not all) studies in public transit literature ignore that the stop level headway was determined (by choice) in response to expected ridership i.e. stops with lower headway were expected to have higher ridership numbers. In traditional ridership studies, this potential endogeneity is often neglected and headway is considered as an independent variable. The approach violates the requirement that the unobserved factors that affect the dependent variable do not affect the independent variable. If this is the case, the estimated impact of headway on ridership would be biased (potentially over-estimated). More importantly, the estimated impact of all other variables (such as land use factors, bus infrastructure) will also be biased (possible under-estimated). In this study, we address these challenges by proposing a simultaneous equation system that considers headway and ridership in a joint framework that accounts for the influence of common unobserved factors that affect headway and ridership. The proposed model is developed employing ridership data from Orlando region from the Lynx bus transit system. The empirical analysis involves estimation of different models: 1) Independent ridership-headway (IRH) model and 2) Trivarite ridership-headway (TRH) model. Prior to discussing the estimation results, we compare the performance of these models in this section. The ridership data includes stop level average weekday boarding and alighting information for 11 four-month time periods from May 2013 to December 2016. The presence of multiple data points for each stop allows us to develop panel models for headway, boarding and alighting. The model estimation results identified that headway, number of the bus stops in the 800m buffer, presence of shelter at the bus stop, bus route length in a 800m buffer, sidewalk length in a 400m buffer, bus stop distance from the central business district (CBD), distance between Sunrail station and bus stop, and automobile ownership are likely to impact the bus ridership in Orlando.

Another study of the dissertation is the cost-benefit analysis for the existing operation SunRail system (Phase I). With regards to cost component, the factors we considered included: (1) capital costs and (2) operation and maintenance costs. In terms of the benefit component, the factors we considered included: (1) personal automobile cost savings, (2) crash cost savings, (3) parking cost savings, (4) energy conservation savings, and (5) assessed property value increase. For cost-benefit analysis, we considered total 21 hypothetical scenarios reflecting the change in ridership and property value increase rate over thirty years. Based on this result, we can conclude that the SunRail commuter system has potential in promoting overall transit oriented development community concept in encouraging sustainable transportation alternatives.

9.2 Research Impact to the society

The dissertation developed several econometric models for enhancing our understanding of factors affecting public transit. While the models make significant methodological contributions, the research also offers significant utility to transit planners and agencies. The models developed for Lynx and SunRail ridership can be utilized for predicting ridership for project expansions and/or modification. For instance, using the SunRail ridership models, transit agencies can generate estimates of ridership at proposed Phase 2 and 3 stations. Further, Lynx agency can employ the transit ridership models to evaluate ridership changes with addition or modification of transit routes in Orlando region. Major recommendations from our research for transit agencies include: (1) increasing bus frequency for high ridership stops, (2) addition of bus shelters, (3) redesign routes to match with land use patterns, and (4) enhance the spatial and temporal connectivity between SunRail and Lynx systems.

With the emergence and deployment of advanced technology including automated vehicles, mobility as a service, real-time transit feeds, there is immense opportunity for increasing ridership across the country. The current study was unable to consider these innovative technologies and their impact on ridership due to lack of data. In the presence of such data, the models developed in the dissertation can be substantially enhanced to offer insights for the future.

REFERENCES

AAA (2014) Your Driving Costs: How Much Are You Really Paying to Drive? Retrieved from http://publicaffairsresources.aaa.biz/wp-content/uploads/2014/05/Your-Driving-Costs-2014.pdf

Abkowitz, M., and J. Tozzi. Transit route characteristics and headway-based reliability control. Transportation Research Record, No. 1078, 1986.

Ang-Olson, J., & Mahendra, A. (2011). Cost/benefit Analysis of Converting a Lane for Bus Rapid Transit: Phase II Evaluation and Methodology, 352. Transportation Research Board.

Banerjee, T., Myers, D., Irazabal, C. Bahl, D., Dai, L., Gloss, D., Raghavan, A., and Vutha, N., 2005. Increasing bus transit ridership: Dynamics of density, land use, and population growth. Final report for METRANS project, School of Policy, Planning, and Development, University of Southern California (accessed on March 13th 2014 from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.409.9824&rep=rep1&type=pdf)

Barber, M. 9 U.S. transportation projects to watch in 2017. https://www.curbed.com/2016/12/29/14060810/us-transportation-construction-projects-2017.

Baum-Snow, N. Changes in transportation infrastructure and commuting patterns in US metropolitan areas, 1960–2000. The American economic review, Vol. 100, No. 2, 2010, p. 378.

Baum-Snow, N. Did highways cause suburbanization? The Quarterly Journal of Economics, Vol. 122, No. 2, 2007, pp. 775-805.

Baum-Snow, N., and M. E. Kahn. The effects of new public projects to expand urban rail transit. Journal of Public Economics, Vol. 77, No. 2, 2000, pp. 241-263.

Baum-Snow, N., M. E. Kahn, and R. Voith. Effects of urban rail transit expansions: Evidence from sixteen cities, 1970-2000 [with comment]. Brookings-Wharton papers on urban affairs, 2005, pp. 147-206.

Beko, J. Estimation of demand functions for services of public railway passenger transportation: evidence from Slovenia. PROMET-Traffic & Transportation, Vol. 16, No. 1, 2004, pp. 5-14.

Bhat, C. R. (2001). "Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model." Transportation Research Part B: Methodological 35(7): 677-693.

BLISS, L. (2017). What's Behind Declining Transit Ridership Nationwide?

Blonn, J., Carlson, D., Mueller, P., & Scott, I. (2006). Transport 2020 Bus Rapid Transit: A Cost Benefit Analysis. Prepared for Susan Devos, Chai, Madison Area Bus Advocates, Madison, Wisconsin

Boarnet, M. G., Giuliano, G., Hou, Y., & Shin, E. J. (2017). First/last mile transit access as an equity planning issue. Transportation Research Part A: Policy and Practice, 103, 296-310. doi:https://doi.org/10.1016/j.tra.2017.06.011

Brinkmann, P. New stats show Orlando grew faster than 30 biggest metros. 2016 March 24.

Brown, J. R. and G. L. Thompson (2008). "The relationship between transit ridership and urban decentralisation: insights from Atlanta." Urban Studies 45(5-6): 1119-1139.

Cervero, R. and E. Guerra, (2011). "Urban Densities and Transit: A Multi-dimensional Perspective."

Chakraborty, A., and S. Mishra. Land use and transit ridership connections: Implications for state-level planning agencies. Land Use Policy, Vol. 30, No. 1, 2013, pp. 458-469.

Chan, S. and L. Miranda-Moreno (2013). "A station-level ridership model for the metro network in Montreal, Quebec." Canadian Journal of Civil Engineering 40(3): 254-262.

Chen, C., D. Varley, and J. Chen. What affects transit ridership? A dynamic analysis involving multiple factors, lags and asymmetric behaviour. Urban Studies, Vol. 48, No. 9, 2011, pp. 1893-1908.

Chiang, W.-C., R. A. Russell, and T. L. Urban. Forecasting ridership for a metropolitan transit authority. Transportation Research Part A: Policy and Practice, Vol. 45, No. 7, 2011, pp. 696-705.

Choi, J., Y. J. Lee, T. Kim, and K. Sohn. An analysis of Metro ridership at the station-to-station level in Seoul. Transportation, Vol. 39, No. 3, 2012, pp. 705-722.

Chu, X. (2004). Ridership models at the stop level. Tampa, Florida: Center for Urban Transportation Research, University of South Florida, Report No. BC137-31.

Currie, G., and J. Phung. Transit ridership, auto gas prices, and world events: new drivers of change? Transportation Research Record: Journal of the Transportation Research Board, No. 1992, 2007, pp. 3-10.

Debrezion, G., E. Pels, and P Rietveld "The impact of railway stations on residential and commercial property value: a meta-analysis." The Journal of Real Estate Finance and Economics 35(2): 161-180 (2007).

Debrezion, G., E. Pels, and P Rietveld. "Modelling the joint access mode and railway station choice." Transportation Research Part E: logistics and transportation review 45(1): 270-283 (2009).

Dill, J., M. Schlossberg, L. Ma, and C. Meyer. Predicting transit ridership at the stop level: The role of service and urban form.In 92nd annual meeting of the Transportation Research Board, Washington, DC, 2013.

Ding, Y., and S. Chien. Improving transit service quality and headway regularity with real-time control. Transportation Research Record: Journal of the Transportation Research Board, No. 1760, 2001, pp. 161-170.

Elhorst, J. P. Spatial panel data models. In Spatial econometrics, Springer, 2014. pp. 37-93

Elhorst, J. P. Specification and Estimation of Spatial Panel Data Models. International regional science review, Vol. 26, No. 3, 2003, pp. 244-268.

Eluru, N., and C.R. Bhat (2007), "A Joint Econometric Analysis of Seat Belt Use and Crash-Related Injury Severity", Accident Analysis & Prevention, Vol. 39, No. 5, pp. 1037-1049

Eluru, N., I. N. Sener, C. R. Bhat, R. M. Pendyala, and K.W. Axhausen. (2009). "Understanding residential mobility: joint model of the reason for residential relocation and stay duration." Transportation Research Record: Journal of the Transportation Research Board (2133): 64-74.

Estupiñán, N. and D. A. Rodríguez (2008). "The relationship between urban form and station boardings for Bogota's BRT." Transportation Research Part A: Policy and Practice 42(2): 296-306.

Faghih-Imani, A., and N. Eluru. Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: A case study of New York CitiBike system. Journal of Transport Geography, Vol. 54, 2016, pp. 218-227.

Fan, J. X, M. Wen, and L. Kowaleski-Jones, (1993). "Sociodemographic and Environmental Correlates of Active Commuting in Rural America".

FDOT. (2017). Assessing the Property Impacts of SunRail Stations. BDV30 TWO 943-20, December 2013, final report. Retrieved From http://www.fdot.gov/transit/Pages/BDV30-943-20FinalReport.pdf.

Ferguson, E. M., Duthie, J., Unnikrishnan, A., & Waller, S. T. (2012). Incorporating equity into the transit frequency-setting problem. Transportation Research Part A: Policy and Practice, 46(1), 190-199. doi:https://doi.org/10.1016/j.tra.2011.06.002

Furth, P. G., & Wilson, N. H. (1981). Setting frequencies on bus routes: Theory and practice. Transportation Research Record, 818(1981), 1-7.

Gomez-Ibanez, Jose A. 1996. "Big-city transit, ridership, deficits, and politics." Journal of the American Planning Association, 62(1), 30-50.

Gallivan, F., Rose, E., Ewing, R., Hamidi, S., & Brown, T. (2015). Quantifying Transit's Impact on GHG Emissions and Energy Use—The Land Use Component (No. Project H-46).

Garrett, T. A. (2004). Light rail transit in America: policy issues and prospects for economic development. Unpublished Manuscript, Federal Reserve Bank of St. Louis, Research Department.

Gkritza, K., M. G. Karlaftis, and F. L. Mannering. Estimating multimodal transit ridership with a varying fare structure. Transportation Research Part A: Policy and Practice, Vol. 45, No. 2, 2011, pp. 148-160.

Godavarthy, R., Mattson, J., & Ndembe, E. (2014). Cost-Benefit Analysis of Rural and Small Urban Transit. University of South Florida, Tampa: National Center for Transit Research.

Gordon, P., & Kolesar, P. E. (2011). A Note on Rail Transit Cost-Benefit Analysis: Do Nonuser Benefits Make A Difference?. Public Works Management & Policy, 16(2), 100-110.

Greene, D. L., & Schafer, A. (2003). Reducing Greenhouse Gas Emissions from US Transportation. Pew Center on Global Climate Change, 68.

Guihaire, V., & Hao, J. K. (2008). Transit network design and scheduling: A global review. Transportation Research Part A: Policy and Practice, 42(10), 1251-1273.

Gutierrez, J., (2001). "Location, Economic potential and daily accessibility: an analysis of the accessibility impact of the high-speed line Madrid-Barcelona-French boarder."

Gutiérrez, J., O. D. Cardoza, and J.C. Garcia-Palomanes, (2011). "Transit ridership forecasting at station level: an approach based on distance-decay weighted regression." Journal of Transport Geography 19(6): 1081-1092.

Gutiérrez, J., O. D. Cardozo, and J. C. García-Palomares. Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. Journal of Transport Geography, Vol. 19, No. 6, 2011, pp. 1081-1092.

Heckman, J. J., & Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. Econometrica, 73(3), 669-738.

Hess, D. B., & Almeida, T. M. (2007). Impact of proximity to light rail rapid transit on station-area property values in Buffalo, New York. Urban studies, 44(5-6), 1041-1068.

Hickey, R. Impact of transit fare increase on ridership and revenue: metropolitan transportation authority, New York city. Transportation Research Record: Journal of the Transportation Research Board, No. 1927, 2005, pp. 239-248.

Huang, Z., M. Zhang, and X. Liu. Estimating light-rail transit peak-hour boarding based on accessibility at station and route levels in Wuhan, China. Transportation Planning and Technology, Vol. 40, No. 5, 2017, pp. 624-639.

I-Jy Chien, S. Optimization of headway, vehicle size and route choice for minimum cost feeder service. Transportation Planning and Technology, Vol. 28, No. 5, 2005, pp. 359-380.

Jaffe, E. The Big U.S. Transportation Infrastructure Projects to Watch in 2016. 2015, December 15.

Johnson, A. (2003). Bus transit and land use: illuminating the interaction. Journal of Public Transportation 6(4): 21-39.

Kain, J. F., and Z. Liu. 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, No. 7–8, 1999, pp. 601-624.

Kain, John F. and Zhi Liu, 1996. Econometric Analysis of Determinants of Transit Ridership: 1960-1990. Prepared for the Volpe National Transit Systems Center, U.S. Department of Transportation, May 25.

Kim, S., G. F. Ulfarsson, and J. Todd Hennessy. Analysis of light rail rider travel behavior: Impacts of individual, built environment, and crime characteristics on transit access. Transportation Research Part A: Policy and Practice, Vol. 41, No. 6, 2007, pp. 511-522.

Kohn, H. Factors affecting urban transit ridership. Bridging the Gaps: Canadian Transportation Research Forum.In Proceedings of the 35th Annual Conference, Charlottetown, Prince Edward Island, Canada, 2000.

Kuah, G. K., and J. Perl. Optimization of feeder bus routes and bus-stop spacing. Journal of Transportation Engineering, Vol. 114, No. 3, 1988, pp. 341-354.

Kuby, M., A. Barranda, and C. Upchurch. Factors influencing light-rail station boardings in the United States. Transportation Research Part A: Policy and Practice, Vol. 38, No. 3, 2004, pp. 223-247.

Lane, B. W. A time-series analysis of gasoline prices and public transportation in US metropolitan areas. Journal of Transport Geography, Vol. 22, 2012, pp. 221-235.

Lane, B. W. Significant characteristics of the urban rail renaissance in the United States: A discriminant analysis. Transportation Research Part A: Policy and Practice, Vol. 42, No. 2, 2008, pp. 279-295.

Lane, B. W. The relationship between recent gasoline price fluctuations and transit ridership in major US cities. Journal of Transport Geography, Vol. 18, No. 2, 2010, pp. 214-225.

Lane, C., M. DiCarlantonio, and L. Usvyat. Sketch models to forecast commuter and light rail ridership: Update to TCRP report 16. Transportation Research Record: Journal of the Transportation Research Board, No. 1986, 2006, pp. 198-210.

Lee, K. S., J. K. Eom, S. Y. You, J. H. Min, and K. Y. Yang. An Empirical Study on the Relationship between Urban Railway Ridership and Socio-economic Characteristics. Procedia Computer Science, Vol. 52, 2015, pp. 106-112.

Lewyn, M. (2018). Why Is Transit Ridership Declining?

Lin J-J, and T-Y. Shin (2008). Does transit-oriented development affect metro ridership? Transportation Research Record 2063: 149-158.

Liu, C., S. Erdogan, T. Ma, and F. W. Ducca. How to Increase Rail Ridership in Maryland: Direct Ridership Models for Policy Guidance. Journal of Urban Planning and Development, Vol. 142, No. 4, 2016, p. 04016017.

Liu, Zhi. 1993. "Determinants of Public Transit Ridership: Analysis of post World War II Trends and Evaluation of Alternative Networks." Cambridge, MA: Harvard University.

Litman T. (2013). Evaluating active transport benefits and costs. Victoria Transport Policy Institute Victoria, BC: Victoria Transport Policy Institute. Retrieved from http://www.vtpi.org/nmt-tdm.pdf.

Litman, T. (2004), Evaluating Public Transit Benefits and Costs, Victoria Transport Policy Institute.

Litman, T. (2012). Evaluating Public Transit Benefits and Costs: Best Practices Guidebook. Victoria Transport Policy Institute.

Litman, T. (2014). A new transit safety narrative. Journal of Public Transportation, 17(4), 8.

Litman, T. (2018). Evaluating Public Transit Benefits and Costs: Guide to Valuing Walking and Cycling Improvements and Encouragement Programs. Victoria Transport Policy Institute.

Mattson, J. W. Effects of rising gas prices on bus ridership for small urban and rural transit systems. Upper Great Plains Transportation Institute, North Dakota State University, 2008.

O'Sullivan, A. Urban economics. Irwin, 1996.

Parks, J., K. Coffel, C. Semler, H. Levinson, and D. Sampson. Estimating Transit Station Ridership: A Method for Calculating Daily Boardings and Evaluating Access Alternatives.In, 2012.

Paulley, N., R. Balcombe, R. Mackett, H. Titheridge, J. Preston, M. Wardman, J. Shires, and P. White. The demand for public transport: The effects of fares, quality of service, income and car ownership. Transport Policy, Vol. 13, No. 4, 2006, pp. 295-306.

Pulugurtha, S. S. and M. Agurla (2012). "Assessment of models to estimate bus-stop level transit ridership using spatial modeling methods." Journal of Public Transportation 15(1): 3.

Ruan, M. Probability-based bus headway regularity measure. IET intelligent transport systems, Vol. 3, No. 4, 2009, pp. 400-408.

Ryan, S. and L. F. Frank (2009). "Pedestrian environments and transit ridership." Journal of Public Transportation 12(1): 3.

Santos, A., N. McGuckin, H. Y. Nakamoto, D. Gray, and S. Liss. Summary of travel trends: 2009 national household travel survey. No. FHWA-PL-ll-022. 2011

Saur, G. J., R. Lee, and C. Gray. New method for transit ridership forecasting.In ITE 2004 Annual Meeting and Exhibit, 2004.

Schrank, D. B. Eisele, T. Lomax and J. Bak (2015), “2015 Urban Mobility Scorecard” Published jointly by The Texas A&M Transportation Institute and INRIX

Schrank, D., B. Eisele, and T. Lomax. TTI’s 2012 urban mobility report. Texas A&M Transportation Institute. The Texas A&M University System, 2012, p. 4.

Schmitt, A. (2017). Transit Ridership Falling Everywhere — But Not in Cities With Redesigned Bus Networks.

Shapiro, R. J., Hassett, K. A., & Arnold, F. S. (2002). Conserving energy and preserving the environment: The role of public transportation. American Public Transportation Association, 9, 1-39.

Shoup, L. (2008). "Ridership and development density: evidence from Washington." DC University of Maryland at College Park, Final paper for URSP 631.

Siddiqui, F. (2018). Falling transit ridership poses an ‘emergency’ for cities, experts fear.

Sun, L.-S., S.-W. Wang, L.-Y. Yao, J. Rong, and J.-M. Ma. Estimation of transit ridership based on spatial analysis and precise land use data. Transportation Letters, Vol. 8, No. 3, 2016, pp. 140-147.

Sung, H. and J.-T. Oh (2011). "Transit-oriented development in a high-density city: Identifying its association with transit ridership in Seoul, Korea." Cities 28(1): 70-82.

Tang, L., and P. Thakuriah. Ridership effects of real-time bus information system: A case study in the City of Chicago. *Transportation Research Part C: Emerging Technologies,* Vol. 22, 2012, pp. 146-161.

Taylor, B. D., D. Miller, H. Iseki, and C. Fink. "Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas." Transportation Research Part A: Policy and Practice 43(1): 60-77 (2009).

TP, 2016. Transport Politic Projects under consideration summary. Accessed online on July 23rd at http://www.thetransportpolitic.com/under-consideration/

Turnquist, M. A. Strategies for improving reliability of bus transit service. 1981.

Vij, A., & Walker, J. L. (2014). Preference endogeneity in discrete choice models. Transportation Research Part B: Methodological, 64, 90-105.

Voith, R. Fares, Service Levels, and Demographics: What Determines Commuter Rail Ridership in the Long Run? Journal of Urban Economics, Vol. 41, No. 2, 1997, pp. 176-197.

Walker, J. L., Ehlers, E., Banerjee, I., & Dugundji, E. R. (2011). Correcting for endogeneity in behavioral choice models with social influence variables. Transportation Research Part A: Policy and Practice, 45(4), 362-374.

Wardman, M. and G. Whelan (1999). "Using geographical information systems to improve rail demand models." Final report to Engineering and Physical science Research council.

Woodcock, J., P. Edwards, C. Tonne, B. G. Armstrong, O. Ashiru, D. Banister, S. Beevers et al. "Public health benefits of strategies to reduce greenhouse-gas emissions: urban land transport." The Lancet 374, no. 9705 (2009): 1930-1943.

Yasmin, S. and N. Eluru (2013). "Evaluating alternate discrete outcome frameworks for modeling crash injury severity." Accident Analysis & Prevention 59: 506-521.

Yasmin. S., N. Eluru, and A. R. Pinjari (2015)"Analyzing the Continuum of Fatal Crashes: A Generalized Ordered Approach," Analytic Methods in Accident Research, Volume 7, July 2015, Pages 1-15

Zhang, D., and X. C. Wang. Transit ridership estimation with network Kriging: a case study of Second Avenue Subway, NYC. Journal of Transport Geography, Vol. 41, 2014, pp. 107-115.

Zhao, J., W. Deng, Y. Song, and Y. Zhu. Analysis of Metro ridership at station level and station-to-station level in Nanjing: an approach based on direct demand models. Transportation, Vol. 41, No. 1, 2014, pp. 133-155.