

Research Article

Examining the Bus Ridership Demand: Application of Spatio-Temporal Panel Models

Moshiur Rahman ¹, Shamsunnahar Yasmin ^{2,3}, Ahmadreza Faghieh-Imani ⁴,
and Naveen Eluru ⁵

¹Marlin Engineering, Marlin, FL, USA

²Queensland University of Technology, Centre for Accident Research & Road Safety—Queensland (CARRS-Q), Brisbane, Australia

³Department of Civil & Environmental Engineering, University of Central Florida, Orlando, FL, USA

⁴Department of Civil and Environmental Engineering, Faculty of Engineering, Imperial College, London, UK

⁵Department of Civil, Environmental & Construction Engineering, University of Central Florida, Orlando, FL, USA

Correspondence should be addressed to Shamsunnahar Yasmin; shams.yasmin@qut.edu.au

Received 8 June 2020; Accepted 22 April 2021; Published 7 May 2021

Academic Editor: Ruimin Li

Copyright © 2021 Moshiur Rahman et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An important tool to evaluate the influence of these public transit investments on transit ridership is the application of statistical models. Drawing on stop-level boarding and alighting data for the Greater Orlando region, the current study estimates spatial panel models that accommodate for the impact of spatial and temporal observed and unobserved factors on transit ridership. Specifically, two spatial models, Spatial Error Model and Spatial Lag Model, 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, and sociodemographic and socioeconomic variables in the vicinity of the stop along with spatial and spatiotemporal lagged variables. The model estimation results are further augmented by a validation exercise. 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.

1. Introduction

The overreliance on private automobile in the US over the last few decades has resulted in various negative externalities including traffic congestion and crashes, air-pollution-associated environmental and health concerns, and dependence on foreign fuel [1]. There is renewed enthusiasm among policy makers and transportation professionals to counter the private automobile reliance. Several urban regions are promoting public transportation and nonmotorized modes of travel through infrastructure investments such as public transit extensions, new commuter rail addition, and bicycle sharing systems (see Jaffe [2] and TPP [3] for public transportation projects under construction or consideration). While nonmotorized 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.

The public transit investments are critical in growing urban regions such as Orlando, Florida. In recent years, the 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 [4]. It is reported that the majority (about 74%) of the population growth in the region is driven by domestic and international migration. The rapid increase in population elevates 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. 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 while also quantifying the magnitude of the impact (e.g., see [5, 6]). These models provide feedback to agencies on the benefits of public transit investments while also providing lessons to improve the investment process.

Orlando, a typical American city in the south, represented by urban sprawl, excessive dependence on automobile, and a captive ridership, provides an ideal test bed to identify factors influencing public transit ridership. 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. In terms of exogenous factors, we consider stop-level attributes (such as headway), transportation infrastructure variables (such as secondary highway length including major and minor arterials and major collectors; railroad length; and local road length and 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 (land use mix, household density, and employment density) and demographic and socioeconomic variables in the vicinity of the bus stop (income, vehicle ownership, and age and gender distribution).

The repeated observation data at a stop level offers multiple dimensions of unobserved factors including stop level and spatial and temporal factors. For instance, it is possible that bus ridership of one bus stop is potentially influenced by the ridership of the neighbouring bus stops. It is also possible that ridership of a bus stop is influenced by the ridership levels of the previous time slots for the stop, while also being interconnected with the ridership of the neighbouring bus stops of the earlier time periods. Neglecting such spatial and temporal interconnections (if present) may result in biased estimates of the underlying ridership mechanism. To that extent, the major objective of this study is to accommodate for spatial and temporal effects (observed and unobserved) for modeling bus stop-level ridership. In our analysis, we apply a framework proposed by Elhost [7] to accommodate for the aforementioned observed and unobserved factors (spatial and temporal effects). Further, to accommodate for the repeated observations of ridership, we employ the spatial panel model in the current study context. The panel models developed include panel spatial error and panel spatial lag formulations (see Faghih-Imani and Eluru [8] for a similar formulation in another context). A validation exercise is conducted to illustrate the applicability of the model framework.

The remainder of the paper is organized as follows. An overview of earlier research is described in Section 2 along with the current study section. In Section 3, the methodology has been outlined. In Section 4, the empirical analysis has

been presented along with the data source and data preparation description for modeling. The model estimation results have been presented in Section 5 along with the discussion on the model results and validation. Finally, Section 6 provides a summary of the findings and concludes the paper.

2. Literature Review and Current Study

Traditional travel demand modeling research has focused on automobile travel. Only recently studies have begun to undertake a 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 [5]. These studies broadly examine macro-level ridership [9, 10], study the impact of financial attributes such as fare, fuel price, and parking cost [11–17], effect of transit attributes, transit level of service [18, 19], and built environment on transit ridership. For the current research effort, the last group of studies are particularly relevant. These studies can be classified by the transit mode of interest such as rail, metro, and bus. As the focus of our current work is bus transit ridership, we limit our review to bus ridership studies. For studies on rail and metro, we refer the reader to [5, 20]. 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.

Ryan and Frank [21] highlighted the value of walkability of an area—computed based on land use mix, street patterns, and density—in determining transit ridership for San Diego. Johnson [22] studied transit boarding's in the Twin Cities region using an ordinary least square approach. The analysis highlighted the value of vertical mixed-use and retail establishments close to the stops. The study also found that population density in the larger vicinity of the stop is more critical to ridership compared to population immediately close to the stop. Pulugurtha and Agurla [6] applied spatial proximity and spatial weighting methods to analyze stop-level ridership data from Charlotte. The models were estimated under various buffer sizes and the authors concluded that 0.25-mile buffer provided adequate model fit. Dill et al. [23], using data from Portland, Eugene–Springfield, and Jackson County, estimated separate log-linear regression models for each region and concluded that improving the transit level-of-service and developing pedestrian friendly environment near the stops positively influenced ridership.

Employing a simultaneous model that accommodates for interaction between transit supply and demand in Bogotá, Estupiñán and Rodríguez [24] concluded that promoting walking and creating barriers to car use are likely to increase ridership. Banerjee et al. [25] examined two corridors in Los Angeles and concluded that several land use and sociodemographic variables affected ridership on rapid bus transit systems. Tang and Thakuriah [26] highlighted that the value of real-time bus information is slightly increasing the bus ridership in Chicago. Chakour and Eluru

[5] employed a composite maximum likelihood approach-based ordered response model to accommodate for common unobserved factor influencing time-period-specific boardings and alightings. The results clearly highlighted the presence of such unobserved dependencies in addition to the impact of land use and urban form variables. More recently, using the same data as adopted in the current paper, Rahman et al. [20] formulated a grouped ordered response model structure that allowed for correlation between daily boardings and alightings at a stop level. The study also accommodated for repeated measures of data available. The study found that transit service affected ridership significantly while the effect of land use and urban form variables was substantially different across various buffer sizes. Further, in their analysis, bus route length, sidewalk length, the presence of low-income population, and the proportion of no vehicle population were likely to increase stop-level ridership.

2.1. Current Study. The review of earlier research 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 (except for Rahman et al. [20]). 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.

In our data, we have average daily boarding and alighting ridership, for weekdays only, for 6 four-month time periods between May 2013 and April 2015. Toward accommodating for spatial factors, we consider the most commonly employed spatial error and spatial lag variants employed for cross-sectional data analysis. The models are developed separately for boardings and alightings. The results from the spatial error and lag models are compared with the results from simple linear regression models to identify the improvement in model fit with accommodation of spatial unobserved effects and panel repeated measures. The model estimation process is conducted employing a host of exogenous variables generated for the study region. The estimated models are validated using a holdout sample.

3. Methodology

In this paper, we considered boarding and alighting data for each bus stop for six time periods. The brief overview of the econometric methodology is presented in this section (see Elhorst [7] for complete econometric model details).

Let $q = 1, 2, \dots, Q$ (in our study $Q = 3,495$) be an index to represent each station (spatial unit) and $t = 1, 2, \dots, T$ (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

$$y_{qt} = \beta' x_{qt} + \mu_q + \varepsilon_{qt}, \quad (1)$$

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 represents 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 a random term that is independently and identically distributed with zero mean and variance σ_μ^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, they are computationally cumbersome for large systems as ours. Therefore, in the current study, we restrict ourselves to a spatial random effects model¹.

In traditional econometric literature, spatial dependency can be incorporated by employing different modeling frameworks, such as spatial lag or spatial autoregressive model (SAR), spatial error model (SEM), geographically weighted regression (GWR), and spatial Durbin model. In the current study, we have considered two different forms of spatial autocorrelations in examining bus ridership: (1) SAR model, which accounts for spatial endogenous interactions by a spatially lagged dependent variable, and (2) SEM model, which accounts for spatial interactions by a spatial autocorrelation process in the error term. Specifically, the first model comprises endogenous interactions effects with dependent variable at other stops and in the second model the spatial interaction is captured through the error term.

A spatial lag model can be written as follows:

$$y_{qt} = \delta \sum_{j=1}^Q w_{qj} y_{jt} + \beta' x_{qt} + \mu_q + \varepsilon_{qt}, \quad (2)$$

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 800 m buffer area of each other and zeros for the rest of the elements. It must be noted that the 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 [7] for more details on W matrix).

A spatial error model may be written as follows:

$$y_{qt} = \beta' x_{qt} + \mu_q + \varphi_{qt}, \quad (3a)$$

$$\varphi_{qt} = \rho \sum_{j=1}^Q w_{qj} \varphi_{jt} + \varepsilon_{qt}, \quad (3b)$$

where φ_{qt} accounts for the spatial autocorrelated 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 [7] for details on likelihood functions). In this paper, we use MATLAB routines provided by Elhorst [7, 27] to estimate pooled spatial lag and error models with spatial specific random effects.

4. Empirical Analysis

The Greater Orlando region with a population of 2.3 million in 2016 is a typical American city in the south with an automobile-oriented transportation system with the following mode shares: 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. The number of bus stops considered for the analysis includes 3,745 stops. Of these, 3,495 stops data are used for model estimation while 250 stops data are set aside for validation. In addition to Lynx, the transit system includes a newly launched commuter rail system—SunRail. The rail line is 31 miles long with 12 stations with average weekday ridership of about 3,800 in 2015. Figure 1 represents the study area along with the Lynx bus route, bus stops, SunRail line, and SunRail station locations.

The ridership data was obtained from Lynx transit authority. For our analysis, weekday boarding and alighting data for the following 6 time periods are considered: May through August 2013, September through December 2013, January through April 2014, May through August 2014, September through December 2014, and January through April 2015. The final sample consists of 20,970 records (3,495 stations \times 6 quarters). The average daily stop-level boarding (alighting) is around 21.03 (20.86) with a minimum of 0 (0) and maximum of 7,022 (6,770). The average daily ridership for January through April 2015 quarter is presented in Figure 2 (Figure 2(a)—boarding and Figure 2(b)—alighting). A summary of the system-level ridership (boarding and alighting) is provided in Table 1. The standard deviation is large as the ridership is different across different bus stops in our analysis.

In our study, we have conducted an extensive literature review and identified factors considered in public transit ridership field for identifying the universal set of attributes. GIS shape files from Lynx were used to generate the number

of bus stops and bus route length. For creating the exogenous variables, we considered various buffer distances (800 m, 600 m, 400 m, and 200 m) from each bus stop. The exogenous variable information was generated based on multiple data sources including 2010 US census data, American Community Survey, Florida Geographic Data Library, and Florida Department of Transportation databases. The exogenous attributes considered in our study can be divided into five broad categories: (1) stop-level attributes (such as headway), (2) transportation and transit infrastructure variables (secondary highway length including major and minor arterials and major collectors, railroad length and local road length, sidewalk length, Lynx bus route length, presence of shelter, and distance of bus stop from CBD), (3) built environment and land use attributes (such as land use mix, household density, and employment density), (4) sociodemographic and socioeconomic variables in the vicinity of the stop (income, vehicle ownership, and age and gender distribution), and (5) temporal and spatiotemporal lagged variables (such as stop boarding (alighting) in the last time period).

Temporal lagged variables were calculated for each bus stop by computing the boarding (alighting) variables from previous time period. Spatiotemporal lagged variables were created based on stops within the buffer (for various buffer sizes including 800 m, 600 m, 400 m, and 200 m). The boardings (alightings) from previous time period for stops within the buffer were generated for spatiotemporal lag variables. The descriptive statistics of exogenous variables are presented in Table 2.

5. Model Estimation Results

5.1. 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 benchmark for advanced models. In this section, we compare SEM and SAR models. 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 [28]. The AIC and BIC for a given empirical model are equal to

$$\begin{aligned} \text{AIC} &= 2K - 2\text{LL}, \\ \text{BIC} &= -2\text{LL} + K \ln(Q), \end{aligned} \quad (4)$$

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



FIGURE 1: Public transit system (Lynx and SunRail) of Orlando.

–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 is –46,094.188 (45,951.073), (2) simple linear regression for alighting is –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 variables is 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 3.

5.2. 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 (800 m, 600 m, 400 m, and 200 m 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.

5.3. Stop Level Variables. The headway between buses at a stop has a significant influence on ridership. The result from all models confirms this. An increase in headway is associated with a significant drop in ridership. The findings are in accordance with the previous literature [20, 29–34].

5.4. Transportation Infrastructures Variables. Several transportation infrastructure variables significantly affect boarding and alighting. Bus route length in a 600 m buffer is associated with an increase in boarding and alighting across all models. The result indicates that an increase in the presence of bus route around the stop results in an increased adoption of public transit for the Greater Orlando region. This is an important finding highlighting how when adequate infrastructure for bus transit exists, it is likely to be used. Sidewalk length in an 800 m buffer is observed to positively influence boarding and alighting in the SEM model. The corresponding coefficient was not significant in the SAR models. It is possible that the presence of sidewalk is serving as surrogate for walkable neighborhoods in the SEM model. The secondary highway length in a 600 m buffer and local road length in an 800 m buffer are positively associated with boarding for SEM and SAR models. However, these variables are statistically insignificant in the alighting models. Railroad length in an 800 m buffer is negatively

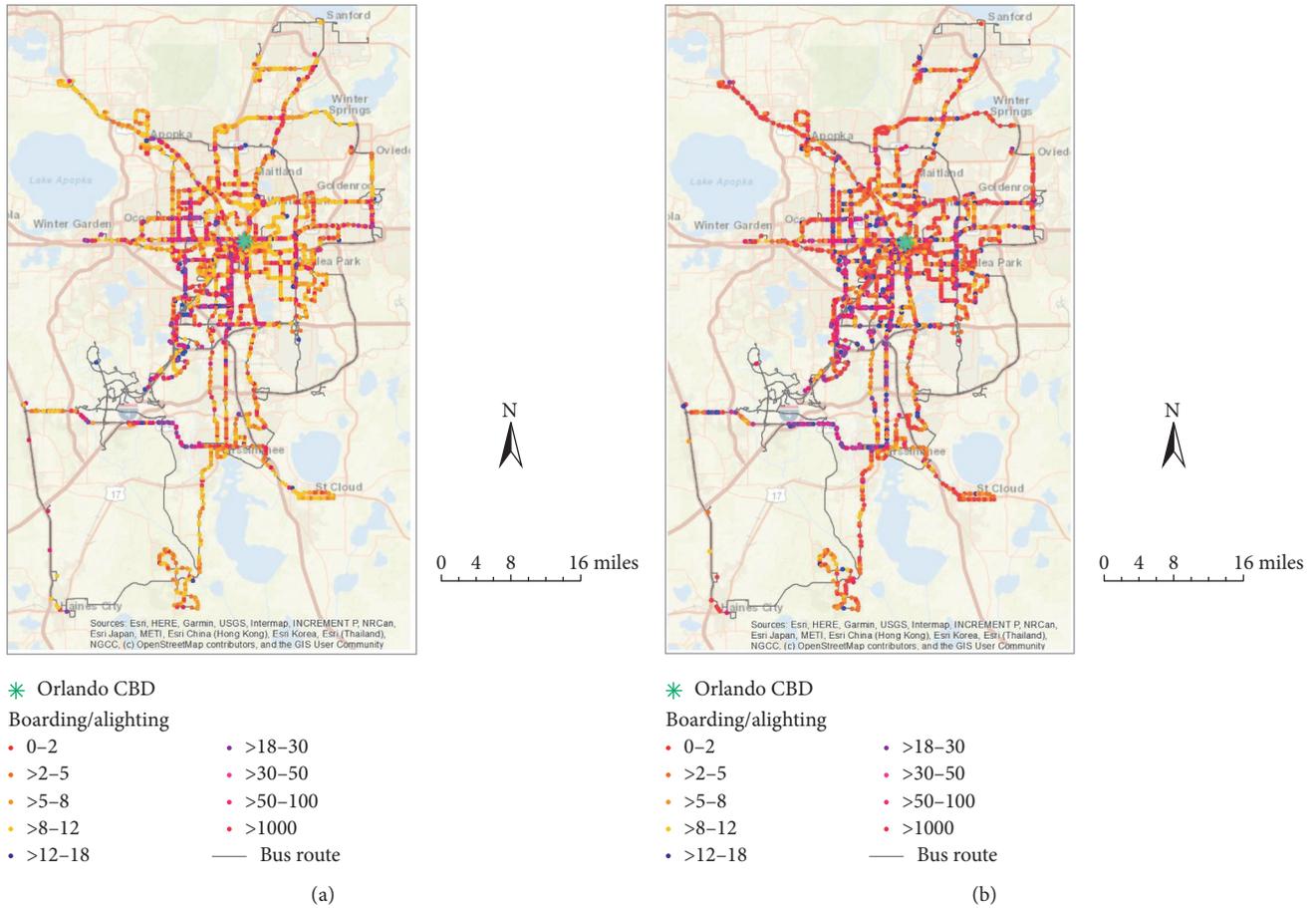


FIGURE 2: Public transit Ridership for April-15 quarter. (a) Boarding and (b) alighting.

TABLE 1: Summary statistics of Lynx bus ridership (August 2013 to April 2015).

Time-period	Quarter name	Boarding		Alighting	
		Mean	Standard deviation	Mean	Standard deviation
1	August-13	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

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. Built Environment Variables. Several built environment variables are found to influence boarding and alighting. Land use mix variable is positively associated with boarding and alighting in SEM and SAR models. The result is quite encouraging toward promoting 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. The result confirms our expectation that large share of transit ridership occurs near the urban core.

5.5.1. 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

TABLE 2: Descriptive statistics of exogenous variables.

Variable name	Variable description	Minimum	Maximum	Mean
<i>Stop-level attributes</i>				
Headway (min)	Ln (headway)	0.105	4.094	3.491
Transportation infrastructure around the stop				
Bus route length in a 600 m buffer	Bus route length in kilometers (bus route length in 600 m buffer/10)	0.105	6.064	0.513
Side walk length in an 800 m buffer	Side walk length in kilometers	0.000	13.270	3.159
Secondary highway length in an 800 m buffer	Secondary highway length in 800 m buffer/total road length in 800 m buffer	0.000	1.000	0.339
Rail road length in an 800 m buffer	Rail road length in kilometers	0.000	6.038	0.307
Local road length in an 800 m buffer	Local road length in 800 m buffer/total road length in 800 m buffer	0.000	1.000	0.653
Presence of shelter in bus stop	(1 = yes and 0 = no)	0.000	1.000	0.227
<i>Built environment around the bus stop</i>				
Land use mix area in an 800 m buffer	Land use mix = $[-\sum_k p_k (\ln p_k) / \ln N]$, where k is the category of land use, p is the proportion of the developed land area devoted to a specific land use, and N is the number of land-use categories within 1 mile buffer of the roadway segment	0.001	0.810	0.501
Household (HH) 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 district (CBD) distance	(Central business area distance in km from bus stop)/10	0.003	5.058	1.178
<i>Sociodemographic and socioeconomic variables in census tract</i>				
Age 0 to 17 years	Ln (people age 0 to 17 years)/census area	-6.584	3.682	-0.282
Age 65 and up	Ln (people age 65 and up)/census area	-6.356	3.231	-1.067
High income (>80 k)	Ln (high-income people (>80 k)/census area)	-8.526	2.740	-1.827
Household rent	Ln (household rent/census area)	-7.210	3.978	-0.528
<i>Spatial and spatiotemporal effect</i>				
Temporal lagged variable for boarding	Ln of temporal lagged variable for boarding	0.000	8.857	1.459
Temporal lagged variable for alighting	Ln of temporal lagged variable for alighting	0.000	8.820	1.490
Spatiotemporal lagged variable for boarding in an 800 m buffer	Ln of spatiotemporal lagged variable for boarding in an 800 m buffer	0.000	9.623	3.811
Spatiotemporal lagged variable for alighting in an 800 m buffer	Ln of spatiotemporal lagged variable for alighting in an 800 m buffer	0.000	9.584	3.815

proportion of people aged between 0 and 17 years is observed to positively influence boarding in both SEM and SAR models. The result is intuitive as an increase in the proportion of young individuals is shown; 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 the proportion of individuals 65 and higher is associated with a reduction in boarding and alighting (except for alighting in SAR model). The result while counterintuitive at 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.

5.5.2. Spatial and Spatiotemporal Effects. The temporal lagged variables are positively associated with boarding and alighting ridership for SEM and SAR models. On the other hand, spatiotemporal 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 represents a system where the same ridership in the urban region is being split across stops.

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 autocorrelated error term. In the SAR model, the spatial autoregressive coefficient indicates a significant impact of unobserved effects.

TABLE 3: Spatial Error Model (SEM) and Spatial Lag Model (SAR) results.

Variable name	Boarding				Alighting			
	SEM		SAR		SEM		SAR	
	Estimates	<i>t</i> -stat	Estimates	<i>t</i> -stat	Estimates	<i>t</i> -stat	Estimates	<i>t</i> -stat
Constant	2.423	19.260	1.723	172.504	3.084	27.137	2.090	182.354
<i>Stop-level attributes</i>								
Headway (Ln of headway)	-0.526	-29.285	-0.403	-3.473	-0.510	-28.956	-0.346	-3.894
<i>Transportation infrastructure around the bus stop</i>								
Bus route length in a 600 m buffer	0.307	7.222	0.208	5.502	0.303	7.623	0.208	5.555
Sidewalk length in an 800 m buffer	0.044	5.360	— ²	—	0.058	7.383	—	—
Secondary highway length in a 600 m buffer	0.769	7.047	0.677	36.325	—	—	—	—
Local road length in an 800 m buffer	0.708	10.919	0.528	-16.331	—	—	—	—
Railroad length in an 800 m 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
<i>Built environment around the stop</i>								
Land use mix area in an 800 m 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 district area distance (km)	-0.110	-5.460	-0.064	-3.920	-0.148	-6.901	-0.055	-3.517
<i>Sociodemographic and socioeconomic variables in census tract</i>								
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 (>80 k)	-0.054	-4.122	—	—	-0.067	-5.178	-0.048	-3.941
HH rent	0.051	2.518	—	—	0.065	3.114	0.056	1.741
<i>Spatial and spatiotemporal effect</i>								
Temporal lagged variable	0.052	13.320	0.050	0.349	0.051	13.513	0.048	0.344
Spatiotemporal lagged variables in an 800 m buffer	-0.032	-12.685	-0.025	-6.305	-0.027	-11.098	-0.023	-6.087
Spatial autocorrelated term	1.617	39.268	—	—	1.710	104.83	—	—
Spatial autoregressive term	—	—	0.336	174.130	—	—	0.374	200.094

¹We restrict ourselves to spatial random effects model as opposed to developing a spatial fixed effects model for multiple reasons. First, in a spatial fixed effects model, several additional parameters are estimated to account for bus-stop-specific effects. In a dataset with over 3000 stops, this would mean estimating a large number of parameters. The presence of such large number of parameters might lead to overfitting of the data. Second, in the presence of bus-stop-specific fixed effects, the impact of other variables that are common across the system is unlikely to be meaningful. Therefore, the results from such an exercise are not transferable to the future or other locations in any meaningful form. Hence, we have not considered spatial fixed effects models. ²“—” means insignificant at 90% confidence interval.

5.7. Model Validation. A holdout sample of 250 stops (250 * 6 = 1500 observations) was set aside for validation purposes. We used both SEM and SAR models to compute predicted boarding and alighting at the stop 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.

6. Conclusion

Toward encouraging a higher level of public transport adoption, it is of utmost importance to examine the critical factors that contributes to public transport ridership. An important tool to evaluate the influence of the critical factors and the future public transit investments opportunities is the application of statistical models. Drawing on stop-level

boarding and alighting data for 6 four-month periods for Greater Orlando region from May 2013 to April 2015, the current study estimated spatial panel models that accommodate for the impact of spatial and temporal observed and unobserved factors.

Two spatial models, (1) Spatial Error Model (SEM) and (2) Spatial Lag Model (SAR), were 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 spatiotemporal lagged variables. The model fit measures clearly confirmed our hypothesis that spatial unobserved effects influence boarding and alighting through the presence of spatial autocorrelated error term in the SEM model and the spatial autoregressive coefficient in the SAR model. Further, the validation exercise results confirmed that the two models performed adequately. The outcomes of the estimated models can be employed to evaluate the changes in the public transport demand due to the changes in the future supply (adding or removing stops in the system). The optimal ridership could be predicted by employing the results of the estimated models while considering the spatial

location of the proposed stops in relation to the existing bus stops (distance matrix). To be sure, the research is not without limitations. In our model, we have considered both boarding and alighting models 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 is a potential avenue for future research. In the future, it would be beneficial to examine how individual-level behavioral preferences for using private vehicle can be incorporated within transit ridership frameworks. There is also a need to accommodate for the endogeneity between transit agency decisions (with regard to headway and new bus routes) and ridership. In the current study, we have estimated SAR and SEM models which account for spatial endogenous interactions and spatial interactions in the error structures. In the future, it might be interesting to estimate a spatial Durbin model which takes into account the advantage of both SAR and SEM models while also allowing for flexible spillover effects.

Data Availability

The dataset used for the study is confidential.

Disclosure

All opinions are only those of the authors. An earlier version of this paper was presented at the 2018 Transportation Research Board (TRB) Annual Meeting.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors acknowledge Mr. William Gant from Lynx and Mr. Jeremy Dilmore from Florida Department of Transportation (FDOT) for helping with data acquisition. The authors also thank FDOT for funding this study.

References

- [1] D. Schrank, B. Eisele, and T. Lomax, *TTI's 2012 Urban Mobility Report*, The Texas A&M University System, College Station, TX, USA, 2012.
- [2] E. Jaffe, "The big U.S. transportation infrastructure projects to watch in 2016," 2015, <https://www.citylab.com/transportation/2015/12/us-infrastructure-projects-2016-transportation/421431/>.
- [3] The Transport Politics (TPP), "Major transit projects under consideration in US and Canada," 2018, <https://www.thetransportpolitic.com/under-consideration/>.
- [4] P. Brinkmann, "New stats show Orlando grew faster than 30 biggest metros," 2016, <https://www.orlandosentinel.com/business/brinkmann-on-business/os-orlando-population-growth-20160324-story.html>.
- [5] V. Chakour and N. Eluru, "Examining the influence of stop level infrastructure and built environment on bus ridership in Montreal," *Journal of Transport Geography*, vol. 51, pp. 205–217, 2016.
- [6] S. S. Pulugurtha and M. Agurla, "Assessment of models to estimate bus-stop level transit ridership using spatial modeling methods," *Journal of Public Transportation*, vol. 15, no. 1, p. 3, 2012.
- [7] J. P. Elhorst, "Spatial panel data models," in *SpringerBriefs in Regional Science Spatial Econometrics* Springer, Berlin, Germany, 2014.
- [8] A. Faghih-Imani 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, pp. 218–227, 2016.
- [9] A. Chakraborty and S. Mishra, "Land use and transit ridership connections: implications for state-level planning agencies," *Land Use Policy*, vol. 30, no. 1, pp. 458–469, 2013.
- [10] B. D. Taylor, 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*, vol. 43, no. 1, pp. 60–77, 2009.
- [11] C. Chen, 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, pp. 1893–1908, 2011.
- [12] G. Currie and J. Phung, "Transit ridership auto gas prices and world events," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1992, no. 1, pp. 3–10, 2007.
- [13] R. L. Hickey, "Impact of transit fare increase on ridership and revenue," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1927, no. 1, pp. 239–248, 2005.
- [14] B. W. Lane, "The relationship between recent gasoline price fluctuations and transit ridership in major US cities," *Journal of Transport Geography*, vol. 18, no. 2, pp. 214–225, 2010.
- [15] B. W. Lane, "A time-series analysis of gasoline prices and public transportation in US metropolitan areas," *Journal of Transport Geography*, vol. 22, pp. 221–235, 2012.
- [16] J. W. Mattson, *Effects of Rising Gas Prices on Bus Ridership for Small Urban and Rural Transit Systems: Upper Great Plains Transportation Institute*, North Dakota State University, Fargo, North Dakota, 2008.
- [17] T. Tsekeris and S. Voß, "Public transport and road pricing: a survey and simulation experiments," *Public Transport*, vol. 2, no. 1–2, pp. 87–109, 2010.
- [18] J. De Oña, R. De Oña, L. Eboli, and G. Mazzulla, "Perceived service quality in bus transit service: a structural equation approach," *Transport Policy*, vol. 29, pp. 219–226, 2013.
- [19] L. Eboli and G. Mazzulla, "Service quality attributes affecting customer satisfaction for bus transit," *Journal of Public Transportation*, vol. 10, no. 3, p. 2, 2007.
- [20] M. Rahman, S. Yasmin, and N. Eluru, "Evaluating the impact of a newly added commuter rail system on bus ridership: a grouped ordered logit model approach," *Transportmetrica A: Transport Science*, vol. 15, no. 2, pp. 1081–1101, 2019.
- [21] S. Ryan and L. F. Frank, "Pedestrian environments and transit ridership," *Journal of Public Transportation*, vol. 12, no. 1, p. 3, 2009.
- [22] A. Johnson, "Bus transit and land use: illuminating the interaction," *Journal of Public Transportation*, vol. 6, no. 4, p. 2, 2003.

- [23] J. Dill, M. Schlossberg, L. Ma, and C. Meyer, "Predicting transit ridership at the stop level: the role of service and Urban form," in *Proceedings of the Paper presented at the 92nd annual meeting of the Transportation Research Board*, Washington, NJ, USA, April 2013.
- [24] N. Estupiñán and D. A. Rodríguez, "The relationship between urban form and station boardings for Bogotá's BRT," *Transportation Research Part A: Policy and Practice*, vol. 42, no. 2, pp. 296–306, 2008.
- [25] T. Banerjee, D. Myers, C. Irazabal et al., *Increasing Bus Transit Ridership: Dynamics Of Density Land Use and Population Growth (No. FHWA/CA/OR-2006/02)*, Division of Research and Innovation, Los Angeles, CA, USA, 2004.
- [26] L. Tang 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, pp. 146–161, 2012.
- [27] J. P. Elhorst, "Specification and estimation of spatial panel data models," *International Regional Science Review*, vol. 26, no. 3, pp. 244–268, 2003.
- [28] K. P. Burnham and D. R. Anderson, "Multimodel inference," *Sociological Methods & Research*, vol. 33, no. 2, pp. 261–304, 2004.
- [29] M. Abkowitz and J. Tozzi, "Transit route characteristics and headway-based reliability control," *Transportation Research Record*, vol. 1078, pp. 11–16, 1986.
- [30] Y. Ding and S. I. Chien, "Improving transit service quality and headway regularity with real-time control," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1760, no. 1, pp. 161–170, 2001.
- [31] S. I-Jy Chien, "Optimization of headway vehicle size and route choice for minimum cost feeder service," *Transportation Planning and Technology*, vol. 28, no. 5, pp. 359–380, 2005.
- [32] G. K. Kuah and J. Perl, "Optimization of feeder bus routes and bus-stop spacing," *Journal of Transportation Engineering*, vol. 114, no. 3, pp. 341–354, 1988.
- [33] M. Ruan, "Probability-based bus headway regularity measure," *IET Intelligent Transport Systems*, vol. 3, no. 4, pp. 400–408, 2009.
- [34] M. A. Turnquist, "Strategies for improving reliability of bus transit service," *Computers and Operations Research*, vol. 1, pp. 201–211, 1974.