Examining Determinants of Rail ridership: A Case Study of the Orlando SunRail System

Moshiur Rahman
PhD Candidate
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: 321-276-7580, Fax: 407-823-3315
Email: moshiur@knights.ucf.edu

Shamsunnahar Yasmin*
Postdoctoral Associate
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: 407-823-4815, Fax: 407-823-3315
Email: shamsunnahar.yasmin@ucf.edu

Naveen Eluru
Associate Professor
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: 407-823-4815, Fax: 407-823-3315
Email: naveen.eluru@ucf.edu

Date: March 2019

* Corresponding author

Abstract

The current study contributes to literature on transit ridership by considering daily boarding and alighting data from a recently launched commuter rail system of Orlando - SunRail. The analysis is conducted based on daily boarding and alighting data for ten months for the year 2015. With the availability of repeated observations for every station the potential impact of common unobserved factors affecting ridership variables are considered. The current study develops an estimation framework, for boarding and alighting separately, 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, transportation infrastructure, transit infrastructure, land use, built environment, 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.

Keywords: SunRail, Boarding, Alighting, Panel linear regression model, Policy analysis.

Introduction

The economic development and the associated growth in household incomes in the United States post Second World War resulted in increased household and vehicle ownership, population and employment decentralization and urban sprawl. While population has increased nearly 72% between 1950 and 1990, aggregate population in central cities declined by 17% (Baum-Snow, 2007). In terms of commute to central cities, only 38% of commute trips in 2000 were destined to central cities; a reduction from 66% in 1960 (Baum-Snow, 2010). These population and employment changes resulted in a drastic reduction in public transit ridership. In fact, in fifty years since 1940, transit ridership in the US decreased by 31% - a drop of about 4 billion trips (Baum-Snow & 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).

In recent years, transportation professionals and policy makers have recognized the potentially beneficial role of public transit in enhancing mobility for urban residents while also 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 (Megan Barber, 2016). The investments include bus and subway system expansions, streetcar additions, light rail and commuter rail system addition (and expansion). With the increasing investments in public transit, federal transit administration and various agencies supporting these initiatives are interested in examining the influence of these investments on transit ridership. A major analytical tool to analyze the impact of these investment 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 from 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 system provides viable transit options for Central Florida residents who live along the Interstate-4 (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. It is beneficial to evaluate factors influencing SunRail ridership to promote a viable and sustainable transit alternative. The analysis is conducted based on daily boarding and alighting data for ten months for the year 2015. 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 and sociodemographic and weather variables on ridership. Further, with the availability of repeated observations for every station, the potential impact of common unobserved factors affecting ridership variables is also considered. The current study develops an estimation framework that accounts for these unobserved effects at multiple levels: station, station-week and station day. 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.

The remainder of the paper is organized as follows. A brief overview of earlier research is described in the literature review section. The methodology section briefly outlines the econometric framework considered. The data section presents data source, data preparation for modeling. In the model estimation results section, we discuss the model results and validation. Policy analysis results to demonstrate the implications of the developed models are discussed in next section. Finally, the conclusion section provides a summary of the findings and concludes our paper.

Literature Review

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; Rahman et al., 2019 for a review). In our review, we focus our attention only on the rail alternative. Table 1 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.

[Table 1 near here]

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 sociodemographics. 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, reliability and land use patterns surroundings the rail station affecting ridership.

Current Study in Context

Based on the literature review, 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, weekday). The most appropriate model structure for the unobserved factors is guided by intuition and data fit metrics. The study is based on ridership data for ten months for the year 2015. Separate models are developed for boarding and alighting. The model developed is validated using hold-out sample data.

Methodology

The focus of our study is to model 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 boardings or alightings. Then, the equation system for modeling boardings/alightings may be written as follows:

$$y_{ir} = (\boldsymbol{\beta_r} + \boldsymbol{\delta_{ir}} + \boldsymbol{\gamma_{qr}})\boldsymbol{x}_{ir} + \varepsilon_q \tag{1}$$

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:

$$L_{qr} = \int_{\Omega} \left(\prod_{q=1}^{Q} \prod_{i=1}^{N} (y_{ir})\right) d\Omega \tag{2}$$

Finally, the log-likelihood function is:

$$LL = \sum_{q} Ln(L_{qr}) \tag{3}$$

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 and 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 3.

Data

Orlando metropolitan region is the 24th largest metropolitan area in the United States. Greater Orlando region has experienced rapid growth in recent years. 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.

Data Description

In our study, the rail ridership analysis is focused on the 12 active stations shown in Figure 1. The main data source of SunRail daily ridership is the SunRail authority. For the purpose of our analysis, 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 2. 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.

[Figure 1 near here]
[Table 2 near here]

Independent Variable Generation

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.

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 around each station. 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. Land use mix is computed as "Land-use mix = $\left[\frac{-\sum_k(\boldsymbol{p_k}(\boldsymbol{lnp_k}))}{\boldsymbol{lnN}}\right]$", 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 in a buffer. Sociodemographic variables considered include number of households with zero, one and two vehicle ownership level. Sociodemographic variables are computed within the influence area of Sunrail stations at census tract level. Finally, weather variables considered include temperature, average wind speed and rainfall.

Table 3 offers a summary of the sample characteristics of the exogenous factors used in the estimation data set. Table 3 represents the definition of variables considered for final model estimation along with the minimum, maximum and average values of the exogenous variables. The final specification of the model development was based on removing the statistically insignificant variables in a systematic process based on statistical confidence (95% confidence level). The specification process was also guided by prior research and parsimony considerations. In estimating the models, several functional forms and variable specifications are explored. The functional form that provided the best result is used for the final model specifications. For determining the appropriate 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.

[Table 3 near here]

Model Estimation Results

Model Specification and Overall Measures of Fit

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 $2[LL_U - LL_R]$, where LL_U and 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 both boarding and alighting models.

Variable Effects

The estimated results for boarding and alighting are presented in Table 4. 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 4 near here]

Temporal and Seasonal Variables

The day of the week variables offer interesting results. Specifically, the result indicates 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. We considered the months of September and October as the base for the month variable. We find that, compared to the base months, the month of March is associated with highest positive impact on boarding and alighting. It is also observed that the association of various months with boarding and alighting are very similar.

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 affects both boarding and alighting ridership. The parking facilities have significantly higher impact on alighting relative to boarding.

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.

Sociodemographic Variables
Several socioeconomic variables 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 influence area around the station at a census tract level is negatively associated with boarding and alighting. While the result is counterintuitive 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.

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.

Station Specific Unobserved Effects
In estimating SunRail daily average ridership models (for boarding and alighting), we estimated several 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 4. 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.

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 372 (31*12 = 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.

Policy Analysis
The parameter effects of exogenous variables in Table 4 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 2 along with the observed daily ridership.

[Figure 2 near here]

Several observations can be made from Figure 2. 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 result 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 as expected because the original variables were 0 for these stations (an increase by percentage does not result in an increase). The elasticity analysis conducted provides an illustration on how the proposed model can be applied for policy evaluation for SunRail ridership.

Conclusions

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 availability 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 computed for 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 model estimation results clearly highlighted how the model that considers unobserved factors offered improved fit. In terms of model estimates, the day of the week variables offer interesting results. Specifically, the result indicates 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 positive impact on boarding and alighting. Local highway length and number of bus stops within a 1500m buffer area around rail stations presents a significant positive impact on boarding and alighting. The availability of free parking space at SunRail stations also positively affected 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. 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.

The study is not without limitations. The data used in the study, while is quite rich with several repetitions per station, would benefit from expanding the time frame of data (to multiple years). As more data becomes available, this would be an appropriate direction for future research.

Acknowledegments
The authors would like to thank SunRail and Mr. Jeremy Dilmore from Florida Department of Transportation (FDOT) for helping with data acquisition. The authors would also like to thank FDOT for funding this study. All opinions are only those of the authors.

References

Baum-Snow, N. (2007). Did highways cause suburbanization? *The Quarterly Journal of Economics, 122*(2), 775-805.

Baum-Snow, N. (2010). Changes in transportation infrastructure and commuting patterns in US metropolitan areas, 1960–2000. *The American economic review, 100*(2), 378.

Baum-Snow, N., & Kahn, M. E. (2000). The effects of new public projects to expand urban rail transit. *Journal of Public Economics, 77*(2), 241-263. doi:http://dx.doi.org/10.1016/S0047-2727(99)00085-7

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

Beko, J. (2004). Estimation of demand functions for services of public railway passenger transportation: evidence from Slovenia. *PROMET-Traffic&Transportation, 16*(1), 5-14.

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

Cervero, R. (2002). Built environments and mode choice: toward a normative framework. *Transportation Research Part D: Transport and Environment, 7*(4), 265-284.

Chakour, V., & Eluru, N. (2016). Examining the influence of stop level infrastructure and built environment on bus ridership in Montreal. *Journal of Transport Geography, 51*, 205-217. doi:http://dx.doi.org/10.1016/j.jtrangeo.2016.01.007

Chen, C., Varley, D., & Chen, J. (2011). What affects transit ridership? A dynamic analysis involving multiple factors, lags and asymmetric behaviour. *Urban Studies, 48*(9), 1893-1908.

Chiang, W.-C., Russell, R. A., & Urban, T. L. (2011). Forecasting ridership for a metropolitan transit authority. *Transportation Research Part A: Policy and Practice, 45*(7), 696-705. doi:http://dx.doi.org/10.1016/j.tra.2011.04.016

Choi, J., Lee, Y. J., Kim, T., & Sohn, K. (2012). An analysis of Metro ridership at the station-to-station level in Seoul. *Transportation, 39*(3), 705-722.

Gkritza, K., Karlaftis, M. G., & Mannering, F. L. (2011). Estimating multimodal transit ridership with a varying fare structure. *Transportation Research Part A: Policy and Practice, 45*(2), 148-160.

Gutiérrez, J., Cardozo, O. D., & García-Palomares, J. C. (2011). Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. *Journal of Transport Geography, 19*(6), 1081-1092.

Huang, Z., Zhang, M., & Liu, X. (2017). Estimating light-rail transit peak-hour boarding based on accessibility at station and route levels in Wuhan, China. *Transportation Planning and Technology, 40*(5), 624-639.

Kain, J. F., & Liu, Z. (1999). Secrets of success: assessing the large increases in transit ridership achieved by Houston and San Diego transit providers. *Transportation Research Part A: Policy and Practice, 33*(7–8), 601-624. doi:https://doi.org/10.1016/S0965-8564(99)00009-9

Kim, S., Ulfarsson, G. F., & Todd Hennessy, J. (2007). 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, 41*(6), 511-522. doi:http://dx.doi.org/10.1016/j.tra.2006.11.001

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

Kuby, M., Barranda, A., & Upchurch, C. (2004). Factors influencing light-rail station boardings in the United States. *Transportation Research Part A: Policy and Practice, 38*(3), 223-247. doi:http://dx.doi.org/10.1016/j.tra.2003.10.006

Lane, B. W. (2008). Significant characteristics of the urban rail renaissance in the United States: A discriminant analysis. *Transportation Research Part A: Policy and Practice, 42*(2), 279-295. doi:http://dx.doi.org/10.1016/j.tra.2007.10.001

Lane, C., DiCarlantonio, M., & Usvyat, L. (2006). Sketch models to forecast commuter and light rail ridership: Update to TCRP report 16. *Transportation Research Record: Journal of the Transportation Research Board*(1986), 198-210.

Lee, K. S., Eom, J. K., You, S. Y., Min, J. H., & Yang, K. Y. (2015). An Empirical Study on the Relationship between Urban Railway Ridership and Socio-economic Characteristics. *Procedia Computer Science, 52*, 106-112. doi:http://dx.doi.org/10.1016/j.procs.2015.05.037

Liu, C., Erdogan, S., Ma, T., & Ducca, F. W. (2016). How to Increase Rail Ridership in Maryland: Direct Ridership Models for Policy Guidance. *Journal of Urban Planning and Development, 142*(4), 04016017.

MEGAN BARBER. (2016, December 29). 9 U.S. transportation projects to watch in 2017. Retrieved from https://www.curbed.com/2016/12/29/14060810/us-transportation-construction-projects-2017

O'Sullivan, A. (1996). *Urban economics*: Irwin.

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

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., . . . White, P. (2006). The demand for public transport: The effects of fares, quality of service, income and car ownership. *Transport Policy, 13*(4), 295-306. doi:http://dx.doi.org/10.1016/j.tranpol.2005.12.004

Rahman M., S. Yasmin, and N. Eluru, (2019), "Evaluating the Impact of a Newly Added Commuter Rail System on Bus Ridership: A Grouped Ordered Logit Model Approach" forthcoming Transportmetrica A: Transport Science

Saur, G. J., Lee, R., & Gray, C. (2004). *New method for transit ridership forecasting.* Paper presented at the ITE 2004 Annual Meeting and Exhibit.

Schrank, D., Eisele, B., & Lomax, T. (2012). TTI's 2012 urban mobility report. *Texas A&M Transportation Institute. The Texas A&M University System*, 4.

Sun, L.-S., Wang, S.-W., Yao, L.-Y., Rong, J., & Ma, J.-M. (2016). Estimation of transit ridership based on spatial analysis and precise land use data. *Transportation Letters, 8*(3), 140-147.

Taylor, B. D., Miller, D., Iseki, H., & Fink, C. (2009). 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. doi:https://doi.org/10.1016/j.tra.2008.06.007

Voith, R. (1997). Fares, Service Levels, and Demographics: What Determines Commuter Rail Ridership in the Long Run? *Journal of Urban Economics, 41*(2), 176-197. doi:http://dx.doi.org/10.1006/juec.1996.1092

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

Zhao, J., Deng, W., Song, Y., & Zhu, Y. (2014). Analysis of Metro ridership at station level and station-to-station level in Nanjing: an approach based on direct demand models. *Transportation, 41*(1), 133-155.

List of Figures

List of Tables

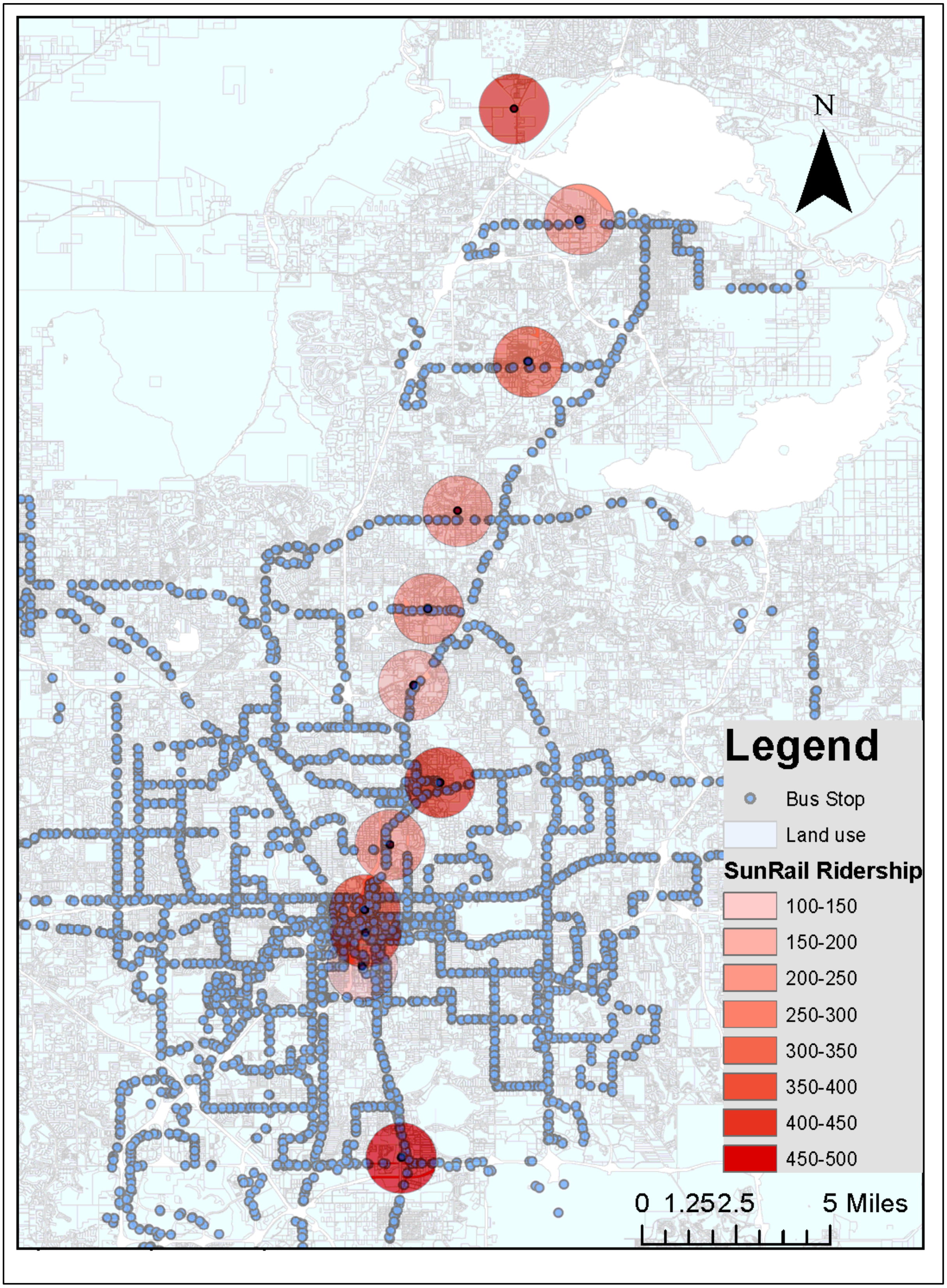


FIGURE 1 SunRail station locations and bus stops

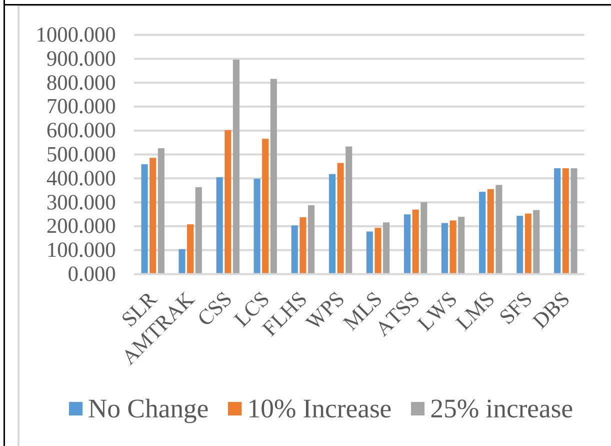
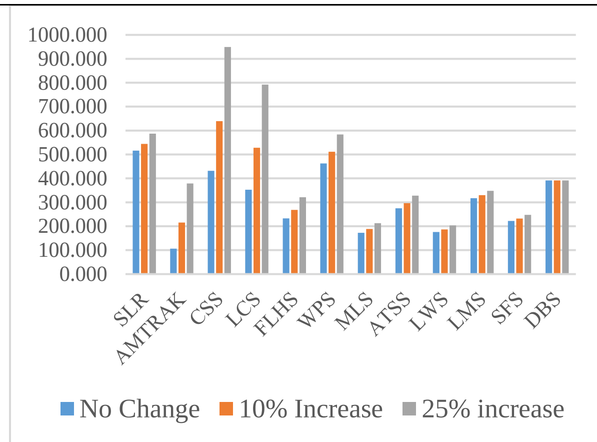
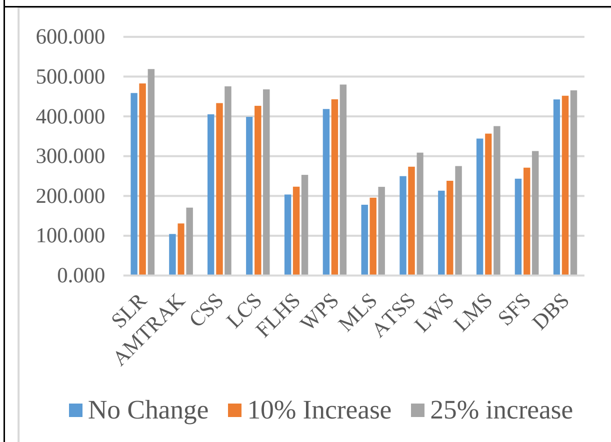
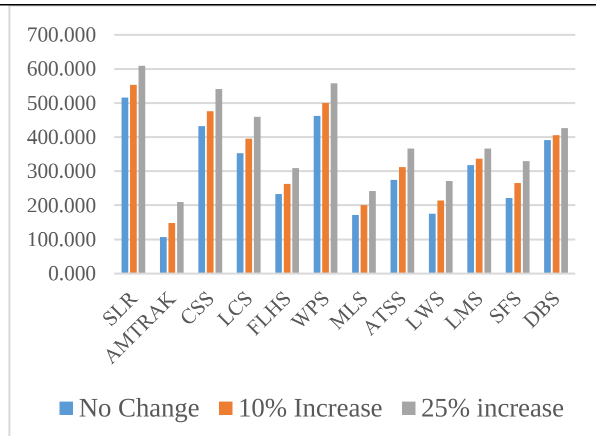
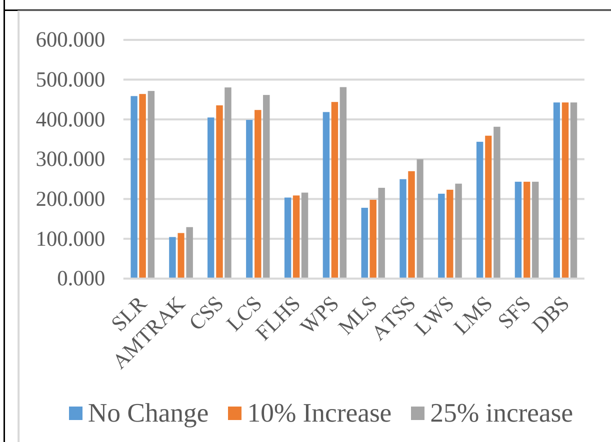
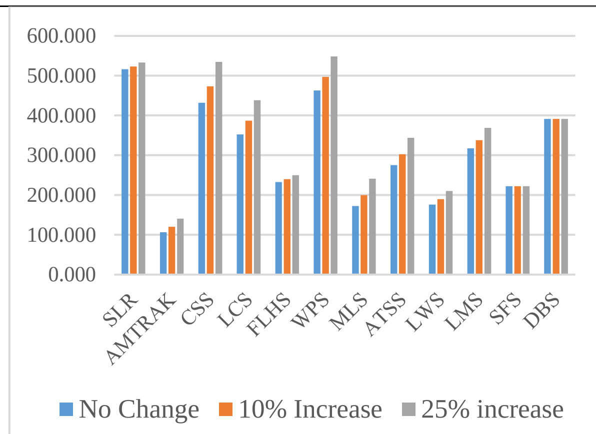
Boarding Ridership	Alighting Ridership
Number of bus stop increased in 1500 m buffer	
	
Land use mix increased in 1500 m buffer	
	
Number of commercial center increased in 1500 m buffer	
	
Note: SLR=Sand Lake Road, AMTRAK Station, CSS=Church Street Station, LCS=Lynx Central Station, FLHS=Florida Hospital Station, WPS=Winter Park Station, MLS=Maitland Station, ATSS=Altamonte Spring Station, LWS=Longwood Station, LMS=Lake Mary Station, SFS=Sanford Station and DBS= BeBary Station	

FIGURE 2 Policy analysis

TABLE 1 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 & Kahn (2000)	Boston, Atlanta, Chicago, Portland, and Washington DC	Multivariate regression	Macro	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Baum-Snow and Kahn (2005)	16 cities of U.S	Regression analysis	Macro	Yes	No	No	Yes	Yes	No	No	Yes
Cervero (2002)	Montgomery County, Maryland	Multinomial mode choice model	Macro	No	Yes	No	Yes	Yes	No	Yes	Yes
Kohn (2000)	Canada	Multiple regression analysis	Macro	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Chen et al. (2011)	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 (1999)	Houston	Cross-section and time series model	Macro	Yes	Yes	Yes	Yes	Yes	No	Yes	No
Kim et al. (2007)	St. Louis Metro Link	Multinomial logit (MNL) model	Macro	No	Yes	Yes	Yes	Yes	No	Yes	Yes
Lane (2008)	35 city of USA	Multiple regression analysis	Macro	No	Yes	Yes	No	No	No	No	Yes
Taylor et al. (2009)	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. (2011)	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. (2011)	Athens, Greece	Generalized least squares method	Macro	Yes	No	No	Yes	Yes	No	Yes	No
Paulley et al. (2006)	Great Britain	Comparison	Macro	No	Yes	No	No	Yes	No	Yes	No

Kuby et al. (2004)	Nine cities in USA	Cross-sectional/Linear regression analysis	Micro, Station level	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Voith (1997)	Southeastern Pennsylvania	Fixed-effects ridership level model	Micro, Station level	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Lee et al. (2015)	Korea	Sketch level ridership models Linear Regression	Micro, Block level	No		No	Yes	Yes	No	No	No
Gutiérrez et al. (2011)	Madrid, Spain	Distance-decay weighted regression model	Micro, Station level	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Huang et al. (2017)	Wuhan, China	Accessibility-weighted ridership model	Micro, Station level	Yes	Yes	No	No	Yes	No	No	Yes
Liu et al. (2016)	Maryland	Direct ridership models (DRM)	Micro, station level	No	Yes	Yes	Yes	Yes	No	No	Yes
Beko (2004)	Slovenia	Multivariate Regression	Micro, Station level	No	No	No	Yes	Yes	No	Yes	No
Saur et al. (2004)	California	Multivariate Regression	Micro, Station level	No	No	Yes	Yes	Yes	No	No	No
Lane et al. (2006)	17 U.S. regions	Multivariate Regression	Micro, Station level	No	No	Yes	Yes	Yes	No	No	Yes
Choi et al. (2012)	Seoul, Korea	Multiplicative model and the Poisson regression model	Micro, Station level	No	Yes	Yes	Yes	Yes	No	No	Yes
Parks et al. (2012)	U.S regions	Linear Regression	Micro, station level	No	Yes	No	Yes	Yes	No	No	Yes
Zhao et al. (2014)	Nanjing, China	Linear, Multiplicative Regression	Micro, station level	No	Yes	No	Yes	Yes	No	No	Yes
Zhang and Wang (2014)	New York	Network Kriging regression	Micro, station level	No	Yes	No	Yes	Yes	No	No	Yes
Sun et al. (2016)	Beijing, China	Direct ridership models (DRM)/Multiple Regression Analysis	Micro, station level	No	No	No	No	No	No	No	Yes

TABLE 2 Summary Statistics for SunRail Average Daily Ridership (January 2015 to October 2015)

Station Name	Boarding		Alighting	
	Mean	**Standard Deviation**	**Mean**	**Standard Deviation**
Sand Lake Station (SLR)	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

TABLE 3 Descriptive Statistics of Exogenous Variables

Variable Name	Variable Description	Minimum	Maximum	Mean
Temporal and Seasonal Variables				
Day of week				
Monday	Rail ridership on Monday	0.000	1.000	0.190
Friday	Rail ridership on Friday	0.000	1.000	0.206
Month of the Year 2015				
January	Rail ridership on January 2015	0.000	1.000	0.094
February	Rail ridership on February 2015	0.000	1.000	0.095
March	Rail ridership on March 2015	0.000	1.000	0.109
April	Rail ridership on April 2015	0.000	1.000	0.105
May	Rail ridership on May 2015	0.000	1.000	0.095
June	Rail ridership on June 2015	0.000	1.000	0.106
July	Rail ridership on July 2015	0.000	1.000	0.111
August	Rail ridership on August 2015	0.000	1.000	0.103
Transportation Infrastructures				
Local roadway length in a 1500 m buffer	Local roadway length in kilometers	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	Number of Commercial centers in a 1500m buffer	0.000	6.000	2.750
Number of Educational centers in a 1500 m buffer	Number of Educational centers in a 1500m buffer	0.000	11.000	4.250
Number of Financial centers in a 1500 m buffer	Number of Financial centers in a 1500m buffer	0.000	55.000	17.833
Land Use mix in a 1500 m buffer	"Land-use mix = $\left[\frac{-\sum_k (p_k(lnp_k))}{lnN}\right]$", 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 in a 1500 m buffer	0.263	0.811	0.638
Sociodemographic Variables				
Vehicle Ownership – No vehicle	Vehicle Ownership – number of HH with No Vehicle in the influence area of station at census tract level	52.000	4532.000	1326.250
Vehicle Ownership – One vehicle	Vehicle Ownership – number of HH with One Vehicle in the influence area of station at census tract level	734.000	15139.000	5425.333

Vehicle Ownership – Two vehicles	Vehicle Ownership – Number of HH with Two Vehicles in the influence area of station at census tract level	2000.000	9189.000	4898.667
Weather Variables				
Average Temperature in air	Average Temperature in air at 2 m height in degree Celsius	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

TABLE 4 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
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	--[1]	--
Land Use mix in a				
1500 m buffer	347.969	20.089	538.002	29.858
Sociodemographic Variables				
Vehicle Ownership - No vehicle	-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

[1] " -- " means insignificant at 95% confidence level

Rainfall	-27.756	-4.028	-25.528	-2.962
Unobserved Effects				
Standard deviation at Station level	2.545	9.689	2.844	14.972