A Joint Panel Binary Logit and Fractional Split Model for Converting Route-Level Transit Ridership Data to Stop-Level Boarding and Alighting Data

Moshiur Rahman
Traffic Engineer
Marlin Engineering
Tel: 321-276-7580
Email: moshiur@knights.ucf.edu
ORCiD number: 0000-0002-3210-6542

Shamsunnahar Yasmin*
Lecturer/Research Fellow
Queensland University of Technology
Centre for Accident Research & Road Safety – Queensland
& Research Affiliate, University of Central Florida
Tel: +61731384677
Email: shams.yasmin@qut.edu.au
ORCiD number: 0000-0001-7856-5376

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
ORCiD number: 0000-0003-1221-4113

---

* Corresponding author

Date: April 2020

ABSTRACT

Detailed ridership analytics requires refined data on transit ridership to understand factors affecting ridership (at the stop and/or route-level). However, detailed data for stop-based boarding and alighting information are not readily available for the entire bus system. Transit agencies usually resort to compiling ridership data on a sample of buses operating on the various routes. We propose an approach to infer stop-level ridership for transit systems that only compile route-level ridership information. A joint model structure of binary logit and fractional split model is proposed to estimate stop-level ridership data sourced from route-level ridership. The model is developed for the Greater Orlando region with ridership data for 8 quadrimesters (four-month time periods) from May 2014 through December 2016. In the presence of repeated data measures, panel version of the joint econometric models for boarding and alighting are estimated. The development of such an analytical framework will allow bus systems with only route-level ridership data to generate stop-level ridership data. The model results offer intuitive results and clearly supports our hypothesis that it is feasible to generate stop-level ridership with route-level ridership data. For transit agencies with ridership data at the stop-level, the proposed model can also be employed to understand how various stops along a route interact with one another toward affecting route-level ridership contributions.

Keywords: *Transit Ridership, Alighting, Boarding, Bus Stop, Route-Level, Joint Model, Binary Model, Fractional Split Model, Panel Joint Model.*

INTRODUCTION

Transit ridership has declined consistently for the last few decades in most of the US metropolitan regions. Specifically, in 2017, transit ridership fell in 31 out of 35 major metropolitan areas in the US transit markets (*1*) while in 2018 public transit ridership reduced about 2% nationally (*2*). Transit agencies are investigating how these declines can be stopped and/or possibly reversed. Several agencies are considering additional investments in transit including adding newer bus routes, commuter/light rail facilities across the nation (*1; 3; 4; 5*). While these investments in transit are encouraging signs, there still needs to be a rigorous analysis of how these investments are influencing ridership and transit accessibility. An important analytical tool for analyzing ridership patterns is the development of statistical and econometric models. Specifically, the emphasis is on developing detailed analytics to understand factors affecting ridership (at stop and/or route-level) and drawing insights to enhance ridership based on these findings. The models developed in this manner have several advantages. First, these studies identify the factors that positively or negatively influence ridership allowing transit agencies to devise strategies to enhance ridership. Second, these frameworks provide a ridership demand prediction platform for newer routes under consideration or modifications to existing routes. Third, these models in simulation mode can be employed to generate estimates of bus occupancy by route in continuous time (*6; 7*). Bus occupancy estimates allow us to identify the ridership peaks and troughs that are useful for determining vehicle fleet allocation and estimation of bus emissions at a fine spatial and temporal resolution.

Not surprisingly, such detailed ridership analytics platform requires refined data on transit ridership. However, depending on the vehicle fleet, ticketing platforms and system size obtaining detailed ridership data for a transit system is far from straightforward. Consider the example of New York City, the largest transit service provider in the US. The transit agency generally provides bus system ridership numbers at the route-level. However, details at a finer resolution of stop-based ridership information (or bus occupancy) are not readily available for the entire bus system. To be sure, the unavailability of such data is typically due to the cost associated with acquiring such data for various bus systems. Some transit systems such as Montreal (in Quebec, Canada) and Orlando (in Florida, USA) do compile stop-level bus ridership. Such data compilation is quite expensive and hence, these transit agencies usually resort to compiling such data on a sample of buses operating on the various routes (as opposed to collecting data for all bus stops across all bus routes). Subsequently, these sampled ridership numbers are weighted to obtain the ridership counts by time-of-day for weekdays and weekends. Montreal bus system, Société de transport de Montréal, conducts a sampling exercise that encompasses 15% of their bus fleet to obtain ridership numbers (*6*). The advances in ticketing technology (such as automated fare collection) and passenger ridership data collection (using automated passenger counts) and their adoption could further enhance ridership data collection across bus systems. However, until all transit systems are upgraded to modern ticketing technology there is a need for computing stop-level ridership for small to mid-range transit systems that typically only estimate route-level ridership.

In this research effort, we propose an approach to infer stop-level ridership for transit systems that only compile route-level ridership information. The stop-level ridership inference will involve the estimation of boarding and alighting at a stop-level using stop-level data compiled in Orlando. The research framework estimates ridership by relating it to various exogenous variables such as headway (or frequency), land-use attributes, transportation and transit infrastructure attributes. The development of such an analytical framework will allow bus systems with only route-level ridership data to generate stop-level ridership data (as an alternative to resorting to changes in current data collection schemes that could be prohibitively expensive). For transit systems that compile stop-level data across a sample of routes, the proposed model can serve as a framework to infer stop ridership across other routes and time periods that are not sampled. Finally, for transit systems that have complete data, the proposed framework can assist in understanding the relation between the contributions from various stops across a route.

The proposed framework relates route-level ridership to stop-level ridership by developing a joint econometric model. In the joint system, the first component identifies stops with non-zero ridership employing a binary logit model. The second component adopts a fractional split structure to estimate the proportion of ridership for each of the non-zero stops (identified in the previous step) along the route. Specifically, we propose to consider a multinomial logit based fractional split formulation to examine the fraction of ridership. In our approach, stops along a route serve as alternatives for that specific route and the outcome to be studied is the fraction of ridership in each of those stops ($\frac{Ridership\ at\ the\ stop\ along\ a\ route}{Total\ route\ level\ ridership}$). The econometric framework recognizes that several common unobserved factors could influence the two components. To accommodate for this, a joint model structure of the proposed binary logit and fractional split model is built. In some studies, such joint models are also referred to as hierarchical models with two levels.

The proposed model framework is estimated using data from Greater Orlando region for 8 quadrimesters (four-month time periods) from May 2014 to December 2016. A total of 58 routes are considered for our analysis. Further, given that we have multiple repetitions of data, a panel joint econometric model is developed. In the panel model estimation, several exogenous variables including route-level variables, stop-level attributes, transportation infrastructure variables, transit infrastructure variables, land-use and built environment attributes and sociodemographic variables in the vicinity of the bus stop are considered. The proposed model can be employed by transit agencies without stop-level data to estimate stop-level ridership metrics.

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 and in model estimation results section, we discuss the model results and validation. The policy analysis results are discussed in next section. Finally, the conclusion section identifies potential applications for the proposed model and discussed potential research avenues for the future.

LITERATURE REVIEW

Earlier Research

As described earlier, the major objective of the proposed research effort is to develop a modeling framework to estimate stop-level ridership from route-level ridership data. However, in our review of earlier work, we did not find any studies focused on this conversion task. Hence, our review is organized along two streams of earlier research that offer useful insights for our study exercise. The first stream of earlier research focuses on studies that analysed ridership at route-level or stop-level to identify factors that are likely to influence the split of route-level ridership to stop-level boarding and alighting. The second stream of earlier research is concentrated on methodological approaches suited for our analysis.

Route/Stop-level ridership

Several research efforts have examined the important factors that affect stop/route-level transit ridership for urban regions. Based on their findings, factors affecting stop/route-level ridership can be classified as stop/route-level attributes, transportation infrastructure variables, land-use and built environment characteristics, and sociodemographic variables. Among stop/route-level attributes, headway (or frequency) is identified as an important variable. Reducing headway typically contributes to higher ridership (*8-17*). Other stop/route-level variables such as presence of other bus stops, bus route length, and presence of shelter affect ridership (*9; 15-18*). Transportation infrastructure variables (such as roadway characteristics, sidewalk length, bike lane length) and road network characteristics also significantly affected bus ridership (*9; 15; 16; 19*).

Earlier research has also found significant influence of sociodemographic variables on ridership. Specific variables such as household (HH) income level, HH car ownership level, gender and age distributions in the vicinity of the stop/route are likely to impact bus ridership of that region (*9; 14-16; 20*). Land-use and built environment attributes surrounding the bus stop (such as land-use mix, residential area, recreational area, institutional area, and office area) also affect bus ridership (*9; 12; 14-17; 19; 20*).

Appropriate Research Methodology

The analysis approach in our research requires determining stop-level boarding and alighting while ensuring that the overall boarding and alighting sums to the total route-level ridership. Thus, the variable to be estimated – stop-level boarding/alighting for all stops on the route – should add up to an observed value (route-level ridership). For such a constrained dependent variable, traditional approaches such as linear (or log-linear) regression or count models are not directly applicable as these approaches do not ensure that stop-level variable estimates sum to the total route-level ridership. Hence, we consider an alternative approach, referred to as the fractional split model for our analysis (*21; 22*). The fractional split modeling approach was proposed by Papke and Wooldrigde (*22*) in 1996. In the fractional split approach, ridership proportions at a stop are directly associated with exogenous variables. For example, a stop with the presence of bus shelter has inclination for a larger proportion of ridership (relative to stop without shelter). The approach has received attention in the transportation field in recent years for analyzing crash frequency, crash severity, vehicle speed analysis and aggregate mode choice analysis (*23-27*). While fractional split models have been developed earlier there has been no attempt to develop a model framework that accommodates for repeated measures for fractional split models with *multiple dependent variables*.

Current Study in Context

As described earlier, the proposed research objective of converting route-level ridership to stop-level ridership has not been examined before. The modeling approach has two steps. In the first step, we examine if the stop will have any ridership or not by using a binary logit model. In the second step, for stops with ridership, a fractional split model is developed to identify the proportion of route-level ridership to be allocated to the stop-level ridership. The allocation is undertaken separately for boarding and alighting dimensions. The independent variables in our model components are drawn from the extensive literature review discussed in the previous section. The model development and estimation are conducted using data from the Greater Orlando region. To develop the model, we selected Orlando region which already has stop-level data. Once stop-level data is available, we can easily reconstruct true route-level ridership data. Once the model is developed for a region, it can be applied for other transit systems without stop-level data by making subtle corrections (described later). The ridership data for 8 quadrimesters between May 2014 to December 2016 for 58 routes in the Orlando region is employed. The estimated model developed is validated using holdout sample data of one time period (September to December 2016). The utility of the framework developed is illustrated through a well-designed policy analysis. It is important to highlight here that the proposed joint panel binary logit-fractional split model implementation in the current paper is the first of its kind in transportation and even in econometric literature.

ECONOMETRIC METHODOLOGY

Model Structure

In this section, we provide a formulation of the proposed econometric structure of joint binary logit – fractional split model (BLFS). The focus of the joint model is to simultaneously model "probability of non-zero ridership" and "proportion of ridership" at each stop for a specific route. In the current study context, zero vs. non-zero ridership is modeled by using a binary logit (BL) formulation. Further, the fractional split component of the joint model is examined by using multinomial logit based formulation to examine proportion of ridership in each stop for a specific route. Let $t (t = 1, 2, \dots, T; T = 7)$ be the index for different time period (quadrimester), q $(q = 1, 2, \dots, Q)$ be an index to represent bus stops and r $(r = 1, 2, \dots, R; R = 58)$ be an index to represent route. Hence, $q_r (q = 1_r, 2_r, \dots, Q_r)$ is an index to represent bus stops specific to route r $(r = 1, 2, \dots, R; R = 58)$.

For the joint approach, the non-zero ridership propensity component takes the following form:

$$v^*_{tq_r} = \{(\boldsymbol{\alpha} + \boldsymbol{\gamma}_{tq_r})\boldsymbol{z}_{tq_r} + \varepsilon_{tq_r} + \psi_{tq}\}, \quad u_{tq_r} = 1 \; if \; v^*_{tq_r} > 0, u_{tq_r} = 0 \; otherwise \tag{1}$$

where, $v^*_{tq_r}$ represents the propensity for non-zero ridership at stop q specific to route r in time period t; u_{tq_r} is 1 if stop q_r specific to route r has non-zero ridership for time period t and 0 otherwise. z_{tq_r} is a vector of attributes associated with stop q_r for time period t. α is the vector of corresponding mean effects. γ_{tq_r} is a vector of unobserved factors on non-zero ridership probability of stop q_r for time period t and its associated characteristics assumed to be realization from standard normal distribution: $\gamma_{tq_r} \sim N(0, \varsigma^2_{tq_r})$. ε_{tq_r} is an idiosyncratic error term assumed to be identically and independently standard logistic distributed. ψ_{tq} term generates the time period specific stop-level correlation between equations for non-zero ridership and fraction of ridership.

In the fractional split component of the joint model, the dependent variable is the proportion of ridership in each stop specific to a route for a time-period instead of the actual ridership. The sum of the proportions across each route for a time-period is equal to unity and the fraction of ridership ranges between zero and one. Let y_{tq_r} be the fraction of ridership for stop q_r specific to route r in time period t. Thus, the functional form of the econometric specification for fractional split model can be expressed as:

$$0 \le y_{tq_r} \le 1, \qquad \sum_{t,q_r=1}^{t,Q_r} y_{tq_r} = 1 \tag{2}$$

$$E\left[y_{tq_r}|d_{tq_r}\right] = G_{tq_r}(.) \tag{3}$$

$$0 < G_{tq_r}(.) < 1 \qquad \sum_{t,q_r=1}^{t,Q_r} G_t(.) = 1 \tag{4}$$

where, the ridership fraction y_{tq_r} be a function of a vector d_{tq_r} of relevant explanatory variables associated with attributes of stop q_r for time period t. $G_{tq_r}(.)$ $(q_r = 1_r, 2_r, \dots, Q_r)$ is a predetermined function. The properties specified in Equation 4 for $G_{tq_r}(.)$ warrant that the predicted fractional ridership will range between 0 and 1; and will add up to 1 for each route over a time period.

In the current study context, we assume a categorical discrete outcome structure for G_{tq_r} in the fractional split model of Equation 4 (following (*28*)). Thus, Equation 4 can be rewritten as:

$$E\left[y_{tq_r}|d_{tq_r}\right] = \left\{\left(\beta + \delta_{tq_r}\right)d_{tq_r} + \xi_{tq_r} \pm \psi_{tq}\right\} \tag{5}$$

where, d_{tq_r} is a vector of attributes, β is the corresponding vector of coefficients to be estimated for ridership fraction. δ_{tq_r} is a vector of unobserved factors assumed to be a realization from standard normal distribution: $\delta_{tq_r} \sim N(0, {\nu_{tq_r}}^2)$. ξ_{tq_r} is the random component assumed to follow a Gumbel type-I distribution. The $\pm$ sign in front of common correlation term ψ_{tq} in Equation 5 indicates that the correlation in unobserved factors between non-zero ridership and fraction of ridership in each route may be positive or negative. A positive sign implies that stop specific to a route with non-zero ridership are intrinsically more likely to incur higher proportions of ridership for that specific stop. On the other hand, negative sign implies that stop specific to a route with non-zero ridership are intrinsically incurring lower ridership for that specific stop. To determine the appropriate sign, one can empirically test the models with both $'+'$ and $'-'$ signs independently. The model structure that offers the superior data fit is considered as the final model.

It is important to note here that the unobserved heterogeneity between non-zero ridership component and ridership proportions component can vary across stops for a given time period. Therefore, in the current study, the correlation parameter ψ_{tq} is parameterized as a function of the observed attributes as follows:

$$\psi_{tq} = \Theta_{tq}\varrho_{tq} \tag{6}$$

where, ϱ_{tq} is a vector of exogenous variables, Θ_{tq} is a vector of unknown parameters to be estimated (including a constant).

Model Estimation

In examining the model structure of non-zero ridership and proportions of ridership, it is necessary to specify the structure for the unobserved vectors γ, δ and Θ represented by Ω. In this paper, it is assumed that these elements are drawn from independent realization from normal population: $\Omega \sim N(0, (\varsigma^2, \nu^2, \aleph^2))$. Thus, the equation system for modeling the probability of non-zero ridership takes the following form (conditional on $\boldsymbol{\gamma}_{tq_r}$ and ψ_{tq}):

$$P_{tq_r} = P(u_{tq_r})|(\boldsymbol{\gamma}_{tq_r}, \psi_q) = \frac{exp\{(\boldsymbol{\alpha} + \boldsymbol{\gamma}_{tq_r})\boldsymbol{z}_{tq_r} + \varepsilon_{tq_r} + \psi_{tq}\}}{1 + exp\{(\boldsymbol{\alpha} + \boldsymbol{\gamma}_{tq_r})\boldsymbol{z}_{tq_r} + \varepsilon_{tq_r} + \psi_{tq}\}} \tag{7}$$

The corresponding probability for zero ridership is computed as

$$Q_{tq_r} = 1 - P_{tq_r} \tag{8}$$

Similarly, the probability for fractional split component takes the form (conditional on δ_{tq_r} and ψ_{tq}):

$$R_{tq_r} = G_{tq_r}(y_{tq_r})|(\delta_{tq_r}, \psi_{tq}) = \frac{exp\{(\beta + \delta_{tq_r})d_{tq_r} + \xi_{tq_r} \pm \psi_{tq}\}}{\sum_{t,q_r=1}^{t,Q_r} exp\{(\beta + \delta_{tq_r})d_{tq_r} + \xi_{tq_r} \pm \psi_{tq}\}} \tag{9}$$

Further, conditional on Ω, the likelihood function for the joint probability can be expressed as:

$$\mathcal{L}_r = \int_{\Omega} \prod_{t=1,r}^{T,r} \left[\prod_{t,q_r=1}^{t,Q_r} \left\{ \left(P_{tq_r} * (R_{tq_r})^{y_{tq_r}}\right)^{u_{tq_r}} * (Q_{tq_r})^{(1-u_{tq_r})} \right\} \right] f(\Omega) d\Omega \tag{10}$$

y_{tq_r} is the fraction of ridership for stop q_r specific to route r in time period t. u_{tq_r} is 1 if stop q_r specific to route r has non-zero ridership for time period t and 0 otherwise, Finally, the log-likelihood function is:

$$\mathcal{LL} = \sum_r Ln(L_r) \tag{11}$$

All the parameters in the model are estimated by maximizing the logarithmic function $\mathcal{LL}$ presented in Equation 11. The parameters to be estimated in the joint model are: α, ς, β, ν and $\aleph$. To estimate the proposed joint model, we apply Quasi-Monte Carlo simulation techniques based on the scrambled Halton sequence to approximate this integral in the likelihood function and maximize the logarithm of the resulting simulated likelihood function across individuals (see (*29; 30*) for examples of Quasi-Monte Carlo approaches in literature). The model estimation routine is coded in GAUSS Matrix Programming software.

EMPIRICAL ANALYSIS

The main public transit service in the Greater Orlando region is provided by the Lynx system that serves an area of approximately 2,500 square miles within Orange, Seminole, Osceola and Polk Counties in Central Florida. The bus system operates 77 daily routes with an average weekday ridership of around 105,000. Out of 77 daily routes, 58 routes are considered for our analysis as they provided complete ridership information. Figure 1 shows the study region and 58 routes network in the Orlando region.

Dependent Variable Generation

For the selected 58 routes, we obtained data from Lynx transit authority for 8 quadrimesters between May 2014 through December 2016. The data provided average daily weekday boarding and alighting information. The reader would note that we are using a transit system that already has stop-level data for our analysis. The final sample consists of 38,432 records (4,804 stops × 8 quadrimesters). We set aside data for one time period (September through December 2016) for model validation.

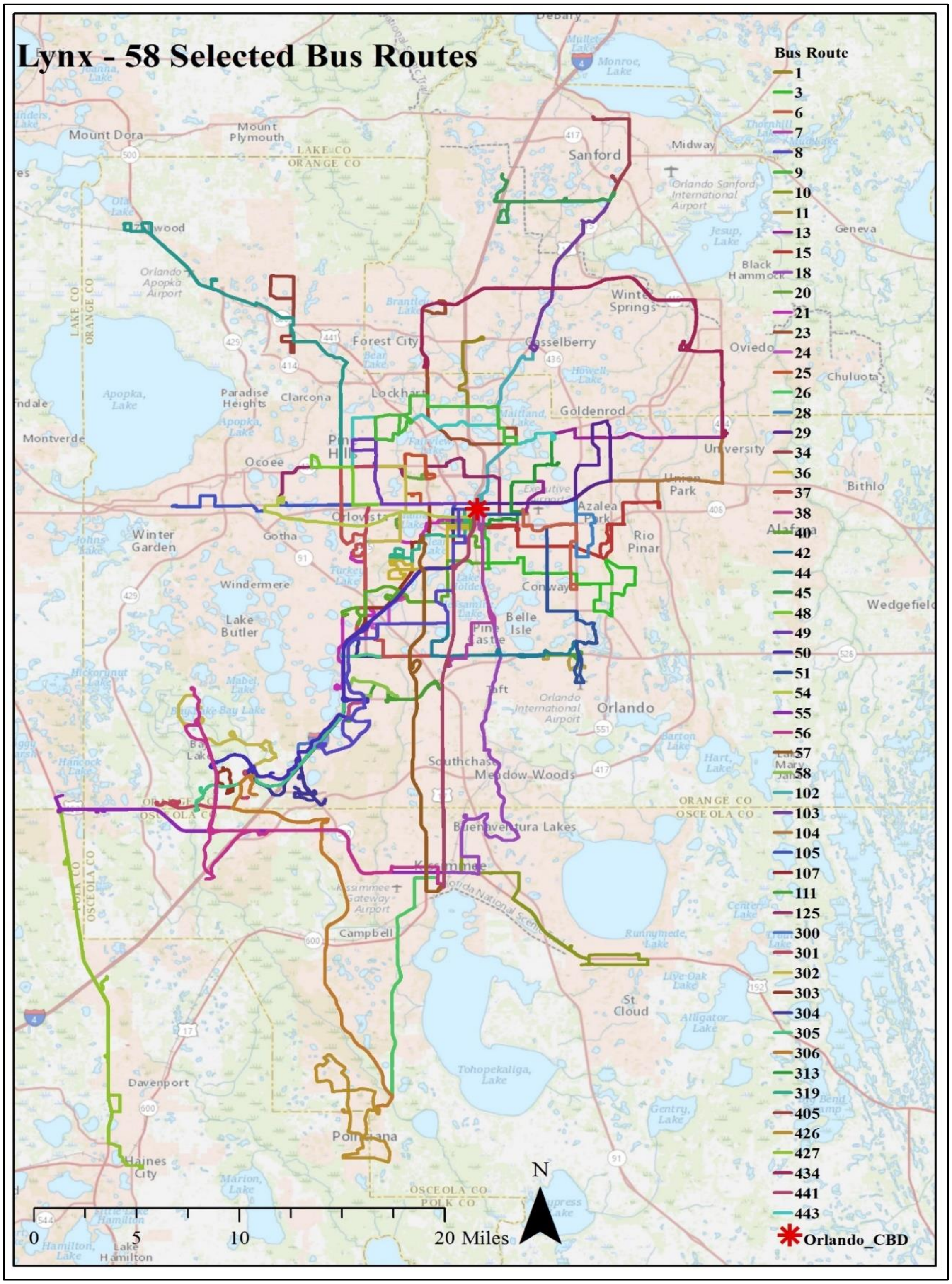


Figure 1: Lynx Bus Network (Selected 58 Routes)

The dependent variable in our analysis is the fraction of ridership at stops along a bus route. The fraction was computed as the ratio of stop-level ridership specific to a route and total route-level ridership. The data we obtained from Lynx provided us total stop-level ridership (boarding and alighting). Employing these data, we determined route-level ridership. Then, the ratio of stop-level ridership variables to the route-level ridership variables was computed to generate the proportion variables. In cases where a stop was part of multiple routes, we allocated stop-level ridership to the route in the ratio of their headway i.e. stops with lower headway were allocated a higher proportion of ridership. The proportion variable for boarding (alighting) ranges from 0 (0) to 1.00 (1.00) across the 58 routes. The reader would note that percentage of zero proportion stops for boarding (alighting) are 7.0% (6.2%). Given the reasonably high share of zero proportion stops, a binary logit model was introduced to identify stops with non-zero ridership prior to developing the fractional split model.

Independent Variable Generation

Several independent variables were compiled for our analysis. The number of bus stops and bus route length was calculated by using Lynx GIS shapefiles. For creating independent variables, we have considered several buffer distances (800m, 600m, and 400m) for each bus stop. The sources of independent variables include 2010 US census data, American Community Survey, Florida Geographic Data Library, and Florida Department of Transportation databases. The 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, rail road length, local road length, sidewalk length, Lynx bus route length, presence of shelter and distance of bus stop from central business district (CBD)), (3) Built environment and land-use attributes (such as institutional area, residential area, recreation area, and office area) (4) sociodemographic variables in the vicinity of the stop (income, vehicle ownership, and age and gender distribution) for each time period and (5) SunRail (commuter rail of Orlando) effect. The descriptive statistics of exogenous variables are presented in Table 1.

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. In determining the appropriate buffer sizes, each variable for a buffer size was systematically introduced (starting from 800m to 400m buffer size) and the buffer variable that offered the best fit was considered in the final specification.

Table 1: Descriptive Statistics of Exogenous Variables

Variable Name	Variable Description	Percentage	Minimum	Maximum	Mean
Stop-Level Attributes					
Headway	Ln (Headway in minutes)	-	-6.908	7.272	4.064
Transportation Infrastructure Around the Stop					
Bus route length in (800m buffer)	Bus route length in kilometers (Bus route length in 800 m buffer/10)	-	0.000	8.710	1.146
Sidewalk length in (400m buffer)	Sidewalk length in kilometers in 400m buffer	-	0.000	20.234	3.844
Secondary highway length (800m buffer)	Secondary highway length in kilometer in 800 m buffer / 10	-	0.000	4.278	0.998
Local road length in (800m buffer)	Local road length in kilometer in 800 m buffer / 10	-	0.000	6.048	2.188
Presence of shelter in bus stop	(1 = Yes and 0 = No)	25.999%	-	-	-
Built Environment Around the Bus Stop					
Residential area (800m buffer)	Proportion of the Residential area = Residential /Total area	-	0.000	0.924	0.410
Institutional area (800m buffer)	Proportion of the Institutional area = Institutional /Total area	-	0.000	0.710	0.043
Recreational area (800m buffer)	Proportion of the Recreational area = Recreational /Total area	-	0.000	0.547	0.011
Office area (800m buffer)	Proportion of the office area = Office/Total area	-	0.000	0.941	0.190
Central business district (CBD) distance (4~8 miles)	(Central business area distance in km from bus stop)/10	23.897%	-	-	-
Sociodemographic Variables in Census Tract					
Female population	Total number of female population in Census Tract/1000	-	0.615	10.526	3.057
High income (>100k)	Percentage of High income HH (>100k)	13.052%	-	-	-
SunRail Effects					
Square distance of SunRail	Square distance of SunRail = (summation of distance of SunRail stations from all bus stops which are within the influence area of SunRail stations in km)^2	-	0.000	1005.098	6.618

MODEL ESTIMATION RESULTS

Model Specification and Overall Measures of Fit

The empirical analysis involves estimation of two different models: 1) an independent binary logit-fractional split (BL-FS) model system, and 2) joint panel BL-FS model with correlation parameterization. These models are estimated for boarding and alighting separately. The independent models (separate BL and FS models) were estimated to establish a benchmark for comparison. Prior to discussing the estimation results, we compare the performance of these models in this section.

We employ the likelihood-ratio (LR) test to determine the best model between independent and joint models. The LR test statistic for a given empirical model is computed by below equation:

$$LR = 2[LL_U - LL_R] \tag{12}$$

where, LL_U and LL_R are the log-likelihood of the unrestricted and the restricted models, respectively.

The log-likelihood values at convergence for the models estimated are as follows: (1) <u>For boarding model</u>: (1.1) Independent BL-FS (with 19 parameters) is -8,649.96 and (1.2) joint panel BL-FS model with correlation parameterization (with 20 parameters) is -8,614.86 and (2) <u>For alighting model</u>: (2.1) Independent BL-FS (with 19 parameters) is -7,794.13 and (2.2) joint panel BL-FS model with correlation parameterization (with 20 parameters) is -7,740.93. The computed value of the LR test is compared with the χ^2 value for the corresponding degrees of freedom (dof). The resulting LR test values for the comparison of independent BL-FS and joint panel BL-FS model is 70.20 (1 dof) and 106.40 (1 dof) for boarding and alighting model, respectively. The LR test values indicate that the joint model outperforms the independent model at any level of significance for both boarding and alighting dimensions. The comparison exercise clearly highlights the superiority of the joint models with the correlation parameterization in terms of data fit compared to the independent models.

Variable Effects

In presenting the effects of the exogenous variables, we will restrict ourselves to the discussion of the joint model with the correlation parameterization. Table 2 presents the estimation results of the joint model of the BL and FS model for boarding and alighting, respectively. Also, the correlation and random component results are shown in the table. The effects of the exogenous variables in the model specification are described in this section for BL and FS components. For the two model components, the parameter estimates are discussed by variable groups. In the BL component, the positive (negative) sign indicates the increasing (decreasing) propensity of non-zero stop-level ridership in Orlando. For the FS model component, the positive (negative) coefficient value indicates increased (decreased) proportion of ridership value categories (i.e. stop-level ridership proportion is increasing (decreasing)). The two models for boarding and alighting offer very similar results for the two model components. Hence, for simplicity we discuss the results together as ridership (as opposed to separately for boarding and alighting).

Table 2 Joint Panel BL-FS Model Results

Variable Name	Alighting Model				Boarding Model			
	BL component		FS Component		BL component		FS Component	
	Estimates	**t-stat**	**Estimates**	**t-stat**	**Estimates**	**t-stat**	**Estimates**	**t-stat**
Constant	6.621	42.520	-	-	6.222	47.202	-	-
Stop-Level Attributes								
Headway (Ln of Headway)	-0.949	-35.717	-	-	-0.876	-40.726	-	-
Presence of shelter in bus stop	-	-	1.061	5.925	-	-	1.283	7.329
Headway*Presence of Shelter	0.121	10.650	-	-	0.099	9.282	-	-
Presence of Shelter*Bus route length in 800m buffer	-	-	-0.210	-1.828	-	-	-0.247	-2.214
Transportation Infrastructures								
Bus route Length in an 800 m buffer	-0.107	-3.728	0.147	3.309	-0.135	-5.631	0.231	6.137
Sidewalk length in an *800 m buffer*	0.041	3.392	-	-	0.035	3.274	-	-
Secondary road length in an *800 m buffer*	0.320	5.925	-	-	0.238	4.981	-	-
Local road length in an *800 m buffer*	-	-	0.245	2.633	-	-	0.189	2.511
Built environment and land use attributes								
Land use area type in an 800m buffer								
Institutional area	-1.113	-2.921	-	-	-1.564	-4.868		
Residential area	1.243	8.792	-	-	1.451	11.074		
Office area	-	-	1.991	4.776	-	-	1.403	3.396
Recreational area	-2.476	-4.212	2.842	2.534	-2.589	-4.897	2.614	2.300
Central business district (CBD) distance								
CBD distance category 2 (4~8 miles)	-0.346	-5.999			-0.352	-6.691		
Sociodemographic variables								
Female Population in the Census Tract	-	-	0.195	6.227	-	-	0.173	5.527
Percentage of high income population	-2.420	-10.892	-	-	-2.598	-13074	-	-
SunRail effects								
Square distance of SunRail	-0.002	-10.152	-	-	-0.003	-13.505	-	-
Correlation Parameters								
Variable	**Estimate**		**t-stat**		**Estimate**		**t-stat**	
Constant (from BL component) and Route Length in 800m buffer (from FS component)	0.401		9.179		9.179		6.507	

Stop-Level Attributes

Transit headway is a significant factor affecting ridership (boarding and alighting). As expected, our model results for the BL component indicate that stops with lower headway are unlikely to have zero ridership. As headway increases, the probability of zero ridership increases. The interaction of headway and presence of shelter variable provides an interesting finding. We find that, in the presence of shelter, the impact of headway reduces i.e. if a stop has a shelter; it is less likely to have non-zero ridership compared to another stop with the same headway without a shelter. The result is quite instructive of how passenger comfort in Orlando with extreme heat and rainfall can influence ridership choice behavior. The reader would note that the main effect of the shelter variable was not significant in the BL component.

For the FS model, the reader would note we cannot employ the route specific headway variable directly for examining the proportion of ridership allocated at the stop-level. The headway variable does not change across the stops in a route and hence it cannot be employed as an independent variable. We tested the headway variable though several interactions (such as headway * presence of shelter). However, none of these variables was statistically significant in our models. Of the other stop-level attributes, the presence of shelter at the bus stop affected proportion of stop-level ridership. Specifically, in the presence of a shelter, stop-level ridership is likely to be higher than stops without a shelter along the same route. This is expected as people are more likely to wait at stops with shelters as they offer protection from the elements. Another variable that was considered in the analysis was the interaction of presence of shelter with route length. The result indicates with increasing length of the route in the buffer, the importance of shelter reduces. This is reasonable because with longer route length in the buffer, more bus routes exist and are likely to result in higher frequency of buses thus ensuring that waiting time is smaller (thus obviating the need for shelter).

Transportation Infrastructure Characteristics

Among transportation characteristics, bus route length, sidewalk length and secondary road length within an 800m buffer affect non-zero ridership propensity. The bus route length negatively impacts the boarding/alighting ridership and increases the zero ridership stop. The result is an indication of potential competition with multiple routes around the stop. The sidewalk length in 800m buffer is positively associated with ridership. The presence of sidewalk encourages pedestrian activity and is possibly surrogate for accessible neighborhoods. The longer length of secondary roads in 800m buffer of the bus stop increases the probability of having less zero ridership stop. Among transportation infrastructure characteristics, bus route length around a stop is likely to increase the ridership proportion in the FS model component. A similar effect is observed with increasing local road length around the stop.

Built Environment Characteristics

Built environment characteristics around a bus stop have a major impact on ridership. The presence of higher institutional and recreational area within 800m buffer of the bus stop significantly increases zero ridership stops. The result is a representation of how individuals visiting institutional and recreational areas prefer automobile in the Orlando region. On the other hand, residential area within 800m buffer of a stop is likely to reduce zero ridership stops. The distance between central business district (CBD) was evaluated in multiple forms in the model. Of these structures, the distance categories provided most intuitive and statistically significant results. The indicator variable for stops between 4 and 8 miles from CBD are likely to have increased propensity for zero ridership.

Among built environment characteristics, office area and recreational area offered statistically significant impact on ridership proportion. Increased presence of office area and recreational area are likely to account for higher proportion of stop-level ridership. The result, particularly for the recreational area is counter intuitive because of the effect of recreational area estimated in the binary logit model. However, it is possible that the effect is meaningful. The results together indicate that not all stops with recreational area attract ridership. However, those stops that are likely to have ridership are more likely to have higher proportion of ridership. Further information of the type of recreational facilities (if available) could offer more insight on these findings.

<u>Sociodemographic Characteristics</u>

Several sociodemographic variables (including income, vehicle ownership, age and gender distribution) were considered in the BL component. Of these variables, only household income distribution affected non-zero ridership propensity. The increase of the proportion of high-income households around a stop are likely to increase zero ridership stops. The result is expected because, high income households have access to automobiles and are unlikely to adopt transit for their mobility in Orlando. Among sociodemographic variables considered, only number of female population at the census tract level affected ridership proportion. In particular, an increase in the female population variable is associated with a corresponding increase in the ridership proportion. The consideration of female population was associated with shorter commuting travel times by public transit in earlier literature (*31*). The current finding could be a manifestation of gender differences among transit riders for proximity to bus stop location.

<u>SunRail Effects</u>

As SunRail was introduced during the analysis time period, we also considered the impact of SunRail in multiple forms. Of these forms, the square distance from all SunRail stations to the bus stop offered interesting result. The result indicates that with increasing distance from SunRail, the likelihood for zero propensity increases perhaps is indicating the effect of transit-oriented development within the vicinity of SunRail. The SunRail variables did not have any impact on ridership proportion.

<u>Unobserved Effects</u>

As discussed earlier, the joint panel model allows for (1) common unobserved factors affecting zero ridership propensity and ridership proportion and (2) common unobserved factors at the stop-level for multiple time periods. In our estimation, we find that correlation between the two components are moderated by constant in the BL component and route length in the FS component for both boarding and alighting models. This supports our hypothesis that BL and FS components are correlated in nature. The correlation parameters are introduced with a " $+$ " sign in the FS component (as described in econometric framework section) since it provided a substantially better fit compared to introducing it with a " $-$ " sign. Overall, the results highlight that accommodating for common unobserved effects across the two model components improves the model fit substantially.

Model Validation

The model developed was validated using a hold-out sample. For this purpose, we generated various measures for the hold-out sample with 4,804 stops (1 quadrimester data, December 2016). We calculated predictive log-likelihood, Bayesian information criterion (BIC), Akaike information criterion (AIC) and Corrected Akaike information criterion (AICc) measures to compare the independent and joint models for both boarding and alighting dimensions separately. The predictive log-likelihood value for the joint model and independent models for boarding (alighting) are -1,091.608 (-1,053.730) and -1,100.400 (-1,065.717), respectively. A similar relationship is observed for all information criterion measures as well. The results clearly highlight that the improvement in the joint model is not a manifestation of over fitting.

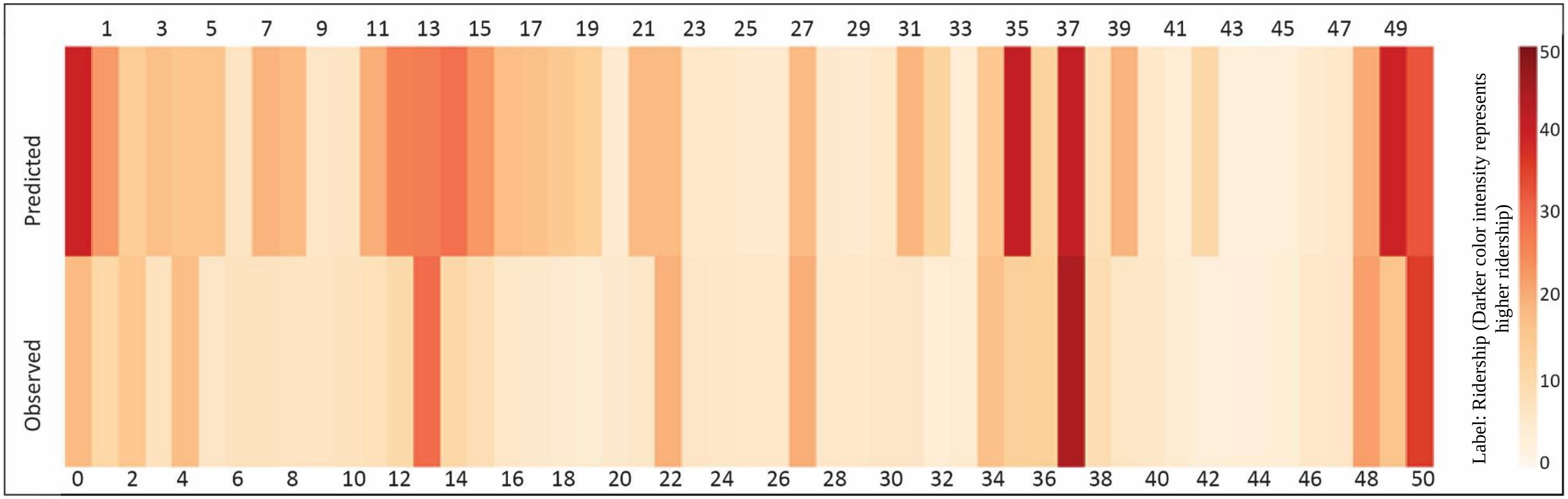
POLICY ANALYSIS

The estimates of the exogenous variables provided in Table 2, do not provide the magnitude of the effects on bus ridership changes over time. For this purpose, we compute the aggregate level "elasticity effects" for an independent variable (see (*32*) for a discussion on the methodology for computing elasticities) to illustrate how the proposed model can be used for policy analysis. In our analysis, we illustrate how the proposed model can be applied at the stop level to evaluate the changes in ridership in response to changes to independent variables. Specifically, we increase the presence of shelter from 25% to 40% of the stops and estimate its impact on ridership. We have generated the predicted fraction of ridership for all stops along the selected 58 routes. Percentage change in ridership (computed as $\left[\left\{\frac{(Predicted-Observed)}{Observed}\right\}*100\right]$) are plotted and presented in Appendix A.1 and A.2. The boarding and alighting percentage changes are plotted by stop for each route. The reader would note that the number of stops vary by route ranging from 2 to 192 stops. To accommodate for the difference in the number of stops across routes, we have used white cells to represent the excess cells for that route. From the figures, we can observe that the percentage increase in boarding is usually slightly higher for a greater number of stops relative to the percentage increase in alighting. Also, the contribution in increased ridership is higher for more stops in boarding compared to alighting. We have also presented the ridership prediction results for enumerating stops where the ridership has increased along a specific route. The results are presented in Table 3. From Table 3, we can observe that presence of shelter has contributed towards higher boarding for more than 53% of stops, while it has contributed towards higher alighting for 29% of stops along these 58 routes. The result is intuitive since people are more likely to be waiting at stops for boarding and the shelter would be rather comfortable for waiting at stops. Further to show the relative change along a route, we have plotted the observed and predicted ridership for two routes: (1) Route 104 – University of Central Florida to Lynx Central Station via Colonial road and (2) Route 107 – Florida mall to Lynx central station (these routes are shown in Figure 1). The results from the policy for boarding and alighting for inbound and outbound directions are presented in Figure 2 (a through d) and Figure 3 (a though d). To be sure, these figures can be generated for all routes. However, we have selected two routes just for the representation purposes. The color palette provided on the right of the figure provides the range of ridership at each stop along the route. For example, In Figure 2a, the ridership at a stop level ranges from 0 to 50 as the color palette becomes dark. In these figures observed and predicted ridership with increased presence of shelter across stops are presented in the lower and upper row panels, respectively. The following observations can be made based on the elasticity effects presented in Figure 2 and Figure 3. Clearly, the addition of shelter for some of the stops results in substantial increase in ridership. The differences in observed and predicted ridership in response to changes to shelter variable are substantial highlighting the importance of shelter at a bus stop in Orlando. The same approach can be employed to evaluate the impact of multiple variables for various routes in the region.

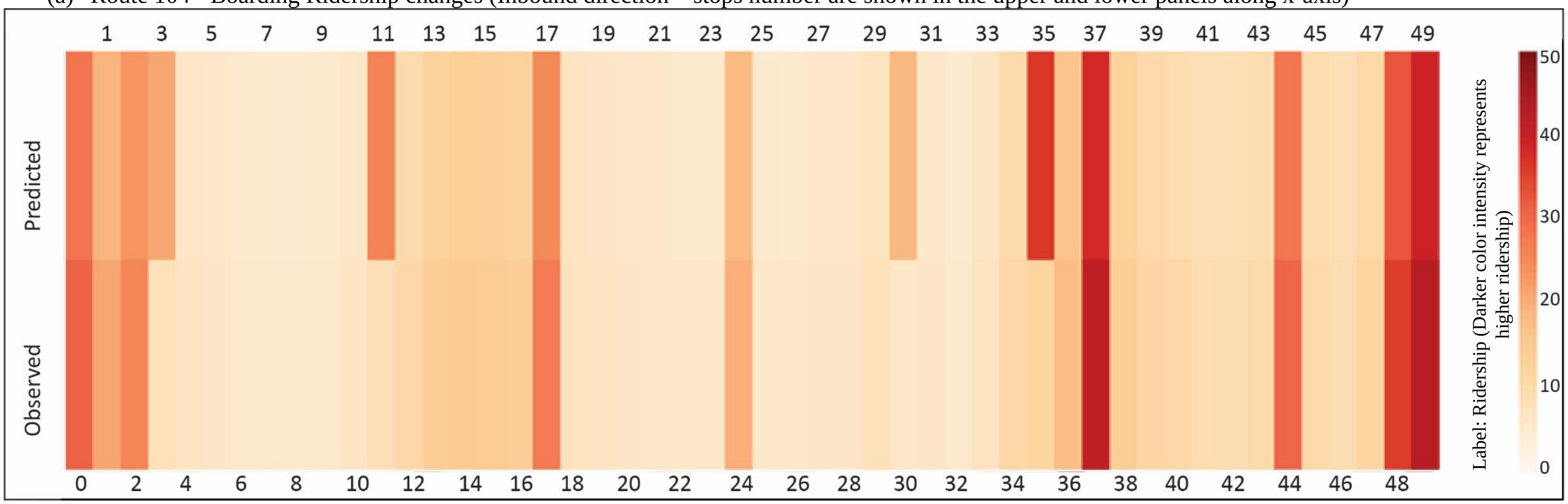
Table 3 Predicted Ridership for the Policy Scenario

Route ID	Total No of Stops	Boarding (Stats for the stops where predicted boarding has increased)		Alighting (Stats for the stops where predicted alighting has increased)	
		% of Stops	% Increase in Boarding	% of Stops	% Increase in Alighting
1	57	5.26	142.60	5.26	192.41
3	157	99.36	2.07	5.10	218.56
6	74	1.35	139.05	1.35	188.37
7	92	97.83	6.88	97.83	5.82
8	183	43.17	44.70	43.17	62.50
9	96	8.33	130.74	8.33	176.32
10	109	7.34	135.61	7.34	183.73
11	95	96.84	12.45	96.84	12.83
13	157	99.36	1.94	4.46	183.94
15	139	98.56	2.86	10.79	191.12
18	192	98.96	6.19	98.96	4.47
20	100	99.00	4.57	99.00	4.10
21	145	98.62	7.27	19.31	118.26
23	107	2.80	132.37	2.80	180.87
24	47	2.13	62.93	2.13	84.72
25	65	98.46	3.47	15.38	69.06
26	66	4.55	164.39	4.55	225.14
28	98	98.98	14.54	16.33	111.60
29	105	99.05	15.05	20.95	120.06
34	8	0.00	0.00	0.00	0.00
36	71	98.59	6.88	98.59	6.35
37	123	19.51	74.58	19.51	103.79
38	2	50.00	115.03	50.00	138.31
40	120	99.17	7.37	15.83	164.01
42	140	18.57	48.39	18.57	68.51
44	126	3.17	119.25	3.17	164.75
45	53	3.77	156.03	3.77	213.03
48	83	98.80	14.59	31.33	157.59
49	91	98.90	16.00	31.87	166.42

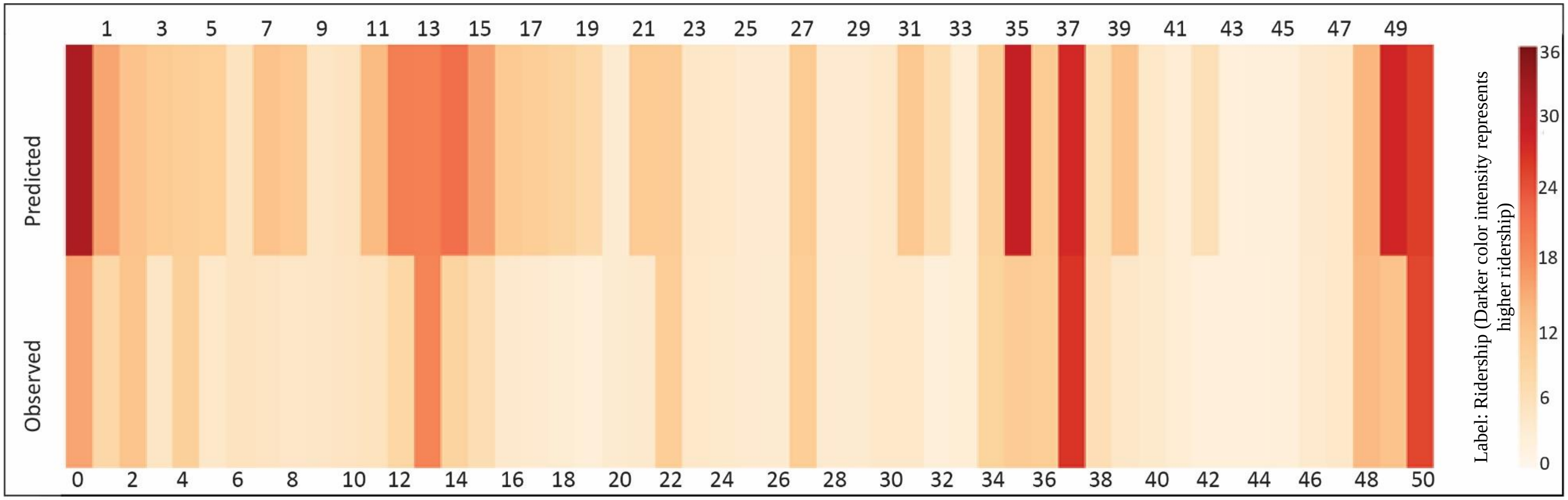
50	31	29.03	81.40	29.03	98.74
51	132	99.24	10.09	99.24	7.24
54	69	8.70	146.41	8.70	193.19
55	73	41.10	86.07	41.10	111.47
56	45	44.44	55.84	44.44	82.48
57	92	9.78	127.02	9.78	169.79
58	17	11.76	70.70	11.76	94.45
102	98	98.98	9.73	18.37	140.36
103	55	5.45	145.06	5.45	199.34
104	105	99.05	16.49	23.81	168.27
105	59	98.31	23.40	33.90	172.87
107	78	42.31	118.01	42.31	134.36
111	53	11.32	69.20	11.32	95.50
125	129	25.58	88.54	25.58	112.95
300	12	0.00	0.00	0.00	0.00
301	55	0.00	0.00	0.00	0.00
302	102	0.00	0.00	0.00	0.00
303	45	0.00	0.00	0.00	0.00
304	102	0.00	0.00	0.00	0.00
305	9	0.00	0.00	0.00	0.00
306	3	0.00	0.00	0.00	0.00
313	80	98.75	2.40	98.75	2.52
319	73	97.26	19.46	97.26	15.98
405	45	11.11	138.55	11.11	181.72
426	68	5.88	154.77	5.88	206.55
427	18	0.00	0.00	0.00	0.00
434	127	3.15	150.40	3.15	204.51
441	8	87.50	20.32	87.50	19.55
443	120	13.33	119.84	13.33	158.65
Total	**4804**	**53.68%**	**--**	**29.45%**	**--**



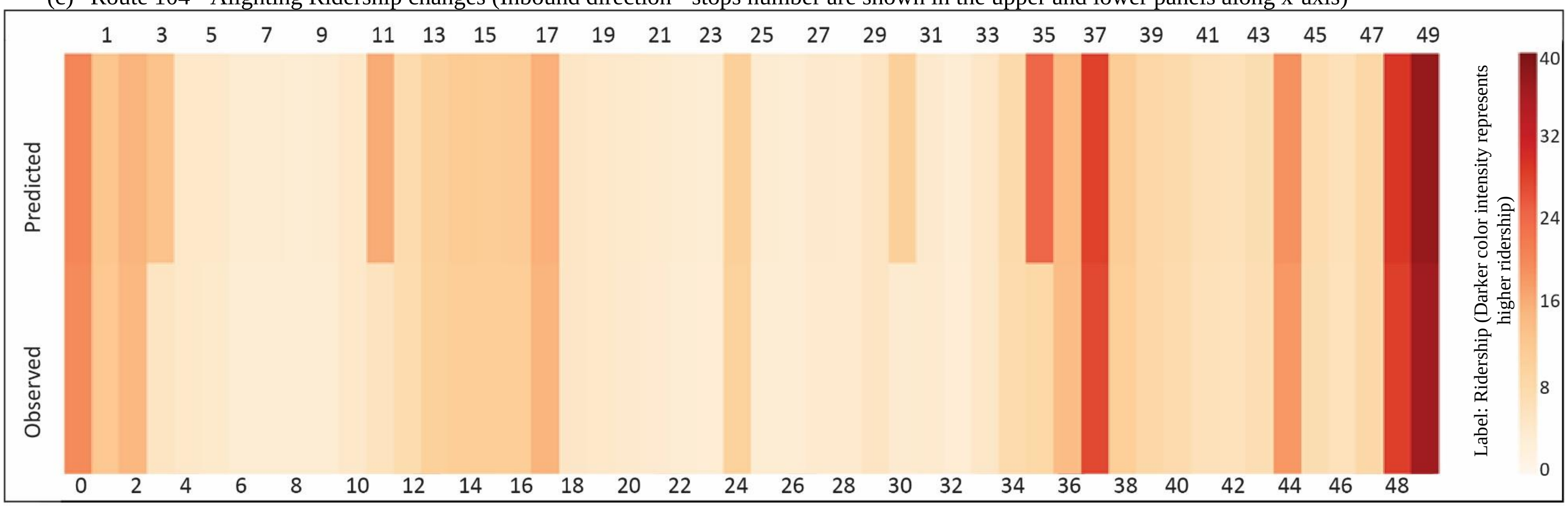
(a) Route 104 - Boarding Ridership changes (Inbound direction – stops number are shown in the upper and lower panels along x-axis)



(b) Route 104 - Boarding Ridership changes (Outbound direction - stops number are shown in the upper and lower panels along x-axis)

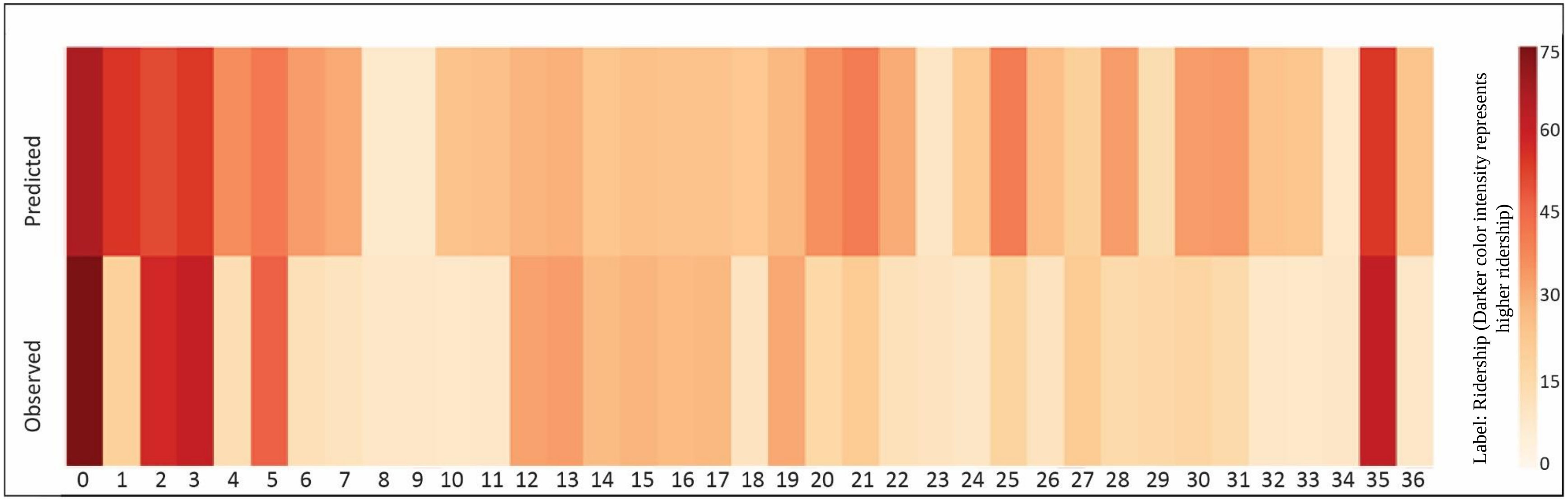


(c) Route 104 - Alighting Ridership changes (Inbound direction - stops number are shown in the upper and lower panels along x-axis)

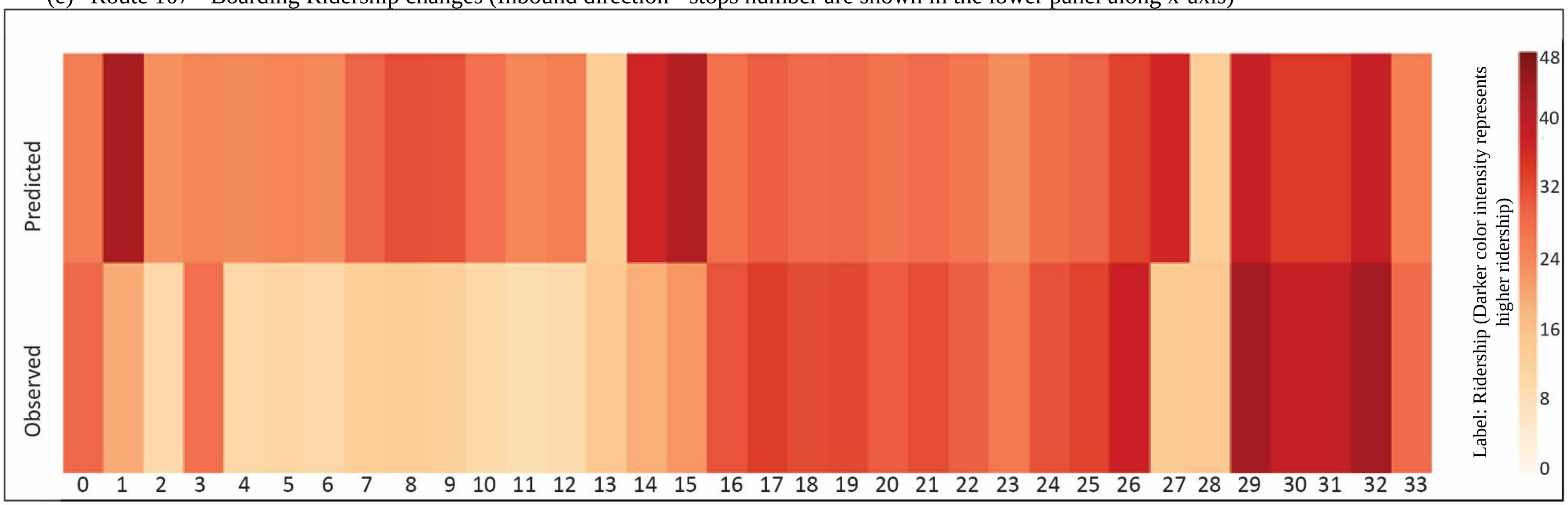


(d) Route 104 - Alighting Ridership changes (Outbound direction - stops number are shown in the upper and lower panels along x-axis)

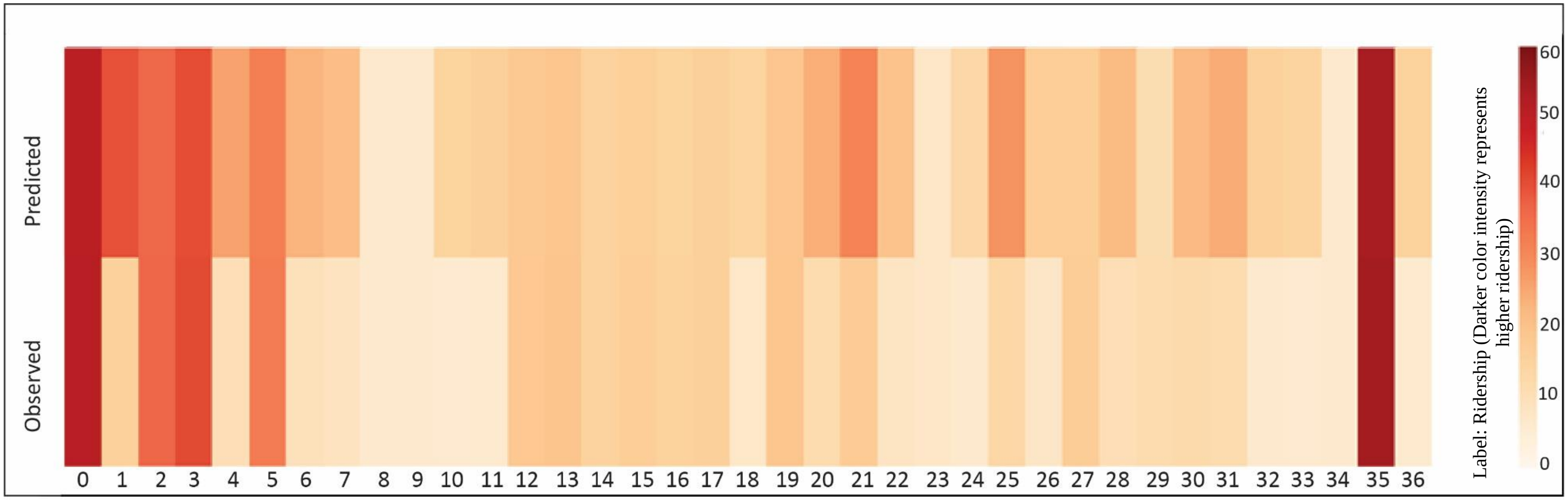
Figure 2: Policy Analysis (Route 104)



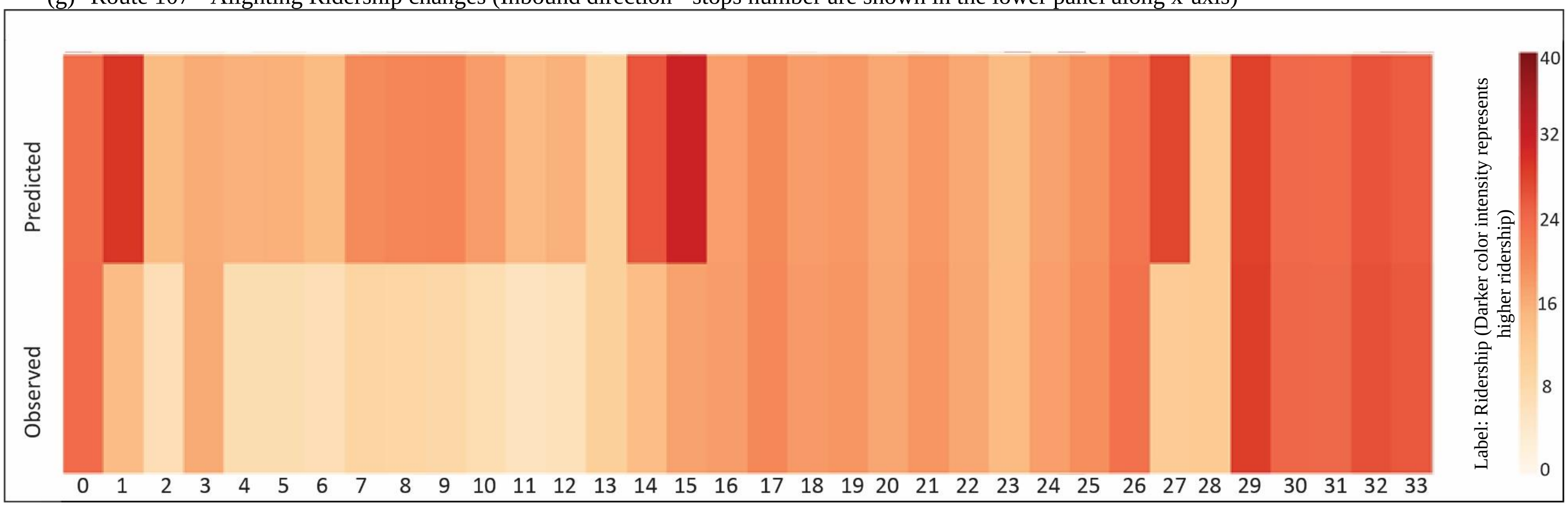
(e) Route 107 - Boarding Ridership changes (Inbound direction - stops number are shown in the lower panel along x-axis)



(f) Route 107 - Boarding Ridership changes (Outbound direction - stops number are shown in the lower panel along x-axis)



(g) Route 107 - Alighting Ridership changes (Inbound direction - stops number are shown in the lower panel along x-axis)



(h) Route 107 - Alighting Ridership changes (Outbound direction - stops number are shown in the lower panel along x-axis)

Figure 3: Policy Analysis (Route 107)

CONCLUSIONS

By developing detailed data-driven and evidence-based analytics to understand factors affecting ridership (at the stop and/or route-level), we can offer insights to enhance ridership. But such detailed ridership analytics platform requires high resolution data on transit ridership. However, finer resolution of stop-based boarding and alighting information are not readily available for the entire bus system. Such data compilation is quite expensive and hence, transit agencies usually resort to compiling such data on a sample of buses operating on the various routes (as opposed to collecting data for all bus stops across all bus routes). In this research effort, we propose an approach to infer stop-level ridership for transit systems that only compile route-level ridership information.

The proposed econometric model had two components each for boarding and alighting dimensions. The first component identified stops with non-zero ridership employing a binary logit model. The second component adopted a fractional split structure to estimate the proportion of ridership for each of the non-zero stops along the route. The econometric framework recognized that several common unobserved factors could influence the two components. To accommodate for this, a joint model structure of the proposed binary logit and fractional split model was built; the proposed model framework was estimated using a system with stop-level ridership data for 8 quadrimesters from May 2014 to December 2016 from the Greater Orlando region. A total of 58 routes were considered for our analysis. The model results offered intuitive results and clearly supported our hypothesis that it is feasible to generate stop-level ridership with route-level ridership data. A validation exercise using a holdout sample also highlighted the superior fit offered by the joint model. The utility of the proposed model is illustrated by generating ridership changes in response to changes to independent variables.

The proposed model offers several advantages to transit policy makers. The main intent of the proposed model development is to employ this model for transit systems without any stop level ridership data. In our paper, the model was estimated in a region with known route level and stop level ridership information. However, given the generic nature of the stop level ridership utility characterization, the proposed approach can easily be transferred across transit systems. To elaborate, the proposed model can be employed to estimate the fraction of stop level ridership using the model estimates from our study analysis. Given the emphasis on the fraction (and not the actual ridership at the stop level), the results are likely to be more transferable across cities. Further, the proposed model system does not have any constants in the utility structure and thus would not be biased by any region-specific preferences. Of course, if the stability of model estimates is a substantial concern, transit agencies can collect stop level data for a small number of bus routes and use that data to calibrate the estimated model from the current paper. Thus, even in cases where calibration is necessary, the amount of data collection efforts is minimal in scope. For transit systems that compile stop level data across a sample of routes, the proposed model can serve as a framework to infer stop ridership across other routes and time periods that are not sampled. Finally, for transit systems that have complete data, the proposed framework can assist in understanding the relation between the contributions from various stops across a route. In this context, the proposed approach would assist in identifying important factors influencing ridership.

While a comprehensive model structure has been considered in the current study, the paper is not without limitations. The proposed approach is unique and does not have an easy counterpart to evaluate model performance. Hence, developing additional measures of effectiveness to compare the proposed model system with log-linear models of proportion variables would be an avenue of research. It might be useful to incorporate the influence of observed and unobserved spatial correlation across the two model components in future analysis. It might also be beneficial to develop an interconnected model system of boarding and alighting within a unified system. Finally, for high public transit usage regions, there might be value in developing these models by time period as the spatio-temporal patterns of ridership distribution are likely to vary substantially across the transit system.

ACKNOWLEDEGMENTS

The authors would like to acknowledge Mr. William Gant from Lynx 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

[1] Siddiqui, F. Falling transit ridership poses an ‘emergency’ for cities, experts fear. 2018.

[2] American Public Transportation Association (APTA). Public Transportation Ridership Report: Fourth Quarter 2010. Washington. Accessed from https://www.apta.com/wp-content/uploads/2018-Q4-Ridership-APTA.pdf. 2011.

[3] Mallett, W. J. Trends in Public Transportation Ridership: Implications for Federal Policy. 2018.

[4] Accuardi, Z. The Stark (and Hopeful) Facts About Bus Ridership. 2018.

[5] Bliss, L. What's Behind Declining Transit Ridership Nationwide? , 2017.

[6] Chakour, V., and N. Eluru. Examining the influence of stop level infrastructure and built environment on bus ridership in Montreal. *Journal of Transport Geography,* Vol. 51, 2016, pp. 205-217.

[7] Waraich, A. S., A. Anowar, T. Tenaglia, T. Sider, A. Alam, N. S. Minaei, M. Hatzopoulou, and N. Eluru. Disaggregate level simulation of public transit emissions in a large urban region.In *95th Annual Meeting of the Transportation Research Board, Washington, DC,* 2016.

[8] Furth, P. G., and N. H. Wilson. Setting frequencies on bus routes: Theory and practice. *Transportation Research Record,* Vol. 818, No. 1981, 1981, pp. 1-7.

[9] Chu, X., and X. Chu. Ridership models at the stop level.In, National Center for Transit Research, University of South Florida, 2004.

[10] Guihaire, V., and J.-K. Hao. Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice,* Vol. 42, No. 10, 2008, pp. 1251-1273.

[11] Watkins, K. E., B. Ferris, A. Borning, G. S. Rutherford, and D. Layton. Where Is My Bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders. *Transportation Research Part A: Policy and Practice,* Vol. 45, No. 8, 2011, pp. 839-848.

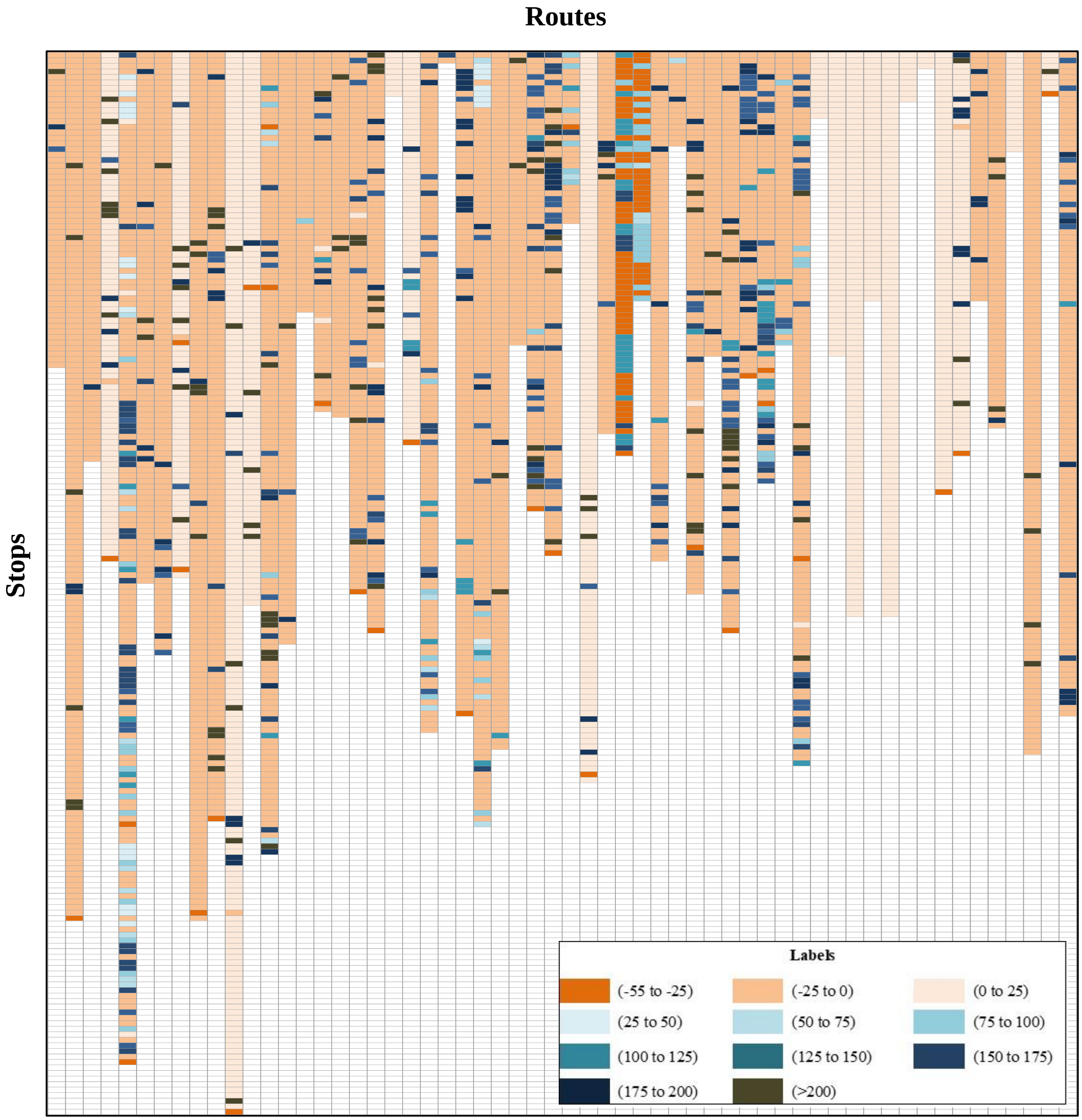
[12] Ferguson, E. M., J. Duthie, A. Unnikrishnan, and S. T. Waller. Incorporating equity into the transit frequency-setting problem. *Transportation Research Part A: Policy and Practice,* Vol. 46, No. 1, 2012, pp. 190-199.

[13] Shimamoto, H., J.-D. Schmöcker, and F. Kurauchi. Optimisation of a bus network configuration and frequency considering the common lines problem. *Journal of Transportation Technologies,* Vol. 2, No. 03, 2012, p. 220.

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

[15] Rahman, M., 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, Issue 2, 2019, pp. 1081-1101.

[16] Rahman, M., S. Yasmin, A. F. Imani, and N. Eluru. Incorporating the Impact of Spatiotemporal Interactions on Bus Ridership. 2018, Technical paper, University of Central Florida.
[17] Ruiz, M., J. M. Segui-Pons, and J. Mateu-Lladó. Improving Bus Service Levels and social equity through bus frequency modelling. *Journal of Transport Geography,* Vol. 58, 2017, pp. 220-233.
[18] Faghih-Imani, A., and N. Eluru. Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: A case study of New York CitiBike system. *Journal of Transport Geography*, Vol. 54, 2016, pp. 218-227.
[19] Taylor, B. D., and C. N. Fink. The factors influencing transit ridership: A review and analysis of the ridership literature. 2003.
[20] Brakewood, C., G. S. Macfarlane, and K. Watkins. The impact of real-time information on bus ridership in New York City. *Transportation Research Part C: Emerging Technologies,* Vol. 53, 2015, pp. 59-75.
[21] Eluru, N., V. Chakour, M. Chamberlain, and L. F. Miranda-Moreno. Modeling vehicle operating speed on urban roads in Montreal: A panel mixed ordered probit fractional split model. *Accident Analysis & Prevention,* Vol. 59, 2013, pp. 125-134.
[22] Papke, L. E., and J. M. Wooldridge. Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of applied econometrics,* Vol. 11, No. 6, 1996, pp. 619-632.
[23] Bhowmik, T., S. Yasmin, and N. Eluru. A joint econometric approach for modeling crash counts by collision type. *Analytic Methods in Accident Research,* Vol. 19, 2018, pp. 16-32.
[24] Ercan, T., N. Keya, N. C. Onat, O. Tatari, and N. Eluru. Sustainability Performance Simulation of the U.S. Urban Mobility Policies. 2018.
[25] Lee, J., S. Yasmin, N. Eluru, M. Abdel-Aty, and Q. Cai. Macroscopic analysis of crash proportion by mode: fractional split multinomial logit modeling approach.In *95th Annual Meeting of the Transportation Research Board (TRB) of the National Academies*, 2016.
[26] Yasmin, S., and N. Eluru. A joint econometric framework for modeling crash counts by severity. *Transportmetrica A: Transport Science,* Vol. 14, No. 3, 2018, pp. 230-255.
[27] Yasmin, S., N. Eluru, J. Lee, and M. Abdel-Aty. Ordered fractional split approach for aggregate injury severity modeling. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2583, 2016, pp. 119-126.
[28] McFadden, D. Conditional logit analysis of qualitative choice behaviour. 1973. *P. Zarembka, New York: Academic Press*, pp. 105-142.
[29] Bhat, C. R. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological,* Vol. 35, No. 7, 2001, pp. 677-693.
[30] Eluru, N., C. R. Bhat, and D. A. Hensher. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis & Prevention,* Vol. 40, No. 3, 2008, pp. 1033-1054.
[31] Kawabata, M., and Shen, Q. Commuting inequality between cars and public transit: The case of the San Francisco Bay Area, 1990-2000. Urban Studies, Vol. 44, No. 9, 2007, pp. 1759-1780.
[32] Eluru, N., and C. R. Bhat. A joint econometric analysis of seat belt use and crash-related injury severity. *Accident Analysis & Prevention,* Vol. 39, No. 5, 2007, pp. 1037-1049.



Appendix A.1: Percentage Changes in Boarding for the Policy Scenario
[Stop numbers vary across routes and white cells are used to represent excess cells for that route]
[Each column panel represents a route in the order of - 1, 3, 6, 7, 8,9,10,11, 13, 15, 18, 20, 21, 23, 24, 25, 26, 28, 29, 34, 36, 37, 38, 40, 42, 44, 45,48, 49, 50, 51, 54, 55, 56, 57, 58, 102, 103, 104, 105, 107, 111, 125, 300, 301, 302, 303, 304, 305, 306, 313, 319, 405, 426, 427, 434, 441, 443 – from left to right]
[Each row panel is a stop in the order of 1 as the top row panel to 192 as the bottom row panel]

Appendix A.2: Percentage Changes in Alighting for the Policy Scenario

[Stop numbers vary across routes and white cells are used to represent excess cells for that route]

[Each column panel represents a route in the order of - 1, 3, 6, 7, 8,9,10,11, 13, 15, 18, 20, 21, 23, 24, 25, 26, 28, 29, 34, 36, 37, 38, 40, 42, 44, 45,48, 49, 50, 51, 54, 55, 56, 57, 58, 102, 103, 104, 105, 107, 111, 125, 300, 301, 302, 303, 304, 305, 306, 313, 319, 405, 426, 427, 434, 441, 443 – from left to right]

[Each row panel is a stop in the order of 1 as the top row panel to 192 as the bottom row panel]