Noname manuscript No.
(will be inserted by the editor)

Estimation of Passenger Wait Time using Automatically Collected Transit Data

Alex Webb[1] · **Pramesh Kumar**[1] · **Alireza Khani**[1]

Received: date / Accepted: date

Abstract Among the many ways to improve a transit system is a reduction in travel time as experienced by the passenger. Hence, passenger wait times remain a topic of interest among transit planners. In this study, the effects of transit vehicle delays on passenger wait time is investigated, as well as the effects of transfer status, boarding location, time of day, and rider travel frequency. The data used in this study were collected using automatic fare collection (AFC) and automatic vehicle location (AVL) technology. A trip chaining algorithm is used to infer the trajectory of each passenger, and as a result produce measures of passenger wait time and vehicle delay. An analysis of an arterial Bus Rapid Transit (aBRT) line in Saint Paul, Minnesota reveals a wait time model consistent with previous literature, a positive relationship between vehicle delay and passenger wait time, and an insignificant relationship between transfer status and passenger wait time. Finally, a simple model relating wait time and vehicle delay is provided for the purpose of transit planning and wait time estimation.

Keywords Automatic Fare Collection (AFC) · General Feed Transit Specification (GTFS) · Wait time · Transit · Trip Chaining Algorithm · Automatic Passenger Count (APC)

1 Introduction

In a commuter landscape dominated by single-occupant vehicles (SOVs), public transit agencies wrestle with fundamental differences that make transit less attractive

Alex Webb
webbx298@umn.edu

Pramesh Kumar
kumar372@umn.edu

Alireza Khani
akhani@umn.edu

[1] Department of Civil, Environmental and Geo-Engineering, University of Minnesota, Twin Cities

than driving alone. Among those differences are the often longer overall travel times faced by transit riders as compared to SOV commuters. While overall travel time is a significant factor in the attractiveness and utility of public transit, the time spent waiting at a public transit station is perceived as the most onerous component of overall travel time [1]. Empirical studies have found that passengers always perceive wait times as longer than actually experienced [1]. This is reflected in transit route choice models where the wait time coefficient is assigned a value more than twice that of the in-vehicle time coefficient [2,3]. Longer wait times reduce the utility of using public transit for travel and hence the modal share. For this reason, transit planners are greatly interested in measuring and minimizing passenger wait time.

This study contributes to the existing literature by using Automatic Fare Collection (AFC) data in conjunction with Automatic Vehicle Location (AVL) data to model passenger wait times. As transit agencies shift towards automated data collection, an ever-expanding body of data will become available for analysis. This study details the process of collecting, organizing, and analyzing data to measure passenger wait times. To illustrate this process, passenger wait times were measured and analyzed for the A-Line Arterial Bus Rapid Transit (aBRT) in Saint Paul, Minnesota. The implications of this study are presented along with a discussion of future research.

2 Background

Early studies on passenger wait time estimation assumed that passengers arrive randomly at transit stops and behave independently of scheduled bus times [4–7]. Under this assumption, the instantaneous rate of passenger arrival is uniform between bus arrivals, resulting in an expected wait time equal to half of the headway. Further, this assumption produces the formula to account for variability in headway,

$$E[W] = \frac{H}{2}(1 + C_v^2) \tag{1}$$

where, C_v is the coefficient of variation in headway, and H is calculated as the mean scheduled headway [8,9]. The assumption that passengers arrive randomly was later found to be invalid [10,11]. Joliffe and Hutchinson challenged the random arrival assumption by considering passengers to be of three behavior types: arriving at a bus station at the same time as a bus by coincidence, intentionally arriving near to a scheduled bus time, or arriving at a bus station randomly [10]. They surveyed ten bus stations for an hour every day for eight days and found that the proportion of passengers who arrive at the optimal time increases with headway. Bowman and Turnquist adapted the previous passenger characterization and found a similar relationship between headway and passenger wait time [11]. They extended the study by including the effect of service reliability and found that passenger wait time was more sensitive to service reliability than to service frequency.

More recently, wait time studies have expanded to regression analysis using characteristics such as gender, ethnicity, location, schedule reliability, and traffic period [15]. Others have used Monte Carlo simulation [16] and probabilistic mixture modelling [17]. In the latter study, passenger wait time is modeled using data from the

Greater Copenhagen Area, including several different train systems, and headways ranging from 2 to 60 minutes. Using AFC data, the study assumes passengers fall in one of two behavior groups: arriving randomly, or trying to minimize wait time by coordinating their arrival with the scheduled train departure time. These behavior groups are modeled using a Uniform and a beta distribution, respectively. Using a mixture model of the two distributions, the study finds that headway is a significant factor in determining the proportion of passengers who arrive randomly. Furthermore, the proportion of passengers who arrive randomly increases as headway decreases. While these results have been shown using AFC data, none of the studies reviewed measure passenger wait time using AVL data. This study contributes to the existing literature by synthesizing AFC and AVL data to analyze vehicle delays and passenger wait times. Ultimately, a generalized linear model will be estimated for passenger wait times on the A Line aBRT in Saint Paul, Minnesota.

3 Methodology

3.1 Data Preparation

The first step in measuring each passenger's wait time is matching their fare transaction to a bus stop. An AFC data set provides geographic coordinates of each transaction, which will have occurred at a bus stop prior to boarding. Often times on a bus route, a north bound station will sit across the street from a south bound station, or an east bound across from a west bound station. The geographic coordinates in an AFC dataset will likely not be accurate enough to indicate which direction a passenger is travelling. For this reason, a pair of transactions, representing an origin and destination, is used to determine which direction each passenger is travelling on a bus route. Once the direction of travel is known for each passenger, the nearest station to a given passenger's fare transaction that goes in the correct direction is identified as their boarding station. Once the boarding station is identified, GTFS and AVL data are used to match each passenger to the bus they most likely boarded. For high frequency routes, matching a passenger's transaction time to a bus purely based on bus arrival time may result in incorrect matching. To avoid this problem, a probabilistic trip chaining algorithm is used to infer each passenger's boarding time and transfer status [18]. The algorithm creates a restricted shortest path problem by considering several trajectories of a passenger with varying boarding location, trip ID, and alighting location. The scheduled boarding and alighting time are assigned based on the trajectory with the highest likelihood value. Finally, the actual arrival time of the bus is inferred by matching the trip ID between GTFS and AVL data, and transfer detection is conducted using previously developed methodology [19]. The matching procedure is described in Figure 1. In order to detect transfers between buses, it is necessary to makes a few assumptions about transfer behavior. First, Metro Transit, the main transit agency in the Twin Cities, defines the transfer period as 2.5 hours after a passengerâs first transaction. In other words, once a passenger has paid for a fare, they may transfer freely between buses with no additional charge. Because 2.5 hours is a relatively long period, a shorter window is used to more accurately capture transfer behavior. Only transactions within 90 minutes of an initial transaction are considered as possible transfers, and the trip chaining algorithm is used to determine whether or not an actual transfer of transit routes has taken place.

3.2 Modelling Wait Time

Previous studies have used the Gamma [20], Beta [17], Lognormal [20], Uniform [4–7], and Johnson SB [12] distributions to model passenger wait time. With this established list of candidate distributions, a series of tests to identify the best fitting distribution is conducted using the R package fitditrplus [21]. The first is a one-sample Kolmogorov-Smirnov test to eliminate any distributions that do not fit the data at the 0.05 significance level. Next, the remaining distributions are compared using the Akaike Information Criteria (AIC), defined as 2k-2ln(L), where k denotes the number of parameters to be estimated and L denotes the maximum value of the likelihood function. Finally, the distribution with the lowest AIC value is used to model wait time. As shown in the results section (Table 2), wait time is best approximated by the Gamma distribution, where k and θ are the shape and scale parameters, respectively [5].

$$f(x;k,\theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}} \tag{2}$$

$$k = \frac{mean^2}{variance}\ ,\ \theta = \frac{variance}{mean} \tag{3}$$

Next, a Generalized Linear Model (GLM) is constructed to relate passenger wait time with bus reliability and several other factors. Bus reliability is represented as the difference between a given actual departure time and the scheduled departure time, and is recorded in our dataset under the name âvehicle delayâ. Given this formulation, it is possible for vehicle delay to be negative. This would indicate that a bus departed early from a station, and thus the actual departure time was earlier than the scheduled departure time.

To construct the GLM, let w_i be the wait time experienced by passenger i, and x_{ij} be the observed value of the explanatory variable j for passenger i. For the proposed model, the explanatory variables are given in Table 2. These variables were selected for analysis due to their ease of availability, practicality for transit planners, and appearance in previous literature [15]. Before the final GLM is constructed, each variable will be tested individually for a significant relationship with wait time. Previous studies have used an Ordinary Least Squares model (OLS) to relate the wait time with explanatory variables (Equation 4), where β is the vector of estimable parameters.

$$w_i = \sum_j \beta_j x_{ij} + \epsilon_i \tag{4}$$

The OLS estimation assumes the error, ϵ_i to be independently and Normally distributed with zero mean and constant variance, i.e. $\epsilon_i \sim Norm(0, \sigma^2)$. However, the Normality assumption may not always hold. To test whether this assumption is appropriate for our data, it is necessary to construct an OLS model and observe the Normality of residuals using the Kolmogorov-Smirnov goodness of fit test:

$$H_0 : \text{The residuals follow the Normal distribution}$$
$$H_a : \text{The residuals do not follow the Normal distribution}$$

In the application of this methodology, the computed test statistic is 0.5134 with a p-value of 2.2e-16, suggesting a rejection of the null hypothesis. Instead, the Gamma distribution better approximates the residuals, justifying the use of a GLM as opposed to an OLS. The Gamma regression specifies the expected value of the wait time as a function of the explanatory variables by using a log-linear function (Equation 5). Moreover, the log transformation provides stability to the variance of the distribution and accommodates negative values of delay.

$$log(E[w_j]) = \sum_j \beta_j x_{ij} \tag{5}$$

As the link function (log) of the transformation is monotonic and differentiable, Equation 5 satisfies the conditions of a generalized linear model. Finally, the iterative least squares estimation (IWLS) algorithm in R is used to estimate the coefficients of the explanatory variables.

4 aBRT in Saint Paul, Minnesota

This study uses AFC and AVL data from the A Line in Saint Paul, Minnesota, which is an aBRT line completed on June 11, 2016. An aBRT line is distinct from other bus lines in that it is designed to reduce dwell time at stations and minimize vehicle delays. The buses have low floors, wide aisles, the ability to board from both doors, and fewer stops per mile than a typical bus route (Figure 1). The most important feature to our study, however, is the presence of pay stations at the bus stops. Instead of paying a bus fare while boarding the bus, all transactions are conducted at the bus stop prior to boarding. The pay stations have card readers, where a passenger may tap their fare card, called a Go-To card. This is crucial to the study, as it allows us to determine the time between when a passenger pays at a station, and when their bus arrives. Using automatically collected data requires the assumption that (1) passengers will tap their Go-To card as soon as they arrive at a station, and (2) passengers board the first A Line bus that arrives at their stop. These assumptions are reasonable, given that A Line buses drive the same route and make the same number of stops. For most of the day, the A Line has a 10 minute headway. In the early morning and the late evening, however, the A Line progresses from 10 minute to 15 and 20 minute headways. In order to only use data from the 10 minute headway period, all transactions used for analysis occurred between 5:45 am and 8:20 pm on September 1st through September 10th of 2016. Vehicle delay is calculated as the difference between the actual and scheduled bus departure time from a stop, and passenger wait time is calculated as the difference between the actual bus departure and the transaction time. After calculating vehicle delay and passenger wait times for the entire data set (n = 2,172), some of the observations must be omitted from the study. First, 3.6% of the vehicle delay values are greater than 10 minutes. While long vehicle delays are possible, vehicle delays over the length of a single headway (10 minutes) are highly unlikely, and would cause problems for our trip-chaining algorithm described in the previous section. As a result, we discarded these entries. Second, 1.3% of our passenger wait time values are less than zero, implying that a transaction occurred after a bus departed. Ultimately, these entries have been discarded because they conflict with the assumption that passengers pay at the station before boarding a bus, and the number of entries is relatively small.

5 Application Results

After excluding negative wait time values from the dataset, the distributions for passenger wait times and vehicle delays are presented in Figures 3 and 4, respectively. Vehicle delays appear to follow the expected behavior - values are largely clustered around zero minutes, and rapidly taper towards the maximum vehicle delay of 9.85 minutes. It should be noted that negative values are acceptable for vehicle delay, as the bus may arrive earlier than the scheduled departure time. Based on the shape of the wait time distribution (Figure 3), several candidate distributions are proposed: Gamma, Exponential, Lognormal, Normal, and Uniform. Table 2 shows the results from the Kolmogorov-Smirnov and AIC tests. At the 0.05 significance level, none of the candidate distributions can be eliminated. The Gamma distribution has the smallest AIC value, and is therefore chosen as the best fitting distribution. The Gamma distribution is parameterized with scale and shape, and yields the fit shown in Figure 3.

Before constructing a GLM for wait time, it is necessary to test the significance of each explanatory variables when modelled in isolation. Table 3 shows the results for each variable from the generalized linear model. First, we observe a highly significant positive relationship between passenger wait time and vehicle delay. This is expected - the longer the bus is delayed after the scheduled arrival, the longer each passenger is expected to wait. While this result exists in previous literature, the exact relationship between passenger wait time and vehicle delay is contested. This is discussed at the end of the section. Moving to the next model, we see some significant differences in wait times between stations. As compared to the reference station, 46th Street Station, three stations do not show significantly longer waits, while the model reveals significantly longer wait times at Snelling & Grand and Snelling & University. Interestingly, the two stations with the shortest expected wait times, 46th Street and Rosedale Transit Center Station, are the endpoints of the A Line. A potential reason for this phenomenon is examined later in the section. While we see significant differences in wait times at different stations, this variable will not be included in the final GLM. This decision is made in an effort to produce a transferable final model. In other words, the objective of this study is to produce results that are easily applied to other transit systems, and including specific stations from this application would conflict with that goal. In place of the station variable, a new variable is created, called "transitway access". On the A Line route, the Snelling & University station is the most heavily used station, likely because it connects to the Green Line light rail. Thus, the new variable aims to capture this distinction with a binary variable, equalling one of a given passenger boarded at Snelling & University, and zero otherwise. The next two models, transfer status and rider frequency, show no significant relationship with wait time. Thus, neither will be included in the final model. Lastly, wait time modelled by time of day produces significant coefficients. As compared to the reference time of day, non-peak hours, passenger's experienced significantly longer wait times during the morning rush period. Due to this significance, Time of Day will be included in the final model.

Based on the results from the individual models, the final GLM is constructed and presented in Table 4. Similar to the previous models, the residuals are fairly skewed, so a GLM with log transformation is used to better approximate the data. While time of day was a highly significant predictor of wait time when modelled in isolation, the variable loses some of its significance when delay is included in the model. This could imply some interaction between the two variables. For example, vehicle delays may occur disproportionately during certain times of day. The transitway access variable is also highly significant, indicating that passenger wait a significantly shorter duration at the Snelling & University station than at any other station. Ultimately, vehicle delay remains a strong predictor of passenger wait time, and should be included in any passenger wait time model. Practically, it can be useful to model passenger wait time as a function of a small number of variables. With the intention of creating a simplified version of the previous GLM, wait time (W) is modeled exclusively by vehicle delay (D), yielding Equation (6).

$$log(E[W]) = 1.279 + 0.16 * D \tag{6}$$

In the absence of a bus vehicle delay, this model finds an expected wait time of 3.59 minutes. Using the average A Line vehicle delay, 1.84 minutes, we find an expected passenger wait time of 4.82 minutes. By eliminating vehicle delays on the A Line route, we find that the average passenger wait time could potentially be reduced by over one minute. As reviewed, previous literature derives the average passenger wait time using headway, a coefficient for variation in headway, and the assumption that passengers arrive randomly [9]. Applying Equation 1 to the data from this study, the expected passenger wait time is found to be 5.24 minutes. While the difference is not large, this result indicates that Equation 1 slightly overestimates expected passenger wait time in the context of this study.

While the variables for Station and Rider Frequency were not included in the final model, they merit further consideration. First, it was observed that passengers boarding at the end point stations on the A Line recorded the shortest average wait times. Figure 5 lends some insight into this observation. The endpoints, 46th Street Station and Rosedale Transit Center, have a significantly shorter average delay when compared to the other stations. It is worth noting this trend, as a wait time model may best be applied to clusters of stations that share similar positions on a route. However, the endpoint stations may also experience some inaccuracies in the AVL data. For example, the A Line may dwell at 46th Street Station for a long period of time until the route is scheduled to begin (layover time), which would undermine the vehicle delay calculation. Future applications of wait time models should be conscious of this phenomenon. Second, the Rider Frequency variable is a fairly crude measure, and may not be showing significance for a variety of reasons. The calculation for Rider Frequency is provided in Table 1, and may not be capturing a very accurate measure of how often each passenger rides the A Line. For example, One of the most frequent riders in the data set only rode the A Line once in the week that this study was conducted. During that single trip, they waited 15 minutes to board the bus. So, this passenger is recorded as a very frequent rider with an average wait time of around 15 minutes. This may be an issue, as the passenger is generally a frequent rider, but was not a frequent rider during the week of the study. This particular outlier illustrates a potential issue with the measure. Future research should aim to produce a more robust representation of rider frequency.

6 Conclusions

This study provides a framework for using automatically collected data to measure passenger wait times, and for making statistical inferences about the data. This framework was applied to the A Line aBRT route in Saint Paul, Minnesota, revealing the distribution of passenger wait times along with the relationship between wait time and several explanatory variables. The demonstrated steps required to use AFC, AVL, and GTFS data in this study should be widely applicable to transit planners interested in measuring wait time with real-time information. For the A Line aBRT, the passenger wait time distribution was found to be best approximated by the Gamma distribution, which is generally consistent with the literature reviewed. While average passenger wait time will differ between bus routes and between transit systems, the methodology developed can be adapted to generate a model for passenger wait time, and doing so would provide an interesting comparison to the results found in this study. Among the results, each explanatory variable was used to model passenger wait time in isolation, and only the station, time of day, and vehicle delay variables were found to be significant. Upon analyzing the significant differences in passenger wait times between stations, it was found that average vehicle delay varied significantly between stations. Further, it was found that the end stations had significantly shorter delays than the mid-route stations. This finding highlights the necessity to understand transit systems within context. While transfer status was found to be insignificant, transit planners are concerned with the burdensome nature of transferring, and therefore the observed effect of transfer status on passenger wait time may warrant further examination. Finally, research should be directed towards testing the assumption that passengers pay their fare as soon as they arrive at a station. Regardless of the outcome, the methodology provided in this study could be modified to accommodate the observed behavior.

Acknowledgements This research is conducted at the http://umntransit.weebly.com/University of Minnesota Transit Lab, currently supported by the following, but not limited to, projects:

- National Science Foundation, award CMMI-1637548
- Minnesota Department of Transportation, Contract No. 1003325 Work Order No. 15
- Minnesota Department of Transportation, Contract No. 1003325 Work Order No. 44
- Transitways Research Impact Program (TIRP), Contract No. A100460 Work Order No. UM2917

The authors are grateful to Metro Transit for sharing the data. Any limitation of this study remains the responsibility of the authors.

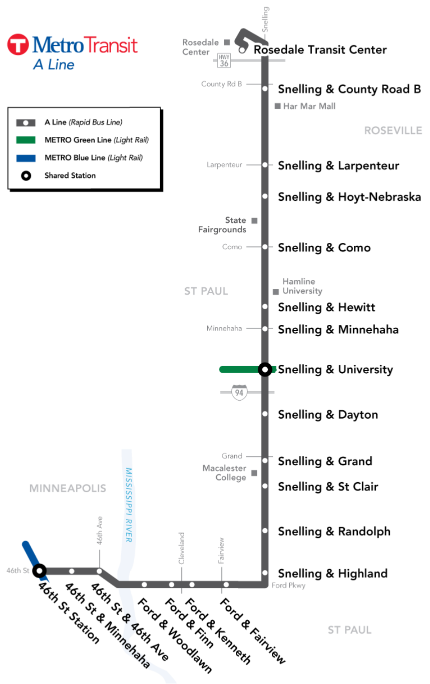

Fig. 1 A Line Route

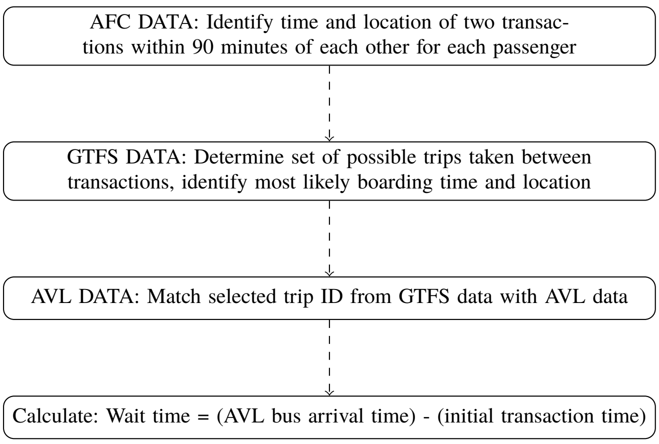

Fig. 2 Data Matching Procedure

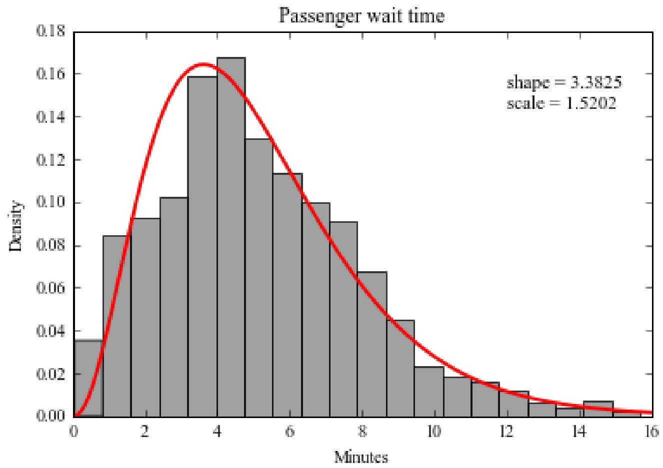

Fig. 3 Passenger Wait Time

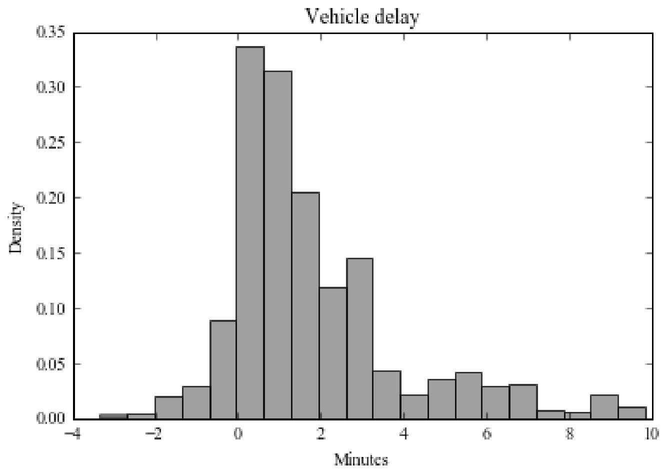

Fig. 4 Vehicle Delay

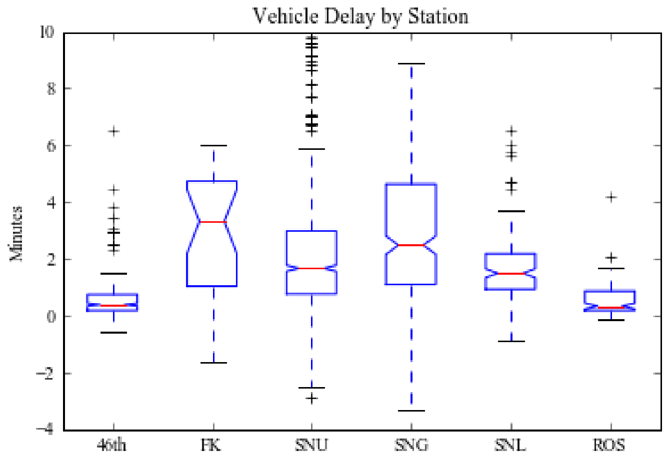

Fig. 5 Vehicle Delay by Station (boxes represent the first quartile, median, and third quartile. '+' represents outliers).

Table 1 Description of Variables

Variable	Description
Wait time	Calculated as: (bus arrival time) - (passenger transaction time)
Vehicle delay	The difference between actual bus arrival and scheduled bus arrival time
Transfer	Takes the value of 1 if the current transaction is a transfer to the A Line, 0 otherwise.
ROS	Takes the value of 1 if the passenger boarded at the Rosedale Transit Center, 0 otherwise.
SNU	Takes the value of 1 if the passenger boarded at Snelling & University, 0 otherwise.
SNG	Takes the value of 1 if the passenger boarded at Snelling & Grand, 0 otherwise.
SNL	Takes the value of 1 if the passenger boarded at Snelling & Larpenteur, 0 otherwise.
FK	Takes the value of 1 if the passenger boarded at Ford & Kenneth, 0 otherwise.
AM	Takes the value of 1 if the transaction time is within 6:00 AM - 9:30 AM, 0 otherwise.
MD	Takes the value of 1 if the transaction time is within 9:30 AM - 3:00 PM, 0 otherwise.
PM	Takes the value of 1 if the transaction time is within 3:00 PM - 6:30 PM, 0 otherwise.
RideFreq	For each passenger, the number of transactions made at an A Line station between August 1st and September 30th, 2016.
Transitway access	takes the value of 1 if passenger boarded at the major transfer station (Snelling & University), 0 otherwise.

Table 2 Goodness-of-Fit Results

	Gamma	Exponential	Lognormal	Normal	Uniform
K-S (Test stat.)	0.0473	0.347	0.0898	0.648	0.223
AIC	9573.465	N/A	9890.027	9757.072	10514.85

Table 3 Single Variable Wait Time Models

		Estimate	Std. Error	t value	Pr(> |t|)	McFadden R^2
Model 1	(Intercept)	1.279	0.0124	102.62	<2e-16 ***	0.354
	delay	0.160	0.00439	36.41	<2e-16 ***	
Model 2	(Intercept)	1.470	0.0279	52.629	2e-16 ***	0.045
	FK	0.324	0.105	3.091	0.00202 **	
	ROS	-0.0836	0.0564	-1.482	0.139	
	SNG	0.386	0.0412	9.359	2e-16 ***	
	SNL	0.107	0.0505	2.118	0.0343 *	
	SNU	0.1793	0.0323	5.547	3.29e-08 ***	
Model 3	(Intercept)	1.638	0.0133	122.929	<2e-16 ***	01.37e-05
	Transfer	-0.00345	0.0329	-0.105	0.917	
Model 4	(Intercept)	1.641	0.0169	97.342	2e-16 ***	5.48e-05
	RideFreq	-0.000127	0.000369	-0.344	0.731	
Model 5	(Intercept)	1.673	0.0166	101.070	2e-16 ***	0.014
	AM	-0.204	0.0348	-5.867	5.19e-09 ***	
	PM	-0.0198	0.0269	-0.739	0.46	

Table 4 Multivariable GLM

	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	1.329	0.0182	73.113	< 2e-16 ***
delay	0.161	0.00442	36.512	< 2e-16 ***
AM	-0.0483	0.0279	-1.732	0.0834 .
PM	-0.00693	0.0212	-0.326	0.744
transitway access	-0.084174	0.019365	-4.347	1.45e-05 ***

Significance codes: "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1
(Dispersion parameter for Gamma family taken to be 0.1802411)
Null deviance: 730.33 on 1992 degrees of freedom
Residual deviance: 467.84 on 1988 degrees of freedom
AIC: 8653.8
McFadden R^2: 0.36

References

1. Psarros, I., K. Kepaptsoglou, and M. G. Karlaftis. *An Empirical Investigation of Passenger Wait Time Perceptions Using Hazard-Based Duration Models*. Journal of Public Transportation, Vol. 14, No. 3, 2011, pp. 109-122.
2. Hunt, J. D. *A Logit Model of Public Transport Route Choice*. ITE Journal, Vol. No. December, 1990, pp. 26-30.
3. Ben-Akiva, E. et al. *Alternative Methods to Estimate Route-Level Trip Tables and Expand On-Board Surveys*. Transportation Research Record 1037, TRB, National Research Council, Washington, D.C., p. 1-11.
4. Barnett, A. *On Controlling Randomness in Transit Operations*. Transportation Science, Vol. 8, No. 2, 1974, pp. 102-116.
5. dePirey Y. A. *Simulation of a Bus Line as a Means to Analyze and Improve Bus Transit Reliability*. Northwestern University, Evanston, Illinois, 1971.
6. Friedman R. R. *Statistical Models of the Mean and Standard Deviation of Passenger Wait Time in Urban Bus Transit*. Northwestern University, Evanston, Illinois, 1976.
7. Jackson W. H. *Bus Passenger Arrival Times*. University of Toronto, Ontario, 1977.
8. Welding, P. I. *The Instability of a Close-Interval Service*. OR, Vol. 8, No. 3, 1957, pp. 133-142.
9. Osuna, E. E., and G. F. Newell. *Control Strategies for an Idealized Public Transportation System*. Transportation Science, Vol. 6, No. 1, 1972, p. 52.
10. Jolliffe, J. K., and T. P. Hutchinson. *A Behavioural Explanation of the Association Between Bus and Passenger Arrivals at a Bus Stop*. Transportation Science, Vol. 9, No. 3, 1975, p. 248.
11. Bowman, L. A., and M. A. Turnquist. *Service Frequency, Schedule Reliability and Passenger Wait Times at Transit Stops*. Transportation Research Part A: General, Vol. 15, No. 6, 1981, pp. 465-471.
12. Luethi, M., U. Weidmann, and A. Nash. *Passenger Arrival Rates at Public Transport Stations*. Institute for Transport Planning and Systems, ETH Zurich, 2007.
13. Csikos, D., and G. Currie. *Investigating Consistency in Transit Passenger Arrivals: Insights from Longitudinal Automated Fare Collection Data*. Transportation Research Record: Journal of the Transportation Research Board, Vol. 2042, 2008, pp. 12-19.
14. Frumin, M., and J. Zhao. *Analyzing Passenger Incidence Behavior in Heterogeneous Transit Services Using Smartcard Data and Schedule-Based Assignment*. Transportation Research Record: Journal of the Transportation Research Board, Vol. 2274, 2012, pp. 52-60.
15. Fan, W., and R. Machemehl. *Do Transit Users Just Wait for Buses or Wait with Strategies?* Transportation Research Record: Journal of the Transportation Research Board, Vol. 2111, 2009, pp. 169-176.
16. Amin-Naseri, M. R., and V. Baradaran. *Accurate Estimation of Average Waiting Time in Public Transportation Systems*. Transportation Science, Vol. 49, No. 2, 2015, pp. 213-222.
17. Ingvardson, J. B., O. A. Nielsen, S. Raveau, and B. F. Nielsen. *Passenger Arrival and Waiting Time Distributions Dependent on Train Service Frequency and Station Characteristics: A Smart Card Data Analysis*. Transportation Research Part C: Emerging Technologies, Vol. 90, No. September 2017, 2018, pp. 292-306.
18. Kumar, P. , A. Khani, and Q. He. *A Robust Estimation of Transit Passenger Trajectories Using Automated Data*. Transportation Research Part C: Emerging Technologies 95 (2018): 731-747.
19. Nassir, N., A. Khani, S. Lee, H. Noh, and M. Hickman. *Transit Stop-Level Origin- Destination Estimation Through Use of Transit Schedule and Automated Data Collection System*. Transportation Research Record: Journal of the Transportation Research Board, Vol. 2263, 2011, pp. 140-150.
20. Gong, H., X. Chen, L. Yu, and L. Wu. *An Application-Oriented Model of Passenger Waiting Time Based on Bus Departure Time Intervals*. Transportation Planning and Technology, Vol. 39, No. 4, 2016, pp. 424-437.
21. Delignette-Muller, M. L., and C. Dutang. *Fitdistrplus: An R Package for Fitting Distributions*. Journal of Statistical Software, Vol. 64, No. 4, 2015, pp. 1-34.
22. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/.