Impact of Carriage Crowding Level on Bus Dwell Time: Modelling and Analysis

Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 6530530, 11 pages
https://doi.org/10.1155/2020/6530530

WILEY | Hindawi

Research Article

Impact of Carriage Crowding Level on Bus Dwell Time: Modelling and Analysis

Yiming Bie,[1] Yunhao Wang,[2] and Le Zhang[3]

[1]*School of Transportation, Jilin University, Changchun 130022, China*
[2]*School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, China*
[3]*Department of Architecture and Civil Engineering, Chalmers University of Technology, Göteborg 412 96, Sweden*

Correspondence should be addressed to Le Zhang; lezh@chalmers.se

Received 21 January 2020; Accepted 23 April 2020; Published 8 May 2020

Academic Editor: Ye Yu

This paper develops two types of estimation models to quantify the impacts of carriage crowding level on bus dwell time. The first model (model I) takes the crowding level and the number of alighting and boarding passengers into consideration and estimates the alighting time and boarding time, respectively. The second model (model II) adopts almost the same regression method, except that the impact of crowding on dwell time is neglected. The analysis was conducted along two major bus routes in Harbin, China, by collecting 640 groups of dwell times under crowded condition manually. Compared with model II, the mean absolute error (MAE) of model I is reduced by 137.51%, which indicates that the accuracy of bus dwell time estimation could be highly improved by introducing carriage crowding level into the model. Meanwhile, the MAE of model I is about 3.9 seconds, which is acceptable in travel time estimation and bus schedule.

1. Introduction

1.1. Background. Bus dwell time is defined as the duration of transit vehicle stopped for serving passengers. It includes the total passenger boarding and alighting time and the time needed to open and close doors. In recent years, the use of advanced traffic detection technique, such as the vehicle navigation system based on GPS and automatic passenger counter, offers a far more convenient and efficient data source, which makes it feasible to conduct the transit travel pattern and reliability analysis reasonably [1–5].

Previous studies have shown that dwell time is an important determinant of transit system performance and service quality in many forms of urban public transportation [6–8]. From the perspective of static bus schedule, bus dwell time at stops is a major component of vehicle travel time, while bus travel time plays an important role in determining the departure frequency and route design for public transit planners and operators. The proportion of bus dwell time in total running time can consume up to 26% for some high-frequency, high-ridership bus transit routes, especially in some high-density areas [9]. Thus, estimating the bus dwell time accurately contributes a lot to planning bus schedule reasonably and efficiently, including the departure headway and the fleet size required to provide service. In terms of dynamic bus schedule, estimating the bus dwell time in advance is of great importance to predict the operating state of vehicles and then determine the punctuality rate or bus headway deviation, avoiding the instability of bus systems or bunching phenomenon. It is also important in determining transit assignment models [10] and reliability analysis of the transit network. Experience also shows that bus arriving irregularly at stops is largely due to the inaccurate estimation of bus dwell time. In addition, while serving passengers at a bus stop, the interaction among buses and passengers may constraint the discharge flows of buses and degrade the bus systems' service quality overall [11, 12]. Consequently, the bus dwell time estimation is essential for improving the service quality as well as reliability of the public transit system [13–16].

There is no doubt that bus dwell time is affected by various factors, including the most significant contributing

factor, that is, the number of boarding or/and alighting passengers, and some secondary contributing factors, for example, the type of bus stops (curbside bus stops or bus bays); payment methods (cash, magnetic stripe tickets, or electronic smart media card); platform height as well as door width. Meanwhile, crowding is also a nonnegligible factor in estimating bus dwell time and longer marginal boarding or/and alighting time is associated with the increase in carriage crowding. It has been well recognized that when buses dwell to load and unload passengers, boarding and alighting occur more slowly when standees are present, especially under door crowding conditions. The standees in the bus or/and around the door impede the boarding (alighting) passengers moving towards the carriage (door) from the bus door (carriage). In other words, the amount of space available decides the circulation speed of passengers within the vehicle. However, higher crowding level does not always result in longer dwell time. For example, when a bus stop is located around a supermarket and also near the terminal station, then fewer passengers are waiting at the stop and most passengers would get off, which accelerates the alighting process accordingly. Consequently, the bus dwells at this stop shorter even if it is highly crowded.

As far as the author knows, the systematical analysis of crowding level on bus dwell time is still missing. This paper aims at studying how carriage crowding level affects dwell time and develops a novel estimation model, taking into account the effect of crowding level for bus dwell time estimation.

1.2. Literature Review. In the past thirty years, much attention has been paid on the subject of bus dwell time estimation model. According to the influence factors, estimation models proposed by previous studies can be categorized into three classes as follows.

1.2.1. Models considering Alighting and Boarding Passenger Volume. The number of boarding and alighting passengers is regarded as the major determinant of bus dwelling. Under given average boarding or alighting speed, the larger the number of passengers disembarking from the bus or waiting at the stop is, the longer buses will dwell [17, 18]. Thus, multiple regression models were developed to analyze the influence of alighting and boarding passenger volume based on data collected from surveys [19–21].

For example, Rajbhandari et al. [9] obtained the data from automatic passenger counter, and the impact of boarding and alighting passengers on dwell time was investigated. Li et al. [22] used data collected from Florida's Broward County Transit system to develop a binary door choice model predicting the proportion of alighting passengers who used the front or rear door to disembark from the bus; then a dwell time estimation model was developed. Li et al. [23] proposed two dwell time models for the BRT station. The first model was a linear model while the second was nonlinear. They introduced the conflict between passengers boarding and alighting into the models. González et al. [24] proposed a bus dwell time model obtained by means of a robust statistical evaluation of boarding passenger data at stops. In the model, dwell time was not increased at a fixed rate of time per passenger. Rashidi and Ranjitkar [25] assessed four different time series based methods (random walk, exponential smoothing, moving average, and autoregressive integrated moving average) to model and estimate bus dwell time based on AVL data collected from Auckland. Rashidi and Ranjitkar [26] proposed a gene expression programming-based approach to model and estimate bus dwell time. The proposed model was calibrated and validated using the data collected from 22 bus stops.

1.2.2. Models considering Secondary Factors. Although dwell time is highly correlated with the number of boarding and alighting passengers, there are also some secondary contributing factors to the bus dwell time, such as crowding, fare type, platform, and bus design [27, 28]. These secondary factors may strongly influence the effectiveness of different strategies used to improve service. Guenthner and Hamat [29] investigated nine different fare types and payment methods and found that they did not affect the dwell time significantly. Levine and Torng [30] evaluated the impacts of low-floor bus design on reducing bus dwell time. Milkovits [31] used the data from the automatic passenger counting, automatic fare counting, and automatic vehicle location systems installed on Chicago Transit Authority buses, then developed a dwell time estimation model, and analyzed the impact of the secondary factors. Currie et al. [32] analyzed the impacts of platform design on streetcar dwell time based on the data collected from Melbourne, Australia, and Toronto, Canada. Jaiswal et al. [33] considered effects of passengers walking on a relatively longer BRT station platform. Results showed that the long BRT platform may lead to the bus experiencing a higher dwell time. Fernández et al. [34] evaluated the impacts of platform height, door width, and fare collection method on bus dwell time based on filed observed data. Tirachini [35] developed multiple regression models to explain the relationship between observed variation in dwell times and different payment methods, the existence of steps at doors, the age of passengers, and the possible friction between users boarding, alighting, and standing.

1.2.3. Models considering Bus Stop Types. All studies mentioned above assumed that bus dwell time is highly related to passenger alighting and boarding volume or velocity, while the third category indicated that it was also affected by bus stop design. Meng and Qu [36] pointed out that the bus dwell time at a bus bay was affected by interactions among buses, arrival passengers, and traffic on the shoulder lane. Bus dwell time at a bus bay possessed a high degree of uncertainty originating from the merging behavior of bus to the vehicles in the shoulder lane. A novel probabilistic methodology was developed to estimate the bus dwell time, including a standard regenerative stochastic process to model the interactions among buses, arrival passengers, and traffic on the shoulder lane.

Besides the three types of studies discussed above, Katz and Garrow [37] investigated how bus design factors influence door crowding and quantified how door crowding is related to operational performance and passenger safety. Results showed that door crowding was affected by multiple bus design factors, including door placement, aisle length, presence of a front seating area, and service type. Increases in door crowding were associated with longer marginal boarding times. Fletcher and El-Geneidy [38] determined the influence of crowding and fare payment on dwell time through manual data collection. Multiple regression dwell time models were performed by using a traditional model and a new expanded model. The expanded model showed that crowding significantly increased dwell time after approximately 60% of bus capacity was surpassed. This was the only study considering the effect of bus crowding on dwell time; however, the quantitative method of crowding level was not given in this paper.

1.3. Research Contribution. In this paper, we propose a quantitative model of crowding level and develop an estimation model of bus dwell time, stressing the influence of carriage crowding level. It will be significant in improving the accuracy of bus dwell time estimation. This research can help public transit planners develop better policies associated with crowding.

This paper is organized as follows. Section 2 presents the data collection method; the quantitative model of crowding level and estimation model of bus dwell time and comparison with the conventional estimation model are proposed in Section 3; and conclusions are presented in Section 4.

2. Data Collection

The data used in this research come from two high-frequency bus lines of Harbin, Route 8 and Route 63, by manual data collection, as shown in Figure 1. Both routes pass through the central area of this city. Route 8 is 13.5 km with 25 bus stops, connecting Harbin Children's Hospital with Minjiang Community, a popular residential suburb. Another 10.1 km bus line, Route 63, has 21 bus stops and terminates at Dajiang Community, which is another high-density residential area. All the passengers are required to board from the front door and alight by the rear door. Thus the dwell time of a bus is determined by the maximum value between passenger boarding time and alighting time. To best capture the effects of crowding, data used in this paper were collected during the morning (7:00 am–9:30 am) and afternoon (4:30 pm–7:00 pm) peak hour periods, 19–23 December, 2016. Meanwhile, if bus stop is close to downstream intersection, then bus dwell time will also be affected by bus queuing state at intersection. To eliminate the disturbance of traffic signals, we examined the distribution of all the stops along Route 8 and Route 63. As the distance from each stop to the corresponding downstream intersection is all no less than 100 meters, the influence of traffic signals can be ignored in this study. In addition, control strategies at stops like bus holding to keep regular headway have not been applied in Harbin, so the effect of holding on bus dwelling does not exist in this study.

Two investigators are arranged for each bus, recording boarding and alighting data individually. The recorder near the front door makes notes of the moment of opening the door, the number of boarding passengers, boarding time, and the moment of closing the door. The other one records the corresponding alighting information of the rear door. The tickets for both routes are one yuan and passengers are required to pay when boarding. As 95% of passengers choose to pay by cash and no change are offered by each bus along these two routes, the effect of payment methods on bus dwell time can be negligible in this paper.

In general, we regard the boarding or alighting time as the difference value between closing and opening moment of the front or rear door. However, it is not always accurate. Previous studies have shown that bus dwell time is also associated with the type of bus stops as well as road traffic status. For example, after the bus loads or unloads all the passengers at a bus bay, it would leave immediately. However, before joining into the traffic, the bus has to spend some time in finding an acceptable gap between consecutive vehicles on the left lane, resulting in longer dwell time. In this case, the real moment for closing door is set as the sum of the moment of the last passenger boarding or alighting the vehicle plus the average duration for closing door, where the average duration for closing a door can be calculated by previous data in general cases; that is, nonpassenger related delays are not included in bus dwell time in our study.

The traffic condition in Harbin is always affected by low temperature snow and ice weather. To eliminate this effect, before the recorders boarded the bus, the weather, temperature, and date were all documented. All the collected data were analyzed preliminarily and abnormal data were removed before estimation, that is, the engine failure caused by severe weather and wheelchair ramp event. We collected 828 groups of bus dwell time at stops. As during peak hour period, the headway of these two lines is short, about 4 minutes, and passenger arrival rate differs from stops. The bus crowding level also varies from each other. We also need to remove the data where no standees are occurred. Finally, we get 640 groups of valid data in total used for Section 3.

3. Model Development and Evaluation

In this section, we develop two estimation models. The first model takes the interaction of bus crowding and number of alighting and boarding passengers into consideration and estimates the alighting time and boarding time in Section 3.1. The second model is indeed a traditional one, only considering the number of boarding and alighting passengers as shown in Section 3.2. Section 3.3 compares the performance of these two models.

3.1. Development of Model I

3.1.1. Crowding Level. Generally, a vehicle is in a crowded state when people on the vehicle impede the individuals

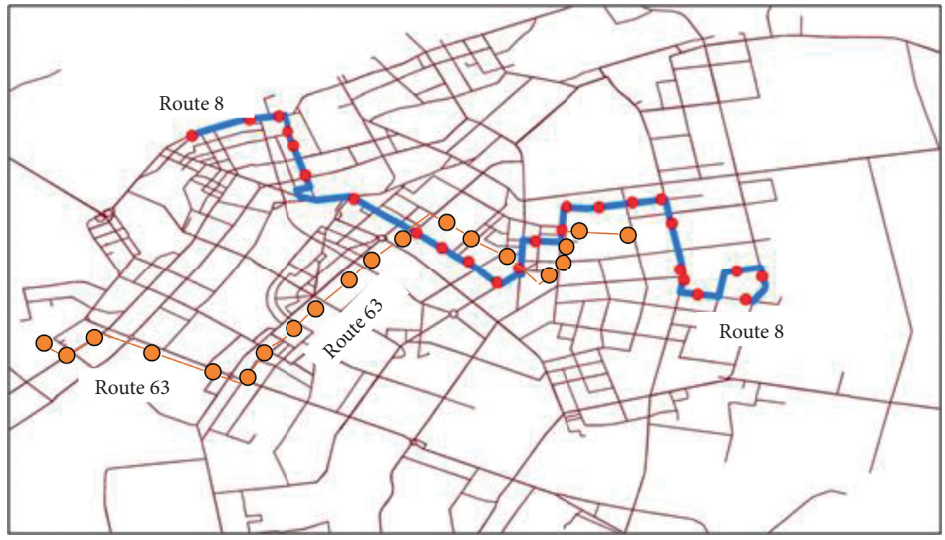

Figure 1: Spatial distribution of bus stop on Route 8 and Route 63.

boarding and alighting. Buses on different routes may have different characteristics, such as the space and capacity (seated and standing) for passengers. To truly measure the effect of crowding on bus stopping duration, a more refined formulaic definition is needed. We define crowding level C quantitatively as follows:

$$C = \frac{N_1}{S_1 \times \lambda}, \tag{1}$$

where N_1 is the number of standees in the vehicle and calculated by subtracting the number of seats from the number of passengers on board; S_1 denotes the available space for standees which is determined by the specialty of buses; and λ is the maximum number of standees in unit space of the bus, passengers/m^2. According to our investigation in Harbin city, unit space of buses can accommodate 7 standees at maximum; that is, $\lambda = 7$ passengers/m^2.

Next, we will estimate alighting time and boarding time in Section 3.1.2 and the longer one plays a decisive role in bus dwelling duration in Section 3.1.3.

3.1.2. Boarding Time and Alighting Time Estimation Models

(1) Boarding Time Estimation Model. Our objective is to estimate the dynamic relationship of passenger boarding time with bus crowding level and the number of boarding passengers. Firstly, 6 types of regression models are developed to fit the collected 640 data sets and their performance measured by the adjusted coefficient of determination (Adj R^2) is shown in Table 1, where independent variables include the number of boarding passengers (X_1) and bus crowding level (C); boarding time (Y_1) is the dependent variable.

Here, we select the Adj R^2 as evaluative criteria, instead of coefficient of determination, R^2. This is because R^2 may increase spuriously when extra explanatory variables are added. Consequently, to improve the accuracy of estimation, some irrelevant variables may be introduced into the model, resulting in over fitting. The Adj R^2 can avoid this phenomenon, as the number of parameters (K) is also taken into consideration as shown in (2).

The Adj R^2 indicates the proportion of the variation explained by the estimated regression model, defined as follows:

$$\text{Adj } R^2 = 1 - \frac{\text{SSE}/(n-K)}{\text{SST}/(n-1)} = 1 - \frac{n-1}{n-K}\left(1 - R^2\right), \tag{2}$$

where n is the size of random sample; SSE and SST are the sum of squared errors and the sum of squares of statistical sample respectively; K indicates the number of variables.

According to Table 1, the fitting result of double-logarithmic regression model is the best, as the model explains 74.7% of the variation, which is higher than other models. The exact double-logarithmic regression model is presented in

$$\ln Y_{1i} = \beta_0 + \beta_1 \ln X_{1i} + \beta_2 \ln C_i + e_i, \quad 0 < C_i \leq 1, \tag{3}$$

where i is the index of observation; β_0, β_1, and β_2 are parameters to be estimated; e_i denotes disturbance of observation i.

Then we use ordinary least squares method (OLS) to estimate the parameters of (3) and the estimation results are shown in Table 2.

Traditional OLS method requires that there is the constant variance in the errors over all values of the explanatory variables (which is also called homoscedasticity). Once this condition is not satisfied, the validity of parameter estimation cannot be guaranteed. If heteroscedasticity occurs, this estimation is invalid even that the regression coefficient of (3) is evident at the 5% significant level. Then, we need to examine whether the variance of errors is constant or not by some statistical tests, for example, White test and Breusch-Pagan test. Here, we use White test to judge homoscedasticity or not.

To test for a constant variance, an auxiliary regression analysis is conducted, which regresses the squared residuals

Table 1: Boarding time models and performances.

Function type	Adj R^2
Linear	0.666
Quadratic	0.670
Cubic	0.670
Exponential	0.605
Semilogarithmic	0.644
Double logarithmic	0.747

Table 2: Parameter estimations of boarding time model based on OLS.

Parameters	β_0	β_1	β_2
Estimated value	0.958	0.929	0.086
t-statistic	8.073	18.489	2.307
p-value	<0.001	<0.001	0.023

F-statistic = 194.334 (p-value <0.001); $R^2 = 0.751$; Adj $R^2 = 0.747$.

from the original regression model onto a set of regressors as shown in (4). The basic logic of the White test is to substitute the squared residuals for the variance of the error term at each observation. The squared residuals from the original model serve as a proxy for the variance of the error term at each observation. If the error term in the original model has a constant variance in fact, then the coefficients in the auxiliary regression should be statistically indistinguishable from zero and F-statistic should be small. Otherwise, evident F-statistic counts against the hypothesis of homoscedasticity:

$$\widehat{e}_i^2 = \gamma_0 + \gamma_1 (\ln X_{1i})^2 + \gamma_2 (\ln C_i)^2, \quad 0 < C_i \le 1, \tag{4}$$

where $\widehat{e}_i$ is the estimated value of e_i in (3), that is, residual.

White test is applied to (3) based on the auxiliary regression in (4); we get F-statistic which is 4.461 and p-value = 0.013. Thus at the 5% significant level, homoscedasticity is not satisfied and we need to modify (3).

Generally, the method of weighted least squares (WLS) can be used when the OLS assumption of constant variance in the errors is violated. It works by incorporating extra nonnegative weights associated with each data point into the fitting criterion. The value of weight indicates the precision of information contained in the associated observation. Usually, an observation with small error variance has a large weight as it contains relatively more information than an observation with large error variance. The objective function of ordinary least squares method is to minimize $\sum_{i=1}^{n} (Y_{1i} - \widehat{Y}_{1i})^2$, where $\widehat{Y}_{1i}$ is the fitted value of Y_{1i} and n denotes the size of random sample. Accordingly, WLS method aims at minimizing $\sum_{i=1}^{n} w_i (Y_{1i} - \widehat{Y}_{1i})^2$ and w_i is the weight of ith observation. Without loss of generality, we define $1/|e_i|$ as the weight, that is, $w_i = 1/|e_i|$, and get the parameter estimations under WLS method as shown in Table 3.

For the modified model, the White test shows that parameter estimations are valid and all the regression coefficients are evident at 5% significant level. Meanwhile, the adjusted coefficient of determination also increases to 0.938, indicating that the accuracy of regression by WLS method is highly improved compared with that estimated by OLS method. Table 3 shows that 1% increase in the number of boarding passengers will lead to 0.926% increase in boarding time. In the meantime, the crowding level also reveals a positive relationship with boarding time in which a 1% increase in crowding level improves boarding time by 0.085%. As a secondary contributing factor, the effect of crowding level reaches up to almost 10% of the influence caused by the most contributing factor, that is, the number of boarding passengers in bus boarding time, which also specifies the significant importance of carriage crowding level on bus dwell time estimation.

Figure 2(a) shows the relationship between the number of boarding passengers (X_1) and the boarding time (Y_1) under given crowding level (C), where different curves represent different crowding levels. The figure shows that boarding time increases as the number of boarding passengers grows, which is as expected. In the meantime, when the number of boarding passengers is fixed, the boarding time also increases as the vehicle becomes more crowded, which is consistent with the previous parameter analysis. The figure also shows that larger crowding level indicates the higher growth rate of bus dwell time. To examine how the crowding level affects the bus boarding time, we plot the relationship between the number of boarding passengers and bus crowding level in Figure 2(b). When the number of boarding passengers is small, boarding time is indeed insensitive to crowding level. However, when X_1 is up to 25 persons, the boarding time is 42.5 seconds under 0.1 crowding level and reaches up to 51.3 seconds under 0.9 crowding level, increased by 20.7%.

(2) Alighting Time Estimation Model. We will use the similar method presented above to estimate the alighting time (Y_2) with independent variables: the number of alighting passengers (X_2) and bus crowding level (C). The double-logarithmic model also performs best among these six types of models with the Adj R^2 equal to 0.678.

The exact alighting time model adopting double-logarithmic approach is presented in

$$\ln Y_{2i} = \beta_0' + \beta_1' \ln X_{2i} + \beta_2' \ln C_i + e_i, \quad 0 < C_i \le 1. \tag{5}$$

The parameter estimation of (5) based OLS method shown in Table 4 is valid. The regression coefficients are significant at 5% level. It also reveals that 1% increase in the number of alighting passengers will lead to 0.848% increase in alighting time. Meanwhile, a 1% increase in crowding level improves alighting time by 0.092%, similar parameter analysis results to boarding process:

$$\widehat{e}_i^2 = \gamma_0 + \gamma_1 (\ln X_{2i})^2 + \gamma_2 (\ln C_i)^2, \quad 0 < C_i \le 1. \tag{6}$$

Comparing the estimation models for boarding time and alighting time, the former presents heteroscedasticity in errors while homoscedasticity is observed in the latter. It is reasonable as the boarding process is disturbed by more various factors, such as the aging of boarding passengers, fare payment method, and the activities of boarding passengers. All the disturbing factors will accumulate and

Table 3: Parameter estimations of boarding time model based on WLS.

Parameters	β_0	β_1	β_2
Estimated value	0.965	0.926	0.085
t-statistic	148.700	298.038	41.784
p-value	<0.001	<0.001	<0.001

F-statistic = 47686.58 (p-value <0.001); R^2 = 0.942; Adj R^2 = 0.938.

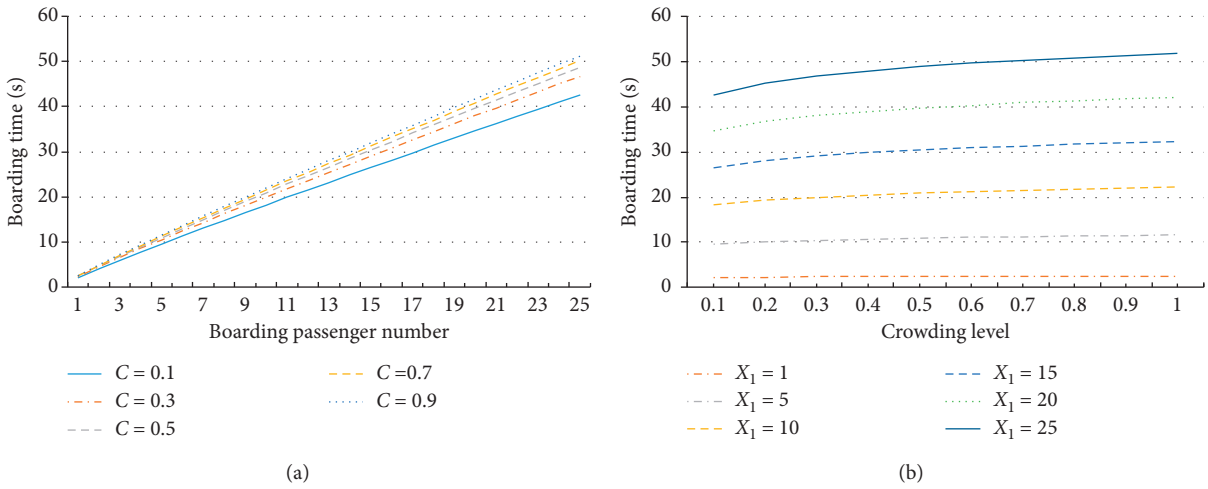

Figure 2: The effect of crowding level and boarding passenger volume on boarding time. (a) Relation between boarding passenger number and boarding time under given crowding level. (b) Relation between crowding level and boarding time under given boarding passengers.

Table 4: Parameter estimations of alighting time model based on OLS.

Parameters	β_0'	β_1'	β_2'
Estimated value	0.635	0.848	0.092
t-statistic	4.996	16.230	2.503
p-value	<0.001	<0.001	0.014

F-statistic = 150.289 (p-value <0.001); R^2 = 0.682; Adj R^2 = 0.678.

become more distinctive as buses load more passengers. However, alighting process is free of such disturbing factors largely.

Figure 3(a) clearly illustrates the effect of alighting passenger number on alighting time under different levels of crowding level. The curves reveal that the alighting time increases as more passengers get off. However, when the number of alighting passengers is larger, higher crowding level indicates longer alighting time but with smaller increasing range. For example, when 24 passengers alight, the alighting time when $C = 0.3$ is much longer than that under 0.1 crowding level, but the difference of alighting time is much smaller between 0.9 crowding level and 0.7 crowding level even that the gaps between the crowding level keep the same. This is because when more passengers alight, a lot of standees will also get off, which therefore does not cause a barrier to alighting passengers.

The relationship between the alighting time and bus crowding level is presented in Figure 3(b). Similar to Figure 2(b), when the number of alighting passengers is small, alighting time is also insensitive to carriage crowding, as compared with the alighting passenger number; the effect of crowding is only secondary contributing to alighting time. However, when the number of alighting passengers is up to 25 persons, the alighting time is 23.4 seconds under 0.1 crowding level and increases to 28.6 seconds under 0.9 crowding level, increased by 22.4%. Further analysis reveals that average alighting time is much shorter than average boarding time under the same crowding level, which is also expected as paying fare is also time-consuming.

3.1.3. Dwell Time Estimation Model. Obviously, bus dwell time is determined by the maximum value between passenger boarding time and alighting time. For each stop, we choose the maximum one between the estimated boarding time and alighting time to fit bus dwell time (Y) linearly as shown in (7) and the parameter estimations are summarized in Table 5:

$$Y_i = \beta_0'' + \beta_1'' \max\{Y_{1i}, Y_{2i}\} + e_i. \quad (7)$$

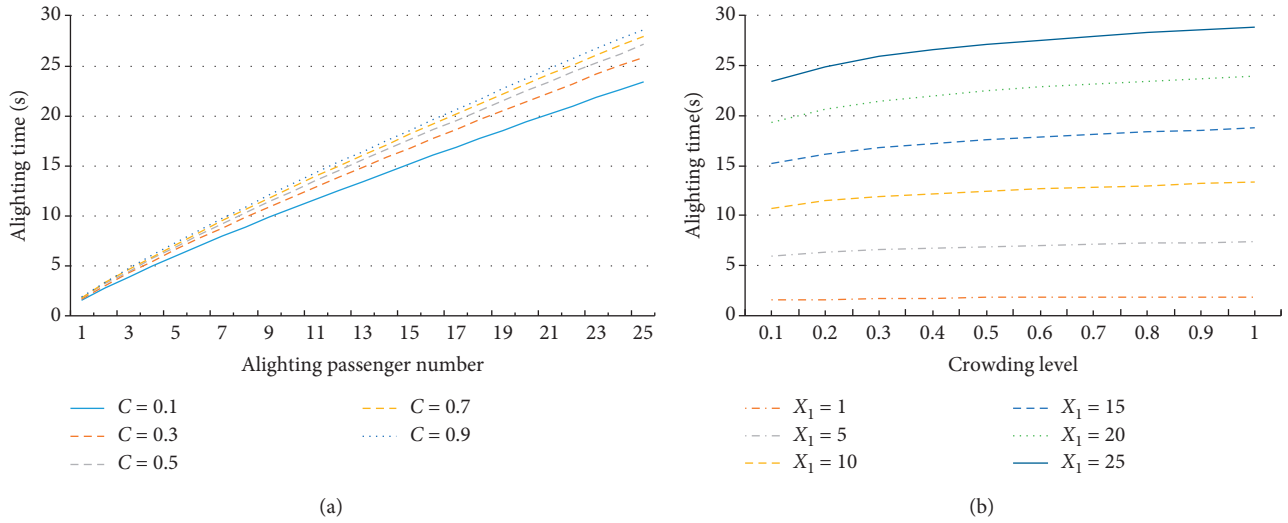

Figure 3: The effect of crowding level and alighting passenger volume on alighting time. (a) Relation between alighting passenger number and alighting time under given crowding level. (b) Relation between crowding level and alighting time under given alighting passengers.

Table 5: Parameter estimations of bus dwell time model based on OLS.

Parameters	β_0''	β_1''
Estimated value	6.819	0.968
t-statistic	8.092	20.320
p-value	<0.001	<0.001

F-statistic = 412.886 (p-value <0.001); $R^2 = 0.728$; Adj $R^2 = 0.727$.

Table 6: Parameter estimations of bus dwell time model based on WLS.

Parameters	β_0''	β_1''
Estimated value	6.936	0.947
t-statistic	65.799	131.663
p-value	<0.001	<0.001

F-statistic = 17335.21 (p-value <0.001); $R^2 = 0.954$; Adj $R^2 = 0.950$.

Similarly, to guarantee the homoscedasticity property, the White test is conducted and an auxiliary regression analysis is developed as shown in

$$\hat{e}_i^2 = \gamma_0 + \gamma_1 \max\{Y_{1i}, Y_{2i}\}. \quad (8)$$

We get F-statistic equal to 7.696 and p-value is 0.006, implying that the homoscedasticity property does not hold at the 5% significant level. It also shows that bus dwell time is significantly influenced by the interference factors along boarding and alighting process.

We set $1/|e_i|$ as weight as Section 3.1.1 and get the parameters estimations by WLS method listed in Table 6.

The coefficient is evident at the 5% significant level and the Adj R^2 is high to 0.954, implying that the novel expanded model can explain 95.4% of the variation by using a sample size of 640 dwells. Model (7) also shows that bus dwell time consists of two parts: one is a constant value, 6.936 seconds, including the time used for opening and closing doors, as well as the preparation time for passengers to alight and board; the other part is the time used for loading and unloading passengers.

To further confirm whether the constant part is 6.936 seconds or not, we collected 640 data recording the duration including the passenger preparation time and door opening and closing time, denoted as Z_i. The mean of this sample is $\overline{Z} = 7.443$ and standard deviation is $\sigma = 5.324$. It is desired to test the simple hypothesis H_0: $\mu = 6.936$. We propose an estimator as $t = \overline{Z} - \mu/(\sigma/\sqrt{\mu})$ and we get $t = 1.118 < t_{0.025}(n-1) = 1.98$. Thus, at the 5% significant level, the hypothesis cannot be rejected, and it is reasonable to get the result that the total amount of time required for bus doors opening and closing as well as passenger preparation is about 6.936 seconds, which can also be approximated as 7 seconds.

3.2. Development of Model II. To clearly show the effect of crowding in bus dwell time, model II is established in almost the same way as described in Section 3.1, except for the crowding level which is not considered in this model; that is, model II describes the relationship between the number of boarding and alighting passengers and bus dwell time.

According to the method described in Section 3.1.1, we get the estimation model of boarding time and alighting time as shown in (9) and (10), respectively:

$$\ln Y_{1i} = 0.736 + 0.973 \ln X_{1i}, \quad (9)$$

$$\ln Y_{2i} = 0.416 + 0.875 \ln X_{2i}. \quad (10)$$

Thus, the bus dwell time estimation model based on model (9) and model (10) is listed as follows:

$$Y_i = 6.936 + 0.968 \max\{Y_{1i}, Y_{2i}\}. \quad (11)$$

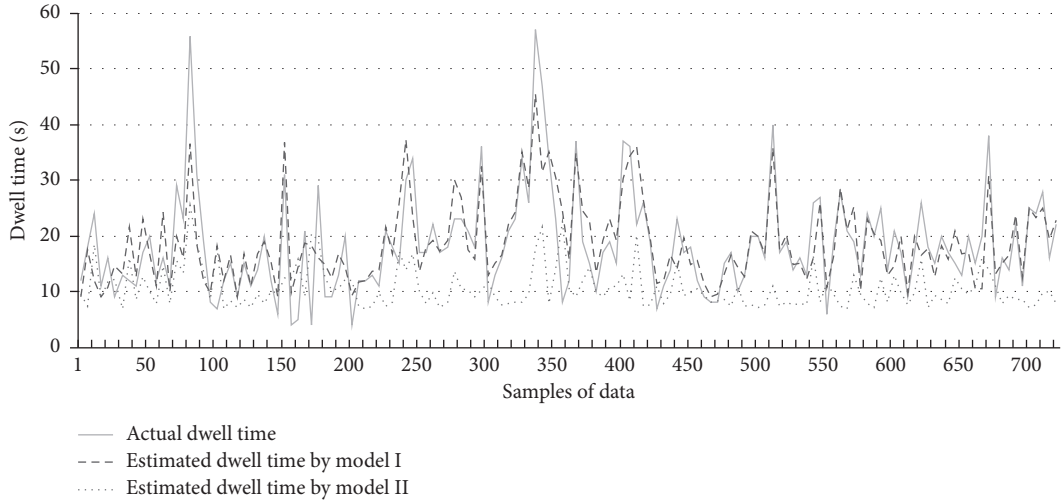

Figure 4: Comparison between real and estimated dwell times based on model I and model II.

3.3. Comparisons of the Two Models. To improve the accuracy of bus dwell time estimation model and analyze the effect of bus crowding level, a comparison is conducted between the extended and the traditional model in this section.

(1) Comparisons of Estimation Errors. Figure 4 unveils the predicted dwell time estimated by model I (described by red curve) and model II (described by blue curve) compared with the actual recorded bus dwell time denoted by the black curve. Among the sample with 640 independent data, the maximum bus dwell time is 57 seconds, while the minimum is only 4 seconds. Figure 4 clearly shows that model I fits the real dwell time better. Next, we compare the absolute error of these two models in Figure 5, which indicates the comparisons of predicted value versus observed value. In Figure 5, red line is the absolute error of model I and the blue curve represents that of model II.

Obviously, we can directly see that model I performs much better than model II, as the absolute error of model I is much smaller than that of model II. To measure the overall performance more accurately, we study the mean absolute error (MAE) of these two models as defined by

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \widehat{Y}_i\right|, \tag{12}$$

where n is the size of random sample; Y_i indicates the actual time; and $\widehat{Y}_i$ is the estimated value of Y_i. Here, Y_i may represent passenger boarding time, alighting time, or bus dwell time.

The MAE of expanded models and traditional models are summarized in Table 7.

Table 7 shows that MAE of dwell time between model II and actual bus dwell time is up to 9.3 seconds, while the MAE caused by model I is only 3.9 seconds averagely, reduced by 137.51%. It unveils that bus crowding level is a nonnegligible factor in bus dwell time estimation and taking the carriage crowding into consideration can highly improve the accuracy of the estimation model for bus dwell time. Meanwhile, we also notice that the MAE of model I (about 3.9 seconds on average) is acceptable in both static and dynamic bus schedule.

In Table 8, we display 6 groups of the collected survey data at stops, including the number of boarding and alighting passengers, carriage crowding level, and actual bus dwell time. It reveals that the carriage crowding level plays a vital role in bus dwell time when the number of boarding and alighting passengers at different stops is the same, which is consistent with finding derived from Table 7.

(2) Analysis on Absolute Errors. In model I, the absolute errors of 391 data sets are less than 3 seconds, taking 61.9% of collected samples. The data with absolute error larger than 5 seconds occupy a fairly small proportion, 22.5%. Among 640 groups of collected data, there are only 21 groups (accounting for 3.3%) with absolute error exceeding 10 seconds. The maximum absolute error (19.46 seconds) occurs when 17 passengers board, 19 passengers alight, and the carriage crowding level is 0.255.

In model II, 18.6% of absolute errors are less than 3 seconds. However, the absolute errors lager than 5 seconds account for 65.9% of total samples, with 31.9% and 6.7% exceeding 10 seconds and 20 seconds, respectively. The maximum absolute error in model II reaches up to 39 seconds, where 20 passengers board, 6 passengers alight, and the carriage crowding level is 0.956.

Further comparison reveals that the maximum absolute errors in both models occur when the numbers of boarding and alighting passengers are large. This is because among all the collected data, 70.9% and 85.4% refer to the case where boarding and alighting passengers are less than 10 people and 15 people, respectively. Meanwhile, these data play a vital role in the parameter regression of the dwell time estimation model. Consequently, both estimation models are less accurate when the numbers of boarding and alighting passengers are relatively large.

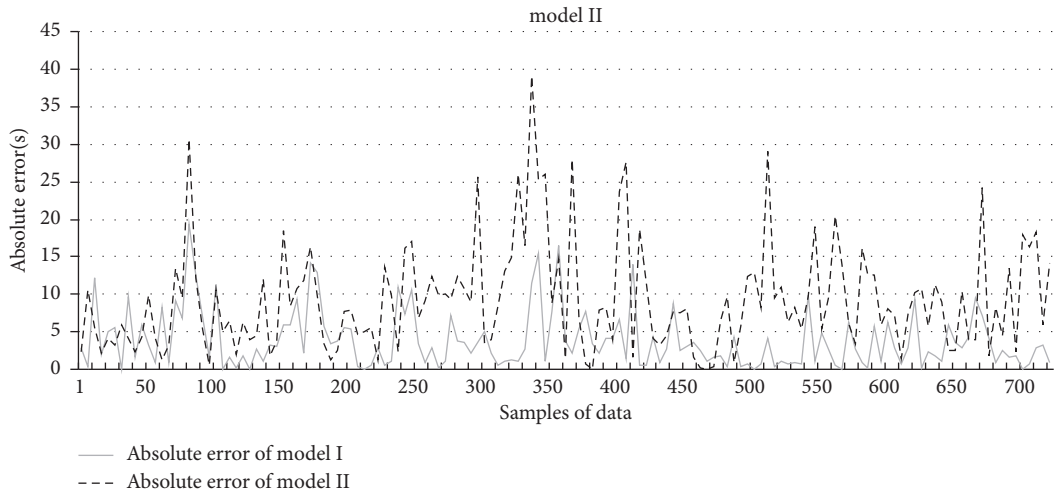

Figure 5: Comparison of the absolute errors between model I and model II.

Table 7: Comparison of MAE between model I and model II (unit: second).

Model	Boarding time	Alighting time	Dwell time
Model I	3.324	3.134	3.904
Model II	8.943	7.917	9.274

Table 8: Part of actual survey data of Route 8.

Data type	Values of the survey data					
Number of boarding passengers	2	2	5	5	8	8
Number of alighting passengers	6	7	8	7	0	0
Carriage crowing level	0.078	0.396	0.299	0.806	0.226	0.986
Actual dwell time (s)	9	13	14	21	21	34

4. Conclusions

The purpose of this research is to study the effect of bus crowding level on bus dwell time. The data is collected along two major bus routes in Harbin by manual data collection. Two statistical models are generated in this paper: one is developed based on the effect of both carriage crowding and passenger volume, while the second model only considers the effect of the number of boarding and alighting passengers on bus dwell time. By comparing these two models, we get some useful insights summarized as follows:

(1) The accuracy of model I is evidently higher than that of model II, with MAE reduced by 137.51%. It displays that the precision of bus dwell time model can be improved significantly by considering bus crowding level.

(2) Except for bus crowding level, bus dwell time is also influenced by the total amount of time used for opening and closing bus doors as well as passenger preparation. According to our analysis, this time is about 7 seconds.

(3) Our model also reveals that a 1% increase in the number of boarding (alighting) passengers will lead to 0.926% (0.848%) increase in boarding (alighting) time. Meanwhile, a 1% increase in crowding level can increase the boarding (alighting) time by 0.085% (0.092%).

In the future, the developed model would be tested on more bus routes. The impacts of other contributing factors can also be considered, such as the route overlap, the usage of mobile payment, and weather conditions. In addition, the developed model is also useful to improve the bus travel time prediction accuracy, which is important in dynamic bus scheduling.

Data Availability

The data supporting the model development of the study can be found in "Data Collection" section.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (nos. 71771062, 71771050, and 71971097), the 13th Five-Year Plan for Scientific Research Project of Department of Education, Jilin Province, China (no. JJKH20180149 KJ), and China Postdoctoral Science Foundation (no. 2019M661214).

References

[1] X. Ma, Y.-J. Wu, Y. Wang, F. Chen, and J. Liu, "Mining smart card data for transit riders' travel patterns," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 1–12, 2013.

[2] Y. Bie, R. Tang, and L. Wang, "Bus scheduling of overlapping routes with multi-vehicle types based on passenger OD data," *IEEE Access*, vol. 8, pp. 1406–1415, 2019.

[3] X. Qu, M. Zhou, Y. Yu, C. T. Lin, and X. Wang, "Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: a reinforcement learning based approach," *Applied Energy*, vol. 257, Article ID 114030, 2020.

[4] M. Zhou, Y. Yu, and X. Qu, "Development of an efficient driving strategy for connected and automated vehicles at signalized intersections: a reinforcement learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 1, pp. 433–443, 2020.

[5] S. Liang, M. Ma, and S. He, "Multiobjective optimal formulations for bus fleet size of public transit under headway-based holding control," *Journal of Advanced Transportation*, vol. 2019, Article ID 2452348, 14 pages, 2019.

[6] D. Arbis, T. H. Rashidi, V. V. Dixit, and U. Vandebona, "Analysis and planning of bicycle parking for public transport stations," *International Journal of Sustainable Transportation*, vol. 10, no. 6, pp. 495–504, 2016.

[7] X. Ma, C. Liu, H. Wen, Y. Wang, and Y. J. Wu, "Understanding commuting patterns using transit smart card data," *Journal of Transport Geography*, vol. 58, pp. 135–145, 2017.

[8] Y. Bie, X. Xiong, Y. Yan, and X. Qu, "Dynamic headway control for high-frequency bus line based on speed guidance and intersection signal adjustment," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 1, pp. 4–25, 2020.

[9] R. Rajbhandari, S. I. Chien, and J. R. Daniel, "Estimation of bus dwell times with automatic passenger counter information," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1841, no. 1, pp. 120–127, 2003.

[10] K. Kepaptsoglou and M. Karlaftis, "Transit route network design problem: review," *Journal of Transportation Engineering*, vol. 135, no. 8, pp. 491–505, 2009.

[11] T. K. Ojo, "Quality of public transport service: an integrative review and research agenda," *Transportation Letters*, vol. 11, no. 2, pp. 104–116, 2019.

[12] A. Ermagun, T. H. Rashidi, and A. Samimi, "A joint model for mode choice and escort decisions of school trips," *Transportmetrica A: Transport Science*, vol. 11, no. 3, pp. 270–289, 2015.

[13] M. Babaei, M. Rajabi-Bahaabadi, and A. Shariat-Mohaymany, "Estimation of travel time reliability in large-scale networks," *Transportation Letters*, vol. 8, no. 4, pp. 229–240, 2016.

[14] Y. Wang, Y. Bie, and Q. An, "Impacts of winter weather on bus travel time in cold regions: case study of Harbin, China," *Journal of Transportation Engineering, Part A: Systems*, vol. 144, no. 11, Article ID 05018001, 2018.

[15] Y. Bie, D. Wang, and H. Qi, "Prediction model of bus arrival time at signalized intersection using GPS data," *Journal of Transportation Engineering*, vol. 138, no. 1, pp. 12–20, 2012.

[16] Y. Bie, X. Gong, and Z. Liu, "Time of day intervals partition for bus schedule using GPS data," *Transportation Research Part C: Emerging Technologies*, vol. 60, pp. 443–456, 2015.

[17] Y. Lin, X. Yang, and N. Zou, "Passive transit signal priority for high transit demand: model formulation and strategy selection," *Transportation Letters*, vol. 11, no. 3, pp. 119–129, 2019.

[18] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.

[19] S. Wang and X. Qu, "Station choice for Australian commuter rail lines: equilibrium and optimal fare design," *European Journal of Operational Research*, vol. 258, no. 1, pp. 144–154, 2017.

[20] S. Wang, X. Qu, and Y. Yang, "Estimation of the perceived value of transit time for containerized cargoes," *Transportation Research Part A: Policy and Practice*, vol. 78, pp. 298–308, 2015.

[21] Z. Liu, Y. Yan, X. Qu, and Y. Zhang, "Bus stop-skipping scheme with random travel time," *Transportation Research Part C: Emerging Technologies*, vol. 35, pp. 46–56, 2013.

[22] M. T. Li, F. Zhao, L. F. Chow, H. Zhang, and S. C. Li, "Simulation model for estimating bus dwell time by simultaneously considering numbers of disembarking and boarding passengers," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1971, no. 1, pp. 59–65, 2006.

[23] F. Li, Z. Duan, and D. Yang, "Dwell time estimation models for bus rapid transit stations," *Journal of Modern Transportation*, vol. 20, no. 3, pp. 168–177, 2012.

[24] E. M. González, M. G. Romana, and O. M. Á. M., "Bus dwell-time model of main urban route stops," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2274, no. 1, pp. 126–134, 2012.

[25] S. Rashidi and P. Ranjitkar, "Estimation of bus dwell time using univariate time series models," *Journal of Advanced Transportation*, vol. 49, no. 1, pp. 139–152, 2015.

[26] S. Rashidi and P. Ranjitkar, "Bus dwell time modeling using gene expression programming," *Computer-Aided Civil and Infrastructure Engineering*, vol. 30, no. 6, pp. 478–489, 2015.

[27] A. Kathuria, M. Parida, C. Sekhar, and M. Pathak, "Examining bus lost time dynamics for a bus rapid transit station," *Journal of Public Transportation*, vol. 19, no. 2, pp. 168–182, 2016.

[28] X. Liu, Y. Yang, M. Meng, and A. Rau, "Impact of different bus stop designs on bus operating time components," *Journal of Public Transportation*, vol. 20, no. 1, pp. 104–118, 2017.

[29] R. P. Guenthner and K. Hamat, "Transit dwell time under complex fare structure," *Journal of Transportation Engineering*, vol. 114, no. 3, pp. 367–379, 1988.

[30] J. C. Levine and G. W. Torng, "Dwell-time effects of low-floor bus design," *Journal of Transportation Engineering*, vol. 120, no. 6, pp. 914–929, 1994.

[31] M. Milkovits, "Modeling the factors affecting bus stop dwell time: use of automatic passenger counting, automatic fare counting, and automatic vehicle location data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2072, no. 1, pp. 125–130, 2008.

[32] G. Currie, A. Delbosc, and J. Reynolds, "Modeling dwell time for streetcars in Melbourne, Australia, and Toronto, Canada," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2275, no. 1, pp. 22–29, 2016.

[33] S. Jaiswal, J. Bunker, and L. Ferreira, "Influence of platform walking on BRT station bus dwell time estimation: Australian analysis," *Journal of Transportation Engineering*, vol. 136, no. 12, pp. 1173–1179, 2010.

[34] R. Fernández, P. Zegers, G. Weber, and N. Tyler, "Influence of platform height, door width, and fare collection on bus dwell time," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2143, no. 1, pp. 59–66, 2010.

[35] A. Tirachini, "Bus dwell time: the effect of different fare collection systems, bus floor level and age of passengers," *Transportmetrica A: Transport Science*, vol. 9, no. 1, pp. 28–49, 2013.

[36] Q. Meng and X. Qu, "Bus dwell time estimation at bus bays: a probabilistic approach," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 61–71, 2013.

[37] D. Katz and L. A. Garrow, "The impact of bus door crowding on operations and safety," *Journal of Public Transportation*, vol. 15, no. 2, pp. 71–93, 2012.

[38] G. Fletcher and A. El-Geneidy, "Effects of fare payment types and crowding on dwell time," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2351, no. 1, pp. 124–132, 2013.