

Statistical Modeling of Bus Dwell Time at Stops

Hooi Ling KHOO

Department of Civil Engineering, Universiti Tunku Abdul Rahman, Kuala Lumpur, 53300, Malaysia. Email: khoohl@utar.edu.my

Abstract: Bus dwell time is one of the important factors that could influence the bus transit system's level of service. It is also an important input for planning and modeling of bus transit system. The objective of this study comes in three-fold. First, it aims to investigate the possible factors that affect the dwell time variability. Second, it intends to find the best statistical distribution that could be adopted to explain and describe the dwell time variability. Third, it aims to develop regression models to understand the degree of influence for each considered factor. Statistical modeling approach is adopted with the aid of statistical software. Dwell time data is collected from 20 bus stops in Klang Valley region, Malaysia using the video recording technique. Results show that the dwell time is influenced by the time of day, payment method, time of the day, and platform crowding level.

Keywords: Dwell time, Regression, Bus, Probability distribution

1. INTRODUCTION

Evaluation of bus transit performance plays an important role in representing its efficiency and as a direct way for better management and improvement. These include accessibility, availability, reliability, safety and comfort issues. Besides convenience and cost, bus transit users are concern in the availability and reliability of the services. Availability defines the provision of bus services within a reasonable distance for the user access while reliability defines the level of punctuality and schedule adherence of the service. Lack of service reliability would result in uncertainty and delays that aggravating anxiety and discomfort for the passengers. Bus transit service reliability is related to the operational characteristics of the system which include bus operating speed, dwell time at stops, drivers' layover time, and the service route length. In fact, Li and Li (2006) stressed that dwell time at stops is one of the most important factors that should be considered in improving bus transit service quality as it is the major delay that is not encountered by the private cars in the network. In addition, Maloney and Boyle (1999) shows that dwell time at stops contributes about 9% -11% of the total bus travel time.

Dwell time at bus stops is a necessary and important input to any bus transit modeling and analysis studies. Many of the existing microscopic traffic simulation models require the specification of empirical distributions to describe the dwell time pattern or functions to estimate dwell time at stops. For example, Paramics (Quadstone, 2013) and Vissim (Visual Solutions, 2013) allow users to specify the dwell time distribution type besides using their default normal distribution. With proper distribution in place, the bus transit services could be simulated closer to reality and improve the results accuracy. Furthermore, dwell time at stops is needed in bus route and schedule planning and design (Fu, 2003).

In general, two approaches are adopted to model and describe the bus dwell time at stops, i.e.

probability distribution and regression modeling. Li and Li (2006) attempted to fit the bus dwell time data collected from three selected bus routes in Florida, USA. They tested sets of data into various distributions and found that lognormal distribution appears to be the best fit. Levinson (1983) developed a simple regression model to predict bus dwell time with the sum of boarding and alighting passengers as the only factor considered. A 5 second is added to the dwell time for bus door open and close. Pretty and Russel (1988) proposed a bus dwell time model. The variables considered are number of alighting/boarding passengers, passengers boarding and alighting time, and time door open and close. York (1993) extended the model by including payment methods while Jaiswal et al. (2008) considered the platform crowding density and walking distance in the dwell time model.

Several factors affect the bus dwell time. Jaiswal et al. (2007) suggested that the platform crowding pattern has a significant effect on dwell time. It affects the passengers' maneuverability and obstructs the clear line of sight to approaching buses. Dorbritz et al. (2008) and Jaiswal et al. (2009) found that the payment method could affect the bus dwell time. Fernandez et al. (2009) showed that the dwell time variability is affected by the platform height, door width, and fare collection method. Results show that by removing on-board ticketing system, the boarding time could decreased by about 15%. Dueker et al. (2004) found that the lift operation would increase the bus dwell time significantly although its occurrence is rare. Bladikas et al. (2009) claimed that the bus dwell time could be affected by the adverse weather as well. They mentioned that adverse weather increases boarding and alighting times.

The objective of this study comes in three-fold. First, it aims to investigate the possible factors that affect the dwell time variability. Second, it intends to find the best statistical distribution that could be adopted to explain and describe the dwell time variability. Third, it aims to develop regression models to understand the degree of influence for each considered factor. Statistical modeling approach is adopted with the aid of statistical software. Dwell time data is collected from 20 bus stops in Klang Valley region, Malaysia with the aid of three students using the video recording technique. Results show that the dwell time is influenced by the time of day, payment method, time of the day, and platform crowding level.

2. METHODOLOGY

In this section, the methodology adopted for the study is explained in details.

2.1 Data Collection

A total of 20 bus stops in Klang Valley region, Malaysia have been identified as the data collection site. The selection criteria are based on the type of location, number of bus routes served, and the estimation of the passenger demand. A site visit is carried out in the pilot study to collect this information for a set of pre-determined bus stops. Then, the bus stops are selected to reflect different level of bus routes and passenger flow level. The bus stop locations and the respective information are shown in Table 1.

A team of 3 student helpers are sent to the bus stops during October 2010 to April 2011 to record the buses dwell time at these stops. The data collection period is categorized into two categories, i.e. peak and non-peak hour. The peak hour period is 8 am – 9 am and 5 pm – 6 pm, while the non-peak hour is 9 am – 10 am and 4 pm – 5 pm. The observers are asked to

record passenger movement, boarding, and alighting activities using the video camera. Figure 1 shows the positions of the observers during video recording. Two cameras are used each held by a person to capture two different angle of the bus stop to obtain different type of data. The camera at position B is to capture the activity of passengers at the platform. When a bus arrives, he speaks out the service route number. The number of passengers on the platform is counted. The camera at position A is to capture the front and rear door of the bus on platform. It captures the number of passenger boarding and alighting. C records manually the payment methods made by each passenger when boarding the bus. He records down the number of passengers that interacts with the bus conductor/driver and those who board without any interaction.

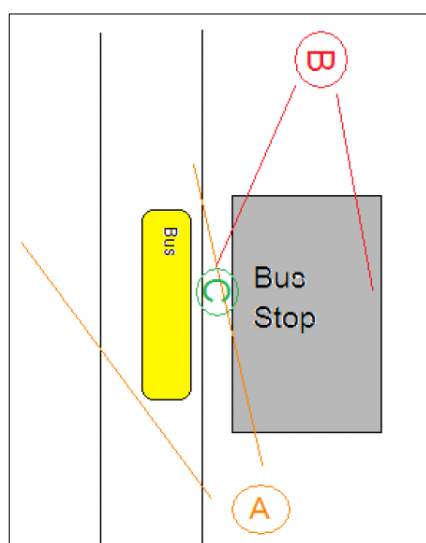


Figure 1. Positioning of observers at bus stops

2.2 Data Extraction

The video footages are played back in the traffic laboratory. The following steps are carried out to extract the required data from the video:

1. Record the plate number and service route number (and its operating company) for each bus which pull-over at the bus stop.
2. Record the time when the bus arrives and comes into a complete halt.
3. Record the number of boarding and alighting passengers at front door.
4. Record the number of boarding and alighting passengers at rear door.
5. Validate the passengers' payment method with the on-site records.
6. Record the number of passengers at the platform.
7. Record the time when the bus closes its doors and departs.

All the data is then recorded in the Excel sheet.

The dwell time is computed as follow:

$$t^i = t_{depart}^i - t_{arrive}^i \quad (1)$$

where t^i : dwell time for bus i

t_{depart}^i : time bus i departs from stop

t_{arrive}^i : time bus i arrives at stop

Table 1. Bus stop locations

Bus stop	Type of location	Location	Remarks
1	City Centre	Jalan Travers	In front of Tiong Nam building
2			In front of Malaysian Institution of Accountants
3		Jalan Raja Chulan	Opposite of Public Bank
4			Opposite of the Weld
5		Jalan Pudu	Opposite of UOB Bank
6		Jalan Raja Laut	NA
7		Jalan Tunku Abdul Razak	At Jamek Mosque
8		Jalan Ampang	Kuala Lumpur Convention Centre Entrance
9		Jalan Bangsar	Opposite of Jalan Riong
10		Jalan Tan Cheng Lock	In Central Market
11	Sub-urban	Jalan Pahang	Near Tawakal Hospital
12			Near Chow Kit Monorail
13			Opposite of Kuala Lumpur General Hospital
14		Jalan Ipoh	Near Sungai Mas Plaza
15			In front of Mutiara Complex
16		Jalan Cheras	Jusco Shopping Complex Main Entrance
17		Jalan Munshi Abdullah	NA
18		Sri Rampai	Near Sri Rampai Lake
19		Jalan Gombak	In front of Chung Hwa School
20		Jalan Genting Klang	Near KFC

2.3 Analysis Method

Two methods are adopted, i.e. statistical distribution fitting and multiple regression modeling to analyse the dwell time data. Statistical distribution fitting is used to find the best probability distribution that could explain the dwell time data pattern. Regression modeling is adopted to investigate the factors that influencing dwell time duration.

2.3.1 Factors

The variables considered in the analysis include time of the day (peak hour/off peak hour), payment method (i.e. cash/card and conductor system), platform crowding level, and boarding/alighting activities. Currently, there are two main bus companies providing the bus transit services, namely RapidKL and Metro Bus. Both operators use different payment system. RapidKL has the electronic payment system (EPS) which allows the passengers to make payment using smart card. However, the EPS is not well received by passengers. Some may still prefer to pay cash by inserting their coins or notes into the payment machine. Metro Bus uses the conductor system in which a conductor is employed to collect fare and issue tickets. The platform crowding level is classified into two categories, i.e. less crowded and

more crowded. A bus stop with less than 15 passengers per hour waiting at the platform is classified as less crowded bus stop and vice versa for more crowded bus stops. The chosen criterion (15 passengers per hour) is 1.5 times higher than the average passenger volume at bus stops during peak period (from the data collected). For the boarding/ alighting variables, the number of passengers boarding at front and rear doors, and alighting from front and rear doors are used as the independent variables.

2.3.2 Statistical distribution fit

The dwell time data is fitted to find the best probability distribution that explains its pattern using StatFit (Geer Mountain Software Corporation, 2013). Its suitability is then analysed with the goodness-of-fit tests. Chi-squared test is adopted to test the level of fit of the probability density function while Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) test is adopted to test the cumulative distribution. Two types of output are expected, i.e. *reject* and *do not reject*. *Reject* output means that the statistic value calculated for the particular distribution is larger than the critical value stated, thus the distribution assumption is rejected as its parameters do not support the estimated value. The parameters considered are such as: accuracy of fit, level of significance and interval types, and the type of estimates. As such, the distribution with *do not reject* output is favored. For a same set of data, there might be more than a suitable distribution that could fit the data. StatFit ranked the distribution between 0 and 100 by combination of the KS test and AD test. The distribution with higher ranking has better fit and more favored. In this study, maximum likelihood method is adopted to estimate the parameters of the distribution. The accuracy and significant level is set at 0.05 and 95% respectively.

2.3.3 Multiple regression modeling

Regression analysis is applied to estimate the factors that determine the dwell time duration of buses at stops. The dependent variable is dwell time (measured in seconds) while the independent variables are number of passengers boarding and alighting, fare collection method (in terms of passenger volume using cash and card payment), and platform crowding level (in terms of passenger volume at stops). Stepwise multiple regressions are performed using statistical software. The best regression with highest R^2 value is chosen. The significant level chosen is 95% confidence level.

3. RESULTS AND DISCUSSIONS

This section presents the findings for the dwell time statistical distribution fitting and regression modeling. The analysis is carried out in a few categories, such as peak-off peak hours, platform crowding level, and bus operating companies. The findings are presented and compared within the same category.

3.1 Statistical Distribution Results

This sub-section presents the findings for dwell time statistical distribution fitting. For each category, the dwell time descriptive statistics is presented first, followed by the elaboration on the best fit distribution.

3.1.1 Time of the day

Table 2 shows the descriptive statistics for the peak and off-peak hour data while Figure 2 presents the results of data fitting. It is observed that a few distributions are qualified to represent the same set of data. The probability density fit for these distributions are about the same and could be acceptable. For example, the density fit for the first two distributions, i.e. Pearson 6 and Weibull shown in Figure 3A have similar curve patterns. However, if the residues (differences between the curve point and the data point) are referred, Pearson 6 distribution has smaller differences compared to Weibull (shown in Figure 3B). As such, the best fit in which the distribution with the highest rank should be chosen, i.e. Pearson 6 (2, 79.3, 1.65, 6.50) with normalized p-value of 0.28 (KS test) for peak hour and Person 6 (2, 59.5, 2.09, 7.17) with normalized p-value of 0.79 (KS test) for off-peak hour respectively.

Table 2. Descriptive statistics for peak and off-peak hour

Item	Peak Hour	Off-peak Hour
Number of data points	275	209
Mean (seconds)	25.2	21.8
Standard deviation (seconds)	23.2	15.9
Coefficient of variation	92.3	72.9

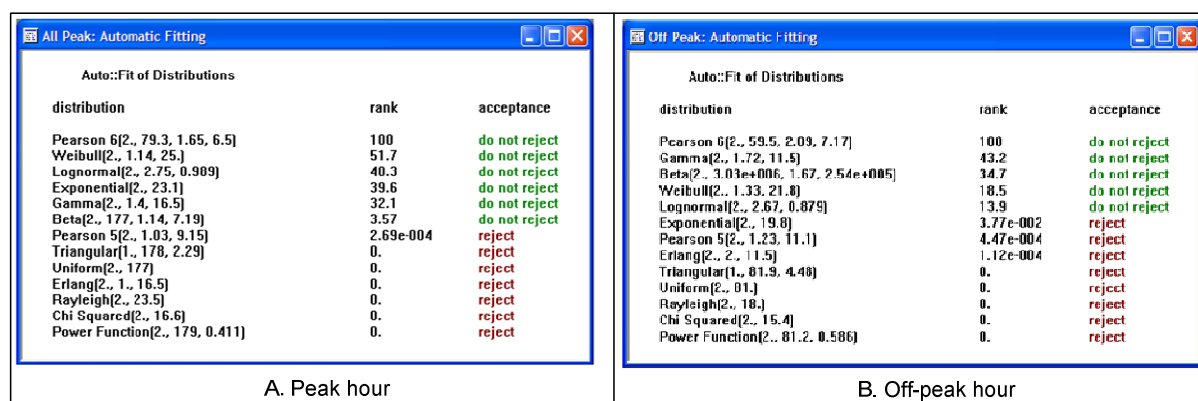


Figure 2. Peak and off-peak hour dwell time distribution

It could be observed from Table 2 and Figure 3A that the peak hour has higher mean and coefficient of variance value which indicates that the dwell time measured during peak hour tend to disperse more compared to off-peak hour. This could be explained by bus bunching effect during peak hour. Traffic congestion on the roadway causes larger bus headway that lead to more passenger accumulation and waiting at the platform. Thus, more boarding activities occur when the first bus arrives that increases the dwell time. Subsequently, lesser number of passengers waits and boards the buses that come later, which leads to shorter dwell time (bus stops only for alighting activity). Whereas during off-peak hour where the traffic condition is smooth, its dwell time tend to concentrate more in the same interval due to bus punctuality within schedule with constant headway, allowing similar number of passengers board that ultimately lead to shorter dwell time.

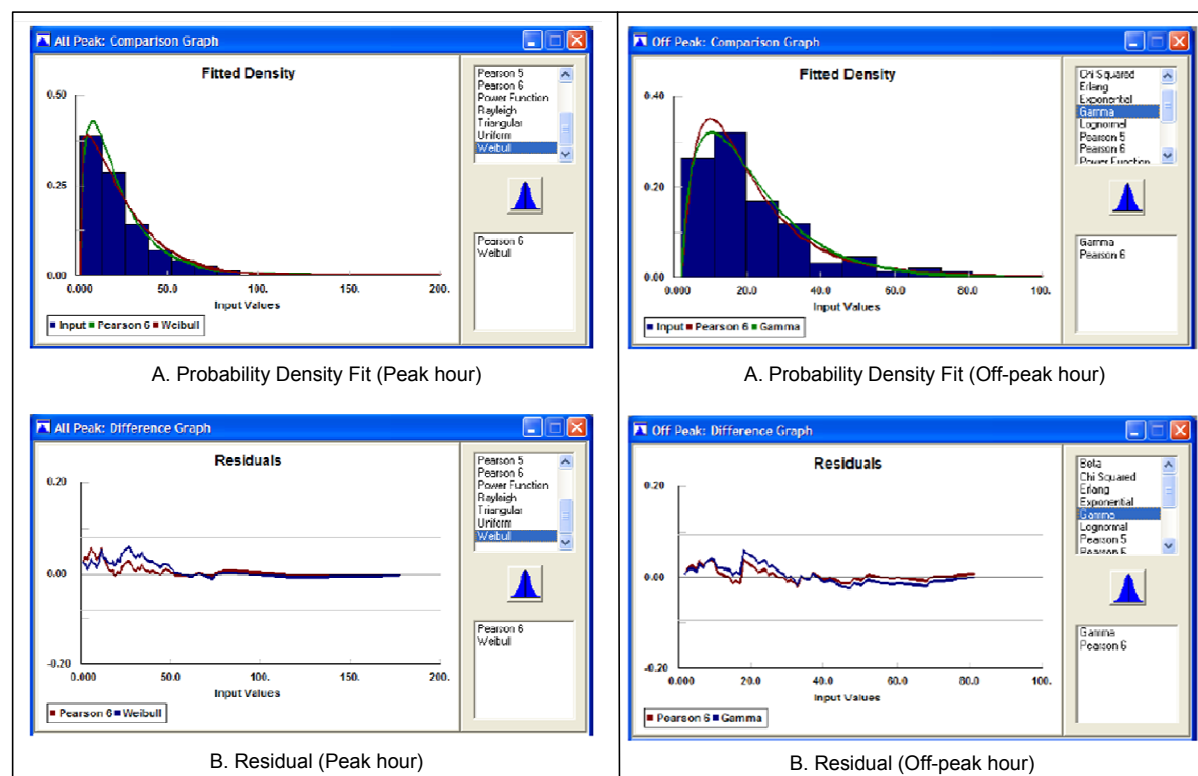


Figure 3. Probability density fit and residuals for peak and off-peak hour

3.1.2 Platform crowding level

It is interesting to investigate whether the crowding level at the platform will affect the dwell time. Table 3 shows the descriptive statistics for dwell time with different crowding level. If the average number of passenger waiting at the platform is lesser than 15 persons, the data obtained is classified as less crowded dwell time data, and vice versa for more crowded category. Figure 4 shows the distribution fitting results. The best fitted distributions are Weibull (2,1.26,19.6) with normalized p-value of 0.182 in KS test and Pearson 6 (6,80.1,1.81,5.65) with normalized p-value of 0.136 in KS test for less and more crowded categories respectively.

It could be observed from Table 3 that the mean dwell time for less crowded category is lower compared to more crowded category. Both data set exhibit different type of distribution pattern. This evidences that the crowding level could influence the dwell time duration. Crowded platform tends to decrease passengers' maneuverability, with over-crowded boarding passengers blocking the smooth alighting activities and vice versa. In addition, with more passengers waiting at the platform, there is higher possibility that more passengers among them will board the same bus, thus creating congestion at bus stops that unavoidably increasing the dwell time and delay. Furthermore, crowded platform could obstruct the clear line of sight to approaching bus. This might result in an increased reaction time for passengers on the arrival of expected bus due to their lower reaction time for passengers on the arrival of expected bus due to their lower readiness and awareness, at the same time increase the overall dwell time.

Table 3. Descriptive statistics for platform crowding level

Item	Less crowded	More crowded
Number of data points	371	118
Mean (seconds)	19.7	36
Standard deviation (seconds)	15.4	22.8
Coefficient of variation	77.9	77.3

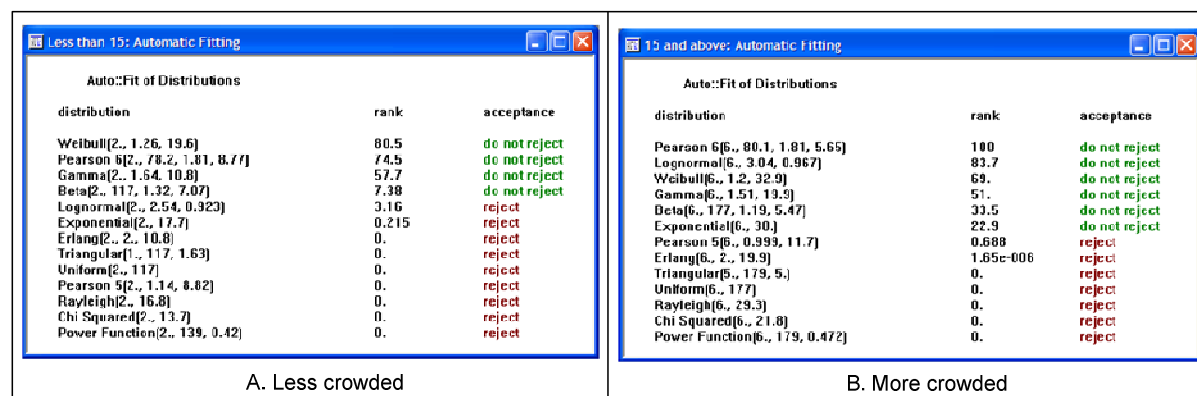


Figure 4. Dwell time distribution for different platform crowding level

3.1.3 Fare collection method

Fare collection method is believed to affect bus dwell time. Currently, two types of payment system, i.e. cash/card system and manual system used by the bus operators. Table 4 shows that the dwell time for cash/card system has higher mean compared to that for the conductor system. This is because the bus with cash/card system has only one driver onboard with no assistant. He manages the payment system at the entrance. Passengers with electronic card could just tap on the machine at the entrance, but those with no card have to make cash payment. The driver himself has to receive payment and issue ticket. Accordingly, passengers have to queue at the entrance which slows down the boarding process. On the other hand, passengers boarding the bus with conductor system could proceed to the rear of the bus or find a seat directly. The conductor will then collect the fare from them. This reduces blockages and delay.

The coefficient of variation is high for the cash/card system as the payment duration is dependent on how individual driver manages the system. Some drivers tend to be more rush and take a few tickets at once from the machine, distribute them to the passengers and let them pay after the bus has departed. Some might want to complete all the payment first before departing. As such, the dwell time variability is high.

The distribution fitting results shown in Figure 5 indicates that Pearson 6 (2., 92.6, 1.69, 7.67) and Pearson 6 (2., 56, 2.6, 6.69) could be adopted to explain the dwell time for both situations. The normalized p-value for cash/card and conductor system is 0.574 and 0.557 in KS test respectively.

Table 4. Descriptive statistics for payment method

Item	Cash/Card System	Conductor System
Number of data points	316	169
Mean	25	21
Standard deviation	22	16.5
Coefficient of variation	88	77.8

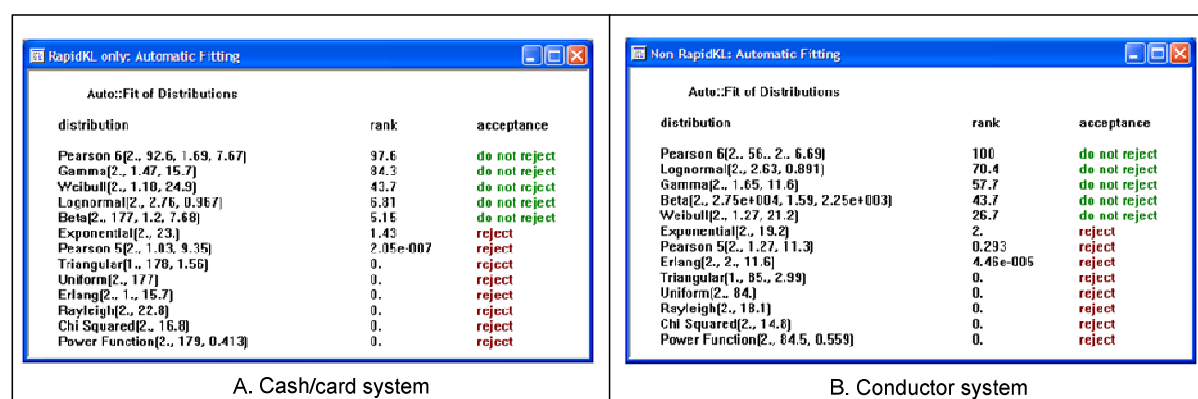


Figure 5. Dwell time distribution for different payment type

3.2 Regression Modeling Results

This section presents the dwell time regression modeling results considering two factors, i.e. time of the day and payment method. Prior to the modeling, a correlation test was carried out to investigate the correlation among the variables used in the modeling. Table 5 shows the Pearson correlation values for all variables are lesser than 0.6 which indicates that there is no strong correlation among them. The P-values show that most of the estimation is statistically significant except three correlation values highlighted in red.

Table 5. Pearson correlation values

	Sum of alighting at front door and boarding at rear door	Sum of boarding at front door and alighting at rear door	Platform crowding level	Card system
Sum of boarding at front door and alighting at rear door	0.433 (0.000)			
Platform crowding level	0.012 (0.849)	0.410 (0.000)		
Card system	0.056 (0.367)	0.437 (0.000)	0.230 (0.000)	
Cash	0.001 (0.984)	0.597 (0.000)	0.553 (0.000)	0.281 (0.000)

Note: Value in bracket is P-value

3.2.1 General

The regression model obtained in general by considering the whole set of data (without differentiating time of day) is presented in eqn. (2). The model shows that all the factors considered could increase the dwell time in which payment method has the most significant impact. Payment by cash contributes about 2.3 times more delay compared to payment by card. This follows by the boarding and alighting activities. More passengers boarding and alighting will increase the dwell time. Lastly, the platform crowding level contributes the least impact on dwell time. All factors presented in eqn. (2) are significant at 95% confidence level and $R^2 = 0.75$.

$$DT_G = 4.93 + 4.27Cash + 1.07(BF + AR) + 0.53PCL + 1.81Card$$

[11.98,0]	[6.4,0]	[4.82,0]	[3.62,0]	(2)
(3.1,3.5)	(4.9,6.4)	(9.5,7.4)	(1.8,2.6)	

Note: [t-statistic, p-value]
(mean, standard deviation)

where DT_G indicates the dwell time (in seconds) for general model; $Cash$ indicates number of passenger who make fare payment by cash (in number of passenger); $(BF + AR)$ indicates the sum of passengers who board at front door and alight at rear door (in number of passenger); PCL indicates Platform Crowding Level, i.e. the passenger volume who wait at stops (in number of passenger); $Card$ indicates the number of passenger who make fare payment by EPS (in number of passenger).

3.2.2 Time of the day

Regression models for peak hour and off-peak hour are developed and shown in eqn. (3) and eqn. (4) respectively. All factors shown are statistically significant at 95% confidence level and the R^2 value for models shown in eqns. (3) and (4) are 0.76 and 0.74 respectively. It could be observed that payment method is still the most significant factor for both peak and off-peak hour. The peak hour model has the significant factors similar to that for general model. Besides, platform crowding level affects the dwell time positively. During peak hour, more passengers are waiting at the platform and this could obstruct the mobility and visibility of the boarding or alighting passengers. However, this is not the case during off-peak period. It is not a significant factor for off-peak dwell time model because there is significantly lesser number of passengers waiting at platform during off-peak hour.

The bus dwell time during off-peak hour is influenced by the payment method and boarding/alighting activities only. It is interesting to note that the factor: number of passengers boarding from the rear and alighting from the front, has more significant impact compared to the factor with usual way of boarding/alighting. And this factor is significant only for off-peak hour model. This shows that during off-peak hour, the passengers have more flexibility to choose their boarding/alighting point according to their convenience. This has caused additional delay. For example, a passenger who sits near the entrance might choose to alight from the front door. This causes the passengers at the platform to wait for him/her to alight

before boarding which incur additional delay.

$$DT_{Peak} = 4.58 + 4.35Cash + 0.99(BF + AR) + 0.56PCL + 1.91Card$$

[9.55,0]	[4.12,0]	[4.06,0]	[2.65,0]	(3)
(3.5,4.0)	(6.2,7.5)	(10.9,8.3)	(1.4,2.7)	

Note: [t-statistic, p-value]
(mean, standard deviation)

where DT_{Peak} indicates the dwell time (in seconds) for peak hour; $Cash$ indicates number of passenger who make fare payment by cash (in number of passenger); $(BF + AR)$ indicates the sum of passengers who board at front door and alight at rear door (in number of passenger); PCL indicates Platform Crowding Level, i.e. the passenger volume who wait at stops (in number of passenger); $Card$ indicates the number of passenger who make fare payment by EPS (in number of passenger).

$$DT_{Off} = 8.2 + 4.17Cash + 0.95(BF + AR) + 1.58(AF + BR) + 2.38Card$$

[6.68,0]	[3.58,0]	[2.29,0]	[3.6,0]	(4)
(2.3,1.8)	(4.5,1.2)	(4.4,2)	(2.5,2)	

Note: [t-statistic, p-value]
(mean, standard deviation)

where DT_{Off} indicates the dwell time (in seconds) for off-peak hour; $Cash$ indicates number of passenger who make fare payment by cash (in number of passenger); $(BF + AR)$ indicates the sum of passengers boarding at front door and alighting at rear door (in number of passenger); $(AF + BR)$ indicates the sum of passengers alighting at front door and boarding at rear door (in number of passenger); $Card$ indicates the number of passenger who make fare payment by EPS (in number of passenger).

4. CONCLUSIONS

This study has successfully developed statistical distributions to describe and explain the dwell time variability. It was found that Person 6 distribution is the best distribution to be adopted for most of the cases, except for dwell time recorded at less crowded bus stops. The regression models developed revealed that payment method is the most significant factor that contributes to the prolonged dwell time. This is especially true for the cash/card payment method used by some operators in the region. Second, the boarding/alighting activities could affect the dwell time variability as well. Furthermore, the boarding/alighting pattern is significantly differing during peak and off-peak hour which has shown to have impact on the dwell time variability. The limitation of the study is that the number of bus line using a stop is not considered as one of the variable that might affect the dwell time. This could be considered in future study.

ACKNOWLEDGMENT

The research study is supported by Science Fund (06-02-11-SF0085) under the Ministry of Science and Technology (MOSTI) Malaysia. The author would like to thank Mr. Gan Wei Huang, Mr. Toh Boon Leong, and Mr. Tan Kok Yong for their assistance in data collection and analysis.

REFERENCES

- Bladikas, A.K., Tsai, F.M., and Chien, I-J. (2009). Evaluation of bus travel time and schedule adherence under adverse weather. Paper presented in the 88th Annual Meeting of Transportation Research Board, Washington D.C., 2009.
- Dueker, K.J., Kimpel, T.J., and Strathman, J.G. (2004). Determinants of bus dwell time. *Journal of Public Transportation*, 7(1), 21-40.
- Fernandez, R., Zegers, P., Weber, G. and Tyler, N. (2010). Platform height, door width and fare collection on public transport dwell time: A laboratory study. Paper presented in the 12th World Conference on Transport Research, Lisbon, Portugal, July 11th-15th 2010.
- Fu, L. (2003). On planning and design of flex-route transit services. *Transportation Research Record*, 1791, 59-66.
- Geer Mountain Software Corporation. StatFit Version 2. <http://www.geerms.com/index.htm>. Accessed 1st January 2013.
- Jaiswal, S., Bunker, J.M., and Ferreira, L. (2007). Operating characteristics and performance of a busway transit station. Paper presented in the 30th Australasian Transport Research Fourm (ATRF), 25 - 27 September 2007, Melbourne, Australia.
- Jaiswal, S., Bunker, J.M., and Ferreira, L. (2009). Effects of fare collection policy on operating characteristics of a Brisbane busway station. In Proceedings of : ATRF 2009 : 32nd Australasian Transport Research Forum : The Growth Engine : Interconnecting Transport Performance, the Economy and the Environment, 29 Sep. - 01 Oct. 2009, Auckland, New Zealand.
- Levinson, H.S. (1983). Analyzing transit travel time performance. *Transportation Research Record*, 915, 1-6.
- Li, M.T. and Li, S.C. (2006). A simulation model for estimating bus dwell time by simultaneously considering numbers of alighting and boarding passengers at stop level. *Transportation Research Record*, 1971, 59-65.
- Maloney, M. and Boyle, D. (1999). Components of travel time on the Glendale Beeline bus network. *Transportation Research Record*, 1666, 23-27.
- Pretty, R.L and Russel D.J. (1988). Bus boarding rates. *Australian Road Research*, 3(18), 145-152.
- Quadstone Ltd. (2013). Quadstone Paramics. www.paramics-online.com. Accessed 1st January 2013.
- Visual Solutions (2013). Vissim: A graphical language for simulation and model-based embedded development. www.vissim.com. Accessed 1st January 2013.
- York, I.O. (1993). *Factors affecting bus-stop times*. Transport Research Laboratory, Project Report PR 2, Crowthorne.