

Statistical modelling of bus travel time with Burr distribution

*Victor Jian Ming Low*¹, *Hooi Ling Khoo*^{1*}, and *Wooi Chen Khoo*²

¹Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Jalan Sungai Long, Bandar Sungai Long, 43000 Kajang, Selangor, Malaysia

²School of Mathematical Sciences, Sunway University, Jalan Universiti, Sunway City, 47500 Selangor Darul Ehsan, Malaysia

Abstract. A better understanding of the travel time distribution shape or pattern could improve the decision made by the transport operator to estimate the travel time required for the vehicle to travel from one place to another. Finding the most appropriate distribution to represent the day-to-day travel time variation of an individual link of a bus route is the main purpose of this study. Klang Valley, Malaysia is the study area for the research. A consecutive of 7 months ten bus routes automatic vehicle location (AVL) data are used to examine the distribution performance. The leading distribution proposed for the research is the Burr distribution. Both symmetrical and asymmetrical distributions that have been proposed in existing studies are also used for comparison purposes. Maximum likelihood estimation is applied for parameter estimation while log-likelihood value, Akaike information criterion (AIC) and Bayesian information criterion (BIC) are applied for performance assessment of the distributions. Promising results are obtained by the leading model in all different kinds of operating environment and could be treated as the preliminary preparation for further reliability analysis.

1 Introduction

Travel time variability (TTV) has always been recognized as the key indicator to evaluate the service quality of the transportation system. It simply means punctuality or the consistency of the transportation system for a certain journey. It can also be explained as the compliance of the public transportation system on the planned schedule. In order to improve the attractiveness of the public transportation system, it is important to maximize its service quality or reliability and improve the performance of on-time arrival. This is always a challenge for the road transport operator as the traffic systems have a stochastic nature. The improvement of the transportation system service quality does not only benefit the transport operator in terms of its attractiveness and revenue, it also benefits the users. Travelers have the expectation of on-time arrival or compliance of the transport system on the planned schedule with some acceptable level of delay. Minimizing the TTV value is

*Corresponding author: khoohl@utar.edu.my

found to be more important than minimizing the average travel time as it could reduce the uncertainty of time travel which leads to stress faced by the users [1-2].

Different types of distributions have been tested for TTV modelling. For example, gamma [3-4], normal [5-6], lognormal [7-8] and more recently, Burr distribution [9-10]. However, some research gaps are observed in these studies. Only a few modelling studies were done on bus TTV. It is noted that none of the studies were conducted on bus TTV modelling with Burr distribution. Besides, only a few TTV studies are done on the link level. Most of the distribution fittings on travel time were done on the route level although studies show independence of travel time between individual links of the same route [11-12]. Most of the TTV studies also did not consider the day of week as an important factor in affecting the travel time variation. Only few studies consider this factor but none of the existing studies fit weekday and weekend travel time values separately with the proposed distribution.

This research study aims to fill the research gaps found in the existing studies on statistical modelling of TTV. The main purpose of the study is to examine the suitability of the proposed distribution in explaining link travel time values. The second purpose of the study is to address the issue of the proposed distribution in fitting different link travel time values as the operating environment of each individual link in a bus route is different.

The focus of this paper is to investigate the most appropriate distribution for day-to-day variability in representing both weekday and weekend bus travel time. Klang Valley, Malaysia is the study area for the research. A consecutive of 7 months ten bus routes automatic vehicle location (AVL) data are used to examine the distribution performance. The leading distribution proposed for the research is the Burr distribution. Other symmetrical and asymmetrical distributions, i.e. normal, Weibull, gamma, generalized Pareto and lognormal are also chosen for comparison purposes. Maximum likelihood estimation is applied for parameter estimation while log-likelihood value, Akaike information criterion (AIC) and Bayesian information criterion (BIC) are applied for performance assessment of the distributions. AIC and BIC are more robust in measuring the statistical model quality as they take into account the maximized log-likelihood value and the number of parameters in the model [13-14].

Promising results are obtained from the leading model in all different kinds of operating environments. Results shown that the Burr distribution is particularly well in representing weekday travel time values. Although the performance of the distribution in fitting weekend travel time values is comparably lower, it is still the “best” representation compared to other distributions. Statistical test conducted in the study also found that the Burr distribution is not a good statistical fit for the individual link weekend travel time values. However, only less than 5% of the cases were rejected.

The remaining section of this paper is presented as follows. The existing studies on TTV modelling are discussed in Section 2. The method and data used for the research are discussed in Section 3. The detailed information of the bus travel time data, the proposed distribution and its characteristics are discussed in this section. Results obtained are discussed in Section 4 while the last section concludes the study and highlights the potential further research.

2 Literature review

The study of statistical modelling on public transport TTV started in the 1980s although research on TTV had started for some decades. The main reason for the limited studies conducted is the lack of accessible to public transport travel time data. The issue is being addressed now with the emergence of advanced technology, i.e. AVL systems and global positioning system (GPS). It was first believed that the symmetrical distribution, i.e. normal

distribution would be a suitable model to represent the vehicular travel time data. However, asymmetrical distribution such as gamma and lognormal distribution was found to be a more appropriate model in recent studies. Table 1 shows some existing studies that have proposed a variety type of distributions to represent different transportation mode travel time data.

Table 1. Selected studies of distribution fitting on various transportation modes.

Source	Type of Variation	Transport Mode	Study Area	Analysis Period	Proposed Distribution
[3]	Day-to-day variability	Individual vehicle	Arterial routes, Chicago, USA	Daily working trip	Gamma distribution
[4]	Day-to-day variability	Bus	Sheridan Road Chicago, USA	Morning peak	Gamma distribution
[5]	Day-to-day variability	Bus & metro	Paris, France	Morning peak	Normal distribution
[6]	Day-to-day variability	Individual vehicle	I-66 Freeway, Virginia, USA	24 hours divided according to day of week	Normal distribution
[7]	Day-to-day variability	Bus	Major arterial streets, Hirakata, Japan	17 hours	Lognormal distribution
[8]	Day-to-day variability	Bus	Brisbane, Australia	About 15 hours weekdays only (excluding Public Holidays & School Holidays)	Lognormal distribution
[9]	Day-to-day variability	Individual vehicle	Glen Osmond Road and South Road, Adelaide, Australia	N/A	Burr distribution
[10]	Day-to-day variability	Individual vehicle	Metropolitan area, Adelaide, Australia	Daily trip from home to work	Burr distribution
[17]	Day-to-day variability	Individual vehicle	I-4 corridor, Orlando, Florida, USA	Evening peak weekdays only	Lognormal distribution
[18]	Day-to-day variability	Bus	Calgary, Canada	Morning peak weekdays only	Lognormal distribution & Normal distribution

It is found that the existing studies have the following limitations:

- a) The data sample size is limited in the studies conducted before the 20th's century. Inaccurate results might be produced if limited data sample size is used to carry out the study.

- b) Most of the studies conducted did not include a variety of routes or sections that have different operating environments. For example, Zhang et al. [7] only used the freeway section to conduct the study.
- c) Some distributions proposed can only be used for approximation. The normal distribution and the lognormal distribution could only approximate travel time values as they produce negative travel time values and very small travel time values respectively [15].

The Burr distribution has been proposed as an appropriate model in several reliability studies. In the actuarial applications studies, it is a popular model in modelling failures [16]. It is well known for its flexibility in capturing various forms of data distribution shape and its mathematical tractability. Recently, it has also been proposed in reliability engineering study, in particular TTV study. The Burr distribution was proposed to fit the individual vehicle travel time values and in overall performance it emerged as a plausible model [9-10]. It is able to capture the strong positive skew patterns and very long upper tails exhibited by the TTV data. However, to the best of the authors' knowledge, the model has not been proposed as an appropriate model to represent the bus travel time values.

3 Data and methodology

The data and the method used for the study are explained in this section. A detailed information of the study area on each route is discussed in the first subsection. The extraction and the filtration of the data used are presented in the second subsection. The last subsection presents the Burr distribution as the leading distribution for the research. Detailed information of the distribution such as the statistical properties of the distribution and the suitability of the model in representing the travel time observations for transportation application study is discussed in this subsection.

3.1 Data descriptions and sample size

The Automatic Vehicle Location (AVL) data which is collected by the local public transport operator is used for the research. The transportation mode and the study area conducted for the research is the bus service system in Klang Valley, Malaysia. Buses equipped with GPS are used to collect for 10 bus routes longitudinal travel time data. A consecutive of 7 months (June 2014 to December 2014) second-by-second GPS data are included. Both weekday and weekend travel time observations are included. The ten sets of bus routes are B115, T626, T629, T634, U32, U40, U48, U62, U76 and U89.

It is important to categorize the bus routes accordingly to have a clearer view on the performance of the distributions for each bus route category. The bus routes examined were categorized according to their route lengths, bus route characteristics or the type of services provided and the total number of stations in the bus route. All important information such as the travel time recorded from one station to another, the stop ID and the direction of each individual station is provided by the AVL system. The travel time used to travel from the preceding stop to the current stop is represented by the stop ID recorded in the AVL system. The AVL system also recorded the travel time taken which represents the departure time difference in any 2 consecutive stops. The detailed information of each bus route is summarized in Table 2 while Fig. 1 shows the plotted route used in the study.

Table 2. Data descriptions of the sample data.

Category*	Distance/ Type	Route Name	Route Number	Operating hours	Number of stations with complete information	Route length (km)**
First Category	Short /CBD	Pasar Seni ⇌ Kompleks Mahkamah Jalan Duta	B115	7:00 am - 7:00 pm (Weekdays only)	19	19.75
	Short /Suburban	Stesen LRT Taman Bahagia ⇌ Damansara Utama	T626	6:00 am - 12:00 am	15	9.91
	Short /Suburban	Stesen LRT Asia Jaya ⇌ Hotel Eastin	T629	6:00 am - 12:00 am	17	12.80
	Short /Suburban	Stesen LRT Bangsar ⇌ Pusat Damansara	T634	6:00 am - 12:00 am	19	10.99
Second Category	Medium /Suburban ⇌CBD	Taman Dagang ⇌ Menara Maybank ~ Bukit Bintang via Jalan Pasar	U32	6:00 am - 12:00 am	49	20.57
	Medium /Suburban ⇌CBD	Desa Petaling ⇌ Hab Lebuh Pudu ~ Bukit Bintang via Jalan Pasar	U48	6:30 am - 11:30 pm	56	27.40
	Medium /Suburban	Stesen LRT Kelana Jaya ⇌ Kota Damansara	U89	6:30 am - 11:00 pm	39	37.68
Third Category	Long /Inter-city	Terminal Kajang ⇌ Hab Lebuh Pudu	U40	6:00 am - 12:00 am	89	55.69
	Long /Inter-city	Klang ⇌ Sunway Pyramid via Hentian Bandar	U62	6:00 am - 12:00 am	31	49.47
	Long /Inter-city	Stesen LRT Subang Jaya ⇌ Pasar Seni	U76	6:30 am - 11:30 pm	72	50.50

Note:

*First category: Short, <20km, <25 stations, shuttle bus services in CBD or suburban area;

*Second category: Moderate, 20-40 km, <65 stations, CBD and suburban area;

*Third category: Long, >40km, >65 stations except U62, Inter-town or Inter-city services.

**Route length is the total distance of the bus travelled on both directions of the route.

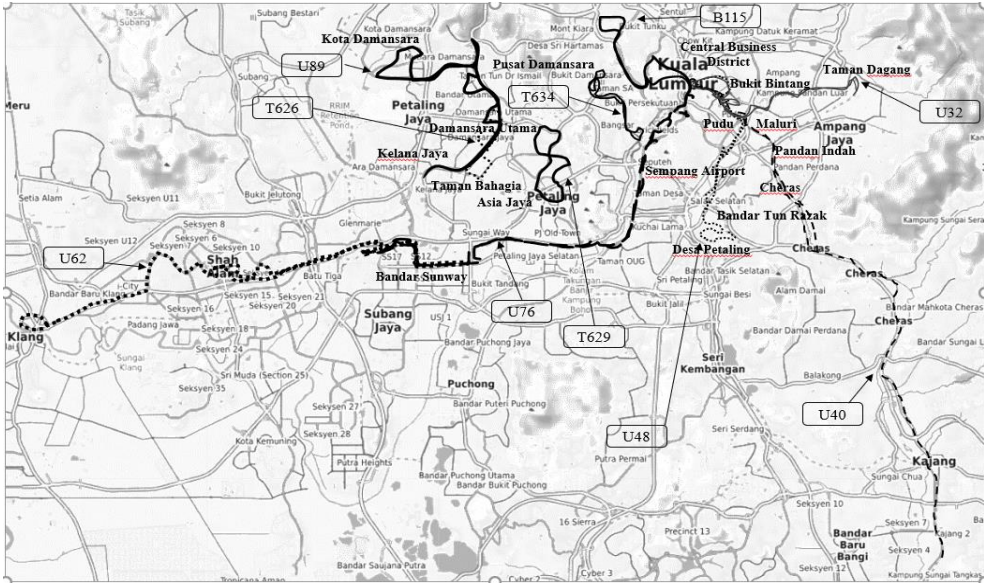


Fig. 1. Configuration of the studied route.

3.2 Burr distribution

Burr [19] first introduced the Burr distribution as a two-parameter family which consists of two shape parameters. Tadikamalla [20] further developed the Burr distribution into a three-parameter family of distribution with an additional scale parameter. The Burr distribution is developed for the flexibility purpose so it has the advantage in fitting various types of frequency data. It is flexible enough to capture a variety form of distribution shapes. Another attractive feature of the distribution is its mathematical and computational characteristics. The cumulative distribution function (cdf) of the distribution can be written in a closed form. Therefore, the computation of its percentile values is simplified. The probability density function (pdf) and the cdf of the distribution are shown in Equation (1) and Equation (2) respectively.

$$f(x,c,k,\alpha) = \frac{ck\alpha^k x^{c-1}}{(\alpha+x^c)^{k+1}} \quad (1)$$

$$F(x,c,k,\alpha) = 1 - \left(\frac{\alpha}{\alpha+x^c}\right)^k \quad (2)$$

where

- x = Travel Time;
- c = Shape parameter;
- k = Shape parameter;
- α = Scale parameter (Estimated median travel time) [21]

There are some interesting statistical properties shown by the three-parameter version Burr distribution. The modal value x_m is given by

$$x_m = \left[\frac{c-1}{ck+1} \right]^{\frac{1}{c}} \quad (3)$$

A L-shaped distribution is formed if $c \leq 1$ while a unimodal distribution is formed if $c > 1$. The r th moment of the Burr distribution is also an interesting statistical property. The r th moment will only exist if $ck > r$. The r th moment of the distribution, $E(X^r)$ is given by

$$E(X^r) = \mu_r = \frac{\alpha^r k \Gamma\left(k - \frac{r}{c}\right) \Gamma\left(\frac{r}{c} + 1\right)}{\Gamma(k+1)} \quad (4)$$

where $\Gamma(\cdot)$ is the mathematical Gamma function. The first moment (also known as the mean or expected value) and the variance of the Burr distribution are given in Equation (5) and Equation (6) respectively.

$$E(X) = \mu = \frac{\alpha k \Gamma\left(k - \frac{1}{c}\right) \Gamma\left(\frac{1}{c} + 1\right)}{\Gamma(k+1)} \quad (5)$$

$$\Sigma^2 = \frac{\alpha^2 k \Gamma\left(k - \frac{2}{c}\right) \Gamma\left(\frac{2}{c} + 1\right)}{\Gamma(k+1)} - \mu^2 \quad (6)$$

The parameters of the Burr distribution are estimated by the maximum likelihood estimation method. Negative log-likelihood function or simply known as log-likelihood value, Akaike information criterion (AIC) and Bayesian information criterion (BIC) were applied in the study to evaluate the performance of the fitness of the distributions on the bus travel time values. The “best” model is the distribution with the maximized log-likelihood of generating the observations with the parameters given by minimizing the negative log-likelihood value or the lowest log-likelihood value, AIC and BIC. Minimizing the negative log-likelihood is the same as minimizing the “loss function” or the distance of two distributions, i.e. the generated result distribution and the target distribution. The formula of AIC and BIC are given in Equation (7) and Equation (8) respectively.

$$AIC = 2k - 2 \ln(L) \quad (7)$$

$$BIC = k \ln n - 2 \ln(L) \quad (8)$$

where

n = number of observations;

k = number of parameter(s) of the respective distribution function;

L = Likelihood function of the estimated distribution

Two ratios which are the Cases_fit ratio and the Cases_top2 ratio are introduced in order to improve the measurement accuracy in model selection. The formulas of the Cases_fit ratio and the Cases_top2 ratio are shown in Equation (9) and Equation (10) respectively.

$$\text{Cases_fit ratio} = \frac{\text{No. of stations that can be fitted with Burr distribution}}{\text{Total number of stations/sections for the respective bus route}} \quad (9)$$

$$\text{Cases_top2 ratio} = \frac{\text{No. of times the model performed as the best or second best}}{\text{Total number of stations/sections for the respective bus route}} \quad (10)$$

Equation (9) shows that the Cases_fit ratio is computed based on the number of Burr distributions that succeed in representing the section travel time values over the number of sections/stations being analysed in the respective bus route. It is observed that the Burr distribution might have convergence issue in estimating the parameters to fit for the travel

time observations as suggested in existing studies. The Cases_fit ratio is thus introduced to compute the number of Burr distributions that succeed in fitting for the travel time values in each route. Besides, Equation (10) shows that the Cases_top2 ratio is computed taking the frequency of the respective distribution emerged as the “best” or second “best” model over the number of sections/stations being analysed in the respective bus route with. This ratio is introduced to improve the accuracy of the measurement when evaluating the performance of each distribution. The goodness-of-fit test values, i.e. the log-likelihood value, AIC and BIC might be difficult to be used for model selection when the difference of these values is small. Therefore, the idea of selecting the “best” two models could prevent bias in evaluating the performance of the distributions while avoiding the possibility of selecting the wrong model.

4 Results and discussions

The results obtained and a detailed interpretation of the results are discussed in this section. The first subsection discusses all the candidature models’ performance and distribution selected as the most outperformed distribution. The statistical tests conducted in the study are presented in this subsection. Parameters estimated for the Burr distribution are presented in the last subsection.

4.1 Distribution fitting

Visualizing the travel time data is the first step of the analysis conducted in this study. The histogram and the fitted Burr distribution on the Station 1001276 in bus route T626 travel time data are shown in Fig. 2. It is observed that both the weekday and weekend travel time data presented are right-skewed and they have very long upper tails. The theoretical curves used to represent the Burr distribution in Fig. 2 show that the Burr distribution might be an appropriate model to fit the travel time values. The flexibility of the Burr distribution and the ability of the Burr distribution in capturing the long tails can be observed in the figure. Further examination in selecting the model to fit for the individual link travel time is conducted by computing the Cases_fit ratio and the Cases_top2 ratio. Table 3 summarizes the results obtained while Table 4 summarizes the results of the best two distributions fitting performance.

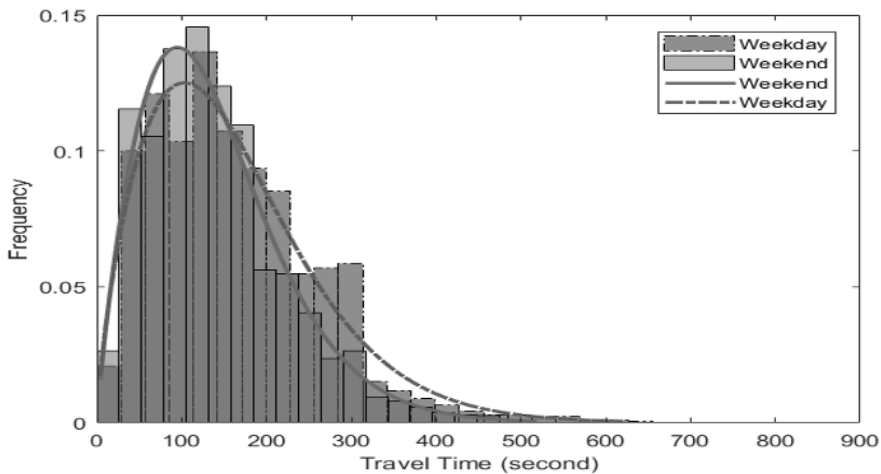


Fig. 2. Histogram and Burr distribution modelling on Station 1001276 travel time data (T626)

Table 3. Comparison of candidature models fitting performance for all categories of bus route.

Category	Bus route number	Cases_fit ratio	Goodness-of-fit test for the respective model (Cases_top2 ratio)					
			Burr	Weibull	Lognormal	Gamma	Normal	GP*
			Mean**	Mean	Mean	Mean	Mean	Mean
First category	B115	0.84	0.72	0.33	0.53	0.42	0.00	0.00
	T626 (D)	1.00	1.00	0.27	0.53	0.20	0.00	0.00
	T626 (E)	0.93	0.93	0.13	0.60	0.33	0.00	0.00
	T629 (D)	0.88	0.84	0.24	0.69	0.24	0.00	0.00
	T629 (E)	0.88	0.82	0.24	0.71	0.24	0.00	0.00
	T634 (D)	1.00	0.95	0.26	0.47	0.32	0.00	0.00
	T634 (E)	0.84	0.79	0.32	0.42	0.47	0.00	0.00
Second category	U32 (D)	1.00	0.90	0.04	0.94	0.12	0.00	0.00
	U32 (E)	0.96	0.82	0.04	0.94	0.20	0.00	0.00
	U48 (D)	0.98	0.90	0.08	0.80	0.21	0.00	0.00
	U48 (E)	0.91	0.80	0.11	0.84	0.23	0.02	0.00
	U89 (D)	0.92	0.85	0.23	0.56	0.36	0.00	0.00
	U89 (E)	0.77	0.69	0.28	0.64	0.38	0.00	0.00
Third category	U40 (D)	1.00	0.99	0.04	0.84	0.13	0.00	0.00
	U40 (E)	1.00	0.97	0.06	0.86	0.12	0.00	0.00
	U62 (D)	0.97	0.96	0.19	0.56	0.29	0.00	0.00
	U62 (E)	0.94	0.86	0.18	0.57	0.39	0.00	0.00
	U76 (D)	1.00	0.94	0.06	0.88	0.12	0.00	0.00
	U76 (E)	0.93	0.85	0.07	0.87	0.22	0.00	0.00

*GP = Generalized Pareto

**Mean is calculated as the total Cases_top2 ratio of the three goodness-of-fit test values over 3.

The bold values indicate the most outperformed model for each bus route

(D) indicates weekday travel time data while (E) indicates weekend travel time data

It is observed that the Burr distribution is the most outperformed model in fitting for the individual links travel time data in the first and third category of the bus routes. The results obtained show that for both weekday and weekend of travel time, the Burr distribution emerged as the most outperformed or the second most outperformed distribution for 80% and 90% of the cases in the first and third category of the bus route respectively. For the second category of bus route, the results are quite contradicting. The Burr distribution and the lognormal distribution both emerged as the most outperformed distribution for half of the conditions evaluated in the second category of bus route as shown in Table 3 and Table 4. However, the cases of bus routes which lognormal distribution emerged as the best distribution are interesting. Although the lognormal distribution emerged as the best distribution, the number of stations showing the Burr distribution as the best distribution or the lowest log-likelihood value, AIC and BIC were higher compared to the lognormal distribution. For example, in bus route U32 (D), the number of stations with Burr distribution performed as the best fitted model in terms of log-likelihood value, AIC and BIC are 29, 29 and 28 respectively compared to the values obtained from the lognormal distribution which are 17, 17 and 16 respectively. This shows that the higher ratio value shown for the lognormal distribution in Table 3 is due to the reason that the lognormal distribution emerged as the second best fitted model for most of the stations in the bus route. This situation not only can be seen in bus route U32 (D), it also can be seen in bus route U32 (E), U48 (E) for the second category of bus route and bus route U76 (E) for the third category of bus route. Therefore, in overall performance, the Burr distribution is still considered as the best fitted model for both weekday and weekend travel time observations.

Table 4. The Burr and lognormal distributions fitting performance for all categories of bus route.

Model Goodness-of-fit test	Cases_top2 ratio					
	Log-likelihood		AIC		BIC	
	Cases	Burr	Lognormal	Burr	Lognormal	Burr
B115	0.74	0.53	0.74	0.53	0.68	0.53
T626 (D)	1.00	0.53	1.00	0.53	1.00	0.53
T626 (E)	0.93	0.60	0.93	0.60	0.93	0.60
T629 (D)	0.88	0.65	0.82	0.71	0.82	0.71
T629 (E)	0.82	0.71	0.82	0.71	0.82	0.71
T634 (D)	0.95	0.47	0.95	0.47	0.95	0.47
T634 (E)	0.79	0.42	0.79	0.42	0.79	0.42
U32 (D)	0.90	0.94	0.90	0.94	0.90	0.94
U32 (E)	0.82	0.94	0.82	0.94	0.82	0.94
U48 (D)	0.91	0.80	0.91	0.80	0.89	0.80
U48 (E)	0.82	0.84	0.82	0.84	0.77	0.84
U89 (D)	0.87	0.56	0.85	0.56	0.82	0.56
U89 (E)	0.69	0.64	0.69	0.64	0.69	0.64
U40 (D)	0.99	0.84	0.99	0.84	0.98	0.84
U40 (E)	0.98	0.85	0.97	0.85	0.96	0.87
U62 (D)	0.97	0.55	0.97	0.55	0.94	0.58
U62 (E)	0.90	0.55	0.84	0.58	0.84	0.58
U76 (D)	0.96	0.88	0.94	0.89	0.93	0.89
U76 (E)	0.86	0.86	0.85	0.86	0.83	0.88

(D) indicates weekday travel time data while (E) indicates weekend travel time data

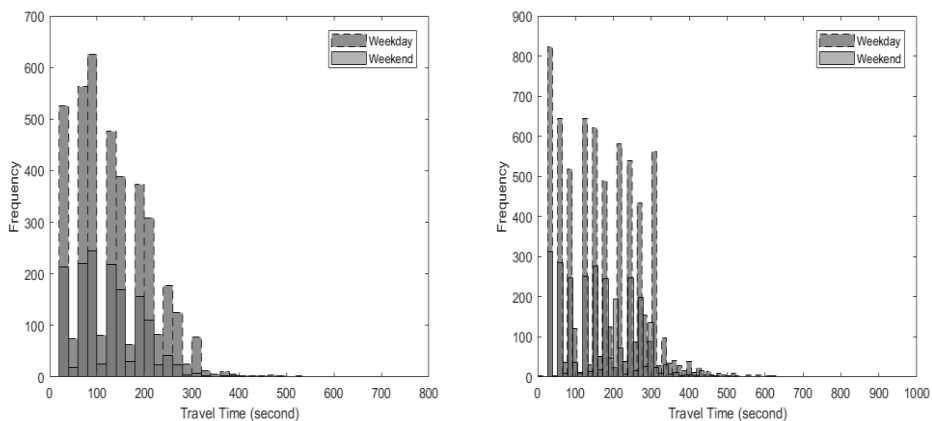


Fig. 3. Histogram for Station 1004155 (left) bus route U89 and Station 1002351 (right) bus route U62 travel time data.

Further analysis is also carried out to compare the performance of the model in fitting between weekday and weekend travel time data. It is obvious that the Burr distribution is more suitable in representing the weekday travel time data. Higher ratio values were obtained for the weekday travel time data of the same bus route. The mean value of the Cases_top2 ratios show that the Burr distribution is best representing the weekday travel time data while the performance of the Burr distribution is comparably lower in representing the weekend travel time.

Statistical tests were also carried out to examine the ability of the Burr distribution in fitting a total of 793 individual links in the study. However, it is found that only 39 cases which accounted for 4.92% or less than 5% of the total cases were unable to be fitted by this distribution. The travel time histogram for Station 1004155 in bus route U89 and Station 1002351 in bus route U62 are shown in Fig. 3. The travel time data in Fig. 3 were unable to be fitted by the Burr distribution except the weekday travel time data of Station 1004155. It is found that the difference between these travel time data is the indications of bimodality or multimodality. More than one peak can be observed in the travel time values that were unable to be fitted with Burr distribution. Besides, it is also noted that weekend travel time observations have a high frequency of failure cases. Therefore, further studies could be conducted on looking into the issue of the poor performance of the Burr distribution in representing the weekend travel time data.

On the basis of the ten datasets, normal, Weibull, gamma and generalized Pareto distributions did not perform well. Although lognormal distribution might be a good statistical fit, the Burr distribution showed a higher performance or promising result in fitting individual links travel time data that are operating in different environments. The log-likelihood value, AIC and BIC computed from the Burr distribution were the lowest compared to other distributions. It is also noted that the convenience brings about by the Burr distribution for statistical descriptors computation such as percentile values. It enhanced the appropriateness of the model in TTV modelling as the traffic management is operating in a dynamic environment. Burr distribution can thus be seen as a plausible model considering its high performance and its statistical characteristics.

4.2 Burr distribution parameters

Parameters estimated for the Burr distribution are discussed in this subsection. The parameters estimated could give an insight on the basic statistical descriptions of the data. A number of selected individual links fitted by the Burr distribution with their respective parameters estimated are shown in Table 5. A higher than average scale parameter values were observed for both weekday travel time or weekend travel time on the selected individual links in the same bus route. The scale parameter of the Burr distribution suggests the most concentrated part of the data or the estimated median travel time based on the fitted distribution as a linear relationship was found in between the scale parameter value of the Burr distribution with the true median travel time value [21]. Therefore, higher travel time values or higher estimated median values were obtained at these stations compared to other stations. This can be explained by the travel time spent by the bus drivers at these stations. The stations shown in the table are either the initial station, transfer station or the last station of a bus route. Therefore, it is reasonable that a higher than average median travel time is observed as the stations are the main bus station in a bus route.

Table 5. The Burr distribution parameters for selected individual links.

Route Number	Bus route category	Stop ID	Type of Station	Burr distribution		
				c	α (median)	k
T626 (D)	1	1001830	Initial	2.667	579.453	3.116
T626 (E)	1	1001830	Initial	2.687	760.543	4.877
T629 (D)	1	1001784	Initial	2.110	1020.698	8.812
T629 (E)	1	1001784	Initial	2.433	765.693	4.465
T629 (D)	1	1000603	Transfer	1.825	4381.617	295.001
T629 (E)	1	1000603	Transfer	1.772	3653.918	228.606
T634 (D)	1	1001786	Initial	1.930	2088.384	16.384
T634 (E)	1	1001786	Initial	2.240	1479.799	12.112
T634 (D)	1	1002917	Initial (nearby)	1.748	2791.055	39.735
T634 (E)	1	1002917	Initial (nearby)	1.744	1326.502	13.842
U32 (D)	2	1003907	Initial (nearby)	1.595	4846.799	31.084
U32 (E)	2	1003907	Initial (nearby)	1.640	3150.797	19.738
U48 (D)	2	1003928	Transfer(nearby)	1.570	456.472	5.977
U48 (E)	2	1003928	Transfer(nearby)	1.472	1511.020	30.142
U48 (D)	2	1002977	Last	2.233	84.643	0.891
U48 (E)	2	1002977	Last	1.710	175.829	1.808
U89 (D)	2	1001804	Initial	1.696	1839.991	19.870
U89 (E)	2	1001804	Initial	1.610	2443.514	30.982
U89 (D)	2	1001508	Initial (nearby)	2.014	1037.816	14.106
U89 (E)	2	1001508	Initial (nearby)	2.048	877.229	11.271
U89 (D)	2	1000371	Last	1.823	242.753	4.052
U89 (E)	2	1000371	Last	1.780	266.296	5.619
U40 (D)	3	1002571	Initial (nearby)	1.500	3480.509	15.661
U40 (E)	3	1002571	Initial (nearby)	1.483	5184.737	22.615
U40 (D)	3	1000948	Transfer (nearby)	1.474	1271.472	9.606
U40 (E)	3	1000948	Transfer (nearby)	1.427	11255.226	214.149
U40 (D)	3	1002080	Transfer	1.510	19614.096	166.391
U40 (E)	3	1002080	Transfer	1.485	13367.732	89.393
U62 (D)	3	1001515	Initial	2.340	609.844	4.235
U62 (E)	3	1001515	Initial	2.326	666.554	4.968
U62 (D)	3	1000941	Initial (nearby)	1.781	1380.795	15.426
U62 (E)	3	1000941	Initial (nearby)	1.755	1446.575	14.693
U62 (D)	3	1003078	Transfer	1.723	408.279	5.216
U62 (E)	3	1003078	Transfer	1.853	314.037	4.582
U62 (D)	3	1001516	Last	1.843	526.492	8.901
U62 (E)	3	1001516	Last	1.882	512.332	9.024
U76 (D)	3	1002351	Last (nearby)	1.872	212.430	2.084
U76 (E)	3	1002351	Last (nearby)	1.695	275.741	2.551

(D) indicates weekday travel time data while (E) indicates weekend travel time data

5 Conclusion

The focus of this study is in finding the most appropriate model to explain each section of weekday and weekend travel time values separately for day-to-day TTV analysis. A total of ten bus routes for a consecutive of 7 months data are used for the research. The bus routes were categorized in three categories according to their characteristics. The Burr distribution was proposed as the leading model while several candidature models, i.e. normal, lognormal, gamma, Weibull and generalized Pareto were also used for comparison purposes.

It is found that the Burr distribution is able to accommodate the shape or pattern, i.e. long upper tails and the right-skewed of the travel time distribution. Results obtained in the study led to the conclusion that the Burr distribution is the best representation of the travel time values in overall performance. The Burr distribution is best in representing weekday travel time values. However, its performance is comparably lower in fitting weekend travel time values. Although there are some failure cases which indicate bimodality or multimodality, it only accounts for less than 5% of the total number of cases examined in the study. It is also important to take into account the statistical properties, i.e. flexibility and mathematical tractability of the Burr distribution which enhanced its suitability for reliability analysis. Therefore, the Burr distribution is considered as the most appropriate model for day-to-day TTV analysis in representing both weekday and weekend travel time values.

This paper could be treated as the preliminary preparation for further reliability analysis. Further research should be conducted on investigating the usage of the distribution in computing the variation in travel time for each individual link of a bus route. It is also important to embed the traffic factors implicitly into the model for TTV statistical modelling study.

This research study is supported by the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme (FRGS), Project No.: FRGS/1/2018/TK08/UTAR/02/1. We would like to thank Rapid KL Sdn Bhd for their kindness in providing the data for us.

References

1. J. Bates, J.W. Polak, P. Jones, A. Cook, *Transp. Res. Pt. e-Logist. Transp. Rev.* **37**, 191 – 229 (2001)
2. C. Sun, G. Arr, R.P. Ramachandran, *Transp. Res. Record* **1826**, 25 – 30 (2003)
3. A. Polus, *Transport* **8**, 141 – 151 (1979)
4. W.C. Jordan, M.A. Turnquist, *Transp. Sci.* **13**, 242 – 268 (1979)
5. M.A.P. Taylor, *Transp. Sci.* **16**, 507 – 521 (1982)
6. W. Zhang, A. Medina, H. Rakha, *Statistical analysis of spatiotemporal link and path flow variability*, IEEE Intelligent Transportation Systems Conference, 30 September-3 October 2007, Washington, United States (2007)
7. N. Uno, F. Kurauchi, H. Tamura, Y. Iida, *J. Intell. Transport. Syst.* **13**, 2 – 15 (2009)
8. M. Kieu, A. Bhaskar, E. Chung, *J. Transp. Eng.* **141**, 1 – 9 (2014)
9. S. Susilawati, M.A.P. Taylor, S.V.C. Somenahalli, *J. Adv. Transp.* **47**, 720 – 736 (2013)
10. M.A.P. Taylor, *Transp. Res. Pt. B-Methodol.* **97**, 50 – 63 (2017)
11. A.J. Richardson, M.A.P. Taylor, *High Speed Ground Transp. J.* **12**, 77 – 99 (1978)

12. J. Eliasson, *The relationship between travel time variability and road congestion*, in Proceedings of the 11th World Conference on Transport Research, 24-28 June 2007, California, United States (2007)
13. G. Schwarz, *Ann. Stat.* **6**, 461 – 464 (1978)
14. H. Akaike, *IEEE Trans. Autom. Control* **19**, 716 – 723 (1974)
15. E. Mazloumi, G. Currie, G. Rose, *J. Transp. Eng.* **136**, 623 – 631 (2010)
16. W.J. Zimmer, J.B. Keats, F.K. Wang, *J. Qual. Technol.* **30**, 386 – 394 (1998)
17. E.B. Emam, H. Al-Deek, *Transp. Res. Record* **1959**, 140 – 150 (2006)
18. M.M. Rahman, S.C. Wirasinghe, L. Kattan, *Transp. Res. Pt. C-Emerg. Technol.* **86**, 453 – 466 (2018)
19. I.W. Burr, *Ann. of Math. Stat.* **13**, 215 – 232 (1942)
20. P.R. Tadikamalla, *Int. Stat. Rev.* **48**, 337 – 344 (1980)
21. S. Susilawati, M.A.P. Taylor, S.V.C. Somenahalli, *Modelling urban travel time variability with the Burr regression technique*, in Proceedings of the Australasian Transport Research Forum, ATRF, 28-30 September 2011, Adelaide, Australia (2014)