


# Factors Affecting Driver Speed Choice Behaviour in a Two-Lane Two-Way Heterogeneous Traffic Environment: A Micro-level Analysis

S. M. Sohel Mahmud<sup>1</sup>  · Luis Ferreira<sup>1</sup> · Md. Shamsul Hoque<sup>2</sup> · Ahmad Tavassoli<sup>1</sup>

Received: 16 December 2018 / Accepted: 20 June 2019 / Published online: 1 July 2019  
© The Institution of Engineers (India) 2019

**Abstract** This paper deals with the speed choice behaviour of driver on a two-lane bidirectional highway in a heterogeneous traffic environment of a developing country. The major contribution of the paper is the identification and quantification of significant attributes influencing speed choice of a driver at micro-level of the road under study. The latter explored the use of ordinary least square (OLS) and random parameter (RPM) regression models as an alternative methodological approach to relate factors such as road geometry, roadside environment, traffic mix and operational characteristics with driver speed choice at the small segmental level. The comparison of the empirical results shows the RPM model outperforms its fixed parameter-based OLS counterpart. The approach justifies the need to take into account the potential heterogeneity in the impact of factors at the micro-level speed choice behaviour analysis. The analysis used data collected from a section of a major national highway in Bangladesh. Speed data extracted second-by-second was analysed for a range

of short road segments. The critical segment length, from the point of view of speed analysis, was identified using different statistical tests. The paper details the data collection method used, as well as the speed-related statistics analysis performed. The results obtained could be used to better understand the speed choice factors of drivers. The findings could also be used to inform policy decisions for managing the appropriate homogeneity of speed among the motorized vehicles of two-lane highways in developing countries, a prerequisite for ensuring safe movement of road users.

**Keywords** Speed behaviour · Speed choice · Affecting factors · Regression model · Random parameter · Heterogeneous traffic

## Introduction

The vehicle speed on a particular road section is known to be an important contributing factor for accidents and their resulting severity. Speed plays a significant role in planning and designing the road geometry, as well as in the setting of safe speed limits. Moreover, the choice of speed by drivers for a prevailing traffic condition is a key factor in models used to monitor traffic operations and to evaluate the performance of traffic system [1].

A driver's choice of speed is affected by many factors and there is a wealth of the literature on this topic. From the perspective of statistical analysis, most of the conventional practices to identify the factors affecting speed choice, as well as to formulate models to predict the speed of a road section, are mainly based on OLS linear regression model (e.g. [1–5]). The OLS regression model has a significant constraint in the degree of uncertainty as it considers the

✉ S. M. Sohel Mahmud  
smsohelmahmud@gmail.com; s.mahmud@uq.net.au

Luis Ferreira  
l.ferreira@uq.edu.au

Md. Shamsul Hoque  
shamhoque84@gmail.com

Ahmad Tavassoli  
a.tavassoli@uq.edu.au

<sup>1</sup> School of Civil Engineering, The University of Queensland (UQ), Advanced Engineering Building (49), Brisbane, QLD 4072, Australia

<sup>2</sup> Department of Civil Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka 1000, Bangladesh

fixed effect of the variables across the entire roadway segments [6, 7]. It is possible that the effect of exogenous factors will be heterogeneous and random in nature raised from the observed and unobserved variables not considered in the data collection process. Therefore, the use of fixed parameter regression approach could lead to the inconsistent and biased identification of the factors. As such, there is a scope for improvement in the performance of the least square model by adding random effects associated with many unaccounted factors. Tarris et al. [8] attempted to incorporate random effects to examine the road geometry and driver effects on operating speeds. Poe and Mason Jr [9] used a mixed-model approach to account for random effect for an individual observation site of horizontal curvature of two-lane highways. Both these two studies used data collected from low-speed urban streets in the USA.

However, most of the literature mainly deals with free-flow speeds at the macro-level, i.e. for a network comprising different road sections, or particular sections of several roads, such as curves or intersections. In addition, most existing models are based on homogenous traffic settings. However, the road-traffic environment, as well as the driving behaviour in developing countries, is very different from those pertaining in developed countries [10, 11]. As a result, there is still much to learn about the influencing factors of speed choice at the micro-level, especially of undivided two-lane rural highways in heterogeneous traffic environments. The main objective of this paper is to identify the factors affecting the speed behaviour of a driver at different speeds level and to develop a model for estimating the impact of different factors on speed choice of a two-lane bidirectional highway in developing countries. The study formulates and estimates a micro-level speed prediction model using the random parameter approach to incorporate individual heterogeneous effect of the exogenous variables. The model is based on the micro-level analysis with the variation of speed being analysed on the basis of short road segments.

To quantify the potential random effects of exogenous factors, the current study compared the performance of OLS and random parameter models, using several aggregated and disaggregated goodness-of-fit measures. With the identification of significant influencing factors, the study also discussed the partial effects to quantify the effect of those individual variables on driver speed choice. The model was validated using data collected from different trips. The results confirm the need to accommodate the random effect of exogenous variables in order to examine the speed choice behaviour under a heterogeneous traffic environment.

The study collected data from a 13 km section of a two-lane bidirectional major national highway (N4) in

Bangladesh. Field observation, including an instrumented floating vehicle, was used to collect speed and related data. This paper reports on the analysis of the data mainly obtained using a vehicle instrumented with three inside cameras and observers. Speed data were extracted from the speedometer reading of the instrumented vehicle. The back camera provided the phasing and opposing traffic data. Detailed geometric and road environmental attributes were gathered using on-site field and video observation. Speed data extracted second-by-second were analysed for a range of short road segments. The critical segment length, from the point of view of speed and overtaking analysis, was identified as 200 m. Speed profiles for each 200 m segment were analysed, and descriptive speed-related statistics were obtained. In addition, the variation of directional speed and its significance was also quantified. Abrupt speed changes were identified to determine the probability of conflict using surrogate safety measures. As speed choice is affected by many factors, the most significant attributes influencing speeding behaviour have been identified using different sensitivity analyses. Those attributes include traffic flow by vehicle composition, road and roadside environment, as well as traffic operational characteristics.

The paper is organized as follows: The following section briefly outlines the overall data collection methodology (Sect. 2). Next section presents some of the key findings of the analysis speed behaviour (Sect. 3). The significant factors affecting speed behaviour are identified in the following section (Sect. 4). Before concluding section, the paper describes the random parameter model estimated and shows the model estimation and validation results (Sect. 5). Finally, the paper summarizes the main findings and provides a discussion of the potential model applications, imitations and areas for future research (Sect. 6).

## Data Collection Methodology

### Study Area

The study area is the Jamuna Multipurpose Bridge approach roads, a major highway section in Bangladesh. The selected road segments are two-lane bidirectional rural highway with a heterogeneous traffic environment, typical of a developing country. The length of the bridge is about 4.8 km in total. The approach roads are 16 km from the east side and 17 km from the west. For the analysis of speed behaviour and for model development, the current study used data collected from the east approach road. Details of the road geometry, environmental and traffic operational characteristics are given in Mahmud et al. [12] and Islam [13].

## Collection Methods

The primary data for the study reported here were collected using mainly the naturalistic driving method using an instrumented vehicle and field observation simultaneously.

### Naturalistic Driving

An instrumented vehicle (microbus) was used to gather data on driving behaviour, as well as to continuously monitor vehicle speed. Two in-vehicle cameras were attached to the front dashboard and rear side of the vehicle to track the movements of the leader and the follower. Another camera was purposefully set to record the speedometer reading. The third camera enabled the continuous monitoring of vehicle speed on a second-by-second basis. Figure 1 shows glimpses of front and speedometer camera views. Simultaneously, two observers recorded all overtaking events and any other risky driver behaviour. The video footage is also included an audio record of any prompt made by the camera operator regarding essential data items, such as the identity of the overtaking vehicle and the details of any manoeuvres deemed to be risky.

The driver was instructed to drive normally at all times during the entire period of data collection. However, to justify the normal behaviour of the floating vehicle driver, some of the driving attributes were compared with those of the general traffic. These attributes include space mean speed (SMS); time mean speed (TMS); maximum and minimum spot speed; travel time; the number of overtaking

events; and average overtaking distance. It is found that the attributes of the test vehicle are not statistically different from those of the corresponding general traffic. The comparative analysis excluded two- and three-wheeler light vehicles (around 10% of total traffic), as these vehicles are significantly slower and their behaviour is different from other mainstream traffic.

### Field Observations

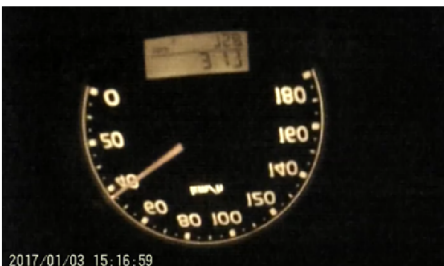
Observational studies were made by group of trained observers continuously, whilst video data were being captured along the entire study section (Fig. 2). Different attributes related to driver behaviour were collected using pre-defined form. Moreover, information related to road geometry, roadway condition, road surface friction, road abutting land use pattern, the degree of access control and spot speeds were also collected by the observers from different segments of the road. All of those data are cross-checked with the video data.

### Selection of Appropriate Segment Length

Figure 3 represents the step-by-step process used to select the most appropriate speed analysis segment length.



(a) Front camera view



(b) Speedometer view

**Fig. 1** Glimpses of front and speedometer camera view



(a) Camera adjustment

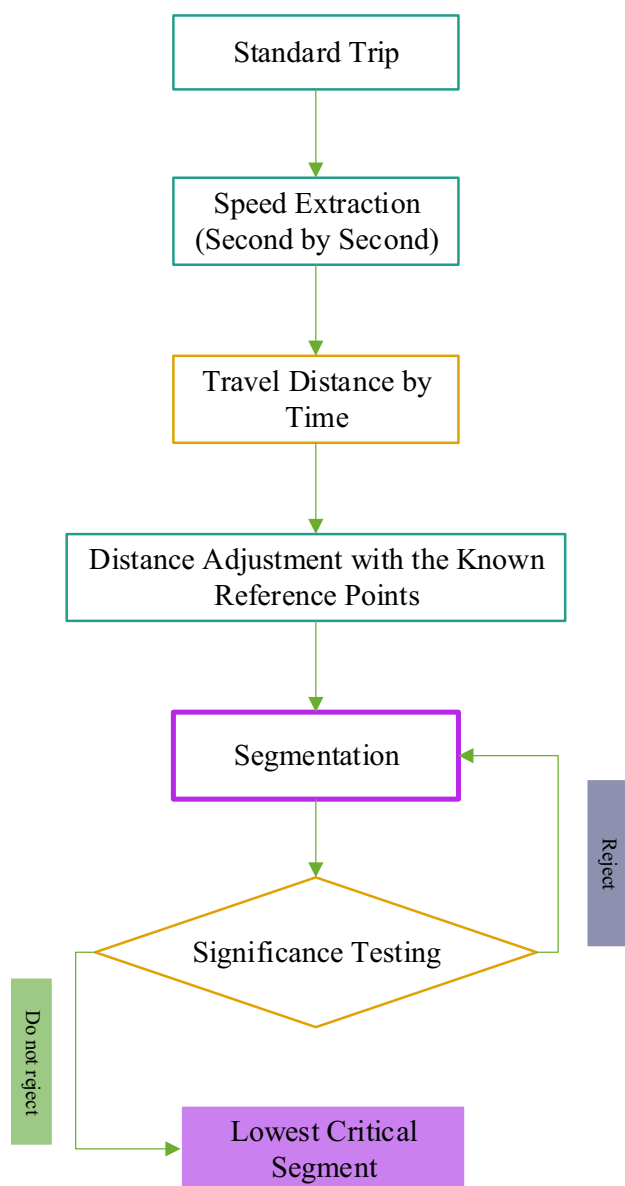


(b) Observation by observers

**Fig. 2** Glimpses of field observation

Firstly, the instrumented vehicle's speeds were extracted from the video footage of speedometer. The speed-related data were verified using a two-step process to minimize the errors. Following a random cross-check of the full dataset, a trip-based time-speed scatter plot was used to identify anomalies and any sudden change in speed. Adjustments were made using alternative video data.

Altogether, 18 reference points along the 13.7 km road section were used to verify and match the speed data and to exact geographic or chainage location. Estimated travel distances were found to have small errors (2–3%) compared with actual on-road distances. Those errors are due mainly to lateral vehicle movements and speedometer reading measurement error.



**Fig. 3** Critical segment identification process

A number of speed-related statistical attributes, such as maximum, minimum, standard deviation, speed difference and average speed for the respective segment length were used in order to determine the most appropriate small segment length for speed analysis purposes which is termed as segmentation. The level of significance in the differences of those attributes for different segment lengths was evaluated using standard sensitivity analysis including ANOVA, *t* test. Segments tested ranged from 100 to 500 m, in steps of 50 m. The results show that a 200-m segment length is the most appropriate, taking into account the statistical analysis and the minimum required overtaking distance.

### Analysis of Speed Behaviour

The study reported here considered 20 typical trips by a floating vehicle (10 trips in each direction), for different times of day. The summation of total travel distance considered for the detailed analysis is around 272 km, equivalent to around 4.3 h of travel. Speed data were extracted second-by-second for each trip. These speed data have been analysed from different perspectives. A brief summary of that analysis is given in the following sub-sections. However, for the factor analysis modelling, the study selected 13 trips, 10 for development and 3 for validation purposes. All of selected trips were in normal traffic conditions under a bright sunny environment.

### Overall Speed Profiles

The posted speed limit for the selected road section is 80 kph. Both the TMS and the SMS were observed to be significantly different from that posted speed. The spot speed or TMS of the floating vehicle reached over 100 kph, and the SMS in different trips ranged from 53 to 73 kph. The mean SMS for the entire study section is 59 kph (60 kph west–east and 58 kph east–west). The summarized overall speed profiles along with different descriptive statistics by direction of travel are presented in Table 1.

**Table 1** Overall speed: descriptive statistics

Items	West–east direction	East–west direction
Mean (kph)	60.2	57.7
Standard error	0.2	0.2
Median	60	57
Mode	45	50
Standard deviation	18.0	16.6
Range	103.5	102
Minimum (kph)	5	0
Maximum (kph)	108.5	102

**Table 2** Distribution of time and distance travelled

Direction Speed range (kph)	West–east direction				East–west direction			
	Time spent		Distance travel		Time spent		Distance travel	
	(s)	%	(m)	%	(m)	%	(m)	%
≤ 50	2381	31	25,971	21	2590	32	28,224	22
50–60	1238	16	18,591	15	1736	21	26,017	20
60–70	1350	18	24,069	19	1653	20	29,332	23
70–80	1372	18	28,203	22	1253	15	25,679	20
80–90	852	11	19,803	16	597	7	13,746	11
90–100	310	4	8013	6	238	3	6168	5
≥ 100	69	1	1993	2	18	0.2	503	0.4

The difference in SMS between the two directions is found statistically significant ( $p$  value is almost zero ( $4.68E-19$ ),  $z > z_{\text{Critical two-tail}} > z_{\text{Critical one-tail}}$ ;  $F > F_{\text{crit}}$ ).

**Overall Distribution of Speeds**

Table 2 shows the total time and distance spent travelling, for each speed range and by direction, for all 20 recorded trips. Although the overall mean speed is well below 80 kph, the survey vehicle spent, on average, 13% of the time and 20% of distance at speeds above 80 kph. As shown in Table 3, around 24 and 17% of total distance travelled at speeds over 80 kph, in the west–east and east–west directions, respectively. Distributions of speeds by direction were found to be normally distributed (Fig. 4).

**Table 3** Deceleration and acceleration frequencies for different ranges of speed change ranges

Ranges of speed change ( $m/s^2$ )	Frequency by direction	
	West–east	East–west
<b>Deceleration</b>		
– 4 to – 3.5	0	2
– 3.5 to – 3	5	4
– 3 to – 2.5	13	8
– 2.5 to – 2	24	21
– 2 to – 1.5	95 (27)*	76 (17)*
– 1.5 to – 1	289 (74)*	237 (48)*
<b>Acceleration</b>		
1–1.5	47	44
1.5–2	6	6
3–3.5	1	1

\*Figure in parenthesis shows the frequencies above 70 kph initial speed

**Acceleration and Deceleration Changes**

The sudden change of speed or acceleration/deceleration behaviour has a significant impact on different traffic operational and road environmental factors [14]. The abrupt change of speed is an indication of disturbance of free flow and indicates a probability of conflict. Table 3 shows the frequency of deceleration and acceleration at different ranges of speed change ( $m/s^2$ ), found in the 20 trips analysed here.

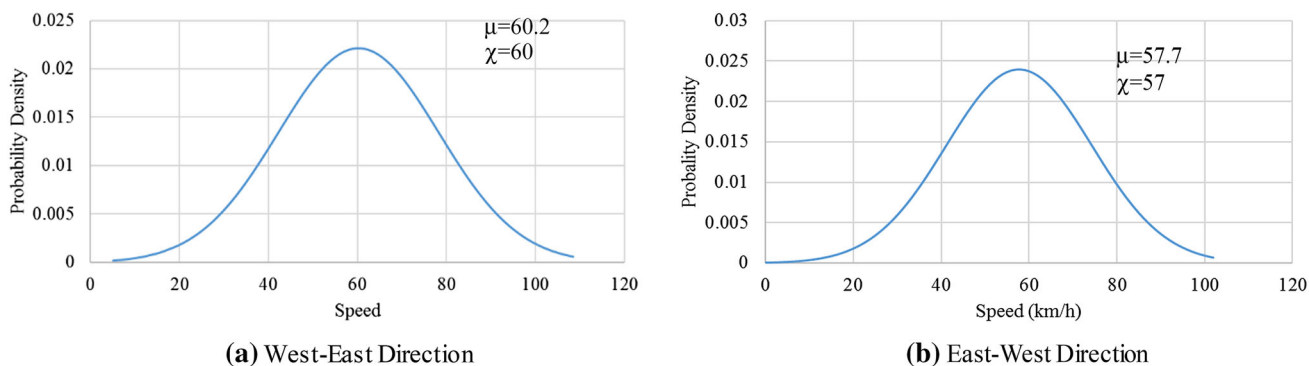
As shown in Table 3, 24% of decelerations between – 2 and – 1  $m/s^2$  occurred from the initial speeds above 70 kph. Observation of this critical deceleration events revealed that certain evasive actions were taken not only to avoid rear-end conflicts but also to avoid severe head-on conflicts with opposing vehicles due to overtaking manoeuvres. Table 3 also shows the number of large acceleration events for different acceleration ranges. Most of those events took place during overtaking manoeuvres, mainly to avoid impending collisions.

**Speed Variation for Small Segment**

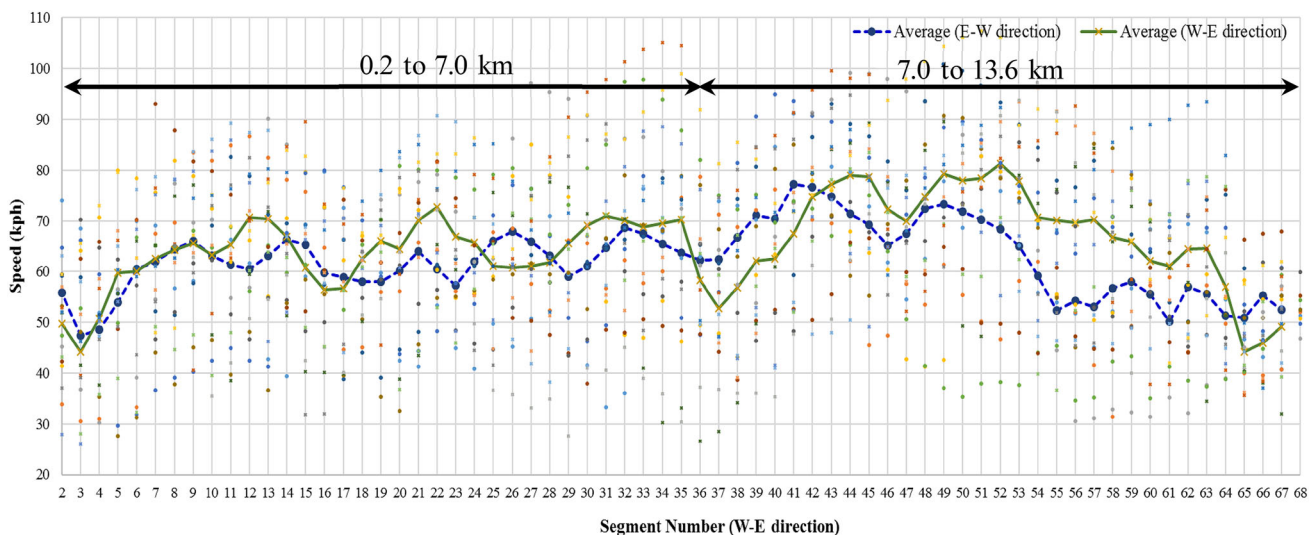
Figure 5 shows the average segmental SMS for each trip. The average segment speed for all trips is also shown. Although there is noticeable variation in speed between different segments, the ranges of speeds are consistently similar.

The SMS overall means are 65 kph and 62 kph in the west–east and east–west direction, respectively. Although the mean speeds show similar profiles in both directions, the differences were found to be statistically significant ( $z > z_{\text{Critical two-tail}} > z_{\text{Critical one-tail}}$ ;  $P(Z \leq z)$  one-tail = 0.017 and  $P(Z \leq z)$  two-tail = 0.035). The differences between the two-directional segmental average SMS are relatively less in the first half of the road segment (segments 2–36) where the average SMS varies between 60 and





**Fig. 4** Normal distribution of speed by direction



**Fig. 5** Segment-wise SMS speed profile

70 kph. In the second half (beyond segment 36), the average SMS varies considerably more. Segments 40–52 (7.8–10.2 km) show the highest SMS in the entire section and it reaches up to 80 kph. The distribution of speed in each segment of the road is shown as box plot in “Appendix B”.

## Factors Affecting Speed Choice

### Selection of Explanatory Variables

The choice of speed is affected by several factors (e.g. [1, 2, 5]). Those factors can be divided into three main categories, namely road geometry; traffic flow; and traffic control. Several studies have dealt with the relationships between speeding behaviour and personal factors, including demography and psychology [15]. Most of such past works relate to free-flow speeds on networks comprising

different types of road. Therefore, psychology, demography of drivers and vehicular characteristics were different for different drivers and roads. Therefore, those attributes were considered and were found significant in speed choice. However, the current study attempted to identify the factors that affect the speed choice of a driver for each small segment of the studied road, where the posted speed limit is the same throughout the section. Moreover, considering the traffic characteristics of the study area, a microbus with a young driver (around 25-year age) has been considered for the data collection to ensure maximum vehicle representation. Hence, those variables are constant for the context of the current study. Indeed, the influencing factors determining speed choice on undivided two-lane rural highways in heterogeneous traffic environment have yet to receive the same level of attention.

A number of external factors have been assessed, such as those related to road geometry, road environment, traffic flow and operations. Special attention has been given to the

different characteristics of road and traffic environment in developing countries. Road geometric factors include alignment, shoulder, presence of bridge, culvert or access road. Road environmental factors include roadside friction in terms of pedestrian activities, roadside trading activities and roadside parking. Factors related to traffic flow and operations include directional flow by different types of vehicles. Segmental traffic flow has been counted during the period of crossing (facing and following and/or overtaking) by the floating vehicle. All of those attributes have been collected for each segment by field observers and instrumented vehicle video. As the study examined the change of speed within small segments, the driver and vehicle-related factors are assumed constant for each segment. A total of 32 independent variables were analysed for four statistical measures of speed levels, namely average; minimum; maximum; and 3rd quartile speed.

For the appropriate variables selection, the study tested variables under different measures (such as indicator, ordinal, scale or continuous), to obtain the best results on a trial and error basis. Separate trials were made in combining and segregating different variables, such as bus and truck. However, segregation of buses and trucks into two different independent variables provided a better result, in terms of goodness of fit. Moreover, geometric and environmental attributes were divided into different categories based on their characteristics. Roadside shoulder was also divided into three different categories according to its geometric conditions, such as good, medium and bad. Finally, each variable was converted to an indicator measure, as it provided better output than the categorical or ordinal measure. A detailed definition of the categories selected is given in “Appendix A”.

### Correlation Test

All of explanatory variables shown in “Appendix A” were considered for analysis of the impact of speed behaviour. Pearson’s correlation test among all of those explanatory variables was initially undertaken (Fig. 6). Those variables which were found to be significantly correlated with others were excluded from the model ( $r > 0.7$ ). Those variables which were found not to be statistically significant were also progressively discarded.

### Identification of Significant Factors

#### Ordinary Least Square Regression (OLS) model

OLS regression was used to identify the significant speed influencing factors. The best model specification for calculating average speed, maximum speed, minimum speed and 75-percentile speed, for a particular road segment, has

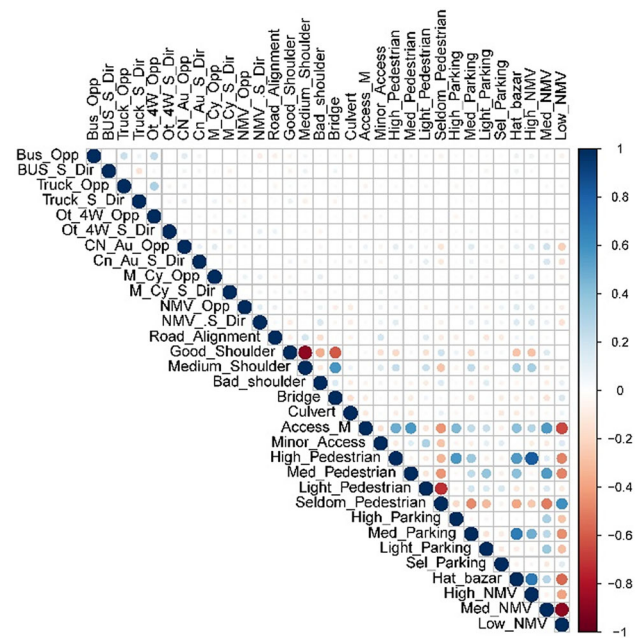


Fig. 6 Correlation among initially considered all of explanatory variables

been developed. Sixteen factors were found to be significant for different speed levels. Among them, nine are related to traffic flow characteristics, such as buses, trucks and other four wheelers from both directions, same directional three wheelers, motorcycles and non-motorized vehicles (NMV) from opposite direction. Five are related to road geometry, including shoulder (good and bad), on road small bridge, culvert and major access; and two are related to roadside environment, such as roadside medium pedestrian activities and low non-motorized vehicle (NMV) along the road side. A list of those factors with coefficient,  $t$  value and  $p$  value are given in Table 4. The constant term is also given to show the full specification of the OLS regression model.

All of the significant factors related to traffic flow influence negatively the choice of speed at all levels (i.e. the higher the flow the lower the speed). Good shoulder, culvert, major access and absence of NMV help to reduce the congestion by increasing the speed choice at the minimum speed level. Culvert and good shoulder also have significant positive impact to maximize the speed level.

### Random Parameter Modelling Approach

Most of the researches deal with the identification of factors affecting speed choice using ordinary least square models. The latter considers the fixed effect of the variables along the sample (e.g. [1, 2, 5, 16, 17]). On the other hand, the effect of an independent variable might be varied

**Table 4** Significant factors affecting speed choice

Speed level	Minimum speed		Average speed		75 percentile speed		Maximum speed	
	Coefficient	<i>t</i> -stat	Coefficient	<i>t</i> -stat	Coefficient	<i>t</i> -stat	Coefficient	<i>t</i> -stat
Constant	63.81	28.19	72.97	39.03	74.02	40.09	62.67	27.83
Traffic flow characteristics								
Bus (opposite direction)	– 1.71	– 3.36	– 2.16	– 4.85	– 2.22	– 5.01	– 1.99	– 3.95
Bus (same direction)	– 12.86	– 9.09	– 11.06	– 8.74	– 10.11	– 8.09	– 12.76	– 8.98
Trucks (opposite direction)	– 2.83	– 9.64	– 2.89	– 11.41	– 2.8	– 11.18	– 2.97	– 10.41
Trucks (same direction)	– 14.65	– 17.3	– 12.68	– 16.76	– 11.52	– 15.42	– 14.52	– 17.05
Other 4 wheelers (opp. direction)	– 1.32	– 2.02	x	x	x	x	x	x
Other 4 wheelers (same direction)	– 7.26	– 4.23	– 6.02	– 3.92	– 5.11	– 3.36	– 7.24	– 4.19
CNG_Auto (same direction)	– 8.39	– 2.76	– 4.83	– 1.79	x	x	x	x
Motorcycle (opposite direction)	– 3.42	– 2.8	– 3.19	– 2.92	– 2.55	– 2.36	– 3.55	– 2.89
NMV (opposite direction)	– 1.85	– 2.75	– 2.07	– 3.44	– 1.5	– 2.51	– 1.87	– 2.76
Road geometric characteristics								
Good shoulder	4.84	3.91	x	x	x	x	4.55	3.65
Bad shoulder	– 6.63	– 2.49	– 7.55	– 3.37	– 6.93	– 3.14	– 6.52	– 2.43
Culvert	3.13	2.15	x	x	2.89	2.25	2.87	1.96
Bridge	x	x	– 4.8	– 3.34	x	x	x	x
Major access	7.92	2.86	6.12	2.49	5.97	2.45	8.71	3.14
Road environmental characteristics								
Medium pedestrian	– 4.35	– 1.82	– 4.87	– 2.28	– 4.79	– 2.26	– 4.33	– 1.8
Low NMV (road side)	10.02	5.3	9.06	5.35	8.13	4.93	x	x

Alpha = 0.05, significant *p* value is shaded

Traffic flow variables were considered as scale variables, and others were considered as binary variables

and random for sites with different influencing attributes. As such, there is scope for model improvement by adding random effects associated with many unaccounted factors. A random parameter model takes into account the effect of heterogeneity due to unobserved factors that may differ from segment to segment [6, 18]. Tarris et al. [8] attempted to incorporate random effects to examine whether the groups comprising individual drivers and time effects vary randomly. Poe and Mason Jr [9] used a mixed-model approach to account for random effect for individual observation sites. Those models are built around the following structural equations:

$$y_{it} = \alpha' w_{it} + x'_{it} \beta_i + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T_i \quad (1)$$

$\varepsilon_{it}$  random error term (representing pure random noise),

$$\beta_i = \beta + \Gamma v_i$$

$$\varepsilon_{it} \sim N[0, \sigma_i^2]$$

The conditional mean function is:

$$E[y_{it}|x_{it}, \beta_i] = \alpha' w_{it} + x'_{it} \beta_i + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T_i \quad (2)$$

The model assumes that parameters are randomly distributed with a possible heterogeneous (across

individuals) distribution. For the random parameter or mixed model, the general form is:

$$\beta_i = \beta + \Delta z_i + \Gamma v_i \quad (3)$$

Where, it is assumed that

$$E[v_i|x_i, z_i] = 0, \quad \text{Var}[v_i|x_i, z_i] = I$$

$$\text{Var}[\beta_i|x_i, z_i] = \Sigma = \Gamma \Gamma'$$

$$\sigma_i^2 = \sigma^2(\text{constant})$$

The detailed methodology for computing the random effect can be seen in Greene [19].

According to the description of the Agbelie [20], the study considered a parameter having random effect if the standard deviation of the variable density is statistically significant. On the other hand, if the estimated standard deviation of the variable is not statistically different from the zero, i.e. not significant, the estimated variable is considered as a fixed across the segments of the highway section.

For the random parameter density functional forms, uniform, triangular and normal distributions were investigated. For a variable with the triangular or uniform distribution, the variance of  $\beta_i$  is  $\sigma_i/6$  or  $\sigma_i/3$ , respectively. It



was found that a normal distribution gave the best statistical result for all random parameters.

### Interpretation of Results

Though separate random parameters models have been developed and validated for four different speed levels, only the mean speed model is discussed here, for reasons of brevity.

#### Model Estimation

Table 5 presents the estimation results for fixed and random parameter average speed models. In the random parameters model, two additional variables were found to be significant, namely medium shoulder and absence of roadside parking. A total of 15 variables were found to be significant; of which six were found to have random effect across the population of segments. The following subsections highlight the results by the type of effect for each parameter.

*Fixed Parameters* Among the nine significant non-random variables, five are related to traffic flow characteristics. The coefficient of all of the fixed parameters variables has shown to have negative sign, which implies that those variables inversely affect the speed choice. In other words, an increase in flow decreases the speed of the vehicle.

However, the impact of the variables and their level of significance for different vehicle types is different. Parameters under the group of traffic flow, operating in the direction of the subject vehicle, have the most significant negative influence in speed choice. Opposite directional motorcycle and non-motorized also have a significantly negative influence.

The impact of the poor shoulder is almost five times more than that of average shoulder. This output confirms that the worsening condition of the road shoulder is one of the most important reasons for the reduction in speed. Sporadic parking, though not significant in the fixed parameter-based OLS model, is significant in the RPL model, where it has a negative influence on the choice of speed.

*Random Parameters* Six variables were found to have a random effect with a normal distribution and significant standard deviation. Buses and trucks travelling in the direction of the subject vehicle are random parameters having a mean of coefficient  $-11.50$  and  $-13.43$ , respectively. Both parameters show a significant influence on the speed choice, and their influences are varied among the segments with the standard deviation of 3.10 and 6.24, respectively. This significant influence of buses and trucks on speed choice is mainly due to their relative low speeds. Most local buses and loaded trucks are very slow and act as a hindrance to maintain normal traffic speeds.

**Table 5** Estimation results for OLS and RPM linear regression models

Variables	Ordinary least squares		Random parameters linear	
	Coefficient	<i>t</i> -stat	Coefficient	<i>t</i> -stat
Constant	72.97	39.03	73.38	46.68
Traffic flow characteristics (continuous variables)				
Bus (opposite direction)	- 2.16	- 4.85	- 1.99	- 5.31
Bus (same direction) (standard deviation of parameter distribution)	- 11.06	- 8.74	- 11.50 (6.24)	- 11.4
Trucks (opposite direction)	- 2.89	- 11.41	- 2.82	- 13.76
Trucks (same direction) (standard deviation of parameter distribution)	- 12.68	- 16.76	- 13.43 (3.10)	- 22.77
Other 4 wheelers (same direction)	- 6.02	- 3.92	- 6.47	- 5.4
CNG_Auto (same direction)	- 4.83	- 1.79	- 5.139	- 1.94
Motorcycle (opposite direction)	- 3.19	- 2.92	- 3.185	- 3.68
NMV (opposite direction)	- 2.07	- 3.44	- 1.96	- 3.93
Road geometric characteristics				
Medium shoulder	-	-	- 1.96	- 1.66
Bad shoulder	- 7.55	- 3.37	- 9.24	- 4.8
Bridge (standard deviation of parameter distribution)	- 4.80	- 3.34	- 3.48 (7.32)	- 2.3
Major access (standard deviation of parameter distribution)	6.12	2.49	6.27 (5.59)	3.06
Road environmental characteristics				
Medium pedestrian (standard deviation of parameter distribution)	- 4.87	- 2.28	- 4.57 (5.45)	- 2.55
Seldom parking	-	-	- 2.71	- 1.81
Low NMV (standard deviation of parameter distribution)	9.06	5.35	9.32 (5.73)	6.62

Among the road geometry parameters, the presence of a small bridge on the segment has a normally distributed random parameter with a mean of  $-3.48$  and as standard deviation of  $7.32$ . The presence of major access roads is also a normally distributed random parameter with a positive effect on speed choice with a mean coefficient of  $6.27$  and a standard deviation of  $5.59$ . Drivers tend to increase speed at major access road locations. This might be due to the segregation of directional traffic using physical devices, since most of the major access segments are channelized.

Two variables were found to have random parameters with respect to road environmental measures. Medium pedestrian activity reduces speed with this parameter having a normally distributed mean coefficient of  $-4.57$  and a standard deviation  $5.45$ . Other random parameter variables related to road environment included the absence or low volume of non-motorized vehicles (NMV). This variable has significant positive impact on speed, with a normally distributed mean of  $9.32$  and standard deviation  $5.73$ . These results imply that the roadside exposure, whether pedestrians or non-motorized vehicles, has a significant influence on the speed choice.

**Model Performance and Validation Analysis**

The performance of both models was assessed using several goodness-of-fit measures, both aggregated and disaggregated. Specifically, the effect of including randomness and heterogeneity when estimating segmental speed choice models was analysed. Akaike Information Criterion (AIC), Corrected Akaike Information Criterion (AICc) and Bayesian Information Criterion (BIC) are commonly used disaggregated goodness-of-fit measures. AIC, AICc and BIC are the relative quality estimator of statistical model based on the log-likelihood function. The model with the lowest BIC is preferred [21–23]. AIC, AICc and BIC values were also calculated using the following equations:

$$AIC = 2K - 2LL \tag{4}$$

$$AICc = 2K - 2LL + \frac{2K(K + 1)}{(n - K - 1)} \tag{5}$$

$$BIC = \ln(n)K - 2LL \tag{6}$$

where  $n$  is the number of observations,  $K$  is the number of parameters estimated by the model and  $LL$  is the log-likelihood value at convergence.

A number of established aggregate level measures were used to evaluate predictive performance of the model, including Pearson correlation ( $R$ ),  $R$ -square ( $R^2$ ), adjusted  $R^2$ , root mean square error (RMSE), mean prediction bias (MPB), mean absolute deviation (MAD); mean squared prediction error (MSPE) and mean absolute percentage error (MAPE).

The Pearson correlation coefficient  $R$  measures the strength of relationship between observed and predicted output estimated from the independent variables. The  $R$  is formulated as:

$$R = \frac{\sum y_i f_i - \sum y_i \sum f_i}{\sqrt{[\sum y_i^2 - (\sum y_i)^2][\sum f_i^2 - (\sum f_i)^2]}} \tag{7}$$

$R^2$  provides the proportion of the variance that is predictable from the independent variables. The  $R^2$  is computed as:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - f_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{8}$$

where  $f_i$  and  $y_i$  are the predicted and observed speed of individual segment population,  $\bar{y}$  is the observed mean speed. Adjusted  $R^2$  is an extension of  $R^2$ , which takes into account the number of explanatory variables and sample size. This is computed as:

$$\text{Adjusted } R^2 = 1 - (1 - R^2) \frac{n - 1}{n - p - 1} \tag{9}$$

where  $p$  is the total number of explanatory variables and  $n$  is the total number of sample.

RMSE provides differences between observed and predicted values and defined as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (f_i - y_i)^2}{n}} \tag{10}$$

where  $f_i$  and  $y_i$  are the predicted and observed speed of individual segment population,

MPB shows the magnitude and direction of average bias in model prediction. MPB is defined as:

$$\text{MPB} = \frac{\sum_{i=1}^n (f_i - y_i)}{n} \tag{11}$$

MAD describes average prediction error of the estimated models. MAD is defined as:

$$\text{MAD} = \frac{\sum_{i=1}^n |f_i - y_i|}{n} \tag{12}$$

MSPE estimates the model prediction error and is defined as:

$$\text{MSPE} = \frac{\sum_{i=1}^n (f_i - y_i)^2}{n} \tag{13}$$

Finally, MAPE expresses error as a percentage and is defined by:

$$\text{MAPE} = \frac{100}{n} \sum \left| \frac{y_i - f_i}{y_i} \right| \tag{14}$$

Table 6 shows the results of the goodness-of-fit measures.

**Table 6** Performance of the models

Measure of fit	OLS model	RM model
Aggregate level		
Pearson correlation coefficient (R)	0.71	0.90
R-square ( $R^2$ )	0.50	0.80
Adjusted $R^2$	0.49	0.79
RMSE	11.56	7.37
MPB	0.03	0.03
MAD	9.05	5.73
MSPE	133.71	54.29
MAPE	13.72	8.60
Disaggregate level*		
AIC	5209.16	5190.98
AICc	5209.81	5191.81
BIC	5272.27	5263.10

\*Model with lower AIC, AICc and BIC value are preferred

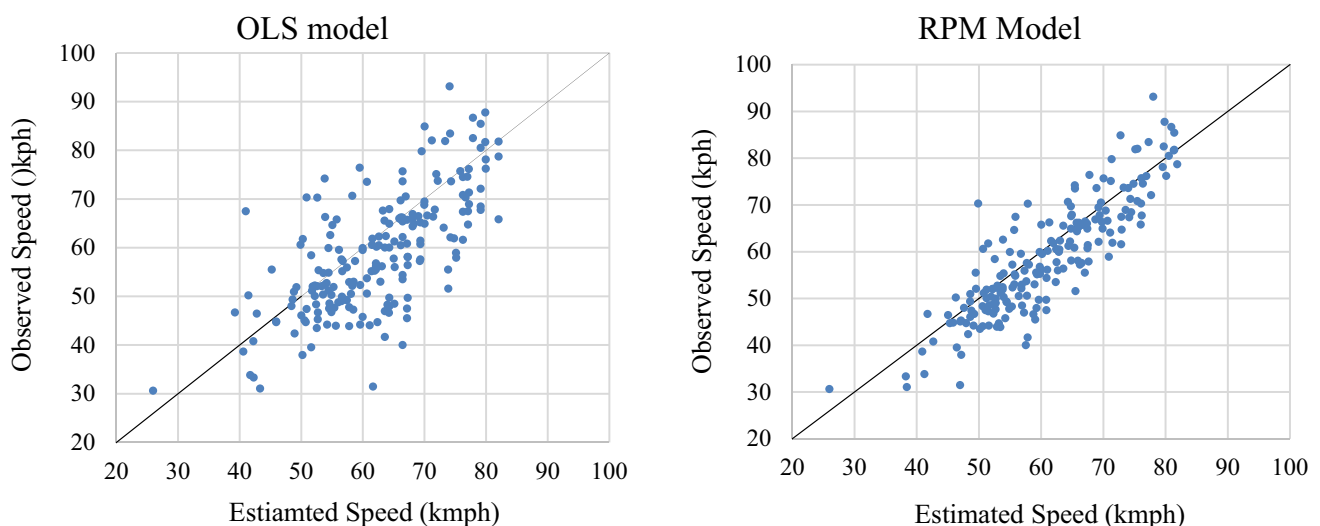
As shown in Table 6, the differences in the goodness-of-fit measures between the models are quite significant. The results clearly suggest that the random parameter model is superior in characterizing driver speed choice across the different road segments in heterogeneous traffic environment of developing countries.

The adjusted  $R^2$  value for the RPM model is 0.79. The model uses the segment mean speed to correlate with the variables for all 200-m segments. However, speed choice of a roadway segment is related to human behaviour, as well as the environment, including weather conditions and the on-road environment. Therefore, the adjusted  $R^2$  of 0.79 indicates that the model has a reasonable fit. The overall statistical assessment shows that the model provides acceptable estimates of segmental speed.

The performance of the models was validated using tenfold cross-validation technique. The observed results for randomly selected three separate trips were compared with model estimates. Figure 7 shows the relationship between the observed and estimated values for each modelling approach. The adjusted  $R^2$  value is 0.50 and 0.78 for OLS and RP model, respectively. This validation test also further established the benefit of accommodating random effects when modelling speed behaviour in heterogeneous traffic environments.

### Summary and Conclusions

This paper summarizes some key features of speed choice behaviour and provides the results of regression modelling used to investigate the factors affecting speed choice behaviour on a two-lane highway in heterogeneous traffic environments of developing countries. Data were collected using different techniques on a 13 km section of a two-lane bidirectional major highway in Bangladesh. The second-by-second choice of driver speeds was extracted using a test vehicle driving normally in traffic. The results show that both the time mean speeds and the space mean speeds are significantly different from the posted 80 kph speed limit. Over-speeding is a common phenomenon, which ranges up to around 110 kph. Vehicles are travelling, on average, 13% of the time and 20% of the distance at speeds above 80 kph. The abrupt change of speed is also a common and concerning issue in this highway section. The analysis of the speed profiles for individual 200 m segments shows that speeds vary considerably across the segments.



**Fig. 7** Performance of the models

The factors influencing speed choice were quantified in detail for four different levels of speeds using traditional ordinary least squares. A comprehensive set of exogenous variables grouped into three main categories, namely traffic flow and operation, road geometry and road and roadside environment, were tested to identify the significant factors and their influences on the speed choice. As the OLS regression model does not consider the randomness and heterogeneity of exogenous variables across the population of segments, the study further formulated random parameter (RPM) regression model. Therefore, this paper contributes to research on driving behaviour by formulating and estimating an RPM regression model to better understand the speed choice behaviour of drivers at micro-level in a developing country two-lane two-way rural highway with a heterogeneous traffic environment.

Though models have been developed for four different speed levels, average speed results have also been put forward. For example, in case of the average speed random parameter model, altogether 15 independent variables were found to have statistically significant impact on the speed choice. Six variables were found to have significant random effects. Those variables include: buses and trucks travelling in the same direction; the presence of a small bridge and major access; medium-level pedestrian activities; and low non-motorized (NMV) activities along the segment. Other significant variables are the presence of four and three wheelers, improper shoulder and parking along the roadside.

In an effort to further assess the predictive performance of the models, eleven goodness-of-fit measures were calculated. The comparative results indicated that the model considering some parameters as random and others as fixed offer a better fit in all aspects. The models were validated using data collected from different trips. These validation results further indicate the considerable potential for the use of random parameter regression when analysing speed behaviour in the heterogeneous traffic environments of developing countries.

The results of the estimated model have important implications for effective speed management to ensure more discipline and safe traffic operation in developing countries. The presence of heavy vehicles, particularly local buses and overloaded trucks, is one of the main reasons for extreme speed differences for different vehicle types. To maintain the homogeneity of the speed throughout the section of the road among different motorized vehicles, the presence of unfit, overloaded heavy vehicle need to be controlled. Long-term policy decision needs to be taken to specify the accepted specification of

freight transport vehicle. Loading and unloading of passengers and freight needs to be restricted to specific locations. Segregation of three-wheeler and non-motorized vehicles from mainstream traffic could be highly effective in improving safety. Road shoulder, presence of a bridge and culvert approach needs to be adequately maintained. The results confirm the need to manage roadside parking, pedestrian and non-motorized activities. Moreover, access management measure has potential to reduce inhomogeneity of speed among the vehicles. Besides, location-specific speed enforcement including police patrolling and installation of speed camera need to be provided to manage speed as well as to efficient and safe traffic operation.

Although the comparative model analysis offers strong evidence of improved performance of the RPM model, further enhancements are possible. For example, more advanced modelling approaches, such as latent class modelling, could be used. Application of machine learning techniques with a larger sample could be a useful improvement in the model and could provide better insight on the factors affecting driver speed behaviour. Discriminant function analysis (DFA) can be used for data classification to investigate the probability of their classification into a certain group, as well as to derive optimal combinations of variables. The current study has some limitations related to data and the consideration of explanatory variables. Due to limited resources and time limitations, only a small number of trips by a particular vehicle operated by a particular driver were used. The collection of data from different roads using different vehicles type and drivers is likely to yield improved models. A further attempt could be made to develop segment-specific micro-level speed prediction models with the incorporation of some other attributes. Those attributes could be time of day, weather, drivers' demographic details, vehicle-related factors, roadside friction profiles and exposure of local traffic. The latter should include pedestrians, informal local para-transit, as well as non-motorized two and three wheelers.

**Acknowledgements** The authors would like to thank the Graduate School and the School of Civil Engineering Transport Group at the University of Queensland for their support during the data collection from Bangladesh.

**Author's Contribution** The authors confirm contribution to the paper as follows: study conception and design: First and Second Author; data collection: First and Third Author; analysis and interpretation of results: First Author; draft manuscript preparation: First author; last three authors also supervised the overall research; all authors reviewed the results, manuscript and approved the final version of the manuscript.

### Appendix A: Summary of Descriptive Statistics of Key Variables

Variables	Description	Mean	SD	Min.	Max.
<b>Dependent variables</b>					
Average speed	Average segmental speed (kph)	63.90	16.44	14	108
Minimum speed	Minimum segmental speed (kph)	58.70	18.40	0	107
Maximum speed	Maximum segmental speed (kph)	69.25	14.77	30	108.5
3rd quartile	75 percentile speed (kph)	66.57	15.81	20	108
<b>Traffic flow characteristics (continuous variables)</b>					
Bus (opposite direction)	Number faces (within the segment during crossing)	0.67	1.04	0	10
Bus (same direction)	Number in front and/or overtake (within the segment during crossing)	0.14	0.36	0	2
Trucks (opposite direction)	Number faces (within the segment during crossing)	1.54	1.84	0	11
Trucks (same direction)	Number in front and/or overtake (within the segment during crossing)	0.51	0.61	0	3
Other 4 wheelers (opposite direction)	Number faces (within the segment during crossing)	0.48	0.83	0	5
Other 4 wheelers (same direction)	Number in front and/or overtake (within the segment during crossing)	0.08	0.30	0	2
CNG_Auto (opposite direction)	Number faces (within the segment during crossing)	0.12	0.39	0	3
CNG_Auto (same direction)	Number in front and/or overtake (within the segment during crossing)	0.03	0.17	0	1
Motorcycle (opposite direction)	Number faces (within the segment during crossing)	0.17	0.42	0	3
Motorcycle (same direction)	Number in front and/or overtake (within the segment during crossing)	0.03	0.16	0	1
NMV (opposite direction)	Number faces (within the segment during crossing)	0.41	0.77	0	4
NMV (same direction)	Number in front and/or overtake (within the segment during crossing)	0.17	0.50	0	4
<b>Road geometric characteristics</b>					
Rd alignment	0 = straight, 1 = otherwise)	0.37	0.48	0	1
Good shoulder	1 = good soft and hard shoulders, 0 = otherwise	0.73	0.44	0	1
Medium shoulder	(1 = good but there is discontinuity, 0 = otherwise)	0.22	0.42	0	1
Bad shoulder	1 = not serving the purpose, 0 = otherwise	0.04	0.21	0	1
Bridge	1 = if there is any small or large bridge in the segment, 0 = otherwise	0.12	0.32	0	1
Culvert	1 = if there is any culvert in the segment, 0 = otherwise	0.15	0.36	0	1
Major access	(1 = if there is access road with motorized movement, 0 = otherwise	0.07	0.26	0	1
Minor access	1 = if there is access but only for pedestrian or NMV, 0 = otherwise	0.22	0.42	0	1
<b>Road environmental characteristics</b>					
High pedestrian	1 = if group of pedestrians with crossing activities (frequent group crossing, more than 5 comprises a group), 0 = otherwise	0.04	0.21	0	1
Medium pedestrian	1 = if couple of pedestrians (random but not group) but no frequent crossing, 0 = otherwise	0.07	0.26	0	1
Light pedestrian	1 = very few pedestrians (scattered on non-random) along the road side, 0 = otherwise	0.18	0.38	0	1
Seldom pedestrian	1 = very rarely pedestrian, 0 = otherwise	0.70	0.46	0	1
High parking	1 = series of parking including MV and NMV, 0 = otherwise	0.01	0.12	0	1
Medium parking	1 = if group parking mainly by NMV, 0 = otherwise	0.12	0.32	0	1
Light parking	1 = if scattered parking mainly NMV, 0 = otherwise	0.13	0.34	0	1
Seldom parking	1 = if no parking activities but rarely, 0 = otherwise	0.07	0.26	0	1
Hat or bazar road side	1 = if yes, 0 = otherwise	0.06	0.24	0	1

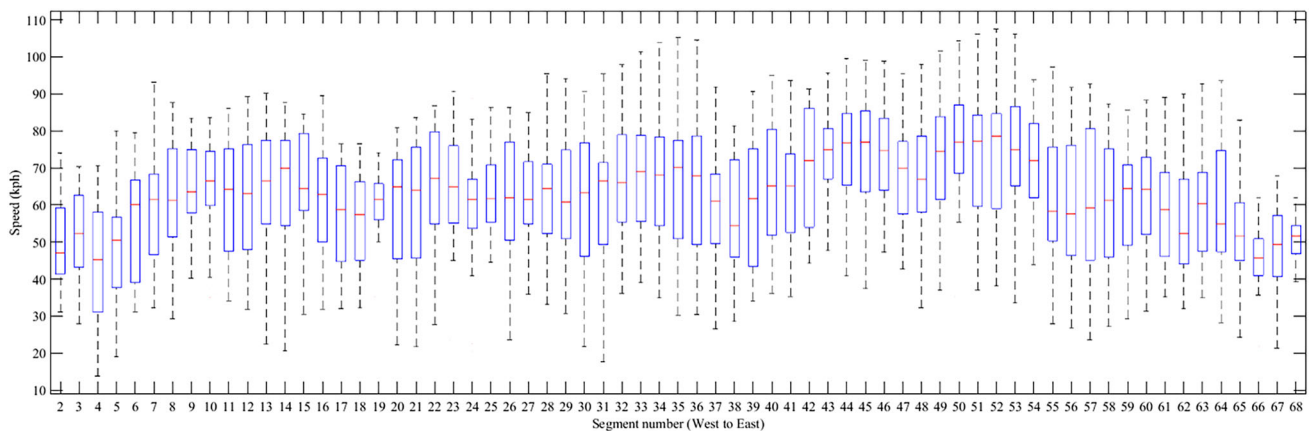


continued

Variables	Description	Mean	SD	Min.	Max.
High NMV	1 = if junk of NMV on on-road or road side, 0 = otherwise	0.03	0.17	0	1
Medium NMV	1 = if couple of NMV on road and road side, 0 = otherwise	0.13	0.34	0	1
Low NMV	1 = if no or very rare NMV, 0 = otherwise)	0.84	0.37	0	1

## Appendix B

See Fig. 8.



**Fig. 8** Speed distribution in different segments of the road

## References

- R. Moses, E. Mtoi, Evaluation of free flow speeds on interrupted flow facilities, *Technical Report*, Department of Civil Engineering, FAMU-FSU College of Engineering, Tallahassee, USA (2013)
- J.N. Ivan, N.W. Garrick, G. Hanson, *Designing Roads that Guide Drivers to Choose Safer Speeds* (Connecticut Transportation Institute, University of Connecticut, Storrs, 2009)
- A. Polus, M. Livneh, B. Frischer, Evaluation of the passing process on two-lane rural highways. *Transp. Res. Rec. J. Transp. Res. Board* **1701**, 53–60 (2000)
- Fitzpatrick K, Eleftheriadou L, Harwood DW, Collins J, McFadden J, Anderson IB, Krammes RA, Irizarry N, Parma KD, Bauer KM (2000) Speed Prediction for Two-Lane Rural Highways (trans: Texas Transportation Institute FHA). USA
- M.A. Figueroa, A. Tarko, Speed factors on two-lane rural highways in free-flow conditions. *Transp. Res. Rec. J. Transp. Res. Board* **1912**, 39–46 (2005)
- P.C. Anastasopoulos, F.L. Mannering, A note on modeling vehicle accident frequencies with random-parameters count models. *Accid. Anal. Prev.* **41**(1), 153–159 (2009)
- M. Park, D. Lee, J. Jeon, Random parameter negative binomial model of signalized intersections. *Math. Problems Eng.* **2016**, 1–8 (2016)
- J. Tarris, C. Poe, J. Mason Jr., K. Goulias, Predicting operating speeds on low-speed urban streets: regression and panel analysis approaches. *Transp. Res. Rec. J. Transp. Res. Board* **1523**, 46–54 (1996)
- C. Poe, J. Mason Jr., Analyzing influence of geometric design on operating speeds along low-speed urban streets: mixed-model approach. *Transp. Res. Rec. J. Transp. Res. Board* **1737**, 18–25 (2000)
- S.M.S. Mahmud, L. Ferreira, M.S. Hoque, A. Tavassoli, Reviewing traffic conflict techniques for potential application to developing countries. *J. Eng. Sci. Technol.* **13**(6), 1869–1890 (2018)
- S. Khan, P. Maini, Modeling heterogeneous traffic flow. *Transp. Res. Rec. J. Transp. Res. Board* **1678**, 234–241 (1999)
- S.M.S. Mahmud, L. Ferreira, M.S. Hoque, A. Tavassoli, M.N. Islam (2017) Some behavioural and safety aspects of Jamuna Multipurpose Bridge (JMB) and its approaches: an analysis using a special dataset, *Paper Presented at the International Conference on Engineering Research, Innovation and Education 2017 ICERIE 2017*, 13–15 January, SUST, Sylhet, Bangladesh, 13–15 January, 2017
- M.N. Islam, *An Investigation of Road Accidents on the Jamuna Bridge and Its Approach Roads* (Department of Civil Engineering, Bangladesh University of Engineering & Technology (BUET), Dhaka, 2004)
- P.S. Bokare, A.K. Maurya, *Acceleration-Deceleration Behaviour of Various Vehicle Types* (RSR Rungta College of Engineering and Technology, Bhilai, 2016)
- H. Wallén Warner, *Factors Influencing Drivers' Speeding Behaviour* (Acta Universitatis Upsaliensis, Uppsala, 2006)

16. S. Yagar, M. Van Aerde, Geometric and environmental effects on speeds of 2-lane highways. *Transp. Res. A General* **17**(4), 315–325 (1983)
17. A. Polus, K. Fitzpatrick, D. Fambro, Predicting operating speeds on tangent sections of two-lane rural highways. *Transp. Res. Rec. J. Transp. Res. Board* **1737**, 50–57 (2000)
18. S.P. Washington, M.G. Karlaftis, F. Mannering, *Statistical and Econometric Methods for Transportation Data Analysis* (CRC Press, Boca Raton, 2010)
19. W.H. Greene, *LIMDEP: Version 11: Econometric Modeling Guide* (Econometric Software Inc, Plainview, 2016)
20. B.R. Agbelie, Random-parameters analysis of highway characteristics on crash frequency and injury severity. *J. Traffic Transp. Eng. (Engl. Ed.)* **3**(3), 236–242 (2016)
21. H. Bozdogan, Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika* **52**(3), 345–370 (1987)
22. G. Schwarz, Estimating the dimension of a model. *Ann. Stat.* **6**(2), 461–464 (1978)
23. S. Yasmin, N. Eluru, C.R. Bhat, R. Tay, A latent segmentation based generalized ordered logit model to examine factors influencing driver injury severity. *Anal. Methods Accident Res.* **1**, 23–38 (2014)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.