How much should a pedestrian be fined for intentionally blocking a fully automated vehicle? A random parameters beta hurdle model with heterogeneity in the variance of the beta distribution

Afghari, Amir Pooyan; Papadimitriou, Eleonora; Li, Xiaomeng; Kaye, Sherrie Anne; Oviedo-Trespalacios, Oscar

DOI
10.1016/j.amar.2021.100186

Publication date
2021

Document Version
Final published version

Published in
Analytic Methods in Accident Research

Analytic Methods in Accident Research 32 (2021) 100186

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

journal homepage: www.elsevier.com/locate/amar

How much should a pedestrian be fined for intentionally blocking a fully automated vehicle? A random parameters beta hurdle model with heterogeneity in the variance of the beta distribution

Amir Pooyan Afghari [a,*], Eleonora Papadimitriou [a], Xiaomeng Li [b], Sherrie-Anne Kaye [b], Oscar Oviedo-Trespalacios [c]

[a] Safety and Security Science Section, Faculty of Technology, Policy and Management, Delft University of Technology, 2628BX Delft, Netherlands
[b] Queensland University of Technology (QUT), Centre for Accident Research and Road Safety-Queensland (CARRS-Q), Kelvin Grove, Queensland 4059, Australia
[c] Strategic Senior Research Fellow, Queensland University of Technology (QUT), Centre for Accident Research and Road Safety-Queensland (CARRS-Q), Kelvin Grove, Queensland 4059, Australia

ARTICLE INFO

Article history:
Received 17 December 2020
Received in revised form 3 August 2021
Accepted 3 August 2021
Available online 11 August 2021

Keywords:
Hurdle regression
Truncated regression
Beta distribution
Fully automated vehicles
Pedestrian receptivity

ABSTRACT

Intentionally blocking the path of fully automated vehicles is an important dimension of pedestrians' receptivity towards these vehicles. The monetary value of this behaviour can be obtained by asking pedestrians about their perception of the "fine" for blocking the path of a fully automated vehicle. Econometric modelling of the reported fine can shed more light on factors influencing pedestrians' receptivity towards fully automated vehicles. However, development of such an econometric model is not straightforward due to the unique characteristics of the dependent variable: it has two fundamentally different states; it is right-truncated; and it may be fat-tailed. Despite fairly extensive methodological advancements in econometric modelling of pedestrian behaviour, there is no model that can adequately explain these characteristics. While a beta distribution in a hurdle setting has the potential to address the above complexities, its applicability in dealing with limited dependent variables in transport applications has remained, by and large, unexplored.

This study aims to fill this gap by developing a new beta hurdle regression model that systematically considers the dual-state of a right-truncated dependent variable representing the fine associated with intentionally blocking a fully automated vehicle. The hypothesized model is empirically tested using data obtained from a survey administered in Queensland, Australia, and the results are compared with truncated lognormal, and truncated lognormal hurdle regression models. Results indicate that the hurdle models are superior to the non-hurdle model. The beta variant of the hurdle model provides a better statistical fit for the data that are near their right limit. In addition, parametrizing the variance of the beta distribution captures the additional heterogeneity in the data. Age, gender, education level, violations, attitudes, behaviours that appease social interactions, and perceived ease or difficulty of interacting with fully automated vehicles influence the likelihood and/or the propensity of the fine and thus are associated with the perceived monetary value of intentionally blocking the path of a fully automated vehicle.

* Corresponding author.
E-mail addresses: a.p.afghari-1@tudelft.nl (A.P. Afghari), e.papadimitriou@tudelft.nl (E. Papadimitriou), xiaomeng.li@qut.edu.au (X. Li), s1.kaye@qut.edu.au (S.-A. Kaye), oscar.oviedotrespalacios@qut.edu.au (O. Oviedo-Trespalacios).

https://doi.org/10.1016/j.amar.2021.100186
2213-6657/

1. Background

Understanding road users' receptivity towards fully automated vehicles, is paramount prior to their adoption in urban transport networks. Many studies have investigated such receptivity by designing hypothetical experiments and applying psychosocial theories such as the Technology Acceptance Model (Davis, 1985; Davis et al., 1989), the Theory of Planned Behaviour (Ajzen, 1991), and the Unified Theory of Acceptance and Usage of Technology (Venkatesh et al., 2012) in order to identify the behavioural factors that are associated with the intention to use fully automated vehicles in the near future. Some examples of these factors are attitudes (favourable attitudes towards fully automated vehicles), social influence or subjective norms (perceptions that an important other would approve the behaviour), perceived usefulness, and perceived ease of use (Buckley et al., 2018; Kaye et al., 2020b; Madigan et al., 2017). In addition, many studies have examined public perceptions and preferences towards fully automated vehicles (e.g., Ahmed et al., 2020; Bansal et al., 2016; Haboucha et al., 2017; Moody et al., 2020; Sheela and Mannering, 2020; Woldeamanuel and Nguyen, 2018). For example, Ahmed et al. (2020) recruited individuals in the United States and explored their perceptions of safety benefits and concerns of fully automated vehicles. Their findings revealed that demographics, opinions, and past crash experiences influence public perceptions towards these vehicles. Further, Bansal et al. (2016) assessed the opinions of another group of respondents in the United States about connected and fully automated vehicle technologies. They found that while most respondents would like to drive automated vehicles along freeways or highways, half of the respondents were very worried about equipment or system failures in these vehicles. They also found that age, gender, number of children, and living in higher-income neighbourhoods are associated with the willingness to pay more money to add automation in vehicles. In a broader study, Haboucha et al. (2017) used a sample of individuals in the United States, Canada, and Israel, and found that older respondents and those who enjoy driving would prefer using their regular vehicle as opposed to an automated vehicle. On the contrary, respondent with a higher education are more likely to report favouring automated vehicles than their regular vehicle. Finally, Moody et al. (2020) surveyed a large number of respondents across 51 countries and found that age, gender, education level, employment status, average household income, and awareness of automated vehicles are associated with safety perceptions towards these vehicles.

However, most of the above studies have focused on drivers' intentions for using fully automated vehicles and only recently has research started to assess pedestrians' receptivity towards these vehicles (Velasco et al., 2019; Deb et al., 2017a; Deb et al., 2017b). In doing so, crossing the road in front of a fully automated vehicle has been used as an example of pedestrians' behavioural interaction with these vehicles (Kaye et al., 2021). Deb et al. (2017b) proposed a conceptual model for such behaviour and found that attitudes, social norms, system effectiveness, trust and personal innovativeness significantly influence pedestrians' intentions to cross in front of a fully automated vehicle. Velasco et al. (2019) also found that pedestrians who reported higher ratings of perceived behavioural control (perceived ease of use) and trust in fully automated vehicles reported higher crossing intentions. Rothenbücher et al. (2016) and Palmeiro et al. (2018) investigated the differences in pedestrians' gap acceptance behaviour in crossing the road in front of a fully automated vehicle with that of a conventional vehicle. While they found no significant difference in the gap acceptance behaviour of participants against the two types of vehicles, their findings were significantly influenced by the differences in the appearance of these vehicles. Finally, Razmi Rad et al. (2020) investigated the yielding behaviour of pedestrians when crossing in front of a fully automated vehicle and found that the probability of crossing decreases with smaller time gaps.

1.1. Intentionally blocking a fully automated vehicle

While the above studies have provided an initial understanding of the psychosocial factors associated with pedestrians' receptivity towards fully automated vehicles, much is still unknown about pedestrians' expectations and behavioural adaptation to the presence of these vehicles on the roads. In particular, intentionally crossing the road in front of a fully automated vehicle is an important dimension of pedestrians' receptivity that has been relatively unexplored. Fully automated vehicles will be risk-averse and thus pedestrians will be able to behave with impunity when interacting with these vehicles due to the secure knowledge that they will yield (Millard-Ball, 2018). This intentional behaviour may represent taking advantage or oppression towards fully automated vehicles (Liu et al., 2020) and may lead to traffic delays, risky events (for the pedestrian and for the vehicle), and ultimately results in substantial societal and monetary costs.

In the conventional traffic environment, legislation and enforcement are routinely applied as effective interventions to deter road users from taking advantage of one another. Legislation communicates the behavioural standards with the road users, whilst enforcement encourages road users to comply with those standards by applying a penalty for non-compliance (Oviedo-Trespalacios and Watson, 2021). Despite these interventions, conflicts and interactions often occur between pedestrians and motorised vehicles. For example, a pedestrian "forcing" a vehicle to yield by leaving the kerb and stepping onto the road. This is a relatively common behaviour on urban roads and is generally not considered a traffic violation in some jurisdictions. This interaction often involves the exchange of human cues between the driver and the pedestrian, which may be visual (e.g. gestures, signs) or auditory (e.g. talking, honking). However, these forms of communication may not be available when interacting with fully automated vehicles without a human driver. The lack of possibility for this type of "negotiation" with a fully automated vehicle, also referred to as "forward incompatibility" (Van Loon and Martens, 2015), may be one reason that would bring the pedestrian in the position of being "accountable" for any delays or risk induced by intentionally blocking the path of the fully automated vehicle. Another reason is the general argument that intentional violations against normal operation of automation, such as deliberate disengagement of autopilot by a human driver, should not be expected to be manageable by automation (Noy et al., 2018); intentionally blocking the path of a fully automated vehicle may be considered as such. However, very little is known about the acceptance of these legislative measures among pedestrians in the context of fully automated vehicles.

While the combination of evidence-based legislation and police enforcement, along with supporting public education, can prevent pedestrians from intentionally blocking the path of fully automated vehicles, it seems difficult to reach a consensus about the monetary value of this behaviour on a theoretical basis. An alternative approach to solve this dilemma can be to empirically ask pedestrians about their perception of the "fine" associated with this behaviour, should these vehicles be on the roads. This approach has been used in the transport policy literature and behavioural science to provide the value of a certain illegal behaviour among road users (Sahebi et al., 2019). Econometric modelling of the reported fine in this approach can then provide deep insight about how pedestrian characteristics influence their behavioural adaptation towards fully automated vehicles and can help develop well-accepted policies by the community and reduce resistance against the policies.

1.2. Econometric modelling of the value of fines

In experimental studies of pedestrian behaviour, the dependent variable representing the fine for intentionally blocking the path of a fully automated vehicle can be derived from a series of questions such as "*do you think the [intentional blocking] behaviour should be fined?*" followed by "*if yes, how much do you think should be the fine?*" in a survey. Econometric modelling of such a dependent variable is, however, confronted with unique complexities as in the following.

First and foremost, the dependent variable extracted from the above questions has two states that are theoretically different: the zero state and the positive state. The zero state is associated with the belief that the intentional behaviour should not be fined whereas the positive state is associated with the propensity for the monetary value of the fine for that intentional behaviour. As a result, the data generating processes behind these two states are fundamentally different. Hurdle regression models have been proposed in the statistical literature to address this complexity (Gurmu, 1997). These models rely on the key assumption that the zero values of a random variable have a separate probability of occurrence than the non-zero values. As a result, the probability of the zero values is obtained using a binary regression and the probability of non-zero values is obtained using a separate generalized regression model.

Hurdle regression models have been widely used in the road safety literature, where they are mostly referred to as zero-inflated models for modelling crash frequency. In this context, zero-inflated models address the implications of excessive zeros in road crash observations, by assuming that the observations include two different states: a 'zero' state (also referred to as 'normal operations', or 'safe' state) in which the probability of crash occurrence is very low and thus indistinguishable from zero, and a 'non-zero' (or 'crash prone') state with a typical Poisson-family crash data generating process e.g. Poisson or Negative Binomial distribution (Katrakazas et al., 2020; Malyshkina and Mannering, 2010). In many applications with varying context-specific formulations, zero-inflated models provide superior statistical fit (Ma et al., 2016; Hong et al., 2019; Yu et al., 2019; Ma et al., 2019; Dong et al., 2014; Cai et al., 2016); however, they are often criticized for certain theoretical and statistical shortcomings. First, it is noted that a zero-crashes state cannot exist truly and constantly over time, and introducing serial correlation among crash observation or random parameters in the model may significantly reduce the probability of being in the 'zero' state (Aguero-Valverde, 2013; Chen et al., 2016). Lord et al. (2005) argue that excessive zero crash observations may be due to other factors, such as low exposure, small spatial scale, and under-reporting. Second, it is noted that the superior fit may be a result of over-fitting, since the zero-inflated specification induces a number of additional model parameters, not always 'penalized' by the used goodness-of-fit measures (Washington et al. 2020; Lord et al., 2007).

Despite numerous applications of the hurdle model in crash frequency analysis, only a few studies have adopted this methodology for crash severity analysis. Malyshkina and Mannering (2009) developed a double state Markov switching model of crash injury severity in order to account for two unobserved states of roadway safety as a means of accounting for potential unobserved heterogeneity. They found that the double state model outperforms the single state model in terms of statistical fit. Xiong et al. (2014) developed a similar model to accommodate time-varying and time-constant unobserved heterogeneity in crash injury severity analysis and also found that the double state model outperforms the single state model. Jiang et al. (2013) developed a double state zero-inflated ordered probit model to account for two distinct sources of injury severity: injury propensity and injury severity. They found that the double state model provides additional insight about the effects of explanatory variables on the injury severity of crashes.

Nonetheless, a relevant conclusion from the road safety literature is that the use of hurdle models is recommended as long as there is "a logical consistency between the dual-state process and the underlying state of causal processes" (Lord et al., 2007). Therefore, they are well justified for modelling the fines in this study.

Another complexity of modelling the fines extracted from the above questions is related to their probability density in the positive state. The dependent variable in this state is bounded below a certain limit. This complexity is due to the experimental nature of the study where a limit has to be set to ask participants' opinion about the fine for the targeted intentional behaviour. The aim of the study is, however, to make inferences about a larger population (those who would have provided any non-negative value of the fine) from a sample that is drawn from a distinct subpopulation (those who responded below the limit defined in the survey). This property which is referred to as truncation in the statistical literature (Greene, 2003) can result in biased parameter estimates if not accounted for in econometric modelling of the fines. Truncated regression models (distinguished from censored regression models[1]) are commonly used to address this complexity by assuming a truncated distribution for the limited dependent variable (Johnson et al., 1995) (more on this will be presented in the next section). For example, truncated lognormal distribution has been widely used in transport applications for modelling positive limited dependent variables such as travel time, vehicle speed, and traffic capacity (Wang et al., 2012; Weng and Yan, 2016). However, the limited dependent variables in the previous applications of this distribution have all been heavy-tailed. Indeed, the lognormal distribution is well suited for the random variables whose probability densities decay as their values increase (Crow and Shimizu, 1987). However, this assumption may not always hold for truncated dependent variables because truncation, by definition, requires that the data are piled up around the limits. As a result, the distribution of the variable may be "fat-tailed" (Ready and Hu, 1995) meaning that the values that are farther from the mean may still have relatively high probability of occurrence.

Alternatively, beta distribution can be used as the distribution of fines in the positive state. The beta distribution has been widely used in various disciplines (Gupta and Nadarajah, 2004) for random variables that are proportions and hence are bounded between zero and one. As such, this distribution has the potential to be used for modelling a limited dependent variable if it is transformed into a proportion. This transformation can be easily achieved by dividing the variable by the maximum value in the experiment. In return and in contrast to the lognormal distribution which has a shape (mean) and a scale (variance) parameter, the beta distribution has two shape parameters with which its mean and variance are obtained. As a result, the beta distribution may be more flexible in covering a wider range of values in the data. In addition, the variance of the beta distribution can be easily parameterized as a function of covariates and in turn can help explain the additional heterogeneity in the data. In spite of these appealing properties, the applicability of the beta distribution for modelling non-negative limited random variables in transport applications has been, by and large, unexplored.

1.3. Study objectives

Econometric modelling of the (perceived) fine associated with a pedestrian intentionally blocking the path of a fully automated vehicle can shed more light on pedestrian interactions with these vehicles in the near future of transport networks. However, development of such an econometric model is not straightforward because of the unique characteristics of the dependent variable: (1) it has two fundamentally different states; (2) it is right-truncated; and (3) it may be fat-tailed. Despite fairly extensive methodological advancements in the econometric modelling of pedestrian behaviour, there is no model that can adequately explain these characteristics. While a beta distribution in a hurdle setting has the potential to address the above complexities, its applicability in dealing with limited dependent variables in transport applications has remained unexplored. This study aims to fill this research gap by developing a new beta hurdle model that systematically considers the dual-state of a right-truncated dependent variable representing the fine associated with intentionally blocking the path of fully automated vehicles. The model is specified with random parameters to address unobserved heterogeneity in data. The hypothesized model is then empirically tested using data obtained from a survey administered in Queensland, Australia, and the results are compared with the state-of-the-art econometric models for handling limited dependent variables including truncated lognormal regression, and truncated lognormal hurdle regression models.

It is worth emphasizing that while the proposed framework in this study, in and of itself, contributes to empirical research about pedestrians' receptivity towards fully automated vehicles, the main focus of the study is methodological development of a new econometric model. We employ the empirical data which have been originally collected for a more general purpose, to show the applicability of the proposed beta hurdle model and compare its performance with that of the other model candidates.

2. Methodology

Let Y be the dependent variable representing the fine associated with intentionally blocking the path of a fully automated vehicle in this study. This dependent variable is bounded between zero and an upper limit. It is hypothesized that a random parameters beta hurdle regression model can adequately capture such a bounded dependent variable and its complexities. This hypothesis will be tested by comparing the performance of this model against empirical data with that of alternative econometric models in the literature (truncated regression models) but prior to that, the specification details of these models are presented in the following.

[1] Censored regression models are also used to address limited dependent variables in the statistical literature (Greene, 2003). However, these models rely on the key assumption that there is partial information about the dependent variable in the data (it is known that the variable is bounded beyond certain limits, but it is not known how far above or below those limits (Koul et al., 1981). However, the dependent variable in this study is not censored because it does not include any observations that are beyond its limits. As such, censored regression models are not within the scope of this study.

2.1. Beta hurdle regression model

The hurdle regression model is constructed by using a hurdle between zero and positive values of the dependent variable and using different probabilities for these two states:

$$f(Y=y|\pi,\theta)=\begin{cases}1-\pi, y=0\\ \pi f(y|\theta), y>0\end{cases} \tag{1}$$

where $0\le\pi\le1$ is the probability that the dependent variable is positive, $f(y|\theta)$ is the probability density of the dependent variable given that it is positive, and θ is the parameters of the model. The log likelihood function of this hurdle model can be presented in two layers and as the sum of two log likelihood functions with disjoint parameters (Ma et al., 2016):

$$LL(y;\pi,\theta)=\sum_{y=0}\log(1-\pi)+\sum_{y>0}\pi f(y|\theta)=\left[\sum_{y=0}\log(1-\pi)+\sum_{y>0}\log(\pi)\right]+\sum_{y>0}f(y|\theta)=LL_1(\pi|y)+LL_2(\theta|y) \tag{2}$$

As a result, the hurdle regression model can be estimated by maximizing the two layers individually.

In the first layer, zero values of the dependent variable are treated as one category (zero) and the remaining positive values are treated as another category (non-zero). A random parameters binary logistic regression model is employed to determine the probability of positive category in the data. Assuming a logit function between the binary dependent variable and explanatory variables, the probability of ith observation being positive can be presented as:

$$\pi=\frac{1}{1+e^{-\left(x_i^t\beta_i\right)}} \tag{3}$$

where β_i is the vector of parameters and x_i^t is the (transposed) vector of explanatory variables. To account for the unobserved heterogeneity in the effects of explanatory variables on the dependent variable, the parameters of explanatory variables are allowed to vary across observations and thus are introduced with a probability density function, $g(\beta_i)$, across the sample (Oviedo-Trespalacios et al., 2020). Maximum simulated likelihood estimation is used to estimate the above random parameters binary logistic regression model in the first layer (Bhat, 2001).

In the second layer, the observed values of the response variable in the non-zero state of the first layer are used and a probability density function, $f(y|\theta)$, is selected based on a distributional assumption for these observed values. In doing so, the values of the dependent variable are transformed into proportions (out of the maximum value) and thus are assigned a beta distribution:

$$f(y|\theta)=\Gamma(a+b)\frac{y^{(a-1)}(1-y)^{(b-1)}}{\Gamma(a)\Gamma(b)} \tag{4}$$

where Γ is the gamma function and a and b are the parameters of the beta distribution. The expectation of beta distribution is equal to $E[Y_i]=\mu=\frac{a}{a+b}$. Using a logit link function, the expectation of the dependent variable can be expressed as a function of exogenous covariates:

$$\mu_i=\frac{1}{1+e^{-\left(z_i^t\lambda_i\right)}} \tag{5}$$

where λ_i is the vector of parameters (also allowed to vary across observations in the same fashion as previously stated) and z_i^t is the (transposed) vector of covariates. In addition, the variance of the beta distribution $VAR[Y_i]=\sigma^2=\frac{ab}{(a+b)^2(a+b+1)}$ is also specified as an exponential function of exogenous covariates to help capturing the heterogeneity of variance:

$$\sigma^2=e^{m_i^t\gamma} \tag{6}$$

where γ is the vector of parameters and m_i^t is the vector of covariates. The second layer is also estimated individually using maximum simulated likelihood estimation approach. The overall model –layer one and layer two– is referred to as the *random parameters Beta hurdle regression model* in this study.

2.2. Truncated lognormal regression model

The truncated regression model can be constructed by assuming a truncated distribution for the dependent variable in a regular generalized linear regression model. This truncated distribution is, indeed, a part of an untruncated distribution which is below (and/or above) a certain value. The dependent variable in this study is right-truncated which means that its corresponding values are less than or equal to a limit –denoted by c. The probability density function of such a right-truncated dependent variable can be obtained as (Greene, 2003):

$$f(Y=y|y\le c)=\frac{f(y\le c)}{F(c)} \tag{7}$$

where $f(.)$ and $F(.)$ are probability density and cumulative probability density functions of the untrunacted dependent variable, respectively. The dependent variable in this study is non-negative and so lognormal distribution is used to construct a truncated lognormal regression model. The choice of the lognormal distribution among all non-negative distributions is motivated by its simplicity in implementation (Embrechts et al., 2013). The probability density function of the lognormal distribution can be stated as:

$$f(y|\theta) = \frac{1}{\sqrt{2\pi}y\sigma}\exp\left[-\frac{(\ln(y)-\mu)^2}{2\sigma^2}\right] \tag{8}$$

where μ and σ^2 are the mean and the variance of the corresponding normal distribution. Using a linear link function, the mean can be linked with explanatory variables:

$$\mu = x_i^t\beta_i \tag{9}$$

where β_i is the vector of parameters and x_i^t is the (transposed) vector of explanatory variables. The parameters are introduced with a probability density function, $g(\beta)$, to account for unobserved heterogeneity in data. Replacing Eq. (9) in Eq. (8), and then Eq. (8) in Eq. (7), the probability density function of the random parameters right-truncated lognormal model can be obtained as:

$$f(Y = y|y \le c) = \frac{\frac{1}{\sqrt{2\pi}y\sigma}\exp\left[-\frac{(\ln(y)-\mu)^2}{2\sigma^2}\right]}{\int_0^c \frac{1}{\sqrt{2\pi}y\sigma}\exp\left[-\frac{(\ln(y)-\mu)^2}{2\sigma^2}\right]dy}g(\beta) \tag{10}$$

The log likelihood function of this model can then be expressed as:

$$LL = \sum\log\left[\int f(y|y \le c)g(\beta)d\beta\right] \tag{11}$$

This log likelihood function does not have a closed form either and is estimated using the maximum simulated likelihood estimation approach. In order to be consistent with the beta hurdle model, the values of the dependent variable in the right-truncated lognormal model are also transformed into proportions (out of the maximum value) and thus the right limit of the data is set to one ($c = 1$) in this model.

2.3. Measures of statistical Goodness-of-Fit

To test the performance of the above models, their statistical fit to empirical data is evaluated and compared with one another. For models that have the same likelihood structure (variants of the hurdle model or variants of the truncated model), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are employed to compare their statistical fit (Washington et al., 2020):

$$AIC = -2LL + 2P \tag{12}$$

$$BIC = -2LL + Pln(N) \tag{13}$$

where LL is the log-likelihood of the estimated model at convergence, P is the number of estimated parameters, and N is the number of observations or sample size. The model with lower AIC and BIC is regarded as a superior model in terms of statistical fit. However, AIC and BIC are not comparable between the models that have different likelihood structures (hurdle versus truncated models). As such, Mean Absolute Deviance (MAD) and Mean Squared Predictive Error (MSPE) are used to compare the performance of the latter models. Suppose, Y_i and $\widehat{Y}_i$ are the observed and the predicted values of the dependent variable for respondent i. The MAD and MSPE are calculated as (Washington et al., 2020):

$$MAD = \frac{1}{N}\sum_{i=1}^{N}|Y_i - \widehat{Y}_i| \tag{14}$$

$$MSPE = \frac{1}{N-P}\sum_{i=1}^{N}\left(Y_i - \widehat{Y}_i\right)^2 \tag{15}$$

where N is the sample size and P is the number of estimated parameters. The model with smaller MAD and MSPE is usually preferred over the other models.

2.4. Principal component analysis

In estimating all of the above models, many behavioural variables are used to predict pedestrians' receptivity towards fully automated vehicles. However, these variables may have high autocorrelation with one another as previously shown in the statistical analysis of behavioural data (Huitema, 1986; Huitema and McKean, 1991). Principal Component Analysis (PCA) is a common approach used in the statistical literature (Tipping and Bishop, 1999) to summarize data and overcome this problem (Henry and Hidy, 1979). The PCA creates a set of new variables, referred to as *principal components (PC)*, each of which is a linear and orthogonal combination of the original variables in such a way that each orthogonal combination captures the maximum variability in the original set of variables and has the minimum autocorrelation with other linear combinations. The principal components can be obtained by applying the orthogonal transformation and finding the Eigenvectors and Eigenvalues of the Spearman correlation matrix of the original set of explanatory variables. The principal components are then arranged based on their decreasing contribution to the total variance of the original set of explanatory variables: the first principal component explains the highest variability in the explanatory variables; the second principal component explains the second highest variability in the explanatory variables, and so forth (the cumulative contribution of all principal components is equal to 1). These principal components can then be used in the analysis as representatives of the original set of variables. The number of principal components to be used in the model depends on the specific research objective, though the common practice is to use all principal components with Eigenvalues greater than one (Tipping and Bishop, 1999). However, in the context of this study, the first principal component is selected as the most appropriate representative of all items within each category of behavioural variables.

3. Empirical data

To test the applicability of the proposed model, empirical data were collected for pedestrians in this study through an online survey administered in Queensland, Australia. The target population was pedestrians aging 18 years or older and residing in Australia. Screening questions (age and residential country) were asked before they started the survey so that the collected sample align with the target population. Participants were asked about their prior knowledge of automated vehicles before they started answering questions related to them. The collected data included participants' demographics, past behaviour on roads, their receptivity towards fully automated vehicles and their perception of fine for intentionally blocking the path of a fully automated vehicle.

The demographic questions included age, gender, education level, walking time and past crash experience. The behaviour questions were adopted from the Pedestrian Behaviour Questionnaire (Deb et al., 2017a) to measure pedestrians' violations (4 items), errors (4 items), lapses (4 items), aggressive behaviours (4 items) and positive behaviours –those behaviours that appease social interactions (4 items). These items were all answered on a 6-point Likert scale, where 1 = "very infrequently or never", 2 = "quite infrequently", 3 = "infrequently", 4 = "frequently", 5 = "quite frequently", and 6 = "very often or always". The receptivity questions investigated the participant's perception about crossing the road in front of a fully automated vehicle. These questions included attitudes (4 items), trust (5 items) and perceived ease of use (2 items). Attitudes is one of the main components of the Theory of Planned Behaviour (Ajzen, 1991), which measures individual's overall positive or negative feelings toward fully automated vehicles. Perceived ease of use is a key construct of the Technology Acceptance Model which measures the individual's perception of the extent to which interacting with a fully automated vehicle will be free of effort (Davis and Venkatesh, 1996). Trust measures the individual's belief that a fully automated vehicle can perform its intended task efficiently, and it is identified as an important influencing factor of automation acceptance in the Automation Acceptance Model proposed by Ghazizadeh et al. (2012).

Attitudes were measured using a 7-point semantic differential scale (Kaye et al., 2020a) whereas trust and perceived ease of use were measured using a 7-point Likert scale (1 = "strongly disagree", 2 = "moderately disagree", 3 = "somewhat disagree", 4 = "neutral", 5 = "somewhat agree", 6 = "moderately agree", and 7 = "Strongly agree"). Lastly, participants were asked whether they "think that pedestrians intentionally blocking the path of a fully automated vehicle should be fined or punished". If the participants answered "Yes", they were asked the following question, "how much should be the fine? (from 1 Australian dollars to 1000 Australian dollars)", and an integer number between 1 and 1000[2] was required to be selected. The complete questionnaire is presented in the Appendix.

The questionnaire was created using the Qualtrics online survey design platform (http://www.qualtrics.com). A convenience sampling method was used, and the survey was distributed through multiple channels. A global online market search firm, Dynata (http://www.dynata.com), was invited to provide survey administration and dissemination, as well as data collection and cleaning services. The online survey was also disseminated using social media (e.g., Facebook, Twitter) and electronic mail through the university mailing lists. The data were collected from October 2019 to December 2019.

A total of 469 individuals participated and completed the survey. The collected sample includes an even distribution of males and females. Participants are minimum 18 years and maximum 85 years old and their average age in the sample is 35.4 years. One third of the participants had a bachelor's degree or higher. More than one third of the participants reported

[2] This value was selected as the upper limit in the survey because a fine of 1000 Australian dollars is considered to be high in Australia and is unlikely to be implemented. Additionally, at the time of the study, the fine for handheld mobile phone use while driving in Australia was set at 1000 Australian dollars followed by a massive public education campaign. So, we suspected that this value was part of the public discourse already. Although higher values of the fine could have been considered in the survey, it is very unlikely that those higher values would have been reported because 1000 Australian dollars is already 82% of the average weekly total earnings of a person in the country. The mean, the standard deviation, and the 97th percentile of the reported fines across the sample are 185.9, 222.6 and 995.1 Australian dollars, respectively. These descriptive statistics show that reporting a value of fine higher than 1000 Australian dollars is very unlikely in the sample data in this study.

Table 1
Descriptive statistics of variables used in this study (sample size = 469, number of records with zero fines = 105, number of records with non-zero fines = 364).

Continuous variables	Mean (Zero/Non-Zero)	S. D. (Zero/Non-Zero)	Min (Zero/Non-Zero)	Max (Zero/Non-Zero)
Fine (Australian Dollars)	185.9 (0, 239.6)	222.6 (0, 225.9)	0 (0, 1)	1000 (0, 1000)
Age (Years)	35.4 (35.095, 35.485)	16.4 (16.049, 16.517)	18 (18, 18)	85 (80, 85)
Pedestrian behaviour on road				
Violations (4 items) (6-Point likert scale)*	2.447 (2.700, 2.374)	1.106 (1.203, 1.067)	1 (1, 1)	6 (6, 6)
Errors (4 items) (6-Point likert scale)	1.707 (1.919, 1.646)	0.734 (0.872, 0.678)	1 (1, 1)	5.25 (5.25, 5.25)
Laps (4 items) 6-Point likert scale)	1.386 (1.445, 1.369)	0.702 (0.758, 0.686)	1 (1, 1)	5 (4.25, 5)
Aggressive behaviours (4 items) (6-point likert scale)	1.421 (1.507, 1.396)	0.654 (0.758, 0.619)	1 (1, 1)	5.25 (4.5, 5.25)
Positive behaviour (4 items) (6-Point likert scale)	4.139 (3.888, 4.211)	1.007 (1.086, 0.973)	1 (1, 1)	6 (6, 6)
pedestrian receptivity towards fav				
attitudes (4 items) (7-point semantic differential scale)**	4.437 (3.838, 4.609)	1.645 (1.558, 1.630)	1 (1, 1)	7 (7, 7)
Trust (5 items) (7-point likert scale)**	3.865 (3.539, 3.959)	1.399 (1.413, 1.383)	1 (1, 1)	6.8 (6.6, 6.8)
Perceived ease of use (2 items) (7-point likert scale)	3.956 (3.814, 3.997)	0.744 (0.725, 0.746)	1 (1, 1)	7 (6, 7)

Binary variables	Sample Share (Zero/Non-Zero)
Gender (1: male, 0: female)	0.544 (0.562, 0.538)
Education: (1: university or more, 0: otherwise)	0.316 (0.476, 0.393)
Walking time in a week (1: more than 5 h per week, 0: otherwise)	0.209 (0.229, 0.203)
Crash experience in the past two years (1: at least one crash, 0: zero crash)	0.333 (0.324, 0.335)
Have heard the term "automated vehicles" before? (1: yes, 0: no)	0.921 (0.895, 0.929)

*: 1 = "very infrequently or never", 2 = "quite infrequently", 3 = "infrequently", 4 = "frequently", 5 = "quite frequently", and 6 = "very often or always".
**: 1 = "strongly disagree", 2 = "moderately disagree", 3 = "somewhat disagree", 4 = "neutral", 5 = "somewhat agree", 6 = "moderately agree", and 7 = "Strongly agree".

a hit or a near-hit by a car while walking on the road in the past two years. About 92% of participants reported they had heard of automated vehicles before while the remaining 8% reported that they had not heard of the term until they joined the study. In addition, about 25% of responses to the trust questions were missing completely at random (MCAR) and thus were replaced by their arithmetic means (Afghari et al., 2019). Lastly, 22.4% (105 out of 469) of participants reported that there should be no fine for intentionally blocking the path of a fully automated vehicle (they responded with a zero fine). The descriptive statistics of the data used in the study is presented in Table 1.

4. Results and discussion

The beta hurdle regression and the truncated lognormal regression models were estimated against the empirical data and their performances were compared to assess the suitability of the hurdle setting. In addition, the right-truncated lognormal distribution was also investigated as the probability density of the positive values of the dependent variable in the second layer of the hurdle regression model and the performance of the resulting *right-truncated lognormal hurdle regression model* was compared with that of the beta hurdle regression model. This comparison can help assess the suitability of the beta distribution in this study.

In all models, explanatory variables were selected using a stepwise variable selection criterion. In addition, explanatory variables were tested for multicollinearity by computing the Pearson correlation coefficients, and the variables with unacceptably high (>0.7) correlation coefficients were excluded from the models. The principal component analysis was applied on the questionnaire items within the same behavioural variable and these items were summarized into orthogonal explanatory variables in the models (Afghari et al., 2020). The results of the principal component analysis, the eigenvalues and the proportion of explained variability are provided in the Appendix Table B1 and B2.

All models were estimated using the maximum simulated likelihood approach with 500 Halton draws. The required number of Halton draws was selected so that further increasing the number of draws does not change the estimates significantly. Although normal, lognormal, Weibull and uniform distributions were tested as the distributions of the random parameters in all of the models, Normal distributions provided a slightly better model performance compared to the other distributions. The performance of model candidates is assessed from two perspectives: (i) statistical goodness-of-fit, and (ii) parameters estimates.

Table 2
Results of goodness-of-fit measures for hurdle regression models versus right-truncated regression model.

	Random Parameters Hurdle Regression Model		Random Parameters Truncated Lognormal Regression Model
	Beta	Truncated Lognormal	
N (sample size)	469	469	469
P (number of estimated parameters)	21	15	5
Log likelihood of the null model	−156.2	−73.3	428.7
Log likelihood at convergence	−106.5	−62.0	446.2
AIC	255.1	154.2	−882.4
BIC	339.1	214.7	−861.7
MAD	0.163	0.159	0.170
MSPE	0.049	0.046	0.051
MAD_Z	0.075	0.075	0.187
$MSPE_Z$	0.019	0.019	0.037

MAD_Z: Mean absolute deviance for zero fines.
$MSPE_Z$: Mean squared predictive error for zero fines.

4.1. Statistical goodness-of-fit

The results of the statistical fit measures for all models are presented in Table 2. Both of the hurdle models have lower MAD and MSPE compared to the non-hurdle truncated lognormal regression model suggesting that the hurdle setting has resulted in an improved statistical fit in comparison with the non-hurdle setting. The hurdle models have substantially lower MAD and MSPE for the zero values of the dependent variable in comparison with the non-hurdle model. This finding is consistent with the primary goal of the hurdle model and confirms our hypothesis that the zero values of the fine are generated by a separate process than the non-zero values.

Between the two hurdle models, the right-truncated lognormal hurdle model has lower AIC and BIC (154.2 and 214.7, respectively) and lower MAD and MSPE (0.159 and 0.046, respectively) compared to the beta hurdle model. This finding indicates that the right-truncated lognormal distribution has a better overall performance than the beta distribution for the limited dependent variable. However, all of the aforementioned measures of fit (AIC, BIC, MAD and MSPE) reflect the overall goodness-of-fit for a model and do not show the longitudinal model performance (model performance across the values of the dependent variable). Such an overall comparison may not reveal the main strength of the beta distribution in capturing part of the data that are not well captured by the lognormal distribution (the fat-tail below the right limit). As a result and to shed more light on the longitudinal performance of the models, kernel densities of fitted values, absolute and squared residuals of models across the non-zero values of the dependent variable are plotted for all three model candidates (Fig. 1 and Fig. 2).

The density plots in Fig. 1 show that both variants of the hurdle model (the blue and the green curves) are superior to the non-hurdle truncated model (the red curve) in fitting the histogram of the dependent variable. This finding is consistent with the previously shown goodness-of-fit results. The new finding, however, is that the kernel density of the beta hurdle model (the green curve) better fits the histogram of the dependent variable than the kernel density of the truncated lognormal hurdle model (the blue curve), particularly within the tail of the histogram where the proportion of fines is higher than 0.45. The

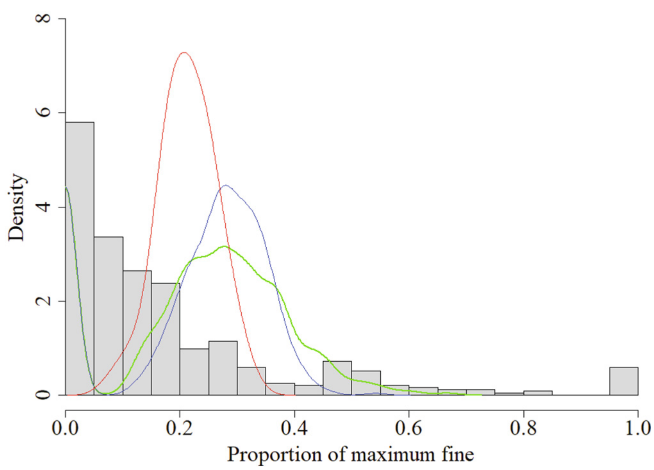

Fig. 1. Kernel density plots of the predictions versus histogram of the dependent variable.

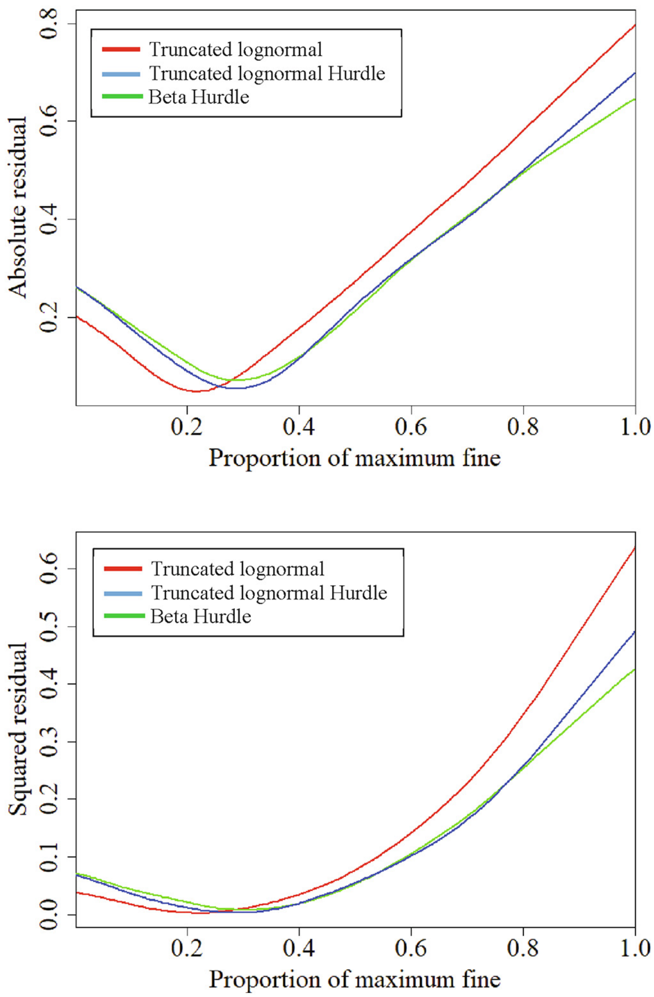

Fig. 2. Residual plots across different values of the dependent variable.

residual plots in Fig. 2 also show that the beta hurdle model results in lower absolute and squared residuals for the values of the dependent variable that are near its right limit. This finding is in line with our hypothesis in that the beta distribution is more flexible to capture the fat-tailed values of the data.

Overall, the results of quantitative measures of fit and visualized residuals indicate that the hurdle setting outperforms the non-hurdle setting and thus the hurdle models are selected as the preferred models in this study. In addition, the choice of the hurdle models over the non-hurdle model is primarily driven by the nature of the dependent variable and its dual data generating process (see Section 1.2).

4.2. Estimated parameters of regression models

It is acknowledged that statistical fit should not be the only criterion for model comparison/selection. As such, the regression results and parameter estimates of all models are also compared in this subsection. These results are shown in Table 3.

The results show that only a few explanatory variables are statistically significant in the non-hurdle right-truncated lognormal model. These variables include violations, attitudes, and positive behaviours. The estimated parameters for these variables indicate that lower scores of violations and higher scores of positive behaviours and attitudes are associated with higher amounts of fine. While the effects of violations and attitudes are fixed across the sample, the effect of positive behaviours varies significantly. These findings are intuitive, but they do not distinguish between factors contributing to the presence/absence of the fine and those contributing to the propensity of the fine.

In contrast, the first layer of the hurdle models provides insight about the factors contributing to the absence/presence of the fine, in particular. Age, education level, violations, attitudes, and positive behaviours are all statistically significant vari-

Table 3
Results of hurdle regression models versus right-truncated regression model.

Explanatory variable	Hurdle models						Right-truncated model	
	First layer (binary fines)		Second layer (positive fines)				Random parameters truncated lognormal	
			Beta		Right-truncated lognormal			
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
Constant	1.606	3.920	–	–	–	–	–	–
Age	0.051 (0.106)	3.850 (7.390)	−0.017	−8.188	−0.027	−8.845	–	–
Gender (male)	–	–	−0.423	−4.205	−0.484	−3.109	–	
Education (university or more)	−0.029 (6.435)	−0.070 (6.650)	–	–	–	–	–	–
Hours walking (more than 5 h)	–	–	−0.370	−2.526	−0.466	−2.254	–	–
Violations – PC1	−0.593	−4.950	−0.172	−4.444	−0.143	−2.574	−0.572	−3.546
Attitudes – PC1	0.635	5.860	0.036 (0.341)	0.786 (6.986)	–	–	0.632	4.532
Positive behaviours – PC1	0.524 (2.370)	4.030 (7.050)	−0.103	−2.273	–	–	0.414 (0.927)	1.985 (2.230)
Perceived ease of use – PC1	–	–	−0.238	−3.531	−0.194	−2.699	–	–
Variance function of beta distribution								
Constant	–	–	1.267	11.269	–	–	–	–
Violations – PC1	–	–	0.224	3.914	–	–	–	–
Positive behaviours – PC1	–	–	0.301	3.372	–	–	–	–
Perceived ease of use – PC1	–	–	0.152	2.711	–	–	–	–
Variance of lognormal distribution	–	–	–	–	1.331	5.260	1.254	32.795

Numbers inside brackets indicate the estimates for the standard deviation of random parameters.
PC1: First principal component.

ables influencing the likelihood of reporting a zero fine for intentionally blocking the path of a fully automated vehicle. While the effects of violations and attitudes are fixed across the sample, the effects of age, education levels and positive behaviours vary significantly. Additionally, the second layer of the hurdle models provides insight about the factors influencing the propensity of the fine among those participants who believe there should be a fine for intentionally blocking the path of a fully automated vehicle. The results of this layer indicate that age, gender, walking hours, violations, and perceived ease of use are significantly associated with the amount of fine. The beta hurdle regression model, in particular, yields statistically significant estimates for the parameters of attitudes and positive behaviours too, with the former having a varied effect across the sample. These findings show that the beta hurdle regression model provides more insight about the effects of explanatory variables on the dependent variable.

Another interesting finding is that the parameters of age and education level are random in the first layer of the hurdle model. The means and standard deviations of these two parameters indicate that their effects are positive for some participants and negative for the others. Interestingly, the parameter of age is negative and the parameter of education level is not statistically significant in the second layer of the hurdle model. This finding is intuitive considering that the first layer is estimated on all data whereas the second layer is estimated on part of the data (only the positive fines). In addition, the parameters of age and education level are not statistically significant in the non-hurdle right-truncated lognormal model suggesting that the positive and negative parameters neutralize one another in this single-layer model, and hence reinforcing the advantage of using the hurdle setting for the data in this study.

As a final note, the lognormal distribution (in hurdle and non-hurdle models) has a fixed estimated variance whereas the beta distribution (in the hurdle model) has a heterogeneous variance linked with statistically significant variables including violations, positive behaviours and perceived ease of use. The parameters of these variables are positive in the variance function of the beta distribution. Interestingly, the same variables have negative parameters in the mean function of the beta distribution. These findings indicate that violations, positive behaviours and perceived ease of use have decreasing effects on the mean but increasing effects on the variance of the reported fines in the sample. These results show the advantage of parameterizing the variance of the beta distribution in capturing the additional heterogeneity in the variance of the dependent variable, which in turn results in a superior statistical fit for a part of the data.

4.3. Model implications

As the results of the goodness-of-fit measures and the parameter estimates do not seem to inform the final model selection between the two variants of the hurdle model, both of the hurdle models are selected in this sub-section for making inferences about the factors associated with the monetary value of pedestrians intentionally blocking the path of fully automated vehicles.

The results of the first layer in these two models indicate that older participants and those participants with higher education levels are, on average, more likely to report that intentionally blocking the path of a fully automated vehicle should be fined or punished. This finding is consistent with the previous findings in the literature indicating that younger and potentially less educated pedestrians are generally overrepresented in groups with more negative attitudes such as disagreeing with policy measures and traffic penalties (Papadimitriou et al., 2013). While our findings suggest that the effects of age and education levels are not homogenous across pedestrians, they imply that younger and less educated pedestrians will be less likely to accept policy measures to prevent bullying behaviour against fully automated vehicles.

Participants who reported more violations are less likely to be in favour of the fine for bullying fully automated vehicles. This finding is also consistent with the previous findings by Papadimitriou et al. (2013) indicating that pedestrians who report higher risk-taking behaviours are more likely to disagree with policy measures and penalties. Similarly, Deb et al. (2017a) found that pedestrians who report higher risky behaviours, whether due to inexperience, stress or aggressiveness, are also more likely to take advantage of fully automated vehicles. These findings imply that risky pedestrians will be less likely to accept legislations imposing a fine to their behaviour. On the contrary, participants engaging more in positive behaviours are more likely to support the fine for intentionally blocking the path of fully automated vehicles. Previous studies have also shown that pedestrians with higher positive behaviours are more concerned with traffic rules and regulations (Deb et al., 2017a).

Another interesting finding from the first layer of the hurdle models is that those participants with more favourable attitudes towards fully automated vehicles are more likely to support the fine for intentionally blocking the path of these vehicles. This finding might indicate that those pedestrians who recognise the potential advantages of fully automated vehicles in terms of safety and efficiency, also appreciate the importance of the rules and regulations that would facilitate the adoption of these vehicles. Past research has also shown that negative attitudes towards automated driving, in general, affect interactions with fully automated vehicles (Reig et al., 2018). However, this is the first time that positive attitudes towards fully automated vehicles have been linked with acceptance of fines targeting pedestrians engaging in bullying behaviours towards vehicles.

The second layer of the hurdle models provides further insight about the factors influencing the propensity of the fine among those participants who believe that there should be a fine for intentionally blocking the path of fully automated vehicles. The results of this layer in both models indicate that age, gender, walking hours, violations, and perceived ease of use are significantly associated with the propensity of the fine. In addition, attitudes and positive behaviours are significantly associated with the propensity of the fine in the beta hurdle model.

More specifically, the results of both variants of the hurdle model show that older participants are more likely to select a lower fine. Note that the first layer showed that this group of participants are more likely to be in favour of the fine. These findings suggest that while older pedestrians may accept the fine as an intervention strategy, they may not approve just any value for the fine. These results also indicate that male participants are more likely to select a lower fine. This finding could be indicative of the higher risk-taking behaviour of males in comparison with females. In an international study of risky walking behaviours, McIlroy et al. (2019) found that males pedestrians are more likely to engage in risky behaviours such as "crossing very slowly to annoy a driver". These findings imply that male pedestrians may not approve higher fines for intentionally blocking fully automated vehicles given they are more likely to engage in risky and aggressive walking behaviour. Furthermore, the results of both models indicate that participants who walk five hours or more per week are more likely to prefer a lower fine, implying that pedestrians with more experience and confidence may not consider bully behaviour as a major problem on the roads. Another possible explanation could be that more experience and exposure to the road environment might result in the perception that fining this behaviour may not be very effective. As this study was conducted in Australia where pedestrians experience a heavy-regulated (King et al., 2009) and car-centric (Flatt and Odinsman, 2015) environment, future research should further explore the reasons behind these findings.

In addition, the results of both models show that those participants who engage in more violations and thus higher risk-taking behaviours are more likely to report a lower fine. This finding is intuitive and implies that this group of pedestrians may not approve high values for the fine. However, the results of the beta hurdle model provides additional insight on the effects of violations on the propensity of the fine, and show that participants who engage in more violations have higher variance in their suggested fines too (in comparison with participants engaging in less violations) indicating that they may belong to heterogeneous groups e.g. groups who under-estimate/over-estimate the fines.

The attitudes towards fully automated vehicles and the positive behaviours are associated with the propensity of the fine in the beta hurdle model, but are not so in the truncated lognormal hurdle model. The results of the former model indicate that attitudes have varied effects on the propensity of fines. While some participants with more favourable attitudes are more likely to report higher fines, others are more likely to be in favour of lower fines. Moreover, the beta hurdle model also indicates that participants who report higher positive behaviours are more likely to prefer a lower fine.

Finally, the results of both models indicate that participants who report higher perceived ease of use are more likely to prefer a lower fine. However, the beta hurdle model again provides additional insight about the effects of perceived ease of use on the propensity of the fine. The result of the variance function of this model indicate that participants who report higher perceived ease of use have also higher variance in their suggested fines, indicating that they have heterogeneous preferences. Overall, these findings clearly show that the current behaviour of pedestrians has a significant impact on the value assigned to bully behaviour towards fully automated vehicles.

5. Conclusions

Intentionally blocking the path of fully automated vehicles is an example of pedestrians' receptivity towards these vehicles which should be investigated prior to their introduction into our urban transport networks. Such an intentional behaviour may represent taking advantage towards fully automated vehicles and may ultimately lead to societal and monetary costs. One way of investigating the monetary value of this behaviour is to empirically ask pedestrians about their perception of the fine associated with this behaviour and econometric modelling of their responses. There are, however, important challenges in modelling this variable including a dual-state data generating process, a truncated and a fat-tailed distribution of the data. This study proposes a beta hurdle regression model to address these challenges.

From the methodological perspective, our findings show that the hurdle setting is superior to the non-hurdle setting for the dual-state of the data and provides insight about the factors associated with the two distinct data generating processes. Moreover, the beta distribution is more flexible than the lognormal distribution for handling truncated data within their fat tail and below their right limit. While the beta distribution is superior to the lognormal distribution only for part of the data in this study, it remains as a canonical distribution for the data bounded between zero and one. In addition, parameterizing the variance of the beta distribution helps decomposing excess variation in the data, captures additional heterogeneity in the data and contributes to revealing the true effects of explanatory variables on the mean of the dependent variable.

From the empirical perspective, our findings suggest that age, gender, education levels, walking experience, attitudes, positive behaviours, violations, and perceived ease of use are all associated with pedestrians' perceived value of intentionally blocking the path of fully automated vehicles. These factors should be accounted for in the development of future policy measures such as new educational tools and road safety programs focused on mitigating maladaptive behaviour among pedestrians. Transport agencies, federal governments and stakeholders should consider these factors when developing adoption strategies that guarantee the safe integration of fully automated vehicles into the urban transport systems. Legislation alone is likely to be ineffective in the prevention of conflicts between pedestrians and fully automated vehicles.

This study is not without limitations. From the methodological perspective, we only compared the performance of the beta distribution with that of the lognormal distribution within the positive state of the proposed hurdle model. Future research should repeat this comparison using other distribution types such as gamma distribution. In addition, we attempted the variance decomposition through exogenous variables only in the beta hurdle model. Future research should consider a similar setting for the truncated lognormal approach to make the comparison more fair. We did not include censored regression models in our study due to the nature of the data, although these models have been widely used in the statistical literature for limited dependent variables. Comparing the performance of the models in this study with that of censored regression models with proper empirical data is thus a worthy research direction. Moreover, we assumed that the effects of explanatory variables on the fine and thus the perceived value of the targeted behaviour is time invariant. However, this assumption may not be totally accurate because of the availability bias in the external factors such as changes in media coverage, social networks and/or global experiences of pedestrians over time (Mannering, 2018). As a result, future research should investigate the temporal instability of pedestrians' receptivity and adaptation towards fully automated vehicles. Finally, we assumed that the behavioural explanatory variables used in this study are exogenous. However, previous research has shown that these variables are usually highly endogenous (Afghari et al., 2018) and may influence one another. Future research should further explore this endogeneity using proper methodological approaches such as latent variable approaches and simultaneous equation models.

From the empirical perspective, potential bias may exist in the collected sample due to the unitary online data collection (convenience sampling) method and the data collection source. As such, the results should be interpreted with caution. A diverse data collection approach such as paper survey dissemination or onsite data collection is recommended in the future studies to increase the representativeness of the sample. In addition, we provided generic and high-level information about fully automated vehicles at the beginning of the survey. However, a better briefing of the respondents, especially for the targeted behaviour in this study where there is lack of public exposure, could help them provide a more informed answer about the value of the fine. Considering that it is quite difficult to provide detailed examples/references for such a behaviour on the roads, future research should explore alternative ways to improve participants' briefing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was funded by the Queensland University of Technology's (QUT) Institute of Health and Biomedical Innovation (IHBI) Injury Prevention Program Seeding Grant, Delft University of Technology, and the Australian Government's Research

Council Discovery Early Career Researcher Award [DE200101079]. The authors would also like to acknowledge the insightful comments by the anonymous reviewers which helped us improve the quality of this manuscript.

Author Statement

A.P. Afghari, E. Papadimitriou, O. Oviedo-Trespalacios: study conception and design; X. Li, S-A. Kaye, O. Oviedo-Trespalacios, A.P. Afghari: data collection; A.P. Afghari, E. Papadimitriou, O. Oviedo-Trespalacios: analysis and interpretation of results; A.P. Afghari, O. Oviedo-Trespalacios, E. Papadimitriou, S-A. Kaye, X. Li: draft manuscript preparation. All authors reviewed the results and approved the final version of the manuscript.

Appendix A. Pedestrian questionnaire

Section 1 Demographic questions

1. What is your age?
2. What is your gender? [ ] Male [ ] Female
3. What is the highest level of education that you have completed?

[ ] Less than Year 12
[ ] Completed Year 12
[ ] Certificate or Diploma
[ ] Bachelor's Degree
[ ] Master's Degree or higher
[ ] Other (please specify)

4. How many hours do you spend walking on footpaths next to roads in an average week? (Please input the number of hours)
5. How many times have you been hit or nearly hit by a car as a pedestrian in the past 2 years?

Section 2 Pedestrian behaviour questionnaire

The following questions relate to your behaviour on roads as a pedestrian. Please estimate how often you did the following when walking over the past two years.

Violations

[V1] I cross the street even though the pedestrian light is red.
[V2] I cross diagonally to save time.
[V3] I cross outside the pedestrian crossing even if there is one (crosswalk) less than 50 m away.
[V4] I take passageways forbidden to pedestrians to save time.

Errors

[E1] I cross between vehicles stopped on the roadway in traffic jams.
[E2] I cross even if vehicles are coming because I think they will stop for me.
[E3] I walk on cycling paths when I could walk on the sidewalk.
[E4] I run across the street without looking because I am in a hurry.

Lapses

[L1] I realize that I have crossed several streets and intersections without paying attention to traffic.
[L2] I forget to look before crossing because I am thinking about something else.
[L3] I cross without looking because I am talking with someone.
[L4] I forget to look before crossing because I want to join someone on the sidewalk on the other side.

Aggressive behaviours

[A1] I get angry with another road user (pedestrian, driver, cyclist, etc.) and I yell at them.
[A2] I cross very slowly to annoy a driver.
[A3] I get angry with another road user (pedestrian, driver, cyclist, etc.) and I make a hand gesture.

[A4] I have gotten angry with a driver and hit their vehicle.

Positive behaviours

[P1] I thank a driver who stops to let me cross.
[P2] When I am accompanied by other pedestrians, I walk in single file on narrow sidewalks so as not to bother the pedestrians I meet.
[P3] I walk on the left-hand (or right-hand) side of the sidewalk so as not to bother the pedestrians I meet.
[P4] I let a car go by, even if I have the right-of-way, if there is no other vehicle behind it.

Section 3 Pedestrian Receptivity Questionnaire

Before today, have you heard of the term "automated vehicle"? [] Yes [] No

Automated vehicles can be categorised as Level 0 (no automation) through to Level 5 (full automation). The focus of this survey is on Level 5: Fully Automated Vehicles (FAVs). Please read the following information on Level 5: Fully Automated Vehicles before continuing with the survey.

Level 5: Fully Automated Vehicle

A FAV is driven by technology instead of by a human. A FAV is equipped with radars, cameras, and sensors which can detect the presence, position, and speed of other vehicles or road-users. With this information, the FAV can then respond as needed by stopping, decelerating and/or changing direction. A FAV has the potential to reduce road crashes and to decrease the possibility of severe injuries by controlling the driving task effectively. The following questions relate to your beliefs and attitude as a pedestrian towards sharing the road with (Level 5) fully FAVs. As a pedestrian, please indicate on the scale provided the extent to which you agree or disagree with each statement about sharing roads with fully automated vehicles in the future, when they are commonly available. It does not matter if you feel you do not know too much about (Level 5) fully automated vehicles, it is your opinion that we are interested in.

Attitudes

As a pedestrian, I would consider crossing the roads in the presence of FAV:
[AT1] Unpleasant [1] [2] [3] [4] [5] [6] [7] Pleasant
[AT2] Unfavourable [1] [2] [3] [4] [5] [6] [7] Favourable
[AT3] Unsafe [1] [2] [3] [4] [5] [6] [7] Safe
[AT4] Negative [1] [2] [3] [4] [5] [6] [7] Positive

Trust

[T1] I would feel comfortable if my child crosses roads in the presence of FAVs.
[T2] I would feel comfortable if my spouse or partner crosses roads in the presence of FAVs.
[T3] I would feel comfortable if my parent(s) crosses roads in the presence of FAVs.
[T4] I would recommend my family and friends to be comfortable while crossing roads in front of FAVs.
[T5] I would feel more comfortable doing other thing (e.g., checking emails on my smartphone, talking to my companions) while crossing the road in front of FAVs.

Perceived Ease of Use

[PEU1] My interaction with FAVs while crossing the road would be clear and understandable.

[PEU2] I would find FAVs difficult to interact while crossing the road (reverse-scaled item).

Section 4 Fine of blocking the path of FAV

Do you think that pedestrians intentionally blocking the path of a FAV should be fined or punished?

[ ] Yes   [ ] No

If yes, how much should be the fine? (from 1 Australian dollars to 1000 Australian dollars)

Appendix B. Results of Principal Component Analysis (PCA)

Table B1
Factor loadings of questionnaire items for the first principal components in the PCA.

Items	Violations	Errors	Laps	Aggressions	Positive behaviour	Trust	Perceived ease of use	Attitudes
V1	0.527							
V2	0.470							
V3	0.524							
V4	0.476							
E1		−0.372						
E2		−0.556						
E3		−0.536						
E4		−0.515						
L1			−0.497					
L2			−0.518					
L3			−0.495					
L4			−0.490					
A1				−0.477				
A2				−0.501				
A3				−0.513				
A4				−0.508				
P1					0.479			
P2					0.530			
P3					0.508			
P4					0.480			
T1						−0.450		
T2						−0.286		
T3						−0.476		
T4						−0.498		
T5						−0.491		
PEU1							0.707	
PEU2							0.707	
AT1								0.494
AT2								0.505
AT3								0.495
AT4								0.506

Table B2
Eigenvalues and proportion of explained variability by the first principal components in the PCA.

	Violations	Errors	Laps	Aggressions	Positive behaviour	Trust	Perceived ease of use	Attitudes
Eigenvalue	2.668	2.322	3.194	2.528	1.913	3.453	1.544	3.504
Proportion of variance in data explained by the first principal component	0.667	0.581	0.799	0.639	0.478	0.691	0.772	0.876

References

Afghari, A. P., Faghih Imani, A., Papadimitriou, E., van Gelder, P., Hezaveh, A. M. 2021. "I didn't wear it, because nobody else did": A Bayesian latent variable multivariate model of seatbelt use choices in multi-occupant vehicles. Journal of Choice Modelling, Forthcoming in.

Afghari, A.P., Washington, S., Haque, M.M., Li, Z., 2018. A comprehensive joint econometric model of motor vehicle crashes arising from multiple sources of risk. Analytic Methods in Accident Research 18, 1–14.

Afghari, A.P., Washington, S., Prato, C., Haque, M.M., 2019. Contrasting case-wise deletion with multiple imputation and latent variable approaches to dealing with missing observations in count regression models. Analytic Methods in Accident Research 24, 100104.

Aguero-Valverde, J., 2013. Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates. Accident Analysis and Prevention 50, 289–297.

Ahmed, S.S., Pantangi, S.S., Eker, U., Fountas, G., Still, S.E., Anastasopoulos, P.C., 2020. Analysis of safety benefits and security concerns from the use of autontomous vehicles: A grouped random parameters bivariate probit approach with heterogeneity in means. Analytic Methods in Accident Research 28, 100134.

Ajzen, I., 1991. The theory of planned behavior. Organizational Behavior and Human Decision Processes 50 (2), 179–211.

Bansal, P., Kockelman, K.M., Singh, A., 2016. Assessing public opinions of and interest in new vehicle technologies: An Austin perspective. Transportation Research Part C 67, 1–14.

Bhat, C.R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. Transportation Research Part B 35 (7), 677–693.

Buckley, L., Kaye, S.-A., Pradhan, A.K., 2018. Psychosocial factors associated with intended use of automated vehicles: A simulated driving study. Accident Analysis and Prevention 115, 202–208.

Cai, Q., Lee, J., Eluru, N., Abdel-Aty, M., 2016. Macro-level pedestrian and bicycle crash analysis: Incorporating spatial spillover effects in dual state count models. Accident Analysis and Prevention 93, 14–22.

Chen, F., Chen, S., Ma, X., 2016. Crash frequency modeling using real-time environmental and traffic data and unbalanced panel data models. International Journal of Environmental Research and Public Health 13 (6), 609.

Crow, E.L., Shimizu, K., 1987. Lognormal Distributions. Marcel Dekker, New York.

Davis, F.D., 1985. A Technology Acceptance Model for Empirically Testing New End-user Information Systems: Theory and Results Doctoral dissertation. Massachusetts Institute of Technology.

Davis, F.D., Bagozzi, R.P., Warshaw, P.R., 1989. User acceptance of computer technology, journal of management science. Journal of Management Science 35 (8), 982–1003.

Davis, F.D., Venkatesh, V., 1996. A critical assessment of potential measurement biases in the technology acceptance model: three experiments. International Journal of Human-Computer Studies 45 (1), 19–45.

Deb, S., Strawderman, L., Carruth, D.W., DuBien, J., Smith, B., Garrison, T.M., 2017a. Development and validation of a questionnaire to assess pedestrian receptivity toward fully autonomous vehicles. Transportation Research Part C 84, 178–195.

Deb, S., Strawderman, L., DuBien, J., Smith, B., Carruth, D.W., Garrison, T.M., 2017b. Evaluating pedestrian behavior at crosswalks: Validation of a pedestrian behavior questionnaire for the US population. Accident Analysis and Prevention 106, 191–201.

Dong, C., Clarke, D.B., Yan, X., Khattak, A., Huang, B., 2014. Multivariate random-parameters zero-inflated negative binomial regression model: An application to estimate crash frequencies at intersections. Accident Analysis and Prevention 70, 320–329.

Embrechts, P., Klüppelberg, C., Mikosch, T. 2013. Modelling Extremal Events: For Insurance and Finance, Springer Science & Business Media, Vol. 33.

Flatt, L., Odinsman, J. 2015. The Institutionalization of Cycling–A Case Study of Sydney and Melbourne. Master's Thesis, Lund University. https://lup.lub.lu.se/student-papers/search/publication/7359696.

Ghazizadeh, M., Lee, J.D., Boyle, L.N., 2012. Extending the Technology Acceptance Model to assess automation. Cognition, Technology & Work 14 (1), 39–49.

Greene, W.H., 2003. Econometric Analysis. Prentice Hall, Upper Saddle River, New Jersey.

Gupta, A.K., Nadarajah, S., 2004. Handbook of Beta Distribution and its Applications. Marcel Dekker, NY, USA.

Gurmu, S., 1997. Semi-parametric estimation of hurdle regression models with an application to Medicaid utilization. Journal of Applied Econometrics 12 (3), 225–242.

Haboucha, C.J., Ishaq, R., Shiftan, Y., 2017. User preferences regarding autonomous vehicles. Transportation Research Part C 78, 37–49.

Henry, R., Hidy, G. 1979. Multivariate analysis of particulate sulfate and other air quality variables by principal components-Part I: Annual data from Los Angeles and New York. Atmospheric Environment (1967), 13(11), 1581-1596.

Hong, J., Tamakloe, R., Park, D., 2019. A comprehensive analysis of multi-vehicle crashes on expressways: a double hurdle approach. Sustainability 11 (10), 2782.

Huitema, B.E., 1986. Autocorrelation in Behavioral Research. Research Methods in Applied Behavior Analysis. Springer.

Huitema, B.E., McKean, J.W., 1991. Autocorrelation estimation and inference with small samples. Psychological Bulletin 110 (2), 291.

Jiang, X., Huang, B., Zaretzki, R.L., Richards, S., Yan, X., Zhang, H., 2013. Investigating the influence of curbs on single-vehicle crash injury severity utilizing zero-inflated ordered probit models. Accident Analysis and Prevention 57, 55–66.

Johnson, N.L., Kotz, S., Balakrishnan, N., 1995. Continuous Univariate Distributions. John Wiley & Sons Ltd, NJ, USA.

Katrakazas, C., Theofilatos, A., Islam, M.A., Papadimitriou, E., Dimitriou, L., Antoniou, C., 2021. Prediction of rear-end conflict frequency using multiple-location traffic parameters. Accident Analysis and Prevention 152, 106007.

Kaye, S.A., Lewis, I., Buckley, L., Rakotonirainy, A., 2020a. Assessing the feasibility of the theory of planned behaviour in predicting drivers' intentions to operate conditional and full automated vehicles. Transportation Research Part F 74, 173–183.

Kaye, S.-A., Lewis, I., Forward, S., Delhomme, P., 2020b. A priori acceptance of highly automated cars in Australia, France, and Sweden: A theoretically-informed investigation guided by the TPB and UTAUT. Accident Analysis and Prevention 137, 105441.

Kaye, S.-A., Li, X., Oviedo-Trespalacios, O., Afghari, A.P. 2021. Getting in the path of the robot: Pedestrians acceptance of crossing roads near fully automated vehicles. Travel Behaviour and Society, Forthcoming in.

King, M.J., Soole, D., Ghafourian, A., 2009. Illegal pedestrian crossing at signalised intersections: incidence and relative risk. Accident Analysis and Prevention 41 (3), 485–490.

Koul, H., Susarla, V.V., Van Ryzin, J., 1981. Regression analysis with randomly right-censored data. The Annals of Statistics 9 (6), 1276–1288.

Liu, P., Du, Y., Wang, L., Da Young, J., 2020. Ready to bully automated vehicles on public roads? Accident Analysis and Prevention 137, 105457.

Lord, D., Washington, S., Ivan, J.N., 2007. Further notes on the application of zero-inflated models in highway safety. Accident Analysis and Prevention 39 (1), 53–57.

Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Analysis and Prevention 37 (1), 35–46.

Ma, L., Yan, X., Wei, C., Wang, J., 2016. Modeling the equivalent property damage only crash rate for road segments using the hurdle regression framework. Analytic Methods in Accident Research 11, 48–61.

Ma, L., Zhang, S., Yan, X., Wei, C., 2019. A hurdle finite mixture lognormal crash rate estimation model for addressing heterogeneous characteristics of influential factors. Journal of Transportation Safety and Security 11, 443–463.

Madigan, R., Louw, T., Wilbrink, M., Schieben, A., Merat, N., 2017. What influences the decision to use automated public transport? Using UTAUT to understand public acceptance of automated road transport systems. Transportation Research Part F 50, 55–64.

Malyshkina, N.V., Mannering, F.L., 2009. Markov switching multinomial logit model: An application to accident-injury severities. Accident Analysis and Prevention 41 (4), 829–838.

Malyshkina, N.V., Mannering, F.L., 2010. Zero-state Markov switching count-data models: An empirical assessment. Accident Analysis and Prevention 42 (1), 122–130.
Mannering, F., 2018. Temporal instability and the analysis of highway accident data. Analytic Methods in Accident Research 17, 1–13.
McIlroy, R.C., Plant, K.L., Jikyong, U., Nam, V.H., Bunyasi, B., Kokwaro, G.O., Wu, J., Hoque, M.S., Preston, J.M., Stanton, N.A., 2019. Vulnerable road users in low-, middle-, and high-income countries: validation of a pedestrian behaviour questionnaire. Accident Analysis and Prevention 131, 80–94.
Millard-Ball, A., 2018. Pedestrians, autonomous vehicles, and cities. Journal of Planning Education and Research 38, 6–12.
Moody, J., Bailey, N., Zhao, J., 2020. Public perceptions of autonomous vehicle safety: An international comparison. Safety Science 121, 634–650.
Noy, I.Y., Shinar, D., Horrey, W.J., 2018. Automated driving: Safety blind spots. Safety Science 102, 68–78.
Oviedo-Trespalacios, O., Afghari, A.P., Haque, M.M., 2020. A hierarchical Bayesian multivariate ordered model of distracted drivers' decision to initiate risk-compensating behaviour. Analytic Methods in Accident Research 26, 100121.
Oviedo-Trespalacios, O., Watson, B. 2021. Navigation apps are becoming a threat to road safety (beyond distraction). Injury Prevention, 27(2), 103-103.
Palmeiro, A.R., van der Kint, S., Vissers, L., Farah, H., de Winter, J.C., Hagenzieker, M., 2018. Interaction between pedestrians and automated vehicles: A Wizard of Oz experiment. Transportation Research Part F 58, 1005–1020.
Papadimitriou, E., Theofilatos, A., Yannis, G., 2013. Patterns of pedestrian attitudes, perceptions and behaviour in Europe. Safety Science 53, 114–122.
Razmi Rad, S., de Almeida Correia, G.H., Hagenzieker, M., 2020. Pedestrians' road crossing behaviour in front of automated vehicles: Results from a pedestrian simulation experiment using agent-based modelling. Transportation Research Part F 69, 101–119.
Ready, R.C., Hu, D., 1995. Statistical approaches to the fat tail problem for dichotomous choice contingent valuation. Land Economics, 491–499.
Reig, S., Norman, S., Morales, C.G., Das, S., Steinfeld, A., Forlizzi, J., 2018. A field study of pedestrians and autonomous vehicles. In: Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pp. 198–209.
Rothenbücher, D., Li, J., Sirkin, D., Mok, B., Ju, W., 2016. Ghost driver: A field study investigating the interaction between pedestrians and driverless vehicles. In: The 25th IEEE International Symposium on Robot and Human Interactive Communication. IEEE, pp. 795–802.
Sahebi, S., Nassiri, H., Van Wee, B., Araghi, Y., 2019. Incorporating car owner preferences for the introduction of economic incentives for speed limit enforcement. Transportation Research Part F 64, 509–521.
Sheela, P.V., Mannering, F., 2020. The effect of information on changing opinions towards autonomous vehicle adoption: An exploratory analysis. International Journal of Sustainable Transportation 14, 475–487.
Tipping, M.E., Bishop, C.M., 1999. Probabilistic principal component analysis. Journal of the Royal Statistical Society 61 (3), 611–622.
Van Loon, R.J., Martens, M.H., 2015. Automated driving and its effect on the safety ecosystem: How do compatibility issues affect the transition period? Procedia Manufacturing 3, 3280–3285.
Velasco, J.P.N., Farah, H., van Arem, B., Hagenzieker, M.P., 2019. Studying pedestrians' crossing behavior when interacting with automated vehicles using virtual reality. Transportation Research Part F 66, 1–14.
Venkatesh, V., Thong, J.Y., Xu, X., 2012. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Quarterly, 157–178.
Wang, Y., Dong, W., Zhang, L., Chin, D., Papageorgiou, M., Rose, G., Young, W., 2012. Speed modeling and travel time estimation based on truncated normal and lognormal distributions. Transportation Research Records 2315 (1), 66–72.
Washington, S., Karlaftis, M.G., Mannering, F., Anastasopoulos, P., 2020. Statistical and Econometric Methods for Transportation Data Analysis. Chapman and Hall/CRC, Boca Raton, FL.
Weng, J., Yan, X., 2016. Probability distribution-based model for work zone capacity prediction. Journal of Advanced Transportation 50, 165–179.
Woldeamanuel, M., Nguyen, D., 2018. Perceived benefits and concerns of autonomous vehicles: An exploratory study of millennials' sentiments of an emerging market. Research in Transportation Economics 71, 44–53.
Xiong, Y., Tobias, J.L., Mannering, F.L., 2014. The analysis of vehicle crash injury-severity data: A Markov switching approach with road-segment heterogeneity. Transportation Research Part B 67, 109–128.
Yu, R., Wang, Y., Quddus, M., Li, J., 2019. A marginalized random effects hurdle negative binomial model for analyzing refined-scale crash frequency data. Analytic Methods in Accident Research 22, 100092.