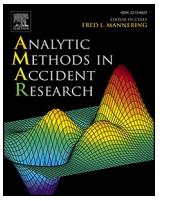
Analytic Methods in Accident Research 41 (2024) 100318

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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



Estimating the effect of proximity to school on cyclist safety using a simultaneous-equations model with heterogeneity in covariance to address potential endogeneity

Shahram Heydari *, Michael Forrest

Transportation Research Group, Department of Civil, Maritime and Environmental Engineering, University of Southampton, Southampton, UK

ARTICLE INFO

Keywords:
Cyclist injury
School siting
Endogeneity
System-equation modelling
Scale mixing
Safety-in-numbers

ABSTRACT

Traffic safety around schools is a major concern for policy makers and as such safety interventions are often targeted near schools. This paper shows the importance of accounting for the potential endogeneity of proximity to school when attempting to estimate its impact on traffic safety. In this research, we use a Bayesian simultaneous econometric approach with heterogeneity in covariance to disentangle the true effect of proximity to school on cyclist injury frequencies at signalised intersections in an urban setting. We assess the robustness of the bivariate normal assumption, using a scale mixing approach. Notably, we found that proximity to school was associated with an increase in cyclist injuries and this association was stronger when endogeneity was accounted for in the model, confirming the importance of considering endogeneity in studies of traffic safety near schools. Our heterogeneity in covariance specification revealed systematic variations in the covariance structure, which would otherwise go unobserved, providing further insights into sources of heterogeneity with the same set of variables available in the data. A safety-in-numbers effect is also found for cyclists in the study area and period. This research offers policy implications based on the findings of the analysis including the need for safety interventions at intersections with high vehicle turning counts and those in proximity to public transport stops, and better informing decision-makers regarding the magnitude of the impact of proximity to school on cyclist safety at intersections.

1. Introduction

Cycling to school is encouraged by most governmental bodies due to the health benefits that come with travelling actively, as well as the positive aspects of reducing car trips for pollution and congestion. On the other hand, the safety of cyclists on the road is a key factor that needs to be considered and advocated. Owing to the lack of physical protection that a cyclist has in the event of a collision, injury rates relative to distance travelled are notably higher for cyclists than motor vehicles and concerns over safety are a barrier to cycling uptake. Furthermore, children have been noted to be susceptible to being involved in traffic crashes due to their underdeveloped cognitive abilities as well as certain behavioural traits. In response to this perceived and observable danger, school areas are often targeted for safety interventions including cycling specific features such as cycle lanes, advanced stop lines and low traffic neighbourhoods.

* Corresponding author.
E-mail addresses: s.heydari@soton.ac.uk (S. Heydari), mgf1g18@soton.ac.uk (M. Forrest).

https://doi.org/10.1016/j.amar.2024.100318
Received 14 April 2023; Received in revised form 23 January 2024; Accepted 24 January 2024
Available online 1 February 2024
2213-6657/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1.1. Previous research

The safety of cyclists at intersections is the subject of a body of research as this is where a high proportion of cyclist crashes occur, 68 % in Great Britain in 2020 for instance (DfT, 2021). The nature of turning vehicles creates more conflicts as, by necessity, more than one road user attempts to use the same space. Many factors have been shown to be significantly associated with cyclist crash frequency at intersections both at an individual level such as gender (Stipancic et al., 2016) and at intersection level such as the type of roads intersecting (Harris et al., 2013). Cycle safety analyses generally draw inferences at either a macroscopic (Osama et al., 2017; Guo et al., 2018) or microscopic (Behnood et al., 2017; Heydari et al., 2017; Fountas et al., 2021) level.

The factors that would generally be assumed to have an effect on safety at an intersection in proximity to a school as opposed to an intersection not in the presence of a school would include: the demographics of the road users and their related behaviours such as distracted children; and the safety features that might be in place at these intersections such as reduced speed limits. The extent to which proximity to a school or the number of schools in an area is associated with pedestrian safety has been studied as part of a number of investigations (Ukkusuri et al., 2012; Bhat et al., 2017; Heydari et al., 2020). However, exploring cyclist safety in relation to school siting is relatively limited in the literature.

Several papers have identified the effect which the volume of children has on crashes (Amoh-Gyimah et al., 2016; Bhat et al., 2017), and the proportion of children is naturally a key factor of intersections in proximity to a school. There is uncertainty regarding the impact of a greater volume of children on safety. On the one hand, children exhibit behavioural traits such as recklessness and inattentiveness that are often found to be associated with a higher crash likelihood (Gitelman et al., 2019); however, some research shows that in certain circumstances other road users take more care around children, which leads to increased safety (Johansson and Leden, 2007). Safety improvement programmes are also often implemented around schools. These are deliberate policy interventions and thus not random, leading to the self-selection problem (Cao et al., 2009). In general, safety improvement programmes have been shown to have a positive effect on walking and cycling safety (Boarnet et al., 2005; Ragland et al., 2014), but there is also the potential for risk compensation (Winston et al., 2006) from road users who assume an increased safety condition exists in such locations and respond by being less attentive in these areas. The issue with estimating the "true" effect of proximity to schools on cyclist safety is that factors which have an association with school siting often themselves have an association with cycling safety; simultaneously, the presence of school has a bearing on cyclist safety. Therefore, issues of confounding would arise within any analysis that seeks to estimate the effect of proximity to school on traffic safety, including cyclist injury frequencies. Also, self-selective sampling, as discussed in Islam et al. (2023), may play a role here. For example, risk taking drivers may avoid intersections in proximity to schools. Therefore, the selective sample of cyclists and motorists in school areas would be distinct from cyclists and/or motorists in other areas in the road network. Due to the abovementioned complexities, treating proximity to school as an exogenous variable may result in misleading inferences. Therefore, statistical methods that accommodate endogenous variables should be considered.

Endogeneity in traffic safety analysis has been explored in the safety literature (Eluru and Bhat, 2007; Oh and Shankar, 2011; Shin and Shankar, 2013; Bhat et al., 2014; Heydari et al., 2020), but it remains an often overlooked factor in safety research despite the fact that biased inferences can be made when failing to account for endogeneity (Mannering et al., 2020). For a review of issues of endogeneity in the safety literature, see Mannering et al. (2020). Despite the importance of traffic safety nearby schools, to our knowledge only one study, Heydari et al. (2020), investigated the issue of endogeneity relating to the presence of a school nearby intersections. The authors investigated this issue in the context of pedestrian safety. In the latter study, the authors found that the presence of a school had a negative effect on pedestrian safety at intersections over the study period, but this was only the case when endogeneity was accounted for.

With respect to theoretical aspects of modelling crash datasets, beyond traditional models that assume a homogeneous dependence structure when modelling correlated outcomes simultaneously, success has been found in modelling multivariate outcomes using heterogeneous covariance structures (Nashad et al., 2016; Heydari et al., 2017; Yasmin et al., 2018). The latter was found to improve model fit and address unobserved heterogeneity more fully. In traffic safety research, heterogeneous covariance structures are however applied rarely and mostly in multivariate settings (modelling the safety of different road user types or differing injury-severity levels jointly).

1.2. Current article

This paper contributes to the road safety literature by handling the endogenous effect of proximity to school on cyclist safety for the first time in the extant traffic safety literature, revealing its true (unconfounded) effect. We show how overlooking endogeneity in this context would lead to the underestimation of the impact of proximity to school, an important risk factor, on cyclist safety. To this end, we use a system-equation (simultaneous-equations) approach, which is considered to be one of the most comprehensive ways of dealing with endogeneity as it explicitly accounts for the correlation between equations representing the main outcome and the endogenous variable of interest (Mannering et al., 2020; Washington et al., 2020; Heydari et al., 2020).

Previously, the simultaneous-equations approach has been applied in traffic safety research. In this paper for our simultaneous-equations model, we specify a heterogeneous covariance structure that allows the covariance to vary across the sample systematically. This approach is particularly appealing as it allows us to account for unobserved heterogeneity (Mannering et al., 2016; Heydari, 2018) more fully when addressing endogeneity using simultaneous-equations models. Our idea is inspired by the heterogeneity in mean/variance approach that models the mean and/or the variance of random parameters as a function of explanatory variables available in the data, thereby providing further empirical insights into sources of heterogeneity in a crash dataset. We employ a Bayesian scale mixing approach to examine robustness regarding the bivariate normal assumption in the covariance structure. We

develop our model under the Bayesian framework and note that the relative merits of Bayesian methods for addressing endogeneity have rarely been explored in the crash literature. In fact, Bayesian methods, relying on simulation-based approaches, can easily handle non-standard specifications. In this regard, for example, Heydari et al. (2020) specified a heterogeneity in mean/variance simultaneous-equations model.

We consider a comprehensive intersection dataset from Montreal that includes a number of explanatory variables based on exposure, built environment, and geometric intersection characteristics. An important feature of the data is that exact cyclist volume and motorised traffic flow by differing turning movements are available. The latter allows us to distinguish between the effect of right-turning, left-turning, and non-turning traffic volume for cyclist safety. This study is expected to provide a more reliable estimation of the effect of proximity to school on cyclist safety at intersections since we not only account for the endogeneity of this variable, but also address unobserved heterogeneity relative to the covariance structure, thereby better informing decision-makers of the risk associated with the presence of a school nearby an intersection.

2. Material and methods

2.1. Data

The data for this research were taken from Montreal, Canada and comprise cyclist injury counts from 647 signalised intersections over the period 2003–2008. Injury counts were included if they occurred within 15 m of an intersection centre point. The injury counts were sourced from the Montreal ambulance services and represent a fairly comprehensive count of individuals who were involved in a crash while riding a bicycle. Data on differing injury severity levels were not available and so are not used in this analysis. Also, included in the analysis were a number of variables relating to exposure of bicycles and motor vehicles at the intersections; built environment factors of the area surrounding the intersections; and geometric measures of the intersections themselves. For brevity, exhaustive details of the data collection and collation process are not included here, but the reader is referred to Strauss et al. (2014) for more details on the acquisition and presentation of this dataset.

Data on exposure of both cyclists and motorised traffic were gathered manually at intersections around Montreal in 2008 and 2009 and as such are assumed to be representative of the counts for the years studied in this analysis. Built environment data goes some way to establishing the nature of the surroundings of an intersection which is useful for comparative purposes. These data were taken from a range of sources including: Statistics Canada, DMTI Spatial Inc., Société de transport de Montréal (STM) and Agence métropolitaine de transport. Finally geometric characteristics were also obtained manually through site visits in the years 2010–2012 by a team from McGill University. Similar to the exposure data, even though this information was gathered in the time following the study period, it was deemed that there would not have been any sufficient change to the intersections in this time to warrant any loss of significance. Summary statistics of the data are reported in Table 1.

2.2. Methodological basis

To conduct the analysis, it was necessary to develop two equations simultaneously, in a way that two correlated outcomes could be

Table 1
Summary statistics of the considered variables.

Variable types	Variables	Mean	Std. Dev.	Min.	Max.
Crash type	Cyclist injury counts	0.63	1.32	0.00	20.00
Exposure measure	Cyclist counts	444.92	717.62	1.66	6433.22
	Total non-motorised volume	2022.99	3792.45	2.96	41541.00
	Left-turning motorised volume	2602.72	2641.86	0.00	23843.00
	Right-turning motorised volume	2668.01	2697.45	0.00	23792.00
	Non-turning motorised volume	19467.96	11084.39	1790.00	76525.00
	Total motorised volume (AADT)	24738.65	12526.06	3751.00	84386.00
	Ratio of non-motorised to motorised	0.13	0.30	0.00	4.01
Built environment	Employment[1] (in 0000 s)	0.58	0.30	0.03	1.49
	Commercial area[1] (in 0000 s m^2)	1.21	1.34	0.00	8.70
	Land use mix[2]	0.67	0.15	0.00	0.92
	Length of cycling facilities[2] (km)	0.54	0.61	0.00	2.96
	Number of schools[2]	1.05	1.24	0.00	6.00
	Number of subway stations[2]	0.44	0.70	0.00	4.00
	Presence of bus stop[3]	0.71	0.46	0.00	1.00
	Presence of school[2]	0.56	0.50	0.00	1.00
	Presence of subway stations[2]	0.34	0.48	0.00	1.00
Geometric/operational	Maximum speed (km/h)	61.82	9.88	50.00	100.00
	Dedicated traffic light for pedestrians	0.25	0.43	0.00	1.00
	Three-leg intersection	0.22	0.42	0.00	1.00
	Presence of raised median	0.48	0.50	0.00	1.00
	Total number of lanes	6.87	2.63	3.00	16.00

(1) buffer of 800 m; (2) buffer of 400 m; (3)buffer of 50 m.

modelled together. One outcome is the propensity for the presence of a school nearby a given intersection (i.e., the endogenous variable of interest) and the other is cyclist injury frequency, both of which are informed by a host of exposure, built environment, and geometric variables. Furthermore, by modelling these two outcomes simultaneously, the injury frequency model is informed by the result of the presence of school model, as proximity to a school may itself have an effect on safety. Such an approach makes it possible to investigate potential endogeneity in the effect of a variable on an outcome of interest (Kozumi, 2002; Heydari et al., 2020; Mannering et al., 2020).

Our models were constructed using the following theoretical framework that is based on the work of Terza (1998) and Kozumi (2002). The presence of a school in proximity to an intersection is modelled as a binary outcome S and is associated with the latent variable s^* that is a function of explanatory variables $\boldsymbol{Z}$ and an error term ε as shown in (1). If s^* is greater than zero then S is equal to one and otherwise is equal to zero as shown in (2). The latent variable s^* is a measure of the propensity of the presence of a school in proximity to a given intersection and it is not directly observable.

$$S = g(\boldsymbol{Z}, \varepsilon) \tag{1}$$

$$S = \begin{cases} 1 & if \quad s^* > 0 \\ 0 & if \quad s^* \leq 0 \end{cases} \tag{2}$$

To model the injury frequency, it can be established that the cyclist injury count y is a function of site characteristics $\boldsymbol{X}$, school proximity S, and an error term ξ as shown in (3).

$$y = f(\boldsymbol{X}, S, \xi) \tag{3}$$

Multiple factors are likely to be affecting both y and S at the same time; and therefore, it is possible that the unobservable factors that comprise the error term ξ have an effect on S as well as y at the same time. This said, it can be seen that only considering inferences from (3) could be misleading since there could be a correlation between ξ and S, which is not taken into account in (3). To attempt to account for this correlation, y and S can be modelled jointly. Fig. 1 shows the potential relationships between the intersection level factors and the presence of a school and cyclist injury frequency, respectively, as well as the potential for a relationship between the presence of a school and cyclist injury frequency.

2.3. Simultaneous-equations model

For the injury frequency element of the simultaneous-equations, let y_i represent the observed cyclist injury frequency and λ_i be the expected cyclist injury frequency with subscript i denoting a given intersection ($i = 1, 2, \ldots, n$). Let $\boldsymbol{\beta}$ be regression coefficients, including a constant term, which correspond to the site characteristic variables $\boldsymbol{X}$. Let γ be a fixed parameter associated with the school proximity S. Note that here we intentionally separate S and $\boldsymbol{X}$ (while these are all site characteristics) as the focus of our study is on estimating the effect of S. Equation (4) shows the form that the cyclist injury frequency model takes, given the above notation.

$$y_i \sim \quad Poisson\,(\lambda_i)$$

$$log(\lambda_i) = \boldsymbol{\beta X_i} + \gamma S_i + \xi_i \tag{4}$$

As well as the injury frequency element, an equation needs to be defined to model the latent variables s^* as described in the previous section. For modelling the presence of school, a binary probit model is a sensible choice here. Let $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_k)$, including a constant term, represent regression coefficients relating to explanatory variables $\boldsymbol{Z}$, then the propensity of the presence of a school in proximity to an intersection can be written as in equation (5) with a larger value of s^* indicating a higher probability that an intersection is in proximity to a school.

$$s_i^* = \boldsymbol{\delta Z_i} + \varepsilon_i \tag{5}$$

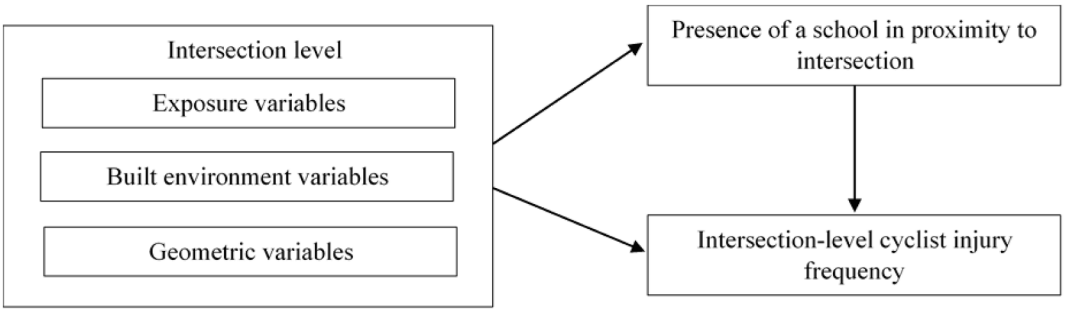


Fig. 1. Schematic view of the relationship between site characteristics, presence of school, and cyclist safety.

The error terms ε and ξ are modelled jointly with an assumption of a bivariate normal (BVN) density with a mean of zero and a covariate matrix as shown in (6) where the variance of ε_i is set to unity for identification. σ^2 represents the variance of ξ_i and ρ gives the correlation between the two error terms, ε and ξ. The equations as laid out here allow for potential endogeneity of S (presence of school) to be accounted for.

$$\Sigma \sim BVN\begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \tag{6}$$

2.4. Robustness regarding the bivariate normal assumption

As discussed above, the error terms of the specified equations are allowed to be correlated following a bivariate normal density, which is a standard assumption commonly used in previous research. However, if this assumption does not hold, statistical inferences will be biased. For example, if there are outlying intersections, a bivariate normal density would result in over-shrinkage towards the overall mean, leading to biased estimates. To assess the robustness of the bivariate normal assumption, we used a Bayesian scale mixing approach that leads to a bivariate t distribution assumption (Chib and Carlin, 1999). The latter, allowing for heavier tails, is more robust compared to a bivariate normal density (Chib and Carlin, 1999; Held and Holmes, 2006) and reduces the impact of untypical intersections if there are any. Scale mixing can be specified easily under the Bayesian framework without involving any additional computational complexities. Given the mathematical notation in Section 2.3, a Bayesian scale-mixture model is specified as follows:

$$\begin{aligned} (\xi_i, \varepsilon_i) &\sim \ BVN\left(0, \sum \Big/ \delta_i\right) \\ \delta_i &\sim \ Gamma(\nu/2, \nu/2) \end{aligned} \tag{7}$$

where δ_i should be positive with mean 1 and ν is equivalent to the degrees of freedom in the Student t density. As indicated in (7), we treat ν as an unknown parameter to be estimated as part of the analysis. Smaller values of ν indicate departures from the bivariate normal assumption.

2.5. Extension to a random parameters simultaneous-equations model with heterogeneity in covariance

To more fully account for unobserved heterogeneity, firstly, it is possible to extend the above simultaneous-equations by allowing the effect of proximity to school or other variables to vary across intersections. As we will discuss in Section 3, we only found the effect of proximity to school to vary across the sample. Therefore, we can rewrite equation (4) by modelling γ as a random parameter and allowing it to vary across observations. For this purpose, γ_i is assumed to follow a density function (usually, a normal density) as shown in (8), where μ_γ and σ_γ represent the mean and the standard deviation of the random parameter, respectively. Secondly, inspired by the idea behind heterogeneity in mean models, we extend the above simultaneous-equations by specifying a heterogeneity in covariance model, which leads to a heterogeneous dependence structure, as described below.

In the covariance matrix specified in (6), the term $\rho\sigma$ represents the covariance between cyclist injury frequencies and the propensity of the presence of a school nearby intersections. With respect to the first component of the covariance, which is the correlation parameter ρ, rather than estimating this parameter in a deterministic way as is commonly the case, we specify a prior distribution for ρ under the Bayesian framework so that ρ is modelled stochastically. This allows us to account for uncertainty associated with the correlation parameter more fully. With respect to the second component of the covariance, we model log-transformed σ as a function of explanatory variables $\boldsymbol{W}_\sigma$ available in the data as specified in (8), where η is a constant term and $\boldsymbol{\alpha}$ are coefficients that are associated with the explanatory variables $\boldsymbol{W}$. In doing so we not only achieve a heterogeneous covariance structure in a simultaneous-equations approach that addresses endogeneity, but also we can explain potential systematic variations in the covariance.

$$\begin{aligned} y_i &\sim \ Poisson\,(\lambda_i) \\ log(\lambda_i) &= \boldsymbol{\beta X_i} + \lambda_i S_i + \xi_i \\ \gamma_i &\sim \ Normal\left(\mu_\gamma, \sigma_\gamma^2\right) \\ s_i^* &= \boldsymbol{\delta Z_i} + \varepsilon_i \\ (\xi_i, \varepsilon_i) &\sim \ BVN\left(0, \begin{pmatrix} {\sigma_i}^2 & \rho\sigma_i \\ \rho\sigma_i & 1 \end{pmatrix}\right) \\ \log(\sigma_i) &= \eta + \boldsymbol{\alpha W} \end{aligned} \tag{8}$$

In our specification, the covariance is not fixed allowing the analyst to better address unobserved heterogeneity as well as to provide further insights using the same set of variables available in the data.

2.6. Prior specification and model estimation

We specified vague normal priors *Normal(0, 100)* for the mean parameter (μ) associated with the random parameter γ and for regression coefficients (β, δ, η, and α). We assigned a *Uniform(-1, 1)* prior distribution to the correlation parameter ρ, guaranteeing realistic values of correlation. We specified a relatively vague *Uniform(0, 10)* prior for σ, the second component of the covariance, and for the standard deviation of the random parameter γ. With respect to the scale mixing approach, as suggested in Bayesian Statistics literature (e.g., see Congdon (2003)), we specified an exponential distribution for ν with parameter k, *Exponential(k)*, and a *uniform (0.01, 0.5)* for k. In Bayesian statistics literature, this prior specification is shown to be a sensible choice for scale mixing that approximates a Student t density and allows the degree of freedom to be within the range 2–100 (Congdon, 2003). We estimated the models using freely available software WinBUGS (Lunn et al., 2000).

2.7. Detecting overdispersion

An important concern in modelling count data is whether the model accounts for overdispersion (which can be caused for instance due to the omission of important variables in the model) satisfactorily. This can be examined through a measure of deviance as in Equation (9), given the above notation. If the estimated deviance is smaller than the number of observations, overdispersion is dealt with (Congdon, 2003).

$$Deviance = 2\sum(y_i{*}\ln(y_i/\lambda_i) - (y_i - \lambda_i)) \tag{9}$$

3. Results and discussion

3.1. Parameter estimates and model comparisons

The result of the simple negative binomial count model that did not account for endogeneity is shown in Table 2. The result of the random parameters endogenous (simultaneous-equations) model that estimated the coefficient of the variables associated both with cycling injury counts and with the presence of a school is shown in Table 3. Based on the model results presented in Table 3, the correlation between these two outcomes was statistically significant: −0.477 (95 % Credible Intervals: −0.786, −0.059). This confirms the need for a simultaneous-equations model and addressing the endogeneity of proximity to school when estimating its effect on cyclist safety at signalised intersections in the study area.

With respect to concerns relating to the bivariate normal assumption, we developed a model using the Bayesian scale mixing approach as discussed in Section 2.4; however, we did not find any evidence to reject the bivariate normal assumption. The estimated weights remained around 1, which is an indication of no outlying intersections, and the model fit deteriorated (DIC increases from 1207.52 to 1214). Therefore, additional parameters introduced by scale mixing does not seem to be justified; and consequently, a bivariate normal density to model the error terms is empirically supported in this study. Based on the model reported in Table 3, a deviance of 536.1 (see Section 2.7), compared to 647 observations, indicates that overdispersion is accounted for properly.

It should be noted that to ensure more robust identification in simultaneous-equations models – which would result in more reliable inferences – it is desirable to avoid significant overlap in explanatory variables in the specified equations. Therefore, exclusion restrictions are often needed (Heckman, 2000; Munkin and Trivedi, 2003). In this regard, at least one explanatory variable in the treatment model should not have any effect on the main outcome (Munkin and Trivedi, 2003). In this paper, we do not have significant overlap in covariates in both equations. Also, the variable land use mix, which is in the presence of school model, did not have any statistically significant effect on cyclist injury frequencies over the study period.

The results of the proposed heterogeneity in covariance simultaneous-equations model are presented in Table 4. The latter specification leads to a decrease in DIC from circa 1208 to 1202 showing that model fit does improve with the addition of heterogeneity in covariance. Note that a DIC difference of more than 5 indicates a relatively important gain in model fit (Spiegelhalter et al., 2003). Also, the deviance, as calculated in Section 2.7, decreases to 501.5 under the heterogeneity in covariance endogenous model compared to 647 data points; therefore, overdispersion is dealt with. According to the latter addition, we found that the covariance varies systematically in the data as a function of one of the explanatory variables available in the data, specifically dedicated traffic lights for pedestrians. A noteworthy observation is that the effect of the presence of a school on cyclist injury frequency is underestimated when endogeneity as well as heterogeneity in covariance are not accounted for. In fact, the impact of proximity to school increases as we

Table 2
Results of the non-endogenous Poisson gamma (negative binomial) model.

Variable	Mean	SD	95 % Credible intervals	
Constant	−7.773	0.695	−9.184	−6.405
Ln(bicycle flow)	0.422	0.062	0.299	0.543
Ln(right turning AADT)	0.264	0.074	0.122	0.408
Ln(left turning AADT)	0.125	0.060	0.008	0.246
Number of subway stations	0.210	0.091	0.031	0.392
Presence of bus stop	0.561	0.164	0.237	0.884
Presence of school	*0.287*	*0.150*	*0.001*	*0.583*

Table 3
Results of the random parameters endogenous (simultaneous-equations) model.

Variable	Mean	SD	95 % Credible intervals	
Injury frequency count model				
Constant	−6.744	0.614	−8.073	−5.601
Ln(bicycle flow)	0.366	0.068	0.236	0.498
Ln(right turning AADT)	0.232	0.063	0.114	0.355
Ln(left turning AADT)	0.155	0.055	0.044	0.256
Number of subway stations	0.207	0.090	0.029	0.378
Presence of bus stop	0.518	0.165	0.194	0.842
Presence of school				
Mean	*0.677*	*0.317*	*0.083*	*1.244*
Variance	*0.271*	*0.217*	*0.004*	*0.754*
Presence of school binary probit model				
Constant	−1.762	0.193	−2.164	−1.409
Employment	2.082	0.221	1.661	2.518
Land use mix	0.677	0.281	0.123	1.220
Presence of a bus stop	0.391	0.123	0.149	0.634
Ratio of non-motorised to motorised volume	1.469	0.388	0.772	2.273
Covariance matrix - see EQ. (6)				
ρ	*−0.477*	*0.192*	*−0.786*	*−0.059*
σ	*0.749*	*0.138*	*0.464*	*0.991*
Model fit measure				
DIC	1207.52	–	–	–

Table 4
Results of the random parameters endogenous (simultaneous-equations) model with heterogeneity in covariance.

Variable	Mean	SD	95 % Credible intervals	
Injury frequency count model				
Constant	−7.512	0.718	−8.881	−6.060
Ln(bicycle flow)	0.361	0.070	0.226	0.500
Ln(right turning AADT)	0.268	0.078	0.118	0.418
Ln(left turning AADT)	0.180	0.067	0.054	0.331
Number of subway stations	0.195	0.102	0.025	0.391
Presence of bus stop	0.491	0.192	0.098	0.862
Presence of school				
Mean	*1.056*	*0.303*	*0.518*	*1.545*
Variance	*0.077*	*0.106*	*0.000*	*0.374*
Presence of school binary probit model				
Constant	−1.751	0.200	−2.157	−1.366
Employment	2.017	0.221	1.603	2.466
Land use mix	0.694	0.285	0.141	1.273
Presence of a bus stop	0.407	0.125	0.162	0.647
Ratio of non-motorised to motorised volume	1.533	0.402	0.803	2.386
Covariance matrix - see EQ. (8)				
ρ	*−0.547*	*0.173*	*−0.803*	*−0.178*
Ln(σ) equation:				
Constant	0.041	0.036	0.001	0.135
Dedicated traffic light for pedestrians	0.071	0.059	0.002	0.219
Model fit measure				
DIC	1202.00	–	–	–

move from the results represented in Table 2 to Table 4. Also, we can observe that the effect of proximity to school is estimated more precisely (the variance decreases) by the heterogeneity in covariance specification (Table 4) compared to the standard simultaneous-equations approach that assumes a homogenous covariance (Table 3). This is perhaps due to the fact that part of the variation in the effect of proximity to school is now captured by a heterogeneous covariance structure under the heterogeneity in covariance specification.

Similarly, the correlation parameter is estimated more precisely (having a smaller coefficient of variation) under the simultaneous-

equations model with heterogeneity in covariance. Another observation is that the simultaneous-equations approach with a homogenous covariance structure slightly underestimates the magnitude of the correlation between the propensity of proximity to school and cyclist injury counts at intersections. Our study thus points out that considering a homogenous covariance structure in simultaneous-equations approaches for addressing endogeneity may lead to erroneous inferences with respect to the statistical significance (importance) of the correlation parameter, delineating the direction of a traffic safety investigation entirely.

We found that a higher bicycle flow at an intersection is associated with an increase in cyclist injury frequency over the study period, though the coefficient value being less than one would elude to a safety-in-numbers effect for cyclists (more on this in section 3.1.1). Both AADT measures appear to have a positive association with cyclist injury frequencies, with right turning volume having a greater effect. Access to public transport, represented by the number of subway stations within 400 m of an intersection and the presence of a bus stop within 50 m, has a degrading effect on cyclist safety. Lastly, the principal focus of this research, the presence of a school in a 400 m buffer around intersections was positively associated with an increase in cyclist injury frequency. A greater propensity for proximity to a school was found to be associated with the amount of employment and the land use mix in an 800 m radius as well as the presence of a bus stop within 50 m and the ratio of non-motorised traffic volume to motorised traffic volume.

3.2. Sensitivity to the definition of proximity to school

In this research, we considered a 400-meter (0.25-mile) radius to define the presence of school as a categorical variable in our analysis. This radius is deemed reasonable in the context of active travel safety nearby schools; see, for example, Clifton and Kreamer-Fults (2007). Obviously, different radius values would result in different inferences and as the radius increases, the effect of the presence of school will reduce until it disappears. Also, the statistical importance of the correlation between the equation modelling proximity to school and the safety equation would change.

It would be interesting to consider different buffers to understand how the effect of proximity to school and its endogenous impact would change. In this study, alternatively, we considered a buffer of 800 m as well to define the proximity to school variable. As expected, inferences change. The correlation between the equations in the alternative simultaneous-equations model contains zero: −0.321 (95 % CI: −0.741, 0.141) although the mean remains negative as before. Also, proximity to school considering a buffer of 800 m does not seem to be statistically significant anymore.

This finding is not surprising as we would expect school-related safety interventions, which are not assigned randomly leading to selection bias, be applied to locations nearby schools only – not, for example, at an 800-meter distance from a school. Apparently, 800 m is far enough for such interventions to have no tangible effect on intersections in our study area. Also, another related reason for which we believe this variable may become exogenous as the distance from schools increases is that special traits in travel behaviour and road user characteristics (e.g., prevalence of children and risk compensations nearby schools) start to disappear.

3.3. Policy implications

3.3.1. Marginal effects

For the ease of interpretation of the impact of explanatory variables on cyclist injury frequencies, it is possible to obtain marginal effects based on the coefficients from the above models. These are presented in Table 5 with the values listed for each model representing the effect which a one unit change in the variable would have on the expected cyclist injury frequency over the study period. We provide the credible intervals of the marginal effects as well (these are rarely provided in the extant road safety literature) to draw a more complete picture. Our final inferences are based on the heterogeneity in covariance model, which provided the best fit to the data.

Bicycle flow is an important factor to examine in the present study as, in a relatively rare example for safety literature, observed counts for bicycle flows were used rather than surrogate cyclist exposure measures or figures from stated preference surveys that are

Table 5
Average marginal effects.

Variable	Mean	SD	95 % Credible intervals	
Random parameters simultaneous-equations model				
Ln(bicycle flow)	0.190	0.043	0.114	0.287
Ln(right turning AADT)	0.121	0.036	0.053	0.191
Ln(left turning AADT)	0.081	0.030	0.022	0.140
Number of subway stations	0.107	0.048	0.015	0.201
Presence of bus stop	0.169	0.050	0.068	0.265
Presence of school	*0.227*	*0.071*	*0.083*	*0.363*
Random parameters simultaneous-equations model with heterogeneity in covariance				
Ln(bicycle flow)	0.164	0.034	0.102	0.235
Ln(right turning AADT)	0.122	0.038	0.051	0.197
Ln(left turning AADT)	0.083	0.034	0.022	0.162
Number of subway stations	0.089	0.048	−0.003	0.187
Presence of bus stop	0.140	0.051	0.033	0.238
Presence of school	*0.252*	*0.075*	*0.125*	*0.388*

more common (Vanparijs et al., 2015). A higher volume of cyclists was associated with a greater frequency of cyclist injuries at a given intersection over the study period with a unit increase in the log value increasing the expected cyclist injuries by 0.164. This is not surprising as, with all other factors held equal, it stands to reason that more cyclists being exposed to injury would result in more injuries. The interesting finding here is that the predicted injury frequency does not rise proportionally with bicycle flow and instead increases at a lower rate. This may well suggest the presence of a safety-in-numbers effect in which, as the volume of a given road user type increases, the rate of injuries for that type increases at a lower rate. This phenomenon has been investigated by several researchers (Elvik and Goel, 2019; Heydari and Elvik, 2023) and the precise causes remain uncertain; however, it has been suggested that greater volume of a given road user type; e.g., bicycles, increases driver awareness of that road user type (Elvik, 2017). There is a case to be made that more people choosing to cycle improves the safety for all cyclists, provided that adequate infrastructure is in place, and therefore policies that encourage this shift should be encouraged.

A greater volume of vehicles being associated with more cyclist injuries is in line with most literature, though it has been posited that, depending on the location, a greater vehicle flow may reduce speeds which can improve cycling safety (Wegman et al., 2012). Crucially in the present research, it is the volume of motor vehicles turning that degrades cyclist safety, with right turning vehicles having a stronger effect than left (marginal effects of 0.122 and 0.083, respectively). This has been found in previous research (Gohl et al., 2019; Heydari et al., 2017) and serves as more evidence for the importance of considering cyclists when designing intersections that have high turning flows.

The presence of a bus stop and the addition of one subway station in proximity to an intersection would be expected to increase cyclist injury frequency by 0.140 and 0.089, respectively. This is broadly the consensus of cycling safety literature, as with walking, areas with higher levels of this type of activity increase crash risk (Heydari et al., 2017). It could also be the case that public transport links attract a higher number of active travellers and in particular, those who are changing between travel modes which could degrade safety. When considering where public transport links are to be situated and how they are to be designed, it is important that practitioners consider how these changes will impact cyclists.

Finally, having the highest marginal effect value, the presence of a school in proximity to an intersection was found to have the largest impact on cyclist injury counts. It increased, on average, the expected cyclist injury by 0.252 over the study period. As previously mentioned, the issue of the effect of a school on the safety of the surrounding area is not a straightforward problem. This research has shown that while accounting for endogeneity, there appears to be a degrading effect of proximity to a school on cyclist safety. With this in mind, it would appear that safety interventions that have been introduced at intersections around schools, if any, have not been effective in overcoming the frequency of injuries for cyclists at these locations, bearing in mind that this frequency is innately more likely to be higher than other locations as outlined in this paper. It is important that practitioners consider this when introducing safety infrastructure and policies as ignoring findings such as those presented here may lead to less efficient allocation of resources.

3.3.2. Implications of the estimated correlation and varying covariance

We found that there is a statistically significant correlation between cyclist injury frequencies and the propensity of the presence of school (see Table 4). Our results show that ignoring this correlation leads to underestimating the effect of proximity to school on cyclist safety; therefore, our study better informs decision-makers about this important risk factor. According to our analysis, the sign of the covariance was negative over the study period, indicating that unknown factors which increase the proximity-to-school propensity, decrease cyclist injury frequencies. Therefore, in our case study intersections nearby schools were, in principal, low cyclist injury frequency (relatively safe) intersections during the study period. This interesting finding implies that road users (here, specifically cyclists and drivers) in general behave cautiously at intersections nearby schools due to their perception of risk at such locations. On the other hand, the results indicate that the presence of a school is positively associated with cyclist injury frequencies; i.e., it decreases the safety of cyclists.

Therefore, other factors might have had a bearing on cyclist safety at these intersections over the study period. For example, it is quite plausible to assume that safety interventions are primarily applied to such intersections because these are considered risky locations (being in proximity to schools) rather than due to their actual safety conditions, causing confusion among road users. As discussed by Bhat et al. (2014) who investigated the impact of flashing lights on intersection safety while addressing endogeneity, safety treatments should be decided based on observed safety conditions (e.g., expected injury frequencies) instead of obvious features such as geometry or perceived risky situations.

Our results not only revealed that the covariance term in simultaneous-equations models might not be fixed (homogenous) across the sample, as it is commonly assumed to be in previous research, but it also can vary systematically. In fact, we found that, in our study, the covariance term varies as a function of a specific control type installed at some intersections; that is, dedicated traffic lights for pedestrians. Specifically, the results imply that the covariance between the proximity-to-school propensity and injury frequency increases in the presence of such a control type. This is an interesting finding that indicates the presence of very complex relationships between safety and intersection characteristics.

4. Conclusions

This research has provided empirical results for an analysis of the extent to which proximity to a school has an effect on the safety of cyclists at intersections while treating this risk factor as an endogenous variable. We employed a simultaneous-equations model, accounting for the covariance between the presence of a school, considering a buffer of 400 m, and cyclist injury counts explicitly. It was found that endogeneity does seem to be present when analysing data on cyclist injury frequency around schools and there is a

confounding effect between the factors that may be associated with cyclist safety and the presence of a school. In fact, the estimated negative correlation in our analysis implies that unobservables that increase cyclist injury frequencies tend to occur with unobservables that decrease the likelihood of the presence of school in proximity to an intersection. When considering a buffer of 800 m to define the presence of a school in proximity of an intersection, the endogenous impact of this risk factor disappears. In this study, the heterogeneity in covariance random parameters simultaneous-equations model provides similar results as the random parameters simultaneous-equations model in terms of coefficient estimates and marginal effects. However, the former model improves the fit as well as provides further insights into sources of heterogeneity in the data.

The factors that were identified as being associated with a higher cyclist injury frequency were bicycle flow, left and right turning motor vehicle flow, the number of subway stations, the presence of a bus stop, and the presence of a school. On the other hand, factors positively associated with a propensity for an intersection to be in proximity to a school were the number of people employed in proximity to the junction, the land use mix, the presence of a bus stop, and the ratio of non-motorised to motorised traffic volume. Due to data limitations (e.g., crash data and site characteristics not being available at a yearly basis; thus being fixed over the study period), caution must be taken in generalising our findings. For instance, had yearly data been available, we would have been able to investigate temporal instability (Mannering, 2018), drawing a more comprehensive and detailed picture.

Our findings reveal a safety-in-numbers effect for cyclists at intersections in proximity to schools over the study period, where an increase in bicycle volume increases injury frequency but at a rate that is less than proportional, suggesting that injury risk per cyclist is lowered with increased volume, contributing to a growing number of studies that have observed this phenomenon. The interventions that can be proposed from this research include encouraging cycling as a transport mode choice, given the above observed advantages of a greater volume of cyclists. Also, it would seem that more needs to be done to improve cycling safety at intersections in proximity to schools as the presence of a school has been shown to have a positive association with an increase in expected crash frequency and the effect is even stronger when endogeneity and heterogeneity in covariance are accounted for.

The analysis presented here can also be used to suggest specific elements that could be targeted to address this fact, such as giving particular consideration to intersections that have high volumes of turning vehicles perhaps by providing advanced stop lines or early traffic signal phases for cyclists; or focusing interventions on intersections that are in proximity to bus stops or a high number of subway stations and considering the impact that these multimodal areas have on cyclists. This study adds more empirical work to the growing body of literature on the importance of considering endogeneity in road safety research. Specifically novel insights are found in issues of endogeneity related to proximity to school and traffic safety, the neglecting of which would result in misinformed policy interventions that aim at increasing traffic safety nearby schools.

CRediT authorship contribution statement

Shahram Heydari: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Writing – original draft, Writing – review & editing. **Michael Forrest:** Investigation, Visualization, Writing – original draft, Writing – review & editing.

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.

Data availability

The authors do not have permission to share data.

Acknowledgments

The authors are grateful to the Editor-in-Chief and anonymous reviewers for their valuable comments.

References

Amoh-Gyimah, R., Saberi, M., Sarvi, M., 2016. Macroscopic modeling of pedestrian and bicycle crashes: A cross-comparison of estimation methods. Accident Analysis and Prevention 93, 147–159.

Behnood, A., Mannering, F., 2017. Determinants of bicyclist injury severities in bicycle-vehicle crashes: A random parameters approach with heterogeneity in means and variances. Analytic Methods in Accident Research 16, 35–47.

Bhat, C., Born, K., Sidharthan, R., Bhat, P.C., 2014. A count data model with endogenous covariates: Formulation and application to roadway crash frequency at intersections. Analytic Methods in Accident Research 1, 53–71.

Bhat, C., Astroza, S., Lavieri, P., 2017. A new spatial and flexible multivariate random-coefficients model for the analysis of pedestrian injury counts by severity level. Analytic Methods in Accident Research 16, 1–22.

Boarnet, M., Anderson, C., Day, K., McMillam, T., Alfonso, M., 2005. Evaluation of the California Safe Routes to School legislation: usrban form changes and children's active transportation to school. American Journal of Preventive Medicine 28, 134–140.

Cao, X., Mokhtarian, P., Handy, S., 2009. Examining the impacts of residential self-selection on travel behaviour: a focus on empirical findings. Transport Reviews 29 (3), 359–395.

Chib, S., Carlin, B., 1999. On MCMC sampling in hierarchical longitudinal models. Statistics and Computing 9, 17–26.

Congdon, P., 2003. Applied Bayesian Modelling. John Wiley & Sons Ltd, England.

Department for Transport (DfT), 2021. *RAS20008: Vehicles involved in reported accidents by vehicle type and manoeuvre, Great Britain.* Road accidents and safety statistics. https://www.gov.uk/government/statistical-data-sets/ras20-drivers-riders-and-vehicles-in-reported-road-accidents.
Eluru, N., Bhat, C., 2007. A joint econometric analysis of seat belt use and crash-related injury severity. Accident Analysis and Prevention 39 (5), 1037–1049.
Elvik, R., 2017. Exploring factors influencing the strength of the safety-in-numbers effect. Accident Analysis and Prevention 100, 75–84.
Elvik, R., Goel, R., 2019. Safety-in-numbers: an updated meta-analysis of estimates. Accident Analysis and Prevention 129, 136–147.
Fountas, G., Fonzone, A., Olowosegun, A., McTigue, C., 2021. Addressing unobserved heterogeneity in the analysis of bicycle crash injuries in Scotland: a correlated random parameters ordered probit approach with heterogeneity in means. Analytic Methods in Accident Research 32, 100181.
Gitelman, V., Levi, S., Carmel, R., Korchatov, A., Hakkert, S., 2019. Exploring patterns of child pedestrian behaviors at urban intersections. Accident Analysis and Prevention 122, 36–47.
Gohl, I., Schneider, A., Nitsch, V., 2019. Assessing the safety criticality of driver behavior toward cyclists at intersections. Traffic Injury Prevention 20 (sup3), 45–50.
Guo, Y., Osama, A., Sayed, T., 2018. A cross-comparison of different techniques for modeling macro-level cyclist crashes. Accident Analysis and Prevention 113, 38–46.
Harris, M.A., Reynolds, C., Winters, M., Cripton, P.A., Shen, H., et al., 2013. Comparing the effects of infrastructure on bicycling injury at intersections and non-intersections using a case–crossover design. Injury Prevention 19, 303–310.
Heckman, J., 2000. Causal parameters and policy analysis in economics: a twentieth century retrospective. Quarterly Journal of Economics 115 (1), 45–97.
Held, L., Holmes, C., 2006. Bayesian auxiliary variable models for binary and multinomial regression. Bayesian Analysis 1, 145–168.
Heydari, S., 2018. A flexible discrete density random parameters model for count data: embracing unobserved heterogeneity in highway safety analysis. Analytic Methods in Accident Research 20, 68–80.
Heydari, S., Elvik, R., 2023. Further compelling evidence for safety-in-numbers: it is more than meets the eye. Accident Analysis and Prevention 179, 106902.
Heydari, S., Fu, L., Miranda-Moreno, L.F., Joseph, L., 2017. Using a flexible multivariate latent class approach to model correlated outcomes: A joint analysis of pedestrian and cyclist injuries. Analytic Methods in Accident Research 13, 16–27.
Heydari, S., Miranda-Moreno, L., Hickford, A.J., 2020. On the causal effect of proximity to school on pedestrian safety at signalized intersections: a heterogeneous endogenous econometric model. Analytic Methods in Accident Research 26, 100115.
Islam, M., Alogaili, A., Mannering, F., Maness, M., 2023. Evidence of sample selectivity in highway injury-severity models: the case of risky driving during COVID-19. Analytic Methods in Accident Research 38, 100263.
Johansson, C., Leden, L., 2007. Short-term effects of countermeasures for improved safety and mobility at marked pedestrian crosswalks in Borås, Sweden. Accident Analysis and Prevention 39 (3), 500–509.
Kozumi, H., 2002. A Bayesian analysis of endogenous switching models for count data. Journal of the Japan Statistical Society 32 (2), 141–154.
Lunn, D., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10 (4), 325–337.
Mannering, F., Shankar, V., Bhat, C., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research 11, 1–16.
Mannering, F., Bhat, C., Shankar, V., Abdel-Aty, M., 2020. Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. Analytic Methods in Accident Research 25, 100113.
Munkin, M., Trivedi, P., 2003. Bayesian analysis of a self-selection model with multiple outcomes using simulation-based estimation: an application to the demand for healthcare. Journal of Econometrics 114, 197–220.
Nashad, T., Yasmin, S., Eluru, N., Lee, J., Abdel-Aty, M., 2016. Joint modeling of pedestrian and bicycle crashes: copula-based approach. Transportation Research Record: Journal of the Transportation Research Board. 2601, 119–127.
Oh, J., Shankar, V., 2011. Corridor safety modeling via segmentation: a selectivity bias perspective to crash count data. Lambert Academic Publishing, p. 84.
Osama, A., Sayed, T., 2017. Investigating the effect of spatial and mode correlations on active transportation safety modeling. Analytic Methods in Accident Research 16, 60–74.
Ragland, D. R., Pande, S., Bigham, J., Cooper, J. F., 2014. *Ten years later: examining the long-term impact of the California safe routes to school program.* Paper presented at the Transportation Research Board 93rd Annual Meeting, Washington, DC.
Shin, S., Shankar, V., 2013. Selection bias and heterogeneity in severity models: some insights from an interstate analysis. Lambert Academic Publishing, p. 132.
Spiegelhalter, D., Thomas, A., Best, N., Lunn, D., 2003. *WinBUGS User Manual*, version 1.4. https://www.mrc-bsu.cam.ac.uk/wp-content/uploads/manual14.pdf.
Stipancic, J., Zangenehpour, S., Miranda-Moreno, L., Saunier, N., Granié, M.-A., 2016. Investigating the gender differences on bicycle-vehicle conflicts at urban intersections using an ordered logit methodology. Accident Analysis and Prevention 97, 19–27.
Strauss, J., Miranda-Moreno, L.F., Morency, P., 2014. Multimodal injury risk analysis of road users at signalized and non-signalized intersections. Accident Analysis and Prevention 71, 201–209.
Terza, J., 1998. Estimating count data models with endogenous switching: sample selection and endogenous treatment effects. Journal of Econometrics 84 (1), 129–154.
Ukkusuri, S., Miranda-Moreno, L., Ramadurai, G., Isa-Tavarez, J., 2012. The role of built environment on pedestrian crash frequency. Safety Science 50 (4), 1141–1151.
Vanparijs, J., Panis, L.I., Meeusen, R., de Geus, B., 2015. Exposure measurement in bicycle safety analysis: a review of the literature. Accident Analysis and Prevention 84, 9–19.
Washington, S., Karlaftis, M.G., Mannering, F., Anastasopoulos, P., 2020. Statistical and econometric methods in transportation. Chapman and Hall/CRC. https://doi.org/10.1201/9780429244018.
Wegman, F., Zhang, F., Dijkstra, A., 2012. How to make more cycling good for road safety? Accident Analysis and Prevention 44 (1), 19–29.
Winston, C., Maheshri, V., Mannering, F., 2006. An exploration of the offset hypothesis using disaggregate data: the case of airbags and antilock brakes. Journal of Risk and Uncertainty 32 (2), 83–99.
Yasmin, S., Momtaz, S., Nashad, T., Eluru, N., 2018. A multivariate copula-based macro-level crash count model. Transportation Research Record 2672 (30), 64–75.