



A Bayesian bivariate hierarchical model with correlated parameters for the analysis of road crashes in Italian tunnels

Ciro Caliendo¹ · Maurizio Guida² · Fabio Postiglione² · Isidoro Russo¹

Accepted: 28 March 2021 / Published online: 13 April 2021
© The Author(s) 2021

Abstract

An analysis of crashes occurring in 252 unidirectional Italian motorway tunnels over a 4-year monitoring period is provided to identify the main causes of crashes in tunnels. In this paper, we propose a full Bayesian bivariate Poisson lognormal hierarchical model with correlated parameters for the joint analysis of crashes of two levels of severity, namely severe (including fatality and injury accidents only) and non-severe (property damage only), providing better insight on the available data with respect to an analysis based on severe and non-severe independent univariate models. In particular, the proposed model shows that for both of severity levels the crash frequency increases with some parameters: the average annual daily traffic per lane, the tunnel length, and the percentage of trucks, while the presence of the sidewalk provides a reduction in severe accidents. Also the presence of the third lane induces a reduction in severe accidents. Moreover, a reduction in the crash frequency of the two crash-types over years is present. The correlation between the parameters might offer additional insights into how some combinations can affect safety in tunnels. The results are critically discussed by highlighting strength and weakness of the proposed methodology.

✉ Fabio Postiglione
fpostiglione@unisa.it

Ciro Caliendo
ccaliendo@unisa.it

Maurizio Guida
mguida@unisa.it

Isidoro Russo
isrusso@unisa.it

¹ Department of Civil Engineering, University of Salerno, Via Giovanni Paolo II, 132-84084 Fisciano, SA, Italy

² Department of Information and Electrical Engineering and Applied Mathematics (DIEM), University of Salerno, Via Giovanni Paolo II, 132-84084 Fisciano, SA, Italy

Keywords Bayesian hierarchical model · Bivariate model · Poisson lognormal model · Correlated parameters · MCMC · Road tunnels crashes

1 Introduction and related work

The World Health Organization (WHO 2018) estimates that about 1.35 million people lose their lives on the roads every year. In the 28 countries of the European Union (EU28 2017) the number of road accident fatalities was 23,500 in 2017. In Italy in the same year, according to the Italian National Institute for Statistics (ISTAT 2017), the deaths and injuries due to road accidents were computed to be 3378 and 246,750, respectively. The magnitude of the number of fatalities and injuries, as well as the associated high social costs due to medical expenses and lost productivity, show that road crashes continue to remain a serious global problem. Therefore, current trends suggest intervening and investing in: (i) technological progress both in vehicles (e.g., increasing electronic control systems on board also to avoid collisions with non-motorized vehicles) and roads (e.g., implementing higher standards of design and maintenance); (ii) more restrictive legislations to mitigate risky behaviours of drivers (e.g., driving under the effect of alcohol and drugs, using mobile phones while driving, exceeding speed limits, unfastening seatbelts); (iii) more advanced safety research (e.g., developing both a greater in-depth knowledge on the methodological frontiers of traditional statistical models used for the analysis of accident data and indicating future directions of statistical modelling). An integrated approach among the three aforementioned issues should be desirable, but unfortunately nowadays this appears still far from being achieved.

In recent years, considerable road safety research efforts have been made for realising a better understanding of the factors that affect the risk of accidents and for modelling the expected crash frequency (number of crashes per year). In this respect, researchers have used several count models of varying complexity, from the classic statistical approach based on the traditional Negative Binomial (NB) regression to more complex models, such as random intercept or correlated parameters ones, or based on different statistical approaches, for instance Bayesian ones. Moreover, multivariate models rather than univariate ones have also been used to model jointly different crash-types. Other statistical models for road crashes are Mannering and Bhat (2014) and Mannering et al. (2016), while an analysis of crash data collected on the Italian roads is presented in Caliendo et al. (2007). However, it is to be stressed that, at the present time, modelling the unobserved heterogeneity in accident data is a major concern for researchers (Aguero-Valverde and Jovanis 2010; Huang and Chin 2010; Mannering and Bhat 2014; Liu and Sharma 2018). Some papers on crashes are based on random parameters models to capture the heterogeneity unexplained by the covariates (e.g., Anastasopoulos and Mannering 2009). Generalized linear models with correlated parameters provide better performance. For example, Yu et al. (2015) show that a correlated parameters Tobit model provides a better fit with respect to the corresponding one assuming

uncorrelated parameters. Coruh et al. (2015), and Hou et al. (2018) also find that the correlated random parameters Negative Binomial model is more appropriate than the one with uncorrelated parameters. Similar results are in Caliendo et al. (2019) for a correlated random parameters Poisson model. Multivariate random parameters models provide a further performance increase in terms of goodness-of-fit of road crashes data with different severity levels, see Han et al. (2018), Dong et al. (2014), Guo et al. (2019). However, very few case-studies are available on road crashes with different levels, adopting multivariate models with correlated parameters.

Moreover, most papers are related to accidents occurring on open roads, while accidents occurring in road tunnels have been less investigated. Driving in road tunnels often causes anxiety in drivers because tunnels are dark, narrow, and monotonous (PIARC 2008). As a result, driving behaviour in road tunnels is different when compared to that on open roads. Moreover, since tunnels are enclosed spaces, the escalation of accidents, in terms of consequences might be more serious than on open roads because more users than those directly involved in the accident can be potentially exposed to the risk of collisions. Also the change in light conditions (e.g., the so-called “black hole effect” when entering the tunnel), the disturbing effects on driving due to the tunnel wall (which is too close to the traffic lane), and the excessive stiffness of the tunnel wall (which in absence of safety barriers with re-directive profile is not able to reduce the severity level of the impact when vehicles collide against it) can contribute to more fatalities and/or injuries when compared to those on open roads. An Italian case-study (Caliendo and De Guglielmo 2012) reports 12 severe accidents/ 10^8 veh.-km for motorway tunnels in contrast to an average rate of 9 severe accidents/ 10^8 veh.-km on the same motorways. Similar results are provided in Zhuang-lin et al. (2009), showing that in China the severity of injuries is higher in freeway tunnels than on open roads. On the other hand, Yeung et al. (2013) report that the traffic accident rates in Singapore expressway tunnels are higher in the transition zones than in the interior zones of tunnels. Therefore, there is evidence that the severity of accidents occurring in tunnels needs further investigations. With reference to Italian tunnels, Caliendo et al. (2013) adopt a bivariate Negative Binomial (NB) model for analysing simultaneously severe and non-severe accidents. To highlight temporal correlations, random effects Negative Binomial (RENB) and Negative Multinomial (NM) models are also developed. Caliendo and Guida (2014) propose a novel bivariate NB model for jointly analysing the severe and total accidents. Both papers apply fixed-effect models with uncorrelated parameters. In Caliendo et al. (2016), instead, univariate random parameters models are proposed both with random effects (i.e., also the intercept was assumed to be random) and without random effects. Also in this case, the parameters are assumed to be uncorrelated. Finally, Caliendo et al. (2019) show that the correlated random parameters model can provide better performance in terms of goodness-of-fit with respect to the corresponding one assuming uncorrelated random parameters, at least for one severity level of crash data.

An alternative to the classical statistical approach for predicting the expected crash frequency may also be the Bayesian method, allowing to include prior information on parameter values in the estimation procedure to obtain a posterior

estimate of values of the parameters and some functions thereof. The Bayesian approach has become increasingly popular in the field of traffic safety research for the accessibility of Bayesian models through the use of Markov Chain Monte Carlo (MCMC) methods (Robert and Casella 2010), and the availability of the software package WinBUGS (Spiegelhalter et al. 2005) has encouraged the application of Full Bayesian (FB) methods also in road crashes studies. For example, Dong et al. (2014) employ the FB method to estimate the parameters of a multivariate random parameters zero-inflated regression model. Lee et al. (2015a) adopt the Bayesian Poisson lognormal (PLN) simultaneous equation spatial error model (BPLSESEM) to identify the significant factors for two target variables (pedestrian crash per crash location, and crash-involved pedestrians per residence). Lee et al. (2015b) develop a multivariate Poisson lognormal conditional autoregressive model, at the macroscopic level, for crashes by different transportation modes such as motor vehicle, bicycle, and pedestrian crashes. Farid et al. (2017) investigate the applicability of informative priors in Bayesian statistics to increase the transferability of safety performance functions. Liu and Sharma (2018) employ a multivariate spatial-temporal Bayesian model for handling the unobserved heterogeneity across space, time, and crash-types. Han et al. (2018) propose a Bayesian hierarchical random parameter approach to investigate the variations on crash frequency across different regions. Afghari et al. (2019) investigate the effects of different types of informative priors using the Bayesian inference. Guo et al. (2019) employ the FB method to develop a Poisson lognormal model accounting for both correlations between crash-types and unobserved heterogeneity across observations. More recently, Huang et al. (2019) implement a Bayesian multivariate random parameters model with mixture components for spatially correlated data. Instead, Zeng et al. (2019) investigate the inclusion of spatio-temporal correlation and interaction in a multivariate random parameters Tobit model, and their influence on fitting a real crash rates with different severity outcomes. Park et al. (2020) develop models incorporating the dependence among the multivariate crash counts by modelling multivariate random effects using copulas.

However, also the Bayesian approach has been employed for analysing the unobserved heterogeneity and/or correlations among crash-types with reference prevalently to those occurring on open roads, while crashes happening in tunnels have scarcely been considered. In this respect, it is to be said that tunnels might not have so much unobserved heterogeneity as open roads because the tunnels are closed structures with confined space where some factors that can cause accidents and/or contribute to unobserved heterogeneity are indeed controlled (e.g., there is absence of variations in the weather conditions within tunnels, except in the entrance and exit zones). Furthermore, no difference exists between night and day because lighting conditions are constant in tunnels (except at the portals in daylight). Moreover, in general there are no hazardous points in tunnels, such as junctions and/or intersections, that might cause sudden variations in driving behaviour in response to their presence.

Given the limited body of research dedicated hitherto to the unobserved heterogeneity in tunnels, the goal of this paper is to apply the Bayesian approach to investigating this issue in greater depth, by assuming a PLN model for crashes. We

consider also that crash counts could be correlated across collision types because of common influencing factors and unobserved effects. To account for the possible correlations among parameters, a multivariate PLN model is adopted.

In details, we study road crashes of two different types, namely severe and non-severe, occurred in Italian road tunnels, by adopting a hierarchical bivariate PLN model and a Bayesian procedure to estimate its parameters. Possible correlations between parameters are also taken into account. It is worth noting that more recent papers (e.g., Guo et al. 2019; Huang et al. 2019) propose bivariate PLN models in which only some parameters are correlated, namely those corresponding to the same covariate but pertaining to different severity levels of crashes, while the others are still assumed uncorrelated.

The paper is organized as follows: the next section describes the data set under analysis. Subsequently, we present the proposed bivariate model. Then the results of statistical analysis are commented on and compared with the outcomes obtained by a simpler procedure that is based on an alternative method exploiting univariate PLN models, in which the random parameters are assumed to be correlated each other. Finally, some conclusions are provided, together with some hints for future research.

2 Data description

The database of crashes collects data from $J = 252$ Italian motorway tunnels (226 two-lane and 26 three-lane) with unidirectional traffic only over a monitoring period of $T = 4$ years (2006–2009), resulting in $N = J \cdot T = 1008$ observations for both severe and non-severe crashes.

Accidents involve only motorized vehicles, since pedestrians and bicycles are forbidden in motorway tunnels. Table 1 gives accident counts observed during the monitoring period. In this study, 2280 accidents are considered, 757 of which are severe accidents. In two-lane tunnels, 1930 accidents are counted (665 of which are severe accidents), while in three-lane tunnels 350 accidents are registered (92 of which are severe crashes). A decreasing trend in the aforementioned three crash-types over time is also observed, as reported in Table 1.

Table 1 Accident count data observed during the 4-year monitoring period

Year	Crashes								
	Two-lane			Three-lane			Two-lane + Three-lane		
	Severe	Non-Severe	Total	Severe	Non-Severe	Total	Severe	Non-Severe	Total
2006	246	418	664	32	87	119	278	505	783
2007	181	364	545	32	75	107	213	439	652
2008	121	263	384	15	56	71	136	319	455
2009	117	220	337	13	40	53	130	260	390
2006–2009	665	1265	1930	92	258	350	757	1523	2280

The length of the tunnels varies from 0.39 to 3.25 km. Unidirectional traffic flow is expressed in terms of the average annual daily traffic *AADT* per lane, and it ranges between 2250 and 20,380 vehicles/day per lane. The percentage of trucks (i.e., vehicles with weight ≥ 30 kN) ranges from 15.0 to 31.2%. The width of each lane is 3.75 m, and the presence of a sidewalk (with a width of about 1.0 m) was observed only in 143 two-lane tunnels. In all the tunnels investigated the emergency lane was absent. The longitudinal slope is variable in 90 tunnels (83 two-lane and 7 three-lane) with positive and negative gradients, while it is constant in the remaining 162 tunnels (143 two-lane and 19 three-lane). The mechanical ventilation system, which is longitudinal with jet-fans, was observed in 55 two-lane tunnels and absent in the remaining 197 tunnels.

It is worth noting that the number of crashes reduces with time between 2006 and 2009, although the average annual daily traffic per lane and the percentage of trucks are stable over time. This outcome is likely to be due to: (i) the increasing use of electronic speed control systems (tutors) on Italian motorways containing the investigated tunnels (e.g., the presence of the tutor might affect driving behaviour making a reduction of speeds possible); (ii) the introduction in 2003 of the driver's license with a demerit points system for violations of the Highway Code (e.g., the point penalties on driving license due to violations of traffic law might also affect driving behaviour and a reduction in accidents might be expected); (iii) the positive effects of the implementation in Italy of the European Directive 2004/54/EC after October 2006, which led to the birth and/or reinforcement of certain safety measures in tunnels.

The database used in the present paper is an extended version of the one analyzed in Caliendo et al. (2019), by the inclusion of the three-lane tunnels data.

2.1 Set of variables

The dependent variables analyzed in the present work are the number of severe and non-severe accidents (property damage only) per year (crash frequency).

The independent variables considered are the most significant for road engineers: the average annual daily traffic *AADT* per lane, the length *L*, the percentage of trucks *%Tr*, the sidewalk *SW*, the longitudinal slope *LS*, the mechanical ventilation *MV*, and the number of lanes *LN*. The variables *SW*, *LS*, *MV*, and *LN* represent qualitative predictors coded as dummy variables: *SW* = 1 if a sidewalk is present, 0 otherwise; *LS* = 1 if the longitudinal slope of tunnel varies with ascending and descending gradient, 0 otherwise; *MV* = 1 if there is the mechanical ventilation, 0 otherwise; *LN* = 1 if there are three lanes, 0 otherwise. Summary statistics of tunnel length, *AADT* per lane and percentage of trucks are given in Table 2.

In order to evaluate the association between pairs of variables, we compute the correlation matrix in Table 3, by using Pearson's correlation coefficient being all the reported variables either real-valued or dichotomous. It is evident that all the independent variables reported in Table 3 are weakly correlated to each other.

Aiming at showing also the year effect on the occurrence of severe and non-severe accidents in tunnels within the monitoring period 2006–2009, the *year* variable is also considered as a predictor. The *year* variable is coded by introducing

Table 2 Summary statistics of variables

Variables		Two-lane			Three-lane			Two-lane + Three-lane		
		Length [km]	AADT per lane / 10,000 [veh./day]	Percentage of Trucks [%]	Length [km]	AADT per lane / 10,000 [veh./day]	Percentage of Trucks [%]	Length [km]	AADT per lane / 10,000 [veh./day]	Percentage of Trucks [%]
Mean	1.20	0.762	21.5	0.770	1.15	22.2	1.16	0.802	21.5	
SD	0.637	0.422	4.25	0.152	0.246	2.02	0.619	0.425	4.09	
Min	0.387	0.225	15.0	0.524	0.572	16.8	0.387	0.225	15.0	
Max	3.25	2.04	31.2	0.981	1.613	23.5	3.25	2.04	31.2	

Note that $SW = 1$ occurrences are 572; $LS = 1$ occurrences are 360; $MV = 1$ occurrences are 220; $LN = 1$ occurrences are 104

Table 3 Correlation matrix of the predictors *AADT*, *L*, *%Tr*, *SW*, *LS*, *MV*, and *LN*

Variables	AADT per lane /10,000	Length	Percentage of trucks	SW	LS	MV	LN
AADT per lane/ 10,000	1.000	-0.367	-0.278	-0.625	0.075	-0.335	0.281
Length		1.000	-0.093	0.195	0.119	0.621	-0.213
Percentage of trucks			1.000	0.395	0.089	-0.153	0.058
SW (1 if present; 0 otherwise)				1.000	-0.085	0.403	-0.388
LS (1 if not constant; 0 otherwise)					1.000	-0.073	-0.062
MV (1 if present; 0 otherwise)						1.000	-0.179
LN (1 if 3 lanes; 0 is 2 lanes)							1.000

the dummy variables *Y2007*, *Y2008*, and *Y2009*, after assuming year 2006 as baseline. We notice that *AADT* and *%Tr* do not show a temporal trend, and the other regressors do not change over time (during the monitoring period).

All the $K = 10$ explanatory variables mentioned above are used to build the road tunnel crash models presented in the next Sections, where independent variables *AADT*, *%Tr* and *L* are considered in logarithmic (base 10) form, in line with the technical literature on tunnel crashes (Caliendo et al. 2016).

3 Modelling approach

Crash data are discrete, non-negative random count data, that are commonly modelled by the Poisson distribution, for which the variance is equal to the mean. This relationship is often violated in crash counts, that shows an empirical variance greater than the sample mean (overdispersion), as a consequence of the strength of some influencing factors. To account for overdispersion, the Poisson lognormal (PLN) model has been proposed in many recent studies to model the variation of the average, by replacing traditional negative binomial (Poisson Gamma) approaches, as detailed in Sect. 1.

As already remarked, the number of severe and non-severe crashes occurring in the same tunnel in different years can not be considered as stochastically independent, since they share the effects of countermeasures implemented in that specific road tunnel over time. More precisely, crashes represent a cluster of observations that share common, even if unknown, random effects. As highlighted in Sect. 1, also the influence of explanatory variables on accidents might vary from tunnel to tunnel, i.e. some heterogeneity on risk factors might be present across tunnels. To cope with this heterogeneity, all the parameters of the model can vary from tunnel to tunnel, and they can be correlated to each other due to the interplay of the risk factors.

A way to account for data-level variability and tunnel-level one is to use a *hierarchical* (often named multilevel) model, where the first level models the overdispersed crash counts and the second level accounts for the variability of parameters across tunnels. Some applications of hierarchical approaches can be found in Furlan (2008), Arima et al. (2012), Han et al. (2018). In order to jointly model all the crash-types, we focus on a multivariate model.

For the aforementioned reasons, a hierarchical multivariate Poisson lognormal model with correlated random parameters is developed in our work for jointly analyzing M different crash levels. It is applied to the available dataset containing $M = 2$ different crashes (severe and non-severe) but can be easily generalized to any M value.

The inferential procedure for the proposed model must estimate a large number of parameters, as detailed in the next Section. To cope with so a complex model, Bayesian inference is preferred (Gelman and Hill 2007), since it provides a reasonable trade-off between computational burden and ease of implementation. Multilevel/hierarchical Bayesian models can be profitably estimated by the MCMC algorithms implemented in many software tools, such as WinBUG or OpenBUG that are also freely available. This feature, together with the possibility to include prior information of experts (if available), lead us to choose a Bayesian approach to the proposed road tunnel crash model.

3.1 Model description

In this Section, we define the proposed Bayesian hierarchical multivariate Poisson lognormal (regression) model with correlated random parameters, by specifying each level.

Level 1: *observed crash counts mode*

Let $\mathbf{Y}_i = (Y_i^1, \dots, Y_i^M)$ denote observation vector i ($i = 1, \dots, N$) containing the crash counts Y_i^m for collision severity $m, \forall m \in \{1, \dots, M\}$, in the same year, where N is still the number of observations for both severe and non-severe crashes (see Sect. 2). The road tunnel j corresponding to observation vector i is denoted henceforth as $j[i]$.

For a given crash-type m , at the first level of the hierarchy, conditional on the mean λ_i^m , vector element Y_i^m of observation vector i is modelled as:

$$Y_i^m | \lambda_i^m \sim \text{Poisson}(\lambda_i^m) \tag{1}$$

and λ_i^m is given by the following link function:

$$\log(\lambda_i^m) = \beta_0^m + \sum_{k=1}^K \beta_{k[j[i]]}^m X_{ik} + \varepsilon_i^m \tag{2}$$

Parameter β_0^m is the intercept of the model for crash-type m , K is still the number of covariates, while $\boldsymbol{\beta}_{j[i]}^m = (\beta_{1[j[i]]}^m, \dots, \beta_{K[j[i]]}^m)^T$ is the K -dimensional random vector containing the (possibly) correlated random parameters in road tunnel $j[i]$ for crash-type m . X_{ik} is the (i, k) entry of the $N \times K$ design matrix \mathbf{X} containing all the

explanatory variables data; \mathbf{X} depends on traffic, road tunnel characteristics and year and does not vary across crash-types. ε_i^m is the so-called “random effect” accounting for overdispersion in crash count data for crash-type m and is an element of the random vector $\boldsymbol{\varepsilon}_i = (\varepsilon_i^1, \dots, \varepsilon_i^M)^T$, assumed as independent of $\boldsymbol{\beta}_{j|i}^m$, that is modelled as a multivariate normal (MVN) with mean vector $\mathbf{0}$ (an M -dimensional zero vector) and $M \times M$ variance–covariance matrix $\boldsymbol{\Sigma}_\varepsilon$:

$$\boldsymbol{\varepsilon}_i | \boldsymbol{\Sigma}_\varepsilon \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon) \quad (3)$$

Level 2: tunnel-level model

In order to model tunnel-level variability, we assume that all the regression coefficients related to the K covariates for the M different crash-types under examination, whose number is $D = M \cdot K$, could be correlated to each other.

Thus, the distribution of vector $\boldsymbol{\beta}_j = (\beta_{1j}^1, \dots, \beta_{Kj}^1, \dots, \beta_{1j}^M, \dots, \beta_{Kj}^M)^T$, containing all those random coefficients, is assumed to be an MVN, viz.

$$\boldsymbol{\beta}_j | \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad j = 1, \dots, J \quad (4)$$

where $\boldsymbol{\mu} = (\mu_1^1, \dots, \mu_K^1, \dots, \mu_1^M, \dots, \mu_K^M)^T$ is a D -dimensional vector containing all the expected values of those random coefficients, i.e. $\mu_k^m = \mathbb{E}(\beta_k^m)$ ($m = 1, \dots, M$, $k = 1, \dots, K$), while $\boldsymbol{\Sigma}$ is the $D \times D$ variance–covariance matrix of those random coefficients.

Level 3: priors

According to the Full Bayesian (FB) paradigm, it is necessary to specify the prior distributions for the intercept vector $\boldsymbol{\beta}_0 = (\beta_0^1, \dots, \beta_0^M)^T$ and for hyperparameters within the models (1)–(4) to obtain Bayesian point and interval estimates of coefficients. Prior distributions reflect the prior belief of an expert about the parameters to be estimated. In the presence of sufficient prior knowledge for the individual parameters of interest, informative prior distributions are usually specified (Bedrick et al. 1996; Schluter et al. 1997). However, when prior knowledge is not available, it is customary to define the so-called non-informative (diffuse) priors expressing vague information on variables. In this study, the following non-informative priors are used for parameters and hyperparameters, in line with those commonly adopted in road accident technical literature.

A diffuse prior distribution is assumed for the intercept vector $\boldsymbol{\beta}_0$ that is modelled to have an MVN distribution:

$$\boldsymbol{\beta}_0 | \mathbf{0}, \boldsymbol{\Sigma}_0 \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}_0) \quad (5)$$

with an M -dimensional mean vector $\mathbf{0}$ and an $M \times M$ variance–covariance diagonal matrix $\boldsymbol{\Sigma}_0$ with large variance values, namely $\boldsymbol{\Sigma}_0 = v_0 \mathbf{I}_M$, where $v_0 = 1000$ and \mathbf{I}_M is the $M \times M$ identity matrix.

A non-informative conjugate prior for variance–covariance matrix $\boldsymbol{\Sigma}_\varepsilon$ in (3) is assumed (Spiegelhalter et al. 1996): an inverse Wishart prior distribution $IW(r_\varepsilon \cdot \mathbf{P}_\varepsilon, r_\varepsilon)$ is specified for $\boldsymbol{\Sigma}_\varepsilon$, where \mathbf{P}_ε is an $M \times M$ matrix representing a

hint for variance–covariance matrix while r_ε is a scalar “degrees of freedom” parameter (Carlin and Louis 2000). Typically, r_ε is an integer $r_\varepsilon \geq M$; the smaller r_ε , the more diffuse the prior distribution of the elements of \mathbf{P}_ε . For $M = 2$ (severe and non-severe crash-types), in road crashes literature (Aguero-Valverde 2013; Huang et al. 2019) common choices are $r_\varepsilon = M = 2$ degrees of freedom and

$$\mathbf{P}_\varepsilon = \begin{bmatrix} 0.1 & 0.005 \\ 0.005 & 0.1 \end{bmatrix} \tag{6}$$

since small values of variances and correlations are usually obtained to model overdispersion in accident data. Accordingly, we select the same values for the inverse Wishart prior of Σ_ε in our Bayesian inferential procedure, namely $r_\varepsilon = 2$ and \mathbf{P}_ε as in (6).

Also for the parameters of the regression coefficients vector β_j in (4), a diffuse prior is specified. The prior distribution of the hyperparameter vector μ in (4) is assumed to be an $MVN(\mathbf{0}, \Sigma_\mu)$, with a D -dimensional mean vector equal to a zero vector, and Σ_μ being a diagonal $D \times D$ variance–covariance matrix with large variance values, namely $\Sigma_\mu = v_\mu \mathbf{I}_D$, where $v_\mu = 1000$.

By following the arguments discussed for the selection of the prior of Σ_ε , we adopt for Σ in (4) a non-informative conjugate prior by specifying an inverse Wishart prior distribution $IW(r \cdot \mathbf{P}, r)$ for Σ , where $r = D$ guarantees a diffuse prior and \mathbf{P} is the $D \times D$ matrix:

$$\mathbf{P} = \begin{bmatrix} 0.1 & \cdots & 0.005 \\ \vdots & \ddots & \vdots \\ 0.005 & \cdots & 0.1 \end{bmatrix}$$

with low variance values and correlations.

3.2 Bayesian estimation procedure

In line with most FB implementations, the posterior distributions were obtained using Markov Chain Monte Carlo (MCMC) sampling techniques available in OpenBUGS (Lunn et al. 2009), which is an open-source statistical software that implements different families of MCMC algorithms, such as Gibbs, Metropolis and slice sampling. For our model, the Gibbs sampler is adopted for the good convergence characteristics and reduced execution times; the MCMC algorithm provides a sample of posterior vectors (containing all the parameters values) that enable us to compute point and interval estimates of each parameter in the model.

4 Model checking and comparison

After model estimation, model fitting and predicting capabilities have to be checked. Cross-validation is considered one of the most useful tools for evaluating model fitting and test error. However, in Bayesian framework the approach to implement cross-validation must be carefully selected. Indeed, leave-one-out cross-

validation (where the model must be re-fit a number of times equal to the cardinality of the sample) can be really highly computationally expensive, because even a single Bayesian model fitting procedure could require a long MCMC computation.

In order to mitigate this problem, we adopt a k -fold cross-validation procedure with $k = 5$, a value guaranteeing a good bias-variance tradeoff (Hastie et al. 2009). The k -fold cross-validation is performed by the data set into k groups (folds), where $k - 1$ are used to fit the model and the remaining one, say group l , is treated as a validation set, by which we compute the Root Mean Squared Error (RMSE) of the model as a validation statistic, viz.

$$RMSE_l = \sqrt{\frac{1}{n_l} \sum_i \frac{1}{M} \sum_{m=1}^M (Y_i^m - \hat{Y}_i^m)^2} \tag{7}$$

where Y_i^m is the crash count i of the accidents of severity level m (here $M = 2$, namely severe or non-severe) in data subset l , \hat{Y}_i^m denotes the value predicted by the model for severity level m in the same subset l , and n_l is the sample size of the group l . The procedure is repeated k times; a different group $l = 1, \dots, k$ is selected as a validation set each time, and an $RMSE_l$ is obtained. The k -fold cross-validation RMSE estimate is then given by:

$$RMSE = CV_{(k)} = \frac{1}{k} \sum_{l=1}^k RMSE_l \tag{8}$$

In general, models with a small RMSE values shows good predicting capabilities.

We use Deviance Information Criterion (DIC), introduced by Spiegelhalter et al. (2002) and implemented in OpenBUGS, to compare the proposed Bayesian hierarchical multivariate PLN with simpler independent hierarchical univariate PLN models that fit single crash-types data, that do not exploit the interdependencies among counts of different severity levels occurring in the same road tunnels. We adopt Bayesian estimation also for those univariate models.

4.1 Hierarchical univariate PLN crash count models with correlated random parameters

Hierarchical univariate PLN crash count models with correlated random parameters, one for each severity level, can be easily derived by considering model presented in Sect. 3.1, where counts Y_i^m are independent of Y_i^ℓ for each severity level $\ell \neq m$. In order to reach this goal, we assume that $\beta_j^m = (\beta_{1j}^m, \dots, \beta_{Kj}^m)^T$ is independent of $\beta_j^\ell \forall \ell \neq m$, and ε_i^m is independent of $\varepsilon_i^\ell \forall \ell \neq m$.

In the first level, Y_i^m ($i = 1, \dots, N$) for each severity level $m = 1, \dots, M$ is modelled according to (1) and (2), but ε_i^m (still assumed independent of $\beta_{j[i]}^m$) is modelled as:

$$\varepsilon_i^m | \sigma_\varepsilon^2 \sim N(0, \sigma_\varepsilon^2) \tag{9}$$

By collecting eqs. (9) $\forall m = 1, \dots, M$, we retrieve (3) in which $\Sigma_\epsilon = \sigma_\epsilon^2 \mathbf{I}_M$.

In level 2, since only coefficients pertaining to the same severity level can be correlated, the distribution of β_j^m is given by:

$$\beta_j^m | \mu^m, \Sigma^m \sim MVN(\mu^m, \Sigma^m) \quad j = 1, \dots, J, m = 1, \dots, M \tag{10}$$

where $\mu^m = (\mu_1^m, \dots, \mu_K^m)^T$ is a K -dimensional vector containing the expected values of random coefficients for crash level m , while Σ^m is the $K \times K$ variance-covariance matrix. Again, by collecting eqs. (10) for $\forall m = 1, \dots, M$, we obtain (4) in which $\Sigma = \text{diag}(\Sigma^1, \dots, \Sigma^M)$.

Regarding the prior distributions, β_0^m in (2) is a diffuse prior following, independently of any other parameter, a normal distribution with zero mean and large variance, $\forall m = 1, \dots, M$. More precisely, prior distribution is:

$$\beta_0^m \sim N(0, \nu_0) \tag{11}$$

where $\nu_0 = 1000$. An improper vague prior is assumed for σ_ϵ^2 in (9), i.e. $p(\sigma_\epsilon^2) \propto 1/\sigma_\epsilon^2$. The prior distribution of the hyperparameter μ^m in (10) is assumed to be an $MVN(\mathbf{0}, \Psi) \forall m = 1, \dots, M$, with a mean vector equal to a K -dimensional zero vector and Ψ being a diagonal $K \times K$ variance-covariance matrix with large variance values, namely $\Psi = \nu_m \mathbf{I}_K$, where $\nu_m = 1000$, like μ in (4). One and the same non-informative conjugate prior is specified for Σ^m in (10) $\forall m = 1, \dots, M$, similar to the prior of Σ in (4): the prior distribution is an inverse Wishart $IW(s \cdot \mathbf{Q}, s)$, where $s = K$ guarantees a diffuse prior and \mathbf{Q} is the $K \times K$ matrix:

$$\mathbf{Q} = \begin{bmatrix} 0.1 & \cdots & 0.005 \\ \vdots & \ddots & \vdots \\ 0.005 & \cdots & 0.1 \end{bmatrix}$$

5 Analysis of results

We implemented the model presented in Sect. 3 in OpenBUGS, for the dataset discussed in Sect. 2, containing crashes with $M = 2$ severity levels (severe and non-severe), thus our model turns out to be a Bayesian hierarchical bivariate Poisson lognormal model with correlated random parameters.

We verified MCMC convergence by launching two separate chains with different starting values, where each chain has 2,000,000 iterations. The first 200,000 iterations were discarded as burn-in period and a thinning interval equal to 20 was applied to guarantee negligible correlations between consecutive values of the MCMC process, thus we obtain a posterior sample of size $n = 90,000$. Convergence to the stationary posterior distribution was assessed using the Brooks-Gelman-Rubin (BGR) statistic (Brooks and Gelman 1998), being its value less than 1.1. Convergence was also verified by visual inspection of the MCMC trace plots of all model parameters, as well as by evaluating their autocorrelation functions.

The sample of posterior vectors (containing parameters values) generated by the MCMC algorithm allows to compute the point and interval estimate for each parameter. The former is the posterior mean of each parameter (the expected value on the posterior distribution), for instance for parameter μ_k^m is given by $\mathbb{E}[\mu_k^m | data] \cong 1/n \sum_{u=1}^n (\mu_k^m)_u$, computed by the sample mean of the corresponding elements $u = 1, \dots, n$ of the posterior sample; the latter is the 90% equal-tailed credible interval of the parameter and is given by the interval between the 5th and 95th percentile of the posterior sample.

Since we are interested in tunnel-level behavior of regression parameters, we estimate the posterior mean of μ_k^m (for each element of the vector β_j) and its 90% equal-tailed credible interval. We also compute the point estimate of the standard deviation of the regression parameters, given by the posterior mean of the square root of the diagonal elements of Σ in (4), to evaluate their variability among tunnels.

Table 4 contains, for severe and non-severe crashes, posterior estimates of the mean and of the standard deviation of all the regression coefficients and the intercept, together with the 90% equal-tailed credible interval of the mean.

In details, the coefficients of *AADT* per lane are 3.317 and 2.657 for severe and non-severe crashes, respectively; in addition, the coefficients of *L* are 0.9256 for severe accidents and 1.709 for non-severe ones. The expected frequency of severe (or non-severe) crashes increases with the average annual daily traffic *AADT* per lane, the percentage of trucks *%Tr*, and the tunnel length *L*. The most likely explanations are the following: as *AADT* per lane and *%Tr* increase, the crash frequency is expected to increase since for keeping the desired speeds, the lane changing and overtaking movements of vehicles are more frequent; by increasing *L*, the crash frequency likely increases for a progressive reduction of the drivers' concentration in long tunnels.

Table 4 shows also that the expected frequency of severe accidents decreases with the presence of the sidewalk *SW* in the tunnel, due to a reduction in disturbing effects on driving behavior that might be caused by the tunnel wall when it is too close to the traffic lane. The modelling results confirmed also, in accordance with the crash data, a reduction in crash frequency (for both crash-types) over years from 2006 to 2009. This decrease may be due to the positive effects of more stringent road-safety policies over time; as well as to the implementation of some countermeasures in the tunnels after October 2006, i.e. the date of the entry into force in Italy of the European Directive 2004/54/EC on minimum safety requirements for tunnels (European Parliament and Council 2004). When the longitudinal slope of tunnel is not constant ($LS = 1$), different signs of the estimation points was obtained in relation to the severity of crashes (i.e., *LS* presents a positive sign for the severe accidents and negative for the non-severe ones). However, we comment on the severe accidents only, since for this crash-type the covariate *LS* is statically significant. One possible explanation might be that the non-constant slope has a minor effect on the reduction in the trucks' average speed (i.e., the average speed of trucks might be more or less the same as in flat tunnels) when compared to upgrade (the grade resistance force hinders the truck motion) or

Table 4 Estimation results of the bivariate Poisson lognormal model with correlated random parameters

Variables	Severe			Non-Severe		
	Posterior mean	Posterior standard deviation	90% equal-tailed credible interval of the mean	Posterior mean	Posterior standard deviation	90% equal-tailed credible interval of the mean
Intercept	0.9098	0.5338	0.0267 1.8000	2.1790	0.3843	1.5280 2.8150
Log of AADT per lane/10,000	3.3170	0.3091	2.8100 3.8210	2.6570	0.2050	2.3160 2.9910
Log of length [km]	0.9256	0.3702	0.3060 1.5280	1.7090	0.2363	1.3210 2.0940
Log of percentage of trucks/ 100	<i>0.8939</i>	<i>0.7248</i>	<i>-0.2866 2.0930</i>	1.9180	0.5228	1.0280 2.7730
SW	-0.2981	0.1783	-0.6008 -0.0081	<i>-0.0148</i>	<i>0.1141</i>	<i>-0.2022 0.1760</i>
LS	0.1915	0.1154	0.0039 0.3811	<i>-0.0795</i>	<i>0.0783</i>	<i>-0.2060 0.0510</i>
MV	<i>0.0497</i>	<i>0.2081</i>	<i>-0.2960 0.3833</i>	<i>-0.1066</i>	<i>0.1361</i>	<i>-0.3320 0.1145</i>
LN	-0.3040	0.1730	-0.5924 -0.0258	0.2792	0.1147	0.0967 0.4722
Y2007	-0.2978	0.0952	-0.4526 -0.1390	-0.1738	0.0650	-0.2792 -0.0657
Y2008	-0.8239	0.1203	-1.0220 -0.6254	-0.5444	0.0783	-0.6716 -0.4155
Y2009	-0.9008	0.1327	-1.1230 -0.6854	-0.6871	0.0866	-0.8298 -0.5445
Number of observations	1008			1008		
RMSE	1.454					
DIC	4638.0					

Regression coefficients showing a credible interval including the zero value are shown in italic font

downgrade (trucks slow down to prevent brake failure) tunnels. This means that the average speed of trucks continues to remain quite high. Also the speed of passenger cars is quite high, since this is not influenced by variations of the longitudinal slope due to the greater ratio between the engine power and vehicle weight compared to that of truck. As a result, an increase in the severity of consequences of the collisions between vehicles (due to high speed) and/or against the tunnel wall might be expected. Also LN regressor was found to be statistically significant for both crash-types, with opposite influence on severe and non-severe accidents. The addition of a lane (e.g., passing from two to three lanes) is generally motivated by the need to obtain better traffic-flow conditions. In these cases, drivers are not generally forced to increase their speed to compensate for the lost time and a decrease in the number of severe crashes (prevalently associated with excessive speeds) is expected. On the other hand, by increasing the number of lanes, drivers have more opportunities for lane change and consequently there might be more traffic conflicts, resulting in an increase in the number of accidents (even if it might cause material damage only).

By looking at standard deviations of the regression parameters with a significant influence on tunnel crashes, some of them are comparable to mean values showing a considerable variability among tunnels. Such parameters are those related to L , SW , LS and LN for severe crashes, while a smaller variability is noticed for parameters related to non-severe crashes.

Additional considerations can be offered by inspecting the correlation matrix of regression coefficients of bivariate PLN model in Table 5, that reports only the main diagonal blocks, being the most significant ones to be analyzed from engineering perspective: some off-diagonal coefficients are reported in italic in case of (i) non-negligible correlation (greater than 0.3) between parameters; (ii) pertaining to covariates showing a relevant influence on crashes, namely having a 90% credible interval on the mean not containing the value 0 (see Table 4). With reference to severe accidents that are prevalently more relevant for causing victims, one may note from the content of Table 5, for example, for some statistically significant covariates that: (i) the non-constant slope ($LS = 1$) is negatively correlated to the length L (-0.3061), which means that it might mitigate the effects of the L on increasing the frequency of this crash-type; (ii) LS presents, instead, a moderate positive correlation with SW (0.5317), and also LN has a moderate positive correlation with SW (0.4060), suggesting further investigations on the presence of sidewalks in tunnels with non-constant slope or three lanes.

5.1 Comparison between bivariate PLN model and two univariate PLN models

The proposed bivariate PLN model is more complex with respect to PLN models that does not exploit the dependencies among coefficients and crash count of different severity levels, resulting in longer execution times of the MCMC algorithm as a consequence of a greater number of parameters to estimate. However, the bivariate PLN model offers some key advantages, such as better fits to data and reduced prediction error.

Table 5 Main diagonal blocks of the correlation matrix of regression coefficients of bivariate PLN model

Variables	Log of AADT per lane / 10,000	Length	Percentage of trucks	SW	LS	MV	LN	Y2007	Y2008	Y2009
Severe										
Log of AADT per lane / 10,000	1.0000	0.2007	0.2047	-0.1759	-0.2458	-0.0208	-0.1569	0.0799	-0.2619	-0.2979
Length	0.2007	1.0000	0.1033	-0.2487	-0.3061	-0.1250	-0.1886	0.0899	-0.3011	-0.3407
Percentage of trucks	0.2047	0.1033	1.0000	-0.0831	-0.1106	0.0382	-0.0239	0.0881	-0.1300	-0.1667
SW	-0.1759	-0.2487	-0.0831	1.0000	0.5317	0.0923	0.4060	-0.1307	0.5665	0.6212
LS	-0.2458	-0.3061	-0.1106	0.5317	1.0000	0.1680	0.4529	-0.1537	0.6514	0.7173
MV	-0.0208	-0.1250	0.0382	0.0923	0.1680	1.0000	0.1429	-0.0275	0.1875	0.1967
LN	-0.1569	-0.1886	-0.0239	0.4060	0.4529	0.1429	1.0000	-0.0904	0.4720	0.5034
Y2007	0.0799	0.0899	0.0881	-0.1307	-0.1537	-0.0275	-0.0904	1.0000	-0.1379	-0.1626
Y2008	-0.2619	-0.3011	-0.1300	0.5665	0.6514	0.1875	0.4720	-0.1379	1.0000	0.7626
Y2009	-0.2979	-0.3407	-0.1667	0.6212	0.7173	0.1967	0.5034	-0.1626	0.7626	1.0000
Non-Severe										
Log of AADT per lane / 10,000	1.0000	0.1504	0.0023	0.1038	-0.0702	-0.0564	0.0065	-0.0113	-0.0857	-0.0781
Length	0.1504	1.0000	-0.2950	-0.2003	-0.4147	-0.3759	-0.1419	-0.1102	-0.4420	-0.4371
Percentage of trucks	0.0023	-0.2950	1.0000	0.4223	0.5796	0.4786	0.3532	0.2612	0.5982	0.5884
SW	0.1038	-0.2003	0.4223	1.0000	0.3997	0.2957	0.2512	0.1159	0.4431	0.4291
LS	-0.0702	-0.4147	0.5796	0.3997	1.0000	0.5607	0.3296	0.2201	0.7362	0.7288
MV	-0.0564	-0.3759	0.4786	0.2957	0.5607	1.0000	0.2890	0.1865	0.6612	0.6541
LN	0.0065	-0.1419	0.3532	0.2512	0.3296	0.2890	1.0000	0.0696	0.3502	0.3451
Y2007	-0.0113	-0.1102	0.2612	0.1159	0.2201	0.1865	0.0696	1.0000	0.2641	0.2605
Y2008	-0.0857	-0.4420	0.5982	0.4431	0.7362	0.6612	0.3502	0.2641	1.0000	0.8508
Y2009	-0.0781	-0.4371	0.5884	0.4291	0.7288	0.6541	0.3451	0.2605	0.8508	1.0000

Note that the off-diagonal coefficients reported in italic refers to parameters with non-negligible correlation (greater than 0.3) and pertaining to covariates showing a relevant influence on crashes (showing a 90% credible interval on the mean not containing the value 0, see Table 4)

PLN models with correlated parameters provides better performance with respect to those assuming independent regression parameters (Anastasopoulos and Mannering 2011). However, it is not clear if the proposed bivariate hierarchical PLN model shows some advantages, in terms of goodness-of-fit and prediction capability, with respect to two independent univariate hierarchical PLN models (see Sect. 4.1) for severe and non-severe accidents with correlated parameters.

In order to analyze tunnel-level behavior of regression parameters, in Table 6 we present the posterior mean of μ_k^m and its 90% equal-tailed credible interval for each parameter computed by the two independent univariate PLN models for severe and non-severe crashes. We also provide the point estimate of the standard deviation of the regression parameters, given by the posterior mean of the square root of the diagonal elements of Σ^m in (10). Table 6 also contains, for both crash-types, the posterior estimates of the mean and of the standard deviation of the intercept β_0^m , together with the 90% equal-tailed credible interval of the mean.

By analyzing standard deviations and means of the regression parameters computed by univariate PLN models, a smaller variability of parameters values among tunnels seems to be present with respect to the variability resulting from the estimates obtained by the bivariate PLN model, being the standard deviation value much less than mean value for all the parameters.

The comparison of Tables 4 and 6 shows that the proposed bivariate model is able to highlight the importance of the presence of the sidewalk to reduce severe accidents (posterior mean significantly less than 0), as expected by road engineers. On the contrary, the univariate models for both crash-types counts are not able to show that the presence of the sidewalk is relevant to reduce accidents, since both SW credible intervals contain the 0 value. This outcome shows that our model can reveal also this weak effect that univariate models can not detect.

Moreover, the DIC value of the bivariate model reported in Tables 4 is 4638, which is much lower than that obtained by the use of two independent univariate PLN models (4725, see in Table 6). The latter DIC value has been computed by summing DIC values of single univariate PLN models, since DIC is additive in the case of independent models and prior distributions (El-Basyouny and Sayed 2009). Therefore, our model provides much better goodness-of-fit performance with respect to the use of two independent PLN models, by showing a drop-off of 87 to the DIC value.

Furthermore, by comparing the fivefold cross-validation RMSE values reported in Tables 4 and 6, it is immediate to notice that the bivariate PLN model has a smaller RMSE value than those provided by both the univariates models (the former has an RMSE = 1.454, while the latter models have RMSE = 1.572 and RMSE = 2.116 for severe and non-severe crashes, respectively), showing that the proposed model has also better prediction performance with respect to single independent univariate PLN models with correlated random parameters.

Table 6 Estimation results of independent univariate Poisson lognormal with correlated random parameters

Variables	Severe			Non-Severe		
	posterior mean	Posterior standard deviation	90% equal-tailed credible interval of the mean	posterior mean	Posterior standard deviation	90% equal-tailed credible interval of the mean
Intercept	1.0430	0.5691	0.1549 2.0320	2.3630	0.3973	1.6810 2.9840
Log of AADT per lane /10,000	3.4880	0.3212	2.9750 4.0270	2.7100	0.1992	2.3850 3.0470
Log of length [km]	0.9297	0.3548	0.3385 1.5190	1.6500	0.2193	1.2860 2.0120
Log of percentage of trucks/100	<i>1.1570</i>	<i>0.7827</i>	<i>-0.0814 2.4980</i>	2.2080	0.5431	1.2750 3.0850
SW	<i>-0.2270</i>	<i>0.1732</i>	<i>-0.5055 0.0557</i>	<i>-0.0185</i>	<i>0.1134</i>	<i>-0.1998 0.1689</i>
LS	0.2347	0.1172	0.0442 0.4294	<i>-0.0548</i>	<i>0.0734</i>	<i>-0.1760 0.0654</i>
MV	<i>0.1598</i>	<i>0.1854</i>	<i>-0.1511 0.4581</i>	<i>0.0091</i>	<i>0.1263</i>	<i>-0.2036 0.2150</i>
LN	-0.3170	0.1787	-0.6261 -0.0276	0.2696	0.1048	0.1002 0.4451
Y2007	-0.3076	0.0927	-0.4601 -0.1553	-0.1647	0.0665	-0.2727 -0.0554
Y2008	-0.7302	0.1064	-0.9089 -0.5570	-0.4581	0.0728	-0.5777 -0.3361
Y2009	-0.7499	0.1114	-0.9324 -0.5707	-0.5977	0.0791	-0.7271 -0.4661
Number of observations	1008			1008		
RMSE	1.572			2.116		
DIC	1957.0			2768.0		
Total DIC (Severe + Non-Severe)	4725.0					

Regression coefficients showing a credible interval including the zero value are shown in italic font

6 Conclusions

This work proposes a Bayesian bivariate Poisson lognormal (PLN) hierarchical model with correlated parameters for the joint analysis of severe (fatality and injury only accidents) and non-severe crashes (property damage only) occurred in Italian road tunnels.

The proposed model provides smaller RMSE and DIC values than those obtained by the use of two independent univariate PLN models with correlated random parameters.

It has been highlighted that the frequency of severe and non-severe crashes increases according to the annual daily traffic *AADT* per lane, the tunnel length *L*, and the percentage of trucks *%Tr*. This might be attributed to the fact that as *AADT* and *%Tr* increase, the changing of lane and overtaking movements are more frequent, so that crashes are expected to increase; by increasing *L*, crashes are also expected to increase for a decrease in the drivers' concentrations. The presence of the sidewalk *SW*, instead, reduces accidents because it mitigates the disturbing effects on driving behavior caused by the tunnel wall. The importance of the presence of the sidewalk is not highlighted by adopting independent univariate PLN models, while our bivariate PLN model reveals that severe accidents can be reduced by introducing a sidewalk and, thus, can offer better insight into crash data collected in tunnel roads.

The modelling results confirmed also, in accordance with the crash data, a reduction in crash frequency over the years from 2006 to 2009. The correlation matrix of the regression parameters shows that the non-constant slope is negatively correlated to the tunnel length, which means that it might mitigate the effects of the *L* on increasing the frequency of severe crashes. The *LS* and *LN* present, instead, a positive correlation with *SW*.

The proposed model allows road engineers to identify the variables on which prevalently to focus data collection, as well as to make more appropriate proposals in order to reduce the probability of accidents in tunnels. Accidents can be lowered by designing tunnels not too long, whose longitudinal slope is possibly non-constant, and by introducing a sidewalk. It is worth remarking that the presence of the sidewalk plays a dual role: (i) it reduces the disturbing effects on driving behaviour caused by the tunnel wall when it is too close to the traffic lane; (ii) the sidewalk can also be used as an escape route by tunnel users during their evacuation process following a disruptive event (e.g., traffic accident or fire). Moreover, our approach can also be adopted by Tunnel Management Agency (TMA) to evaluate probable variations in the frequency of severe and non-severe crashes that might be due to changes in traffic flow by day of the week. Our model can also help to define a control system on traffic flows entering the tunnel, able also to suggest an alternative itinerary if necessary.

Funding Open access funding provided by Università degli Studi di Salerno within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Afghari AP, Haque MdM, Washington S, Smyth T (2019) Effects of globally obtained informative priors on Bayesian safety performance functions developed for Australian crash data. *Accident Anal Prev* 129:55–65
- Aguero-Valverde J (2013) Multivariate spatial models of excess crash frequency at area level: Case of Costa Rica. *Accident Anal Prev* 59:365–373
- Aguero-Valverde J, Jovanis PP (2010) Spatial correlation in multilevel crash frequency models effects of different neighboring structures. *Transp Res Record* 2165:21–32
- Anastasopoulos PC, Mannering FL (2009) A note on modelling vehicle accidents frequencies with random-parameters count models. *Accident Anal Prev* 4:153–159
- Anastasopoulos PC, Mannering FL (2011) An empirical assessment of fixed and random-parameter logit modes using crash-and non-crash specific injury data. *Accident Anal Prev* 43:1140–1147
- Arima S, Cretarola L, Jona Lasinio G, Pollice A (2012) Bayesian univariate space-time hierarchical model for mapping pollutant concentrations in the municipal area of Taranto. *Stat Method Appl* 21:75–91
- Bedrick EJ, Christensen R, Johnson W (1996) A new perspective on priors for generalized linear models. *J Am Stat Assoc* 91(436):1450–1460
- Brooks S, Gelman A (1998) General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* 7(4):434–455
- Caliendo C, De Guglielmo ML (2012) Accident rates in road tunnels and social costs evaluation. In: *SIIV–5th International Congress-Sustainability of Road Infrastructures*. *Procedia-Social and Behavioral Sciences* 53: 166–177
- Caliendo C, De Guglielmo ML, Guida M (2013) A crash-prediction model for road tunnels. *Accid Anal Prev* 55:107–115
- Caliendo C, De Guglielmo ML, Guida M (2016) Comparison and analysis of road tunnel traffic accident frequencies and rates using random-parameter model. *J Transp Saf Secur* 8(2):177–195
- Caliendo C, De Guglielmo ML, Russo I (2019) Analysis of crash frequency in motorway tunnels based on a correlated random-parameters approach. *Tunn Undergr Sp Tech* 85:243–251
- Caliendo C, Guida M (2014) A new bivariate regression model for the simultaneous analysis of total and severe crashes occurrence. *J Transp Saf Secur* 6(1):78–92
- Caliendo C, Guida M, Parisi A (2007) A crash-prediction model for multilane roads. *Accident Anal Prev* 39(4):657–670
- Carlin BP, Louis TA (2000) *Bayes and empirical Bayes methods for data analysis*, 2nd edn. Chapman & Hall/CRC, Boca Raton
- Coruh E, Bilgic A, Tortum A (2015) Accident analysis with aggregated data: the random-parameters negative binomial panel count data model. *Anal Method Accid Res* 7:37–49
- Dong C, Clarke DB, Yan X, Khattak A, Huang B (2014) Multivariate random-parameters zero-inflated negative binomial regression model: An application to estimate crash frequencies at intersections. *Accid Anal Prev* 70:320–329
- El-Basyouny K, Sayed T (2009) Collision prediction models using multivariate Poisson lognormal regression. *Accident Anal Prev* 41(4):820–828

- European Parliament and Council (2004) Directive 2004/54/EC. Official Journal of the European Union. L.167, Bruxelles, 30 April. European Commission (EU28). Road Safety 2017. https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/vademecum_2018.pdf. Accessed 9 June 2020
- Farid A, Abdel-Aty M, Lee J, Eluru N (2017) Application of Bayesian informative priors to enhance the transferability of safety performance functions. *Accid Anal Prev* 62:155–161
- Furlan C (2008) Hierarchical random-effect models for coastal erosion of cliffs in the Holderness coast. *Stat Method Appl* 17:335–350
- Gelman A, Hill J (2007) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York
- Guo Y, Li Z, Liu P, Wu Y (2019) Exploring risk factors with crashes by collision type at freeway diverge areas: Accounting for unobserved heterogeneity. *IEEE Access* 7:11809–11819
- Han C, Huang H, Lee J, Wang J (2018) Investigating varying effect of road-level factors on crash frequency across regions: A Bayesian hierarchical random-parameter modeling approach. *Anal Method Accid Res* 20:81–91
- Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning. Springer, New York
- Hou Q, Tarko AP, Meng Z (2018) Analyzing crash frequency in freeway tunnels: a correlated random-parameters approach. *Accident Anal Prev* 111:94–100
- Huang H, Chang F, Zhou H, Lee J (2019) Modeling unobserved heterogeneity for zonal crash frequencies: A Bayesian multivariate random-parameters model with mixture components for spatially correlated data. *Anal Method Accid Res* 24:100–105
- Huang H, Chin HC (2010) Modeling road traffic crashes with zero-inflation and site-specific random-effects. *Stat Method Appl* 19:445–462
- Huang H, Zhou H, Wang J, Chang F, Ma M (2017) A multivariate spatial model of crash frequency by transportation modes for urban intersections. *Anal Method Accid Res* 14:10–21
- Italian National Institute for Statistics (2017) Road Accidents in Italy, year 2017. <https://www.istat.it/en/archivio/219652>. Accessed 9 June 2020
- Lee J, Abdel-Aty M, Choi K, Huang H (2015a) Multi-level hot zone identification for pedestrian safety. *Accid Anal Prevention* 76:64–73
- Lee J, Abdel-Aty M, Jiang X (2015b) Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. *Accid Anal Prevention* 78:146–154
- Liu C, Sharma A (2018) Using the multivariate spatio-temporal Bayesian model to analyze traffic crashes by severity. *Anal Method Accid Res* 17:14–31
- Lunn D, Spiegelhalter DJ, Thomas A, Best N (2009) The BUGS project: Evolution, critique and future directions (with discussion). *Stat Med* 28(25):3049–3082
- Mannering F, Bhat C (2014) Analytic methods in accident research: Methodological frontier and future directions. *Anal Method Accid Res* 1:1–22
- Mannering FL, Shankar V, Bhat CR (2016) Unobserved heterogeneity and the statistical analysis of highway accident data. *Anal Method Accid Res* 11:1–16
- Park ES, Oh R, Ahn JY, Oh MS (2020) Bayesian analysis of multivariate crash counts using copulas. *Accid Anal Prevent* 149:105431
- PIARC Technical Committee C3.3, Road tunnel operation, Report R17 (2008) Human factors and road tunnel safety regarding users, Paris. <https://www.piarc.org/en/order-library/6173-en-Human%20factors%20and%20road%20tunnel%20safety%20regarding%20users.htm?catalog&catalog-offset=7>. Accessed 9 June 2020
- Robert CP, Casella G (2010) Introducing Monte Carlo methods with R. Springer, New York
- Schluter P, Deely J, Nicholson A (1997) Ranking and selecting motor vehicle accident sites by using a hierarchical Bayesian model. *J Roy Stat Soc D-Sta* 46(3):293–316
- Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A (2002) Bayesian measures of model complexity and fit. *J Roy Stat Soc B* 64(4):583–639
- Spiegelhalter DJ, Thomas A, Best N (1996) Computation on Bayesian graphical models. *Bayesian stat* 5(5):407–425
- Spiegelhalter DJ, Thomas A, Best N, Lunn D (2005) WinBUGS User Manual. MRC Biostatistics Unit, Cambridge
- World Health Organization, WHO (2018) Global status report on road safety 2018. Geneva. https://www.who.int/violence_injury_prevention/road_safety_status/2018/GSRRS2018_Summary_EN.pdf. Accessed 9 June 2020
- Xie K, Ozbay K, Yang H (2019) A multivariate spatial approach to model crash counts by injury severity. *Accident Anal Prev* 122:189–198

- Yeung JS, Wong YD (2013) Road traffic accidents in Singapore expressway tunnels. *Tunn Undergr Sp Tech* 38:534–541
- Yu R, Xiong Y, Abdel-Aty M (2015) A correlated random-parameter approach to investigate the effects of weather conditions on crash risk for a mountainous freeway. *Transp Res C-Emer* 50:68–77
- Zeng Q, Guo Q, Wong SC, Wen H, Huang H, Pei X (2019) Jointly modeling area-level crash rates by severity: a Bayesian multivariate random-parameters spatio-temporal Tobit regression. *Transportmetrica A* 15(2):1867–1884
- Zhuang-lin M, Chun-fu S, Sheng-rui Z (2009) Characteristics of traffic accidents in Chinese freeways tunnels. *Tunn Undergr Sp Tech* 24:350–355

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