Spatiotemporal and random parameter panel data models of traffic crash fatalities in Vietnam

Article in Accident Analysis & Prevention - May 2016
DOI: 10.1016/j.aap.2016.05.028

CITATIONS
20

READS
108

3 authors, including:

Long Truong
La Trobe University
36 PUBLICATIONS 315 CITATIONS

Minh Le Kieu
University of Auckland
28 PUBLICATIONS 268 CITATIONS

Some of the authors of this publication are also working on these related projects:

Project
Utilising big transit data for transfer coordination View project

All content following this page was uploaded by Long Truong on 10 June 2016.
The user has requested enhancement of the downloaded file.
Spatiotemporal and random parameter panel data models of traffic crash fatalities in Vietnam

Long T. Truong¹,²*, Le-Minh Kieu³, Tuan A. Vu⁴

¹Institute of Transport Studies, Department of Civil Engineering, Monash University, Melbourne, Australia
²Directorate for Roads of Vietnam, Hanoi, Vietnam
³Smart Transport Research Centre, School of Civil Engineering and Build Environment, Science and Engineering Faculty, Queensland University of Technology, Brisbane, Australia
⁴Vietnamese-German Transport Research Centre, Vietnamese-German University, Binhduong, Vietnam

* Corresponding author: Email: long.truong@monash.edu, Phone: +61 3 9905 1851, Fax: +61 3 9905 4944,

Abstract

This paper investigates factors associated with traffic crash fatalities in 63 provinces of Vietnam during the period from 2012 to 2014. Random effect negative binomial (RENB) and random parameter negative binomial (RPNB) panel data models are adopted to consider spatial heterogeneity across provinces. In addition, a spatiotemporal model with conditional autoregressive priors (ST-CAR) is utilised to account for spatiotemporal autocorrelation in the data. The statistical comparison indicates the ST-CAR model outperforms the RENB and RPNB models. Estimation results provide several significant findings. For example, traffic crash fatalities tend to be higher in provinces with greater numbers of level crossings. Passenger distance travelled and road lengths are also positively associated with fatalities. However, hospital densities are negatively associated with fatalities. The safety impact of the national highway 1A, the main transport corridor of the country, is also highlighted.

Keywords: crash, province-level, random parameter, negative binomial, spatiotemporal, conditional autoregressive

1. Introduction

Similar to other developing countries, rapid economic growth in Vietnam has been accompanied by an enormous increase in motorisation and a high level of traffic crashes. Between 2006 and 2014, the number of motorcycles increased with an annual growth rate of 16.1% whereas the number of cars increased with an annual growth rate of 22.6% (JICA, 2009; NTSC, 2015). By 2014, there were 43.3 million registered vehicles, of which 94.6% were motorcycles. Unsurprisingly, the number of traffic crashes increased significantly from 9,470 in 1992 to 27,993 in 2002 while the number of fatalities increased from 3,077 in 1992 to 13,186 in 2002 (NTSC, 2015). Fortunately, since 2007, there has been a small, but steady reduction in the number of fatalities, which would be attributable to stronger traffic safety programs and measures implemented by authorities (Passmore et al., 2010; Ngo et al., 2012; Nguyen et al., 2013a; Nguyen et al., 2013b). In 2014, Vietnam had 25,322 reported traffic crashes and 8,996 fatalities (NTSC, 2015). It is not surprised that motorcycles accounted for around 70% of traffic crashes (NTSC, 2015; Truong et al., 2016). The World Health Organisation (WHO) estimated that the traffic fatality rate in Vietnam was nearly 24.5 per 100,000 population, which is 44% higher than the average fatality rate in South East Asia (WHO, 2015).

Traffic crashes are one of the leading causes of deaths and disabilities in Vietnam (Nguyen et al., 2012; Tran et al., 2012). In addition, their economic impact is profound. It was estimated that the cost of traffic crashes is between 2.5% and 2.9% of the country’s gross domestic product (GDP) (ADB, 2005; JICA, 2009). Traffic crashes can also cause a significant economic burden at individual and family levels. For example, a study in Thaibinh province found that the average cost of a traffic injury during hospitalisation was greater than 6 months’ average salary (Nguyen et al., 2013a).

Forecasting safety impacts of alternative transport planning schemes is essential for proactive safety planning. During the last decade, there has been a growing body of research
on macro-level safety analyses. In macro-level safety studies, safety performance measures, e.g. crash counts, spatially aggregated at a certain spatial unit are modelled against area-wide variables. A wide range of spatial units have been investigated, e.g. block groups (Dumbaugh and Rae, 2009), grid structure (Kim et al., 2006), census tracts (Wier et al., 2009), wards (Noland and Quddus, 2004), cantons (Aguero-Valverde, 2013), counties (Traynor, 2008; Huang et al., 2010), provinces (Erdogan, 2009; Tolón-Becerra et al., 2012), cities (Moeinaddini et al., 2014; Coruh et al., 2015), multiple provinces (Torre et al., 2007), states (Noland, 2003), countries (Kumara and Chin, 2004), and traffic analysis zones (TAZs) (Ng et al., 2002; Hadayeghi et al., 2003; Lovegrove and Sayed, 2006; Pirdavani et al., 2012; Wang et al., 2013). Effects of spatial units on modelling results have been discussed in few studies (Abdel-Aty et al., 2013; Lee et al., 2014b; Xu et al., 2014).

A variety of area-wide variables have been considered in previous macro-level safety analyses: socioeconomic variables, e.g. population density (Hadayeghi et al., 2003; Noland and Quddus, 2004; Huang et al., 2010; Tolón-Becerra et al., 2012; Lee et al., 2014b), age groups (Noland, 2003; Noland and Oh, 2004; Noland and Quddus, 2004; Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Huang et al., 2010; Aguero-Valverde, 2013; Lee et al., 2014a), income (Noland, 2003; Traynor, 2008; Pirdavani et al., 2012), GDPs (Kumara and Chin, 2004; Tolón-Becerra et al., 2012), and employment (Siddiqui et al., 2012); land use variables (Ng et al., 2002; Lovegrove and Sayed, 2006; Pulugurtha et al., 2013; Wang et al., 2013; Lee et al., 2014b); healthcare variables (Ng et al., 2002; Coruh et al., 2015); road infrastructure variables, e.g. road density, intersection density, road length (Amoros et al., 2003; Hadayeghi et al., 2003; Noland, 2003; Lovegrove and Sayed, 2006; Pirdavani et al., 2012; Tolón-Becerra et al., 2012; Jiang et al., 2016), roads with different functions (Lovegrove and Sayed, 2006; Huang et al., 2010), and road network structures (Wang et al., 2013; Moeinaddini et al., 2014); traffic pattern variables, e.g. vehicle kilometres travelled (VKT) (Dumbaugh and Rae, 2009; Abdel-Aty et al.,
2013; Aguero-Valverde, 2013), highway usage (Traynor, 2008), traffic volume (Quddus, 2008; Wier et al., 2009), speed (Quddus, 2008), volume to capacity ratios (Hadayeghi et al., 2003), and trip generation and distribution (Abdel-Aty et al., 2011); and environmental variables, e.g. rainfall (Coruh et al., 2015) and snowfall (Aguero-Valverde and Jovanis, 2006).

Spatial effects, i.e. spatial dependence or correlation and spatial heterogeneity (Anselin, 1988), have been considered in macro-level safety analyses. For example, Bayesian spatial approaches have been used to account for possible spatial correlation between areas (Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Siddiqui et al., 2012; Wang et al., 2013; Xu et al., 2014; Dong et al., 2015; Lee et al., 2015; Song et al., 2015; Siddiqui and Watkins, 2016). To consider spatial heterogeneity, previous macro-level safety studies have adopted the geographically weighted regression (GWR) models (Hadayeghi et al., 2003; Erdogan, 2009; Hadayeghi et al., 2010; Li et al., 2013) and random parameter models (Coruh et al., 2015; Xu and Huang, 2015). Space-time interaction has also been considered by Aguero-Valverde and Jovanis (2006).

A number of studies have focused on macro-level safety analyses, which however is mainly Western-based. There is a need to understand the safety effects of area-wide characteristics in the context of developing countries, including South East Asia countries and Vietnam specifically. In addition, the issue of level crossings are not considered in existing macro-level safety studies. In Vietnam, most of railway crashes occurred at level crossings (JICA, 2009). Level crossing fatalities represent about 1.5% of all traffic fatalities in Vietnam (NTSC, 2015), which is higher when compared to Australia and the European Union with 0.5% and 1.3% respectively (ITSR, 2011; ERA, 2014). Note that Vietnam has one of the densest level crossing systems and the lowest percentages of protected level crossings in Asia (ESCAP, 2000). This paper investigates the relationships between various area-wide variables and traffic crash fatalities in Vietnam at the province level. Random effect and random parameter
negative binomial panel data models are utilised to account for spatial heterogeneity. In addition, a conditional autoregressive (CAR) model is utilised to account for spatiotemporal autocorrelation. The number of level crossings in each province is included in the analysis. A better understanding of safety effects of area-wide variables is critical to safety planning and policy in Vietnam.

2. Data

The number of traffic crash fatalities in 63 provinces from 2012 to 2014, was obtained from the National Transportation Safety Committee. Level crossing data were collected from the Vietnam Railway Administration (VNRA, 2015). Road network data, consisting of lengths of national highways, provincial roads, district roads, commune roads, and urban roads, were obtained from the Directorate for Roads of Vietnam. Expressways are not considered in this study due to the limited length of the current network, i.e. only 700km of the planned 6,000km expressway network have been in operation recently (VEA, 2015). A dummy variable is added to investigate the safety effect of the national highway 1A, which is the main transport corridor of the country with over 2,200km long travelling through 30 provinces (DRVN, 2015). Note that level crossing and road length data were only reported in 2015. This study reasonably assumes that changes in the number of level crossings and road lengths from 2012 to 2015 were minor and could be ignored.

Socio-demographic panel data, e.g. population, population density, residential area ratio, and passenger distance travelled (PDT), in each province from 2012 to 2014 were downloaded from the website of the General Office of Statistics of Vietnam. In addition, medical-related panel data (e.g. the number of hospitals) were also downloaded, which is then used to calculate the hospital density in each province.

Fig. 1a shows the number of traffic crash fatalities by province in 2014. Hanoi and Ho Chi Minh City, two major cities, had the highest numbers of traffic crash fatalities. It is clear
that high values were clustered around Ho Chi Minh City. Moreover, results of the Moran’s I test (Moran, 1950; Truong and Somenahalli, 2011) indicated evidence of spatial autocorrelation with the Moran’s I statistic of 0.28 and p-value<0.001. Fatality rates by million passenger km travelled by province in 2014 are presented in Fig. 1b. It can be seen that high values were clustered in north-west provinces, which were in mountainous areas and had low PDTs. Spatial autocorrelation was also evident with the Moran’s I statistic of 0.16 and p-value<0.01.

![Map showing traffic crash fatalities and fatality rate by province in 2014](image)

**Fig. 1** The number of traffic crash fatalities and fatality rate by million passenger km travelled by province in 2014

Descriptive statistics of variables are presented in Table 1. Two dummy variables, i.e. year 2013 and year 2014, were included to compare with the base year (2012). The number of
traffic crash fatalities was selected as the dependent variable while other variables were considered as independent variables. Multicollinearity was checked by calculating the variance inflation factors (VIFs) for all independent variables. As a common rule of thumb, a VIF value of larger than 5 indicates high multicollinearity. As a result, population, population density, and residential area ratio were omitted. VIF values of ten variables included in the models are presented in Table 1.

Table 1 Descriptive statistics of variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Notations</th>
<th>Mean</th>
<th>Std.</th>
<th>Min</th>
<th>Max</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traffic crash fatalities</td>
<td></td>
<td>148.63</td>
<td>123.82</td>
<td>18.00</td>
<td>787.00</td>
<td></td>
</tr>
<tr>
<td>Population (1000 persons)</td>
<td></td>
<td>1,424.86</td>
<td>1,236.05</td>
<td>303.00</td>
<td>7,981.90</td>
<td></td>
</tr>
<tr>
<td>Population density (persons per km²)</td>
<td></td>
<td>476.06</td>
<td>578.67</td>
<td>43.80</td>
<td>3,809.00</td>
<td></td>
</tr>
<tr>
<td>Residential area ratio (%)</td>
<td></td>
<td>3.37</td>
<td>3.07</td>
<td>0.40</td>
<td>12.20</td>
<td></td>
</tr>
<tr>
<td>Passenger distance travelled (million passenger km)</td>
<td>PDT</td>
<td>1,392.40</td>
<td>2,033.27</td>
<td>22.90</td>
<td>13,137.90</td>
<td>4.37</td>
</tr>
<tr>
<td>Hospital density (number of hospitals per 1000km²)</td>
<td>HDEN</td>
<td>5.07</td>
<td>4.92</td>
<td>0.99</td>
<td>24.81</td>
<td>1.90</td>
</tr>
<tr>
<td>Length of national highways (km)</td>
<td>LNH</td>
<td>337.11</td>
<td>225.40</td>
<td>0.00</td>
<td>1,146.00</td>
<td>2.18</td>
</tr>
<tr>
<td>Length of provincial roads (km)</td>
<td>LPR</td>
<td>423.17</td>
<td>200.55</td>
<td>0.00</td>
<td>999.00</td>
<td>1.51</td>
</tr>
<tr>
<td>Length of district and commune roads (km)</td>
<td>LDCR</td>
<td>3,688.34</td>
<td>2,431.18</td>
<td>110.74</td>
<td>14,375.20</td>
<td>1.77</td>
</tr>
<tr>
<td>Length of urban roads (km)</td>
<td>LUR</td>
<td>339.90</td>
<td>378.63</td>
<td>42.34</td>
<td>2,320.65</td>
<td>3.52</td>
</tr>
<tr>
<td>On national highway 1A (dummy)</td>
<td>NH1A</td>
<td>0.48</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.45</td>
</tr>
<tr>
<td>Number of level crossings</td>
<td>LX</td>
<td>23.75</td>
<td>32.19</td>
<td>0.00</td>
<td>181.00</td>
<td>1.81</td>
</tr>
<tr>
<td>Year 2013 (dummy)</td>
<td>Y13</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.34</td>
</tr>
<tr>
<td>Year 2014 (dummy)</td>
<td>Y14</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>1.34</td>
</tr>
</tbody>
</table>

Table 2 presents correlation coefficients for variables selected for modelling. There was a positive correlation between passenger distance travelled and length of urban roads (correlation coefficient = 0.797). However, the VIF results suggested no evidence of multicollinearity, in which, passenger distance travelled had the highest VIF value of 4.37, followed by length of urban roads with a VIF value of 3.52. Nevertheless, models with and without passenger distance travelled were compared in the following analysis.
Table 2 Pearson correlation matrix for variables selected for modelling

<table>
<thead>
<tr>
<th></th>
<th>PDT</th>
<th>HDEN</th>
<th>LNH</th>
<th>LPR</th>
<th>LDCR</th>
<th>LUR</th>
<th>NH1A</th>
<th>LX</th>
<th>Y13</th>
<th>Y14</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDT</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDEN</td>
<td>0.49</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LNH</td>
<td>-0.17</td>
<td>-0.50</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LPR</td>
<td>-0.10</td>
<td>-0.36</td>
<td>0.53</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDCR</td>
<td>0.28</td>
<td>-0.23</td>
<td>0.45</td>
<td>0.24</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LUR</td>
<td>0.797</td>
<td>0.29</td>
<td>0.11</td>
<td>0.12</td>
<td>0.35</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NH1A</td>
<td>0.24</td>
<td>0.09</td>
<td>0.03</td>
<td>0.01</td>
<td>0.03</td>
<td>0.12</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LX</td>
<td>0.39</td>
<td>0.11</td>
<td>0.19</td>
<td>-0.02</td>
<td>0.38</td>
<td>0.38</td>
<td>0.48</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y13</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y14</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.50</td>
<td>1.00</td>
</tr>
</tbody>
</table>

3. Models

3.1 Random effect and random parameter negative binomial models

Since data contain province-specific characteristics and annual traffic crash fatalities in each province are likely to be correlated, panel count models should be employed. To account for spatial heterogeneity, random effect negative binomial (RENB) and random parameter negative binomial (RPNB) models have been applied in recent macro-level safety studies (Coruh et al., 2015; Xu and Huang, 2015). Therefore, RENB and RPNB models for panel data were used in this paper.

The negative binomial (NB) model has been widely used for crash frequency analysis, particularly for over-dispersed data (Lord and Manering, 2010; Washington et al., 2011). Let \( y_{it} \) denote the observed number of traffic crash fatalities in province \( i \) and year \( t \), \( X_{itk} \) is the \( k \)th variable for province \( i \) and year (time period) \( t \), \( \beta_k \) is the coefficient to be estimated, \( p \) is the number of variables, \( n \) is the number of provinces (zones), and \( T \) is the number of years (time periods). The NB model is derived by assuming:

\[
y_{it} \sim \text{Poisson}(\lambda_{it})
\]

\[
\lambda_{it} = \exp(\beta_0 + \sum_{k=1}^{p} \beta_k X_{itk} + \epsilon_{it})
\]

where \( \lambda_{it} \) is the Poisson parameter, which is the expected number of fatalities in province \( i \) and...
year $t$, $\exp(\varepsilon_{it})$ is a gamma-distributed error term with mean one and variance $\alpha$. With the addition of this term, the variance can be different to the mean as $\text{VAR}(y_{it}) = \lambda_{it} + \alpha\lambda_{it}^2$.

To account for heterogeneity across individuals, e.g. provinces, the RPNB model can be written as:

$$\lambda_{it} = \exp \left( (\beta_0 + \omega_{i0}) + \sum_{k=1}^{p} (\beta_k + \omega_{ik})X_{itk} + \varepsilon_{it} \right)$$

(3)

where $\omega_{ik}$ is a randomly distributed term, e.g. a normally distributed term with mean zero and variance $\sigma_k^2$ (Washington et al., 2011; Greene, 2012). In practice, a random parameter will be used if its standard deviation is significantly larger than zero; otherwise, the parameter is fixed across individuals.

It is noteworthy that a RENB model is equivalent to a RPNB model with the intercept term being the only random parameter (Anastasopoulos and Mannering, 2009; Washington et al., 2011; Chen and Tarko, 2014). In this paper, RENB and RPNB models are estimated using NLOGIT 5 (Econometric Software, 2012).

To compare models, the Likelihood Ratio (LR) test is used. The LR test statistic is calculated as:

$$LR = -2[LL_0 - LL_1]$$

(4)

where $LL_0$ and $LL_1$ are the log likelihood at convergence for null and alternative models respectively. This statistic is $\chi^2$-distributed with degrees of freedom equal to the difference in the numbers of parameters between these models.

3.2 Spatiotemporal model with conditional autoregressive priors (ST-CAR)

In a previous study, Aguero-Valverde and Jovanis (2006) extended the model proposed by Bernardinelli et al. (1995), which specified a space-time interaction term, to include covariates. However, these models strictly assume linear temporal trends. In this paper, the model proposed by Rushworth et al. (2014) is utilised to account for spatiotemporal autocorrelation:
$y_{it} \sim Poisson(\lambda_{it})$ \hspace{1cm} (5)

$\lambda_{it} = \exp(\beta_0 + \sum_{k=1}^{p} \beta_k X_{itk} + \phi_{it})$ \hspace{1cm} (6)

$\phi_{i1}|\phi_{-i1} \sim N\left( \frac{\rho_S \sum_{j=1}^{n} w_{ij} \phi_{j1}}{\rho_S \sum_{j=1}^{n} w_{ij} + 1 - \rho_S}, \frac{\tau^2}{\rho_S \sum_{j=1}^{n} w_{ij} + 1 - \rho_S} \right)$ \hspace{1cm} (7)

$\phi_{t}|\phi_{t-1} \sim N(\rho_T \phi_{t-1}, \tau^2 Q(W, \rho_S)^{-1}) \hspace{0.5cm} t = 2, ..., T$ \hspace{1cm} (8)

where $\phi_{it}$ are random effects that account for residual spatiotemporal autocorrelation in the data after the effects of covariates have been removed, $\phi_{-i1}$ is the vector of random effects for time period 1 except for $\phi_{i1}$, $\phi_t$ is the vector of random effects for time period $t$, $W = \{w_{ij}\}$ is the $n \times n$ adjacent matrix ($w_{ij}=1$ if provinces $i$ and $j$ are adjacent or 0 otherwise), $\rho_S$ is the spatial parameter, $\rho_T$ is the temporal parameter, and $\tau^2$ is the parameter controlling the variance of random effects. The precision matrix $Q(W, \rho_S)$ corresponds to the conditional autoregressive (CAR) prior proposed by Leroux et al. (2000) and is given by $Q(W, \rho_S) = \rho_S (\text{diag}(W \mathbf{1}) - W) + (1 - \rho_S) I$, where $\mathbf{1}$ is the $n \times 1$ vector of ones, $I$ is the $n \times n$ identity matrix.

Eq. (7) corresponds to the intrinsic CAR prior (Besag et al., 1991) for strong spatial correlation if $\rho_S=1$. On the contrary, $\rho_S=0$ suggests independent random effects with constant mean and variance. In Eq. (8), spatial and temporal autocorrelation are induced by the variance and mean respectively. Strong temporal autocorrelation is suggested by $\rho_T=1$ whereas temporal independence is indicated by $\rho_T=0$.

Model parameters were estimated in a Bayesian setting using Markov Chain Monte Carlo (MCMC) simulation. A burn-in period was set as 10,000 iterations. Model estimates are then based on 40,000 samples. Convergence of the model was examined by visual diagnostic and Geweke convergence diagnostic. Deviance Information Criteria (DIC) (Spiegelhalter et al., 2002) was used to provide a measure of model fit. Data analyses were performed using the CARBayesST package in the R Statistical Environment (R Development Core Team, 2015).
3.3 Measures of model prediction performance

To compare model prediction performance, the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) are adopted.

\[
MAE = \frac{1}{n_o} \sum_{j=1}^{n_o} |O_j - P_j| 
\]

\[
RMSE = \sqrt{\frac{1}{n_o} \sum_{j=1}^{n_o} (O_j - P_j)^2} 
\]

\[
MAPE = \frac{1}{n_o} \sum_{j=1}^{n_o} \left| \frac{O_j - P_j}{O_j} \right| 
\]

Where \(O_j\) is the observed value, \(P_j\) is the predicted value from the model, and \(n_o\) is the number of observations.

4. Results

4.1 RENB and RPNB models

Given the correlation between PDT and length of urban roads, four models, including random effect with PDT (RENB-1), random effect without PDT (RENB-2), random parameter with PDT (RPNB-1), and random parameter without PDT (RPNB-2), were compared. Estimation results using 200 Halton draws are presented in Table 3.

In the RENB-1 and RENB-2 models, all variables were significant at \(p<0.01\), except for the year 2013 variable. The standard deviation of the intercept distribution was significantly different to zero, indicating that modelling the intercept as random parameter was appropriate.

Similarly, in the RPNB-1 and RPNB-2 models, all variables were significant at \(p<0.05\), apart from the variable for year 2013. In addition, the intercept, length of urban roads, and the number of level crossings resulted in random parameters. Length of provincial roads resulted in a random parameter in the RPNB-1 model, but its standard deviation of parameter distribution was not significantly different to zero in the RPNB-2 model. The signs of parameters were consistent among the models. Dispersion parameters for these models were significantly
different to zero, suggesting the use of the negative binomial model over the Poisson model was appropriate.

Similar parameters and associated significance levels between the RENB-1 and RENB-2 models confirmed that there is no multicollinearity when both PDT and length of urban roads were considered. Compared to the RENB-2 model, the RENB-1 model had better log likelihood; however, the LR test was not significant (LR = 2.32, df = 1, p-value = 0.127). Likewise, estimates were consistent between the RPNB-1 and RPNB-2 models, suggesting no multicollinearity when both PDT and length of urban roads were considered. Moreover, the RPNB-1 model was significantly better than the RPNB-2 model as shown by the LR test (LR = 4.67, df = 1, p-value = 0.031) and AIC. In addition, the RENB-1 model had slightly better AIC compared to the RPNB-1 model, which could be attributed to fewer parameters in the RENB-1 model. Although the RPNB-1 model had the best log likelihood, results of the LR test showed it was not significantly better than the RENB-1 model (LR = 5.15, df = 3, p-value = 0.161) or the RENB-2 model (LR = 7.47, df = 4, p-value = 0.112).

Fig. 2 presents estimates of local parameters for variables associated with length of provincial roads and length of urban roads, obtained from the RPNB-1 model. It can be seen that the parameters for these variables varied across provinces. These variations may be attributed to different levels of traffic composition, traffic congestion, enforcement, and safety measures among provinces. Spatial variation patterns were however different between these two variables. Table 3 suggests coefficients of variations of local parameters for length of provincial roads and length of urban roads were 12% (0.00004/0.00034) and 50% (0.00037/0.00074) respectively. Fig. 2b shows that the effect of length of urban roads seemed to be smaller in major cities, i.e. Hanoi and Ho Chi Minh City. This may be due to the fact that urban roads in major cities have better traffic safety systems, e.g. modern signal controls and car/motorcycle lane separation, and are strongly enforced.
### Table 3: Estimation results for RENB and RPNB models

<table>
<thead>
<tr>
<th>Variables</th>
<th>RENB-1</th>
<th>RENB-2</th>
<th>RPNB-1</th>
<th>RPNB-2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>z value</td>
<td>Estimate</td>
<td>z value</td>
</tr>
<tr>
<td>Intercept</td>
<td>4.29200 ***</td>
<td>132.73</td>
<td>4.24300 ***</td>
<td>130.70</td>
</tr>
<tr>
<td>Standard deviation of parameter distribution</td>
<td>0.46823 ***</td>
<td>47.11</td>
<td>0.46752 ***</td>
<td>47.41</td>
</tr>
<tr>
<td>Passenger distance travelled (million passenger km)</td>
<td>0.00007 ***</td>
<td>8.62</td>
<td>0.00009 ***</td>
<td>10.32</td>
</tr>
<tr>
<td>Hospital density (number of hospitals per 1000km$^2$)</td>
<td>-0.01320 ***</td>
<td>-5.56</td>
<td>-0.00887 ***</td>
<td>-3.91</td>
</tr>
<tr>
<td>Length of national highways (km)</td>
<td>-0.00055 ***</td>
<td>-10.86</td>
<td>-0.00068 ***</td>
<td>-14.00</td>
</tr>
<tr>
<td>Length of provincial roads (km)</td>
<td>0.00055 ***</td>
<td>10.86</td>
<td>0.00042 ***</td>
<td>8.39</td>
</tr>
<tr>
<td>Standard deviation of parameter distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length of district and commune roads (km)</td>
<td>0.00001 ***</td>
<td>2.81</td>
<td>0.00003 ***</td>
<td>7.95</td>
</tr>
<tr>
<td>Length of urban roads (km)</td>
<td>0.00061 ***</td>
<td>15.19</td>
<td>0.00098 ***</td>
<td>40.16</td>
</tr>
<tr>
<td>Standard deviation of parameter distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>On national highway 1A (dummy)</td>
<td>0.14345 ***</td>
<td>6.87</td>
<td>0.17940 ***</td>
<td>8.88</td>
</tr>
<tr>
<td>Number of level crossings</td>
<td>0.00561 ***</td>
<td>17.25</td>
<td>0.00439 ***</td>
<td>13.51</td>
</tr>
<tr>
<td>Standard deviation of parameter distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 2013 (dummy)</td>
<td>-0.01278</td>
<td>-0.54</td>
<td>-0.00392</td>
<td>-0.16</td>
</tr>
<tr>
<td>Year 2014 (dummy)</td>
<td>-0.10533 ***</td>
<td>-6.04</td>
<td>-0.09603 ***</td>
<td>-5.60</td>
</tr>
<tr>
<td>Dispersion parameter for negative binomial distribution</td>
<td>217.868 ***</td>
<td>4.82</td>
<td>214.795 ***</td>
<td>4.84</td>
</tr>
<tr>
<td>Number of observations</td>
<td>189</td>
<td>189</td>
<td>189</td>
<td>189</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-882.29811</td>
<td>-883.22039</td>
<td>-879.72426</td>
<td>-882.07340</td>
</tr>
<tr>
<td>Akaike information criterion (AIC)</td>
<td>1790.6</td>
<td>1790.4</td>
<td>1791.4</td>
<td>1794.1</td>
</tr>
</tbody>
</table>

Note: * p<0.1; ** p<0.05; *** p<0.01
Fig. 2 Parameters obtained from the RPNB-1 model for length of provincial roads and length of urban roads by province

4.2 ST-CAR model

Estimation results for the ST-CAR model are presented in Table 4. Note that the year 2013 and year 2014 variables were excluded since temporal autocorrelation was already considered in the ST-CAR model. It is clear that 95% Bayesian credible intervals (BCIs) of all parameters had the same sign (or did not contain zero). As the variance parameter $\tau^2$ was 0.16 (95%BCI: 0.12 – 0.21), spatial dependence parameter $\rho_S$ was 0.55 (95%BCI: 0.31 – 0.79), and temporal parameter $\rho_T$ was 0.95 (95%BCI: 0.86 – 0.998), the spatiotemporal autocorrelation in the data was evident. In general, the high temporal parameter is consistent with a decreasing trend of fatalities during the three-year study period suggested by the RENB and RPNB models. For example, in the RENB and RPNB models, both year 2013 and year 2014 variables resulted in
negative coefficients where the year 2014 variable was significant at p<0.01. The effect of including the temporal parameter was further examined by estimating a model with spatial autocorrelation only ($\rho_T$ was set as 0). The DIC of the model with spatial autocorrelation only was considerably larger than that of the ST-CAR model with spatiotemporal autocorrelation (1650.892 versus 1586.463). This suggests the ST-CAR model was the better model and considering spatiotemporal autocorrelation improved model fit. In addition, the signs of parameters in the ST-CAR model were in accordance with those in the RENB and RPNB models.

Table 4 Estimation results for the ST-CAR model

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>s.d.</th>
<th>Bayesian credible interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.22784</td>
<td>0.06447</td>
<td>4.10487 - 4.36023</td>
</tr>
<tr>
<td>Passenger distance travelled (million passenger km)</td>
<td>0.00007</td>
<td>0.00001</td>
<td>0.00005 - 0.00010</td>
</tr>
<tr>
<td>Hospital density (number of hospitals per 1000km²)</td>
<td>-0.00308</td>
<td>0.00065</td>
<td>-0.00435 - 0.00178</td>
</tr>
<tr>
<td>Length of national highways (km)</td>
<td>-0.00034</td>
<td>0.00012</td>
<td>-0.00059 - 0.00013</td>
</tr>
<tr>
<td>Length of provincial roads (km)</td>
<td>0.00024</td>
<td>0.00001</td>
<td>0.00002 - 0.00005</td>
</tr>
<tr>
<td>Length of district and commune roads (km)</td>
<td>0.00004</td>
<td>0.00001</td>
<td>0.00002 - 0.00005</td>
</tr>
<tr>
<td>Length of urban roads (km)</td>
<td>0.00042</td>
<td>0.00007</td>
<td>0.00030 - 0.00056</td>
</tr>
<tr>
<td>On national highway 1A (dummy)</td>
<td>0.24537</td>
<td>0.02983</td>
<td>0.18438 - 0.29819</td>
</tr>
<tr>
<td>Number of level crossings</td>
<td>0.00259</td>
<td>0.00071</td>
<td>0.00115 - 0.00397</td>
</tr>
<tr>
<td>$\tau^2$</td>
<td>0.15937</td>
<td>0.02350</td>
<td>0.11801 - 0.21046</td>
</tr>
<tr>
<td>$\rho_S$</td>
<td>0.55164</td>
<td>0.12338</td>
<td>0.31036 - 0.78930</td>
</tr>
<tr>
<td>$\rho_T$</td>
<td>0.95150</td>
<td>0.03766</td>
<td>0.85880 - 0.99827</td>
</tr>
<tr>
<td>DIC</td>
<td>1586.463</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.3 Model comparison

A summary of model comparison statistics is shown in Table 5. NB is the fixed parameter negative binomial model. All RENB and RPNB models provided similar MAE, RMSE, and MAPE, which were significantly better than those in the NB model. This suggests the consideration of random effects or random parameters improved model fit considerably. It is clear that the ST-CAR model, which accounted for spatiotemporal autocorrelation,
outperformed other models as its MAE, RMSE, and MAPE were significantly lower. The performance of the ST-CAR model is further illustrated in Fig. 3 that compares predicted versus observed values by NB, RENB-1, RPNB-1, and ST-CAR models. In short, the ST-CAR model was the most favourable model.

Table 5 Model comparison statistics

<table>
<thead>
<tr>
<th></th>
<th>NB</th>
<th>RENB-1</th>
<th>RENB-2</th>
<th>RPNB-1</th>
<th>RPNB-2</th>
<th>ST-CAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>51.96</td>
<td>9.57</td>
<td>9.55</td>
<td>9.56</td>
<td>9.54</td>
<td>3.23</td>
</tr>
<tr>
<td>RMSE</td>
<td>78.78</td>
<td>13.56</td>
<td>13.68</td>
<td>13.56</td>
<td>13.69</td>
<td>4.53</td>
</tr>
<tr>
<td>MAPE</td>
<td>46.63%</td>
<td>7.44%</td>
<td>7.45%</td>
<td>7.43%</td>
<td>7.42%</td>
<td>2.74%</td>
</tr>
</tbody>
</table>

Fig. 3 Predicted versus observed traffic crash fatalities

5. Discussion

The discussion in this section is focused on estimates of the ST-CAR model. Passenger distance travelled had a positive parameter of 0.00007 (95%BCI: 0.00005 – 0.0001), suggesting greater passenger distance travelled values are associated with increasing traffic fatalities. The positive
effect of passenger distance travelled is expected, given that it is considered as a measure of exposure to safety risk. Previous studies also indicated the positive effect of VKT, another exposure measure, on traffic crashes (Dumbaugh and Rae, 2009; Abdel-Aty et al., 2013; Aguero-Valverde, 2013).

Hospital density resulted in a negative parameter of -0.00308 (95%BCI: -0.00435 to -0.00178). This indicates provinces with higher hospital densities are more likely to have lower fatalities. The availability of hospitals related to hospital density is essential to post crash care since a faster access to medical care could prevent deaths. This result is in line with findings of previous research, which found that reduced crash fatalities are associated with higher hospital bed densities (Castillo-Manzano et al., 2013) and the availability of Magnetic Resonance Imaging (MRI) scans (Torre et al., 2007). An implication of this result is that enhancing local emergency medical services, e.g. ambulances and first-aid training, could also contribute to limiting the severity of traffic crashes.

Length of national highways was negatively associated with fatalities with a parameter of -0.00034 (95%BCI: -0.00059 to -0.00013). This should be carefully interpreted since important factors such as intersection density were not included in the model. Previous studies showed that higher intersections densities are associated with increasing crashes (Hadayeghi et al., 2003; Pirdavani et al., 2012). However, provinces in mountainous areas tend to have longer national highways, but lower intersection densities. The impact of length of national highways should be investigated in future work by enhancing the variety of area-wide factors. Nevertheless, fatalities in a province would increase if it contains the national highway 1A, indicated by a positive parameter of 0.24537 (95%BCI: 0.18438 – 0.29819). This is expected given that the national highway 1A is the backbone (north-south) transport corridor with heavy traffic travelling through 30 provinces. This finding supports authorities’ ongoing programs to upgrade and expand the national highway 1A.
Length of provincial roads, length of district and commune roads, and length of urban roads resulted in positive parameters of 0.00024 (95%BCI: 0.00002 – 0.00045), 0.00004 (95%BCI: 0.00002 – 0.00005), and 0.00042 (95%BCI: 0.0003 – 0.00056) respectively. These results suggest that increases in lengths of district and commune roads, provincial roads, and urban roads are associated with increases in fatalities, which is in accordance with findings of previous studies (Quddus, 2008; Tolón-Becerra et al., 2012).

Number of level crossings resulted in a positive parameter of 0.00259 (95%BCI: 0.00115 – 0.00397), suggesting that provinces with a greater number of level crossings tend to have more traffic fatalities. This is an important result that should be considered by transport planners and authorities. Countermeasures might include removing level crossings at critical intersections, reducing the number of unprotected level crossings, and enhancing enforcement and safety education.

In the RENB and RPNB models, the year 2014 variable was significant at p<0.01 with a negative parameter. Although the variable for year 2013 was not significant, but it also resulted in a negative parameter. In general, a decreasing trend in fatalities during the period from 2012 to 2014 was evident.

6. Conclusion

This paper has explored factors associated with traffic crash fatalities in 63 provinces of Vietnam during the period from 2012 to 2014. The RENB and RPNB panel data models were adopted to consider spatial heterogeneity across provinces, which can arise from observed and unobserved factors. In addition, the ST-CAR model was utilised to account for spatiotemporal autocorrelation in the data. The statistical comparison indicated that the signs of estimated parameters were consistent among these models and the ST-CAR model outperformed the RENB and RPNB models.
Estimation results provide several significant findings. For example, traffic crash fatalities were positively associated with the number of level crossings, passenger distance travelled, length of provincial roads, length of district and commune roads, and length of urban roads. In addition, traffic crash fatalities were positively associated with the presence of the national highway 1A. However, hospital densities were negatively associated with traffic crash fatalities.

Although the models performed well in exploring effects of area-wide factors on traffic crash fatalities, there are several areas that can be improved. For example, other important factors, e.g. urbanisation rate, motorcycle to car ratio, intersection density, and education level, should be investigated. Expressways may also need to be considered in future as the length of the expressway network is growing. Future work should also investigate factors associated with fatal and injury crashes at different spatial levels, e.g. districts and suburbs. Nevertheless, this study gives an important contribution towards understanding safety effects of area-wide factors in the context of South East Asia countries and Vietnam specifically.

Acknowledgements

The authors would like to thank Hoang Anh Tuan, Vu Duc Phuc, Tran Minh Thu, and Cao Hoang Can for their support in data collection. The authors are also grateful to the three anonymous reviewers for their constructive comments to improve the manuscript.

References


