IATSS Research 41 (2017) 123–128

Contents lists available at ScienceDirect

IATSS Research

Research article

Crash rates analysis in China using a spatial panel model

Wonmongo Lacina Soro [a,*], Yiwei Zhou [b], Didier Wayoro [c]

[a] School of Transportation, Southeast University, No. 2 Sipailou, Nanjing 210096, China
[b] Precision Systems, Inc., 80 M Street, SE, Suite 180, Washington, DC 20003, USA
[c] Department of Economics, University of Massachusetts Amherst, Thompson 200 Hicks Way, MA, 01003, USA

ARTICLE INFO

Article history:
Received 29 July 2016
Received in revised form 26 October 2016
Accepted 9 November 2016
Available online 12 November 2016

Keywords:
China
Crash rates
Spatial panel
Spatial lag
Error lag
Maximum likelihood

ABSTRACT

The consideration of spatial externalities in traffic safety analysis is of paramount importance for the success of road safety policies. Yet, the quasi-totality of spatial dependence studies on crash rates is performed within the framework of single-equation spatial cross-sectional studies. The present study extends the spatial cross-sectional scheme to a spatial fixed-effects panel model estimated using the maximum likelihood method. The spatial units are the 31 administrative regions of mainland China over the period 2004–2013. The presence of neighborhood effects is evidenced through the Moran's I statistic. Consistent with previous studies, the analysis reveals that omitting the spatial effects in traffic safety analysis is likely to bias the estimation results. The spatial and error lags are all positive and statistically significant suggesting similarities of crash rates pattern in neighboring regions. Some other explanatory variables, such as freight traffic, the length of paved roads and the populations of age 65 and above are related to higher rates while the opposite trend is observed for the Gross Regional Product, the urban unemployment rate and passenger traffic.

1. Introduction

Traditional modeling methods of road accidents are based on the assumption that the variables are random and the disturbance terms are unrelated. These precepts have been questioned since the paper of [1] stating that a closer spatial examination of movements and interactions as well as of neighborhood and community will provide hard evidence on which to base road safety policies. Road networks being interconnected, traffic counts are interrelated making systematic spatial and serial autocorrelation between observations on the different networks [2–4]. Neighboring spatial units are likely to share common infrastructures resulting in similar traffic regulations and road users' behaviors, and definitely in comparable road safety levels. Ignoring the spatial dependence in traffic crash modeling will certainly produce bias and inconsistent estimates [5]. The spatial dependence relies on Tobler's [6] first law of geography which says 'everything is related to everything else but near things are related more than distant things' p. 236.

The incorporation of spatial dependence in traffic safety analysis can be traced back to [7]. They used a spatial lag model (SLM) to investigate the spatial patterns of motor vehicle accidents in Honolulu, Hawaii, during the year 1990. They found that spatial location is a key element in road crash analysis. Papadimitriou et al. [8] also used another SLM to model the spatial variation of crash rates in Greece for the year 2002. The rates in each county were found to significantly depend on some factors that are due to similarities with neighboring counties. LaScala et al. [9] used a spatial error model (SEM) model to map locations of pedestrian injuries in San Francisco, California, for the year 1990. The analysis supported that the geographical proximity has an impact on the traffic safety of spatial units. Quddus [10] explored the 2001-crash data of the census wards in London. Using a SEM, the author reported a spatial dependence in crash observations. Hong et al. [11] used both a SLM and a SEM to inspect crash occurrence in the administrative zones of Seoul in South Korea for the year 2010. The study revealed a high spatial correlation between accidents. Wang and Kockelman [12] used a panel seemingly unrelated regression SEM to evaluate crash rates in some Chinese cities over the period 1999–2002 and highlighted the necessity of incorporating spatial effects in crash rates analysis.

All these studies support the inclusion of spatial dependence in road crash modeling. Nevertheless, excepted [12], the remaining studies are performed within the framework of a single-equation scheme and have limitations in the sense that they are cross-section analyses [13–15]. Cross-section data ignore the dynamics of the changes and the heterogeneity of spatial units. This study is an attempt to overcome these limitations by using a spatial panel method. Panel data are more efficient in that they provide more information, more variability, less collinearity among the covariates and more degree of freedom [13–15]. The spatial units are the administrative regions of mainland China. Road accidents and the resulting fatalities are still prevalent in

* Corresponding author.
E-mail addresses: sorow@seu.edu.cn (W.L. Soro), zywadam@gmail.com (Y. Zhou), dwayoro@econs.umass.edu (D. Wayoro).

Peer review under responsibility of International Association of Traffic and Safety Sciences.

https://doi.org/10.1016/j.iatssr.2016.11.001

China compared to many other countries in the world. Many efforts are needed to make a substantial progress [16] because road traffic accidents are still the leading cause of injury deaths in China [17,18]. In 2010, there were 207 million vehicles in China against 60 million in 2000 and only 2 million in 1980 [19]. This steep growth rate of motorization is expected to be correlated with more road traffic crashes and fatalities.

The remaining paper is organized as follows. Section 2 introduces the fundamentals about spatial effects and spatial panels. Section 3 presents the data. Section 4 describes the modeling technique while Sections 5 and 6 are respectively devoted to the results and their discussion. The conclusions are given in Section 7.

2. Estimation techniques for spatial panel models

Spatial dependence among observations across space is investigated through the so-called spatial weight matrix W that highlights the organization of geographical entities in the space [20].

Three types of spatial effects exist in spatial econometric: the endogenous interaction effects among the explained variable, the exogenous interaction effects among the regressors and the interaction effects among the error terms [14,21]. Much of the literature has been devoted to the interaction among the explained variable known as spatial lag, the interaction among the error terms known as error lag and the combination of both [14]. Following [22], the model structure for the spatial lag is the spatial autoregressive (SAR) model; for the error lag it is the spatial error model (SEM) and for the combination of both it is the spatial autoregressive model with autoregressive disturbances (SARAR).

These models have been traditionally estimated using cross-section data. To ease empirical investigations the development of the estimation techniques has been extended to spatial panel models [20,23]. The traditional ordinary least squares (OLS) regression is inappropriate for models incorporating spatial effects [20,24]. Due to the presence of the spatial weight matrix, the OLS estimator is biased and inconsistent for SAR models because of the quadratic form in the error term and inconsistent for SEMs because of the nondiagonal form of the error term variance matrix [20]. The maximum likelihood (ML) and the quasi-maximum[1] likelihood (QML) methods have been viable alternatives to overcome these shortcomings of the OLS estimator since a first trial by [25] and mainly since the textbook by [20].

Much of the literature on spatial panels is devoted to dynamic spatial panel models [26–29]. However, dynamic spatial models are beyond the scope of this paper; only static spatial panels are dealt with. Kapoor et al. [30] extended Kelejian and Prucha's [22] spatial cross-section generalized method of moments (GMM) to panel SEMs with random effects. Mutl and Pfaffermayr [31] proposed the GMM for SARAR panel models with fixed and random effects and designed the Hausman test to discriminate between these effects. Lee and Yu [32] established the asymptotic properties of the QML estimators for SARAR panel models with fixed effects. Yang et al. [33] investigated the finite sample properties of these fixed effects estimators. The ML and QML methods are so far the most appropriate methods to fit spatial panels [32,34–36].

3. Data description

The spatial units are the administrative regions of mainland China. Altogether, there are 31 regions. The study covers the period 2004–2013. The data for each year are obtained from the corresponding Chinese Statistical Yearbook published by the National Bureau of Statistics of China.[2] The study considers two dependent variables representing crash rates: the annual number of traffic accidents per 100,000 population and the resulting number of injuries also per 100,000 population. These crash rates are computed based respectively on the number of traffic accidents and the associated number of injuries.

The set of covariates included in the analysis are related to populations, traffic, infrastructures and the economic prosperity represented by the Gross Regional Product (GRP) as a proxy for the income of each region. The summary statistics of the variables are given in Table 1.

A preliminary analysis in the spatial dependence analysis is the visualization of the scheme of the dependent variable in order to have an initial idea about its dynamics. This step is done here through the quantile map where the observations are divided into six categories Fig. 1 shows the quantile map of the average injury rates over the study period.

The map for the average accident rates is dropped because it shows similar structures to the first one. The map shows evidence of spatial autocorrelation as most of the neighboring regions tend to display similar configurations. This spatial connection will be further analyzed through appropriate tests in the next section.

4. Methodology

Road networks between spatial units are interconnected and road users in neighboring units are likely to have similar behavioral schemes. Thus, the different spatial units are expected to influence each other either directly and/or through unobserved factors. Both spatial and error lags can presumably be included in the model and later checked by an appropriate diagnostic test.

4.1. Model specification

Following [32], the regression model is specified with individual fixed effects in Eqs. (1) and (2) as follows:

$$Y_{it} = X_{it}\beta + \lambda W Y_{it} + \alpha_i + u_{it} \tag{1}$$

$$\begin{aligned} &u_{it} = \rho M u_{it} + \varepsilon_{it} \\ &i = 1, \dots, N \\ &t = 1, \dots, T \end{aligned} \tag{2}$$

where i and t respectively refer to the spatial unit and the year, Y is the $N\times 1$ matrix of the observed endogenous variables, X is the $N\times k$ matrix of the observed exogenous variables, W and M are the $N\times N$ spatial weighting matrices defining the dependence across the spatial units, α is the $N\times 1$ matrix of individual fixed effects, β is the $k-$ dimensional matrix of the regression parameters, λ and ρ are respectively the spatial autoregressive and the spatial autocorrelation parameters, u is the $N\times 1$ matrix of the disturbance terms, and ε is the $N\times 1$ matrix of innovation error terms. WY and Mu are respectively defined as the spatial and the error lags. They respectively express the impact of the observed and the unobserved effects of the injury rate of spatial unit i on the injury rates of neighboring spatial units.

A positive (negative) sign of the coefficient of each of these lags implies that higher injury rates in a given region correspond to higher (lower) injury rates in neighboring regions. This positive (negative) sign therefore indicates similarities/dissimilarities of crash rates between neighboring spatial units.

4.2. Model assumptions

The estimation of the model in Eqs. (1) and (2) lays on the existence of its log-likelihood function Ln(L). This function is specified as follows:

Plugging Eq. (2) into Eq. (1) gives Eq. (3)

$$Y_{it} = (I-\lambda W)^{-1} X_{it}\beta + (I-\lambda W)^{-1}\alpha_i + (I-\lambda W)^{-1}(I-\rho M)^{-1}\varepsilon_{it} \tag{3}$$

[1] The term "quasi" covers specifications where the actual distribution is permitted to differ from the normal distribution.

[2] Website: http://www.stats.gov.cn/english/statisticaldata/.

Table 1
Summary statistics of the variables.

Variables	Description	Min	Max	Mean	SD
ACC	Number of traffic accidents per 100,000 population	3.543	147.882	23.743	16.213
INJ	Number of injuries per 100,000 population	4.063	103.037	25.904	15.778
GRP	Gross regional product (100 million Yuan current price)	211.54	62,163.97	11,930.42	11,324.64
POP14	Population of age between 0 and 14 years	584,503.9	162,000,000	7,829,427	10,100,000
POP64	Population of age between 15 and 64 years	1,865,140	594,000,000	32,700,000	37,400,000
POP65	Population of age 65 years and above	146,323.8	100,000,000	4,237,699	6,065,621
UUNEMP	Percentage of registered urban unemployment	1.21	6.5	3.668	0.665
Passenger	Passenger traffic on highways (100 million passenger-1000 m)	14.8	2470.106	411.916	370.291
Freight	Freight traffic on highways (100 million ton-1000 m)	23.1	7266.771	952.697	1321.881
PAVEDR	Length of paved roads (10^3 m)	313	42,875	9014.138	8474.244

Letting $A=(I-\lambda W)^{-1}(I-\rho M)^{-1}$, Eq. (3) results in Eq. (4)

$$Y_{it}=(I-\lambda W)^{-1}X_{it}\beta+(I-\lambda W)^{-1}\alpha_i+A\varepsilon_{it} \tag{4}$$

The variance-covariance Ω of the regression is given in Eq. (5) as $\Omega=\mathrm{E}[(A\varepsilon_{it})(A\varepsilon_{it})']=AE[\varepsilon_{it}\varepsilon_{it}']\,A'=\sigma_\varepsilon^2 A\,A'$, therefore

$$\Omega=\sigma_\varepsilon^2(I-\lambda W)^{-1}(I-\rho M)^{-1}\left(I-\rho M'\right)^{-1}\left(I-\lambda W'\right)^{-1}=\sigma_\varepsilon^2 AA' \tag{5}$$

Then, the log-likelihood function Ln(L) of the regression is defined in Eqs. (6)–(9)

Letting $V_{it}=A\varepsilon_{it}=[Y_{it}-(I-\lambda W)^{-1}(X_{it}\beta+\alpha_i)]$,

$$\mathrm{Ln}(\mathrm{L})=-\frac{NT}{2}Ln\left(2\pi\sigma_\varepsilon^2\right)+\mathrm{TLn}|I-\lambda W| + T\mathrm{Ln}|I-\rho M|-\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\frac{1}{2\sigma_\varepsilon^2}\left(\frac{V_{it}}{A}\right)'\left(\frac{V_{it}}{A}\right)\right] \tag{6}$$

$$=-\frac{NT}{2}Ln\left(2\pi\sigma_\varepsilon^2\right)+\mathrm{TLn}|I-\lambda W| + T\mathrm{Ln}|I-\rho M|-\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\frac{1}{2\sigma_\varepsilon^2}(V_{it})'A^{-1'}A^{-1}(V_{it})\right] \tag{7}$$

$$=-\frac{NT}{2}Ln\left(2\pi\sigma_\varepsilon^2\right)+\mathrm{TLn}|I-\lambda W| + T\mathrm{Ln}|I-\rho M|-\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(V_{it})'\sigma_\varepsilon^{-2}(AA')^{-1}(V_{it})\right] \tag{8}$$

$$\mathrm{Ln}(\mathrm{L})=-\frac{NT}{2}Ln\left(2\pi\sigma_\varepsilon^2\right)+\mathrm{TLn}|I-\lambda W| + T\mathrm{Ln}|I-\rho M|-\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(V_{it})'\Omega^{-1}(V_{it})\right] \tag{9}$$

The log-likelihood function in Eq. (9) is based on the five following assumptions:

Assumption 1. The diagonal elements of W and M are all equal to zero.

Assumption 2. The innovation error terms are independently and identically distributed with mean 0, variance σ_ε^2 and their moments of higher than the fourth order exists.

Assumption 3. The matrices $(I-\lambda W)$ and $(I-\rho M)$ are invertible for $\lambda<1$ and $\rho<1$.

Assumption 4. W and M are uniformly bounded in absolute value in column and row sums.

Assumption 5. The matrix X has full column rank and its elements are uniformly bounded in absolute value.

By convention, self-neighbors (also called self-influence) are excluded in spatial analysis. Assumption 1 is in line with this convention as it implies that no spatial unit is its own neighbor. Assumption 2 provides independently and identically distributed regularity assumptions for the innovation term. Assumption 3 specifies the completeness of the model because it makes possible the computation of Y and u. Also, the inferiority of the spatial parameters to the unity is to ensure that the process is stationary. Assumption 4 restrains the spatial connection

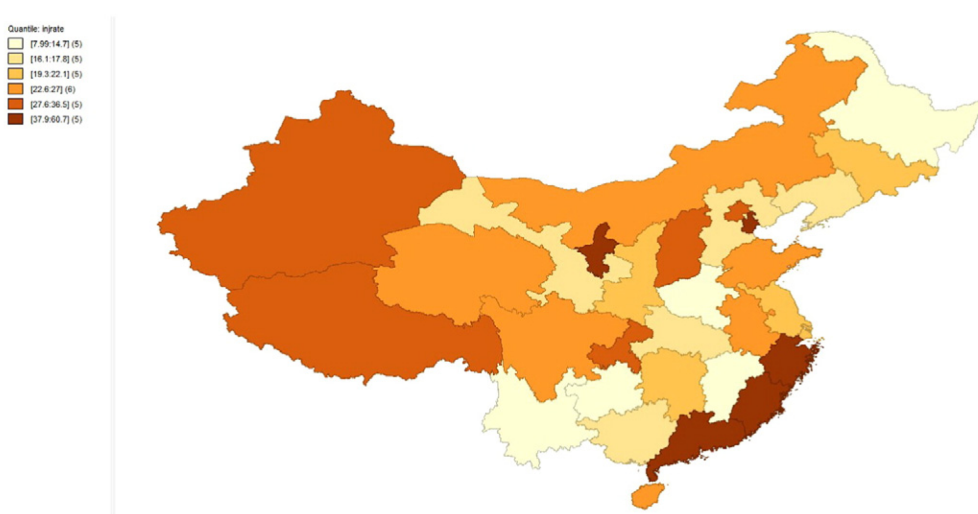

Fig. 1. Quantile map of the average injury rates over 2004–2013.

between Y and u to a controllable degree. Assumption 5 prevents perfect multicollinearity.

4.3. Tests for spatial effects

It is necessary to check the presence of the spatial effects prior to their incorporation in the model [20]. The most familiar statistic for this kind of test is the Moran's I score under the null hypothesis of absence of spatial dependence [20]. However the Moran's I cannot discriminate between spatial and error lags. The appropriate alternative is the Lagrange multiplier index and its robust version under the null hypothesis of no lag (either spatial or error, depending on the test) dependence [37–39]. All these statistics are based on the Ordinary least squares (OLS) estimates of the dependent variable and a spatial weight matrix.

However, both lags can be integrated together in the SARAR as proposed by [22] for cross-section studies and by [32] for panel studies. Belotti et al. [40] proposed the ML method to estimate spatial panel models and the AIC score to discriminate between the SEM, the SAR model and the SARAR.

In practice the spatial weight matrices W and M may or may not be equal [32]. In this study they are supposed to be equal. In spatial analysis, it is commonly intuitive to test the robustness of the results using more than one spatial weight matrix. However, because one of the regions (Hainan Province) in mainland China is an island, the contiguity matrices are out of choice restricting the analysis to a row-normalized inverse distance-based matrix.

Following [38], the inverse distance matrix W is a based on the specification of a minimum distance (d_{min}) to ensure that every spatial unit has at least one neighbor. Each element w_{ij} of W is defined as $w_{ij} = {}^{1}/_{d_{ij}}$ where $d_{ij} \geq d_{min}$, $i,j = 1, \ldots, N$, $i \neq j$ and d_{ij} is the distance between the centroids of locations i and j. As can be seen, the spatial effects of a location on another one decrease as the distance between them increases. The row-normalized form $\tilde{W}$ of W is $\tilde{w}_{ij} = {}^{w_{ij}}/_{\sum_{j=1}^{N} w_{ij}}$; by convention, $w_{ij} = \tilde{w}_{ij} = 0$ for $i = j$ due to the exclusion of self-neighborhood as explained in the assumptions section.

The Moran's I statistic for the two dependent variables accident rate and injury rate was found to be positive and statistically significant at the 5% significance level meaning respectively that there are similar crash rates in neighboring spatial units and the null hypothesis of absence of spatial dependence is rejected. This rejection suggests that the OLS model needs to be improved by incorporating the spatial effects in the analysis. The new estimations and their discussions are presented in the subsequent section.

5. Results

Altogether, four models were estimated using the Stata software [41]. Model 1 is a non-spatial one unlike the others. The Hausman test rejected the random effects in each of the models. So, all the models are fixed-effects estimations. The spatial models are fitted using the ML method by [40]. In the presence of fixed-effects, the ML provides consistent estimations for the regression parameters [32]. Model 2 is a SAR model, Model 3 is a SEM while Model 4 is combination of Models 2 and 3, say a SARAR model. The estimation results are given in Table 2.

The coefficients of population of age between 0 and 14 years (POP14) and urban unemployment rate (UUNEMP) are positive and not statistically significant in the non-spatial model (Model 1). These coefficients become negative and statistically significant when the spatial effects are incorporated in the model. It is believed that the non-spatial model has given the wrong results as the unemployment generates fewer trips meaning lower crash rates. The same explanation holds for children because they are mostly less likely to be exposed to traffic risks due to their inactivity in the economic activities. The obtained

Table 2
Fixed effects estimation results for crash rates.

Variables	Model 1 non-spatial panel		Model 2 panel SAR		Model 3 panel SEM		Model 4 panel SARAR	
	ACC	INJ	ACC	INJ	ACC	INJ	ACC	INJ
GRP	−0.002**	−0.0014**	−0.0011**	−0.001**	−0.0012**	−0.001**	−0.0012**	−0.0011**
	(−9.62)	(−11.42)	(−7.86)	(−8.90)	(−7.12)	(−8.31)	(−7.74)	(−8.89)
POP14	5.67e-07	1.80e-07	−1.04e-06*	−7.85e-07*	−1.65e-06**	−1.33e-06**	−1.47e-06**	−1.13e-06**
	(0.79)	(0.35)	(−1.78)	(−1.78)	(−2.90)	(−3.07)	(−2.47)	(−2.49)
POP64	−9.15e-07**	−9.02e-07**	−8.98e-07**	−7.94e-07**	−1.08e-06**	−8.97e-07**	−1.04e-06**	−8.62e-07**
	(−3.33)	(−4.54)	(−4.14)	(−4.83)	(−4.97)	(−5.46)	(−4.67)	(−5.15)
POP65	4.50e-06**	5.09e-06**	7.02e-06**	6.01e-06**	9.10e-06**	7.51e-06**	8.59e-06**	6.98e-06**
	(3.14)	(4.91)	(6.10)	(6.98)	(7.60)	(8.30)	(6.70)	(7.34)
UUNEMP	1.064	0.045	−3.871**	−3.897**	−2.856*	−2.667**	−2.984*	−3.028**
	(0.59)	(0.03)	(−2.62)	(−3.38)	(−1.75)	(−2.16)	(−1.88)	(−2.47)
Passenger	−0.014**	−0.012**	−0.007*	−0.009**	−0.007	−0.010**	−0.008*	−0.0103**
	(−2.70)	(−3.27)	(−1.77)	(−2.89)	(−1.48)	(−2.77)	(−1.71)	(−2.94)
Freight	0.002*	0.002**	0.002**	0.003**	0.002**	0.002**	0.002**	0.003**
	(1.94)	(2.64)	(3.00)	(4.42)	(2.44)	(4.21)	(2.63)	(4.37)
PAVEDR	0.002**	0.001**	0.0013**	0.0005*	0.001**	0.0004*	0.001**	0.0005*
	(5.15)	(3.51)	(4.25)	(1.93)	(4.15)	(1.84)	(4.35)	(1.96)
Spatial lag			0.722**	0.622**			0.538**	0.495**
			(12.25)	(9.52)			(3.66)	(4.51)
Error lag					0.818**	0.797**	0.521**	0.435**
					(18.92)	(16.37)	(2.91)	(2.69)
Constant	31.954**	43.667**						
	(4.26)	(8.06)						
Variance (e)			53.211**	30.511**	51.361**	29.438**	58.204**	33.306**
			(12.33)	(12.38)	(12.27)	(12.27)	(13.73)	(13.76)
Log-likelihood				−974.243	−1061.141	−973.873	−1059.476	−971.337
AIC	2235.279	2033.95	2143.802	1966.486	2140.281	1965.745	2138.953	1962.674
Panel length	10	10	10	10	10	10	10	10
Groups	31	31	31	31	31	31	31	31
Observations	310	310	310	310	310	310	310	310

In parenthesis are the t-statistics.
** Statistically significant at the 0.05 level.
* Statistically significant at the 0.10 level.

expected results in the spatial model are surely due to the consideration of spatial interactions which are inevitable between regions. Therefore, ignoring the spatial effects would have produced misleading estimates. The performances of all models are compared using the AIC scores. For each of the two dependent variables, the non-spatial model (Model 1) offers the highest AICs meaning that this model is better than none of the spatial ones. Then, the SEM is better than the SAR model. Model 4, the SARAR model provides the lowest AICs. Thus, it is the best model for explaining the variations in crash rates in the different spatial units. Therefore, the discussions of the results of this model are given in what follows.

6. Discussion

In Model 4, each variable is statistically significant and has the same sign in both equations. The coefficient of the spatial lag is positive indicating that neighboring regions tend to displays similar patterns in terms of accident and injury rates. The coefficient of the error lag is also positive in both equations suggesting the existence of common unobserved factors affecting accident and injury rates in neighboring regions. These findings are consistent with previous studies.

The Gross Regional Product and the urban unemployment rate are connected to lower crash rates. This is intuitive because regions with higher incomes are more likely to have more resources for investment in road safety in order to get facilities such as roads of good quality, safer vehicles, efficient regulatory agencies and emergency medical facilities. There is a large clustering of employment in cities generating more trips, so fewer trips thus fewer accidents are expected from the unemployment in these areas.

Passenger traffic is associated with lower crash rates unlike freight traffic. The bus is the most common mode used for passenger transport and is traditionally involved in fewer accidents compared to other modes of passenger transport. It also helps to improve road safety by decreasing traffic congestion. However, it should also be noted that because of its large capacity, the bus is associated with substantial property losses during crashes. As far as freight traffic is concerned, freight vehicles are life threatening than the other vehicles. Similar scheme is likely to be present in China where commercial vehicles are still great perpetrators of crashes [16] and the majority of the victims are the vulnerable road users (pedestrians, bicyclists, and motorcyclists).

The length of paved roads is correlated with higher crash rates. This can be attributed to the fact that good infrastructures foster faster driving and therefore increase the probability of more crashes and injuries. In line with this result, speeding has been reported among the major causes of crashes in China [16,42,43].

The population aged between 0 and 14 years negatively influences crash rates while the one of age 65 years and above positively influences these rates. Although an increased involvement in accidents of old populations may be unexpected, these accidents may results in more injuries because of the fragility of these populations [44]. Populations in the age range of 15–64 years are correlated with lower crash rates. This result is unexpected because populations in this age group are the most exposed to traffic risks due to their high involvement in economic and social activities. They are exposed as pedestrians, cyclists, motorcyclists, passengers, drivers and drinking-prone road users. For example, those in the age range 40–49 are vulnerable because of their high car ownership rates [45] while the range 16–25 is at higher risks as motorcyclists [18]. A split of the group based on the gender and/or into at least 4 standard categories, such as 15–24, 25–34, 35–44 and 45 and above would allow a closer inspection of the features of each sub-group [46].

7. Conclusions

Spatial analysis of road crash rates is widespread. However, the evaluation of the neighborhood effects (spatial and error lags) of these rates is still embryonic. The quasi-totality of these evaluations is performed within the framework of single-equation spatial cross-sectional studies. The contribution of this study is the extension of these studies to a spatial panel analysis with application to the 31 administrative regions of mainland China. A spatial autoregressive model with autoregressive disturbances (SARAR) is estimated using the maximum likelihood method. The findings support that ignoring the spatial spillovers in the analysis of road traffic crash data produces bias estimates. A possible extension of this study is the replication at lower scales or the consideration of dynamic spatial panels.

The presence of neighborhood effects is evidenced through the Moran's I statistic. Models incorporating neighborhood effects outweigh those that do not and the model incorporating both types of effects is found to be the best to fit the data. The coefficient of the spatial lag is positive and statistically significant suggesting a similarity of crash rates configuration between neighboring regions. The coefficient of the error lag is also positive and statistically significant implying the existence of common unobserved factors affecting crash rates in neighboring regions. These findings are consistent with previous studies. A policy implication of the findings is that neighboring regions should cooperate in the design of their road safety policies so that to mitigate the resulting negative spatial effects and benefit from the positive ones.

The analysis also provides the impact of other covariates on crash rates. The Gross Regional Product and the urban unemployment rate are correlated to lower crash rates. Populations of age 65 and above are positively associated with higher rates. Passenger traffic and freight traffic are respectively related to lower and higher rates. The length of paved roads is unexpectedly the source of higher rates; speeding is probably the reason.

There are however some limitations to this study due to the underreporting gaps which are well-known in traffic crash data and are likely to be more severe in China [43,47]. The Chinese official crash statistics are collected by the police. Hu et al. [19] and Qiu et al. [48] compared data based on police reports to those from the hospitals. Their studies revealed large inconsistencies due to the underreporting of data from the police.

Funding

None declared.

Conflicts of interest

None declared.

Ethical approval

None sought.

Acknowledgments

The authors are very thankful to the anonymous reviewers for their invaluable comments and suggestions to improve the content and presentation of the manuscript. However, the authors remain responsible for the facts and accuracy of the data presented herein.

References

[1] J. Whitelegg, A geography of road traffic accidents, Trans. Inst. Br. Geogr. (1987) 161–176.
[2] J. Aguero-Valverde, P. Jovanis, Analysis of road crash frequency with spatial models, Transp. Res. Rec. (2008) 55–63.
[3] X. Wang, K. Kockelman, Maximum simulated likelihood estimation with spatially correlated observations: a comparison of simulation techniques, RSAI's 53rd Annual Meeting, Toronto (in 2006), and Forthcoming in Transportation Statistics, 2008.
[4] C. Wang, M. Quddus, S. Ison, A spatio-temporal analysis of the impact of congestion on traffic safety on major roads in the UK, Transp. Sci. 9 (2013) 124–148.

[5] B. Fingleton, J. Le Gallo, Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: finite sample properties, Pap. Reg. Sci. 87 (2008) 319–339.
[6] W.R. Tobler, A computer movie simulating urban growth in the Detroit region, Econ. Geogr. 46 (1970) 234–240.
[7] N. Levine, K.E. Kim, L.H. Nitz, Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators, Accid. Anal. Prev. 27 (1995) 675–685.
[8] E. Papadimitriou, V. Eksler, G. Yannis, S. Lassarre, Modelling the spatial variation of road safety in Greece, Proceedings of the Institution of Civil Engineers-Transport 2013, pp. 49–58.
[9] E.A. LaScala, D. Gerber, P.J. Gruenewald, Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis, Accid. Anal. Prev. 32 (2000) 651–658.
[10] M.A. Quddus, Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data, Accid. Anal. Prev. 40 (2008) 1486–1497.
[11] J. Hong, S. Lee, J. Lim, J. Kim, Application of spatial econometrics analysis for traffic accident prediction models in urban areas, Proceedings of the Eastern Asia Society for Transportation Studies, 2013.
[12] X. Wang, K.M. Kockelman, Specification and estimation of a spatially and temporally autocorrelated seemingly unrelated regression model: application to crash rates in China, Transportation 34 (2007) 281–300.
[13] B. Baltagi, Econometric Analysis of Panel Data, John Wiley & Sons, 2008.
[14] J.P. Elhorst, Spatial panel data models, Spatial Econometrics, Springer 2014, pp. 37–93.
[15] C. Hsiao, Analysis of Panel Data, Cambridge University Press, 2014.
[16] Y. Li, J. Zhang, G. Zhang, Situations and challenges of road safety in China, Road Safety on Four Continents: 16th International Conference, 2013.
[17] S. Wang, Y. Li, G. Chi, S. Xiao, J. Ozanne-Smith, M. Stevenson, et al., Injury-related fatalities in China: an under-recognised public-health problem, Lancet 372 (2008) 1765–1773.
[18] X. Zhang, H. Xiang, R. Jing, Z. Tu, Road traffic injuries in the People's Republic of China, 1951–2008, Traffic injury prevention 12 (2011) 614–620.
[19] J. Qiu, J. Zhou, L. Zhang, Y. Yao, D. Yuan, J. Shi, et al., Chinese traffic fatalities and injuries in police reports, hospital records, and in-depth records from one city, Traffic injury prevention 16 (2015) 565–570.
[20] L. Anselin, Spatial Econometrics: Methods and Models, Kluwer Academic Publishers, Dorddrecht, 1988.
[21] L. Anselin, Spatial externalities, spatial multipliers, and spatial econometrics, Int. Reg. Sci. Rev. 26 (2003) 153–166.
[22] H.H. Kelejian, I.R. Prucha, A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, J. Real Estate Financ. Econ. 17 (1998) 99–121.
[23] J.P. Elhorst, Specification and estimation of spatial panel data models, Int. Reg. Sci. Rev. 26 (2003) 244–268.
[24] J. LeSage, R.K. Pace, Introduction to Spatial Econometrics, CRC Press, New York, 2009.
[25] K. Ord, Estimation methods for models of spatial interaction, J. Am. Stat. Assoc. 70 (1975) 120–126.
[26] L. Su, Z. Yang, QML estimation of dynamic panel data models with spatial errors, J. Econ. 185 (2015) 230–258.
[27] H.H. Kelejian, G. Piras, AJ test for dynamic panel model with fixed effects, and nonparametric spatial and time dependence, Empir. Econ. (2016) 1–25.
[28] P. Čížek, J.P. Jacobs, J. Ligthart, H. Vrijburg, GMM Estimation of Fixed Effects Dynamic Panel Data Models with Spatial Lag and Spatial Errors, 2015.
[29] W. Xie, Spatial Panel VAR and Application to Forecast Influenza Incidence Rates of US States, 2015 (Available at SSRN 2646870).
[30] M. Kapoor, H.H. Kelejian, I.R. Prucha, Panel data models with spatially correlated error components, J. Econ. 140 (2007) 97–130.
[31] J. Mutl, M. Pfaffermayr, The Hausman test in a Cliff and Ord panel model, Econ. J. 14 (2011) 48–76.
[32] L.-f. Lee, J. Yu, Estimation of spatial autoregressive panel data models with fixed effects, J. Econ. 154 (2010) 165–185.
[33] Z. Yang, J. Yu, and S. F. Liu. (April 22nd, 2015). Bias correction for fixed effects spatial panel data models. Available: http://ink.library.smu.edu.sg/soe_research/1754.
[34] M.H. Pesaran, E. Tosetti, Large panels with common factors and spatial correlation, J. Econ. 161 (2011) 182–202.
[35] L.f. Lee, J. Yu, Spatial panels: random components versus fixed effects, Int. Econ. Rev. 53 (2012) 1369–1412.
[36] B.H. Baltagi, P. Egger, M. Pfaffermayr, A generalized spatial panel data model with random effects, Econ. Rev. 32 (2013) 650–685.
[37] L. Anselin, A.K. Bera, R. Florax, M.J. Yoon, Simple diagnostic tests for spatial dependence, Reg. Sci. Urban Econ. 26 (1996) 77–104.
[38] L. Anselin, Exploring spatial data with GeoDaTM: a workbook, Urbana 51 (2004) 61801.
[39] L. Anselin, N. Lozano-Gracia, Errors in variables and spatial effects in hedonic house price models of ambient air quality, Empir. Econ. 34 (2008) 5–34.
[40] F. Belotti, G. Hughes, A.P. Mortari, XSMLE-A command to estimate spatial panel models in Stata, Material From the 2013 German Stata Users Group Meeting, 2013.
[41] StataCorp, Stata Release 14. Statistical Software, StataCorp LP, College Station, TX, 2015.
[42] S. Zhao, Road traffic accidents in China, IATSS Research 33 (2009).
[43] X. Zhang, H. Yao, G. Hu, M. Cui, Y. Gu, H. Xiang, Basic characteristics of road traffic deaths in China, Iranian J. Public Health 42 (2013) 7.
[44] J.I. Castillo-Manzano, M. Castro-Nuño, X. Fageda, Are traffic violators criminals? Searching for answers in the experiences of European countries, Transp. Policy 38 (2015) 86–94.
[45] C.-W. Wang, C.L. Chan, Estimated trends and patterns of road traffic fatalities in China, 2002–2012, Traffic Inj. Prev. (2016) 1–6.
[46] W. Odero, P. Garner, A. Zwi, Road traffic injuries in developing countries: a comprehensive review of epidemiological studies, Tropical Med. Int. Health 2 (1997) 445–460.
[47] T. Alcorn, Uncertainty clouds China's road-traffic fatality data, Lancet 378 (2011) 305–306.
[48] G. Hu, T. Baker, S.P. Baker, Comparing road traffic mortality rates from police-reported data and death registration data in China, Bull. World Health Organ. 89 (2011) 41–45.