Accident Severity Levels and Traffic Signs Interactions in State Roads: A Seemingly Unrelated Regression Model in Unbalanced Panel Data Approach

Xuecai Xu^a, Željko Šarić^{b*}, Feng Zhu^a, Dario Babić^c

^a School of Civil and Environmental Engineering, Nanyang Technological University, Singapore

^b Department of Traffic Accident Expertise, Faculty of Transport and Traffic Sciences, University of Zagreb, Croatia

^c Department of Traffic Signaling, Faculty of Transport and Traffic Sciences, University of Zagreb, Croatia

*Corresponding author

Abstract: This study intended to investigate the interactions between accident severity levels and traffic signs in state roads located in Croatia, and explore the correlation within accident severity levels and heterogeneity attributed to unobserved factors. The data from 410 state roads between 2012 and 2016 were collected from Traffic Accident Database System maintained by the Republic of Croatia Ministry of the Interior. To address the correlation and heterogeneity, a seemingly unrelated regression (SUR) model in unbalanced panel data approach was proposed, in which the seemingly unrelated model addressed the correlation of residuals, while the panel data model accommodated the heterogeneity due to unobserved factors. By comparing the pooled, fixed-effects and random-effects SUR models, the random-effects SUR model showed priority to the other two. Results revealed that (1) low visibility and the number of invalid traffic signs per km increased the accident rate of material damage, death or injured; (2) average speed limit exhibited a high accident rate of death or injured; (3) the number of mandatory signs was more likely to reduce the accident rate of material damage, while the number of warning signs was significant for accident rate of death or injured.

Keywords: Accident Severity Level; Traffic Sign; Seemingly Unrelated Regression; Panel Data

Introduction

With the development of economy and society, a variety of roadside signs, e.g. advertising signs, neon lights, and gaudy billboards are emerging and dominating the visual landscape in many urban and suburban areas, which makes the roadside environment changeable and complicated. Although there are some studies focusing on the impact of such development, limited studies have investigated the relationship between the signs and traffic safety in Croatia. Among these, traffic signs, primary means of communication between the road authorities and traffic participants, play an important role as a principal causative factor by affecting drivers' reaction time and road safety.

Currently a number of studies concentrate on the traffic sign detection, e.g. the color, the size, and lettering of road signs (Villalón-Sepúlveda et al. 2017; Vilchez, 2017), and although there are some studies (Carson and Mannering, 2001; Porter et al., 2007; Wu and Wang, 2013) attempting to establish the relationship between accidents and traffic signs, merely one type of specific traffic signs, e.g. warning signs, speed-limit signs, was investigated, which might not completely reflect the effectiveness of traffic signs and safety problem due to different categories of traffic signs.

During the last decade a variety of different approaches and perspectives (Lord & Mannering, 2010; Savolainen et al., 2011; Mannering & Bhat, 2014) have been presented in safety evaluation, involving accident frequency, accident severity or both as the function of influencing factors. However, different severity levels might be related, as possible correlations among different severity levels may exist. If dependent variables are co-dependent, and independent variables include endogenous and exogenous ones, a simultaneous equation model (SEM) is appropriate, whereas the variables are simply correlated in their regression error terms, a seemingly unrelated regression (SUR) approach becomes more reasonable (Wang and Kockelman, 2007; Washington et al. 2010). The seemingly unrelated regression estimation approach proposed by Zellner (1962) employs the correlation between disturbances in different equations to improve the efficiency of joint estimation (You and Zhou, 2014).

Moreover, various studies have indicated that heterogeneity issue can be addressed through finite mixture regression models (Park and Lord, 2007), and random parameter model (Anastasopoulos and Mannering, 2009; El-Basyouny and Sayed, 2009), in which the heterogeneity from the data or locations caused by unobserved factors was accommodated, and the estimation results and the statistical inferences were improved (Kim et al. 2010; Anastasopoulos and Mannering, 2011, Anastasopoulos et al. 2012; Abay 2013; Aziz et al. 2013; Mannering et al. 2016; Anastasopoulos, 2016; Zeng et al. 2017; Fountas and Anastasopoulos, 2017). One way to deal with the heterogeneity issue is to adopt panel data, permitting more efficient estimation of parameters (You and Zhou, 2014), which has attracted considerable research interest in transportation safety.

To take advantage of the SUR approach and panel data modeling, a panel data SUR model was proposed combining the SUR approach with panel data modeling, and some scholars (Avery, 1977; Biørn, 2004; You and Zhou, 2014) have subsequently studied the statistical inference and applications of the panel data SUR model. Thus, the purpose of this study is to apply panel data SUR model to analyze different severity levels as functions of a series of traffic signs, in which the correlations between disturbances in different equations can be addressed, as well as the heterogeneity issue among different roadways.

Literature Review

There have been a variety of studies about the relationship between accidents and traffic signs during the last thirty years, but they are investigated from different perspectives. Initially Loo (1979) discussed the role of primary personality factors in the perception of traffic signs, driver violations and accidents. The findings indicated that the fast decision-time component of personality factors carried the relationships with accidents and the ability to perceive embedded traffic signs in the verbal message and symbolic message types. As for the warning signs, Jørgensen and Wentzel-Larsen (1999) optimized the use of warning signs in traffic. It was concluded that warning signs would increase safety and have a greater positive impact on total driving costs than on accident costs. Similarly, Carson and Mannering (2001) investigated the effect of ice warning signs on ice accident frequencies and severities. It was found that an ice warning sign would be one way of understanding the impact of warning signs, and ice-warning signs were more likely to be placed at locations with high numbers of ice-related accidents. From the freeway point of view Wu and Wang (2013) explored the impact of speed-limit signs on freeway work zones. The speed-limit model was established by considering the relevant factors

1 of driver's field of vision and transportation distance of visual cognition. The results showed the
2 proposed speed-limit model within warning area played an important role for freeway traffic
3 safety, while Xuan and Kanafani (2014) evaluated the effectiveness of accident information on
4 freeway changeable message signs (CMS) by comparing different aggregate analysis
5 methodologies. The findings showed that CMS accident messages had no significant effect on
6 driver diversion but visible congestion is an important factor. On the other side, if too many signs
7 are installed, the effect may be opposite. Strawderman et al. (2015) investigated the effect of
8 signs on driver behavior and accident frequency in school zones. It was found that sign saturation
9 had a significant effect on vehicle speed, compliance and accident frequency. All of the
10 literatures abovementioned have not indeed built up the actual relationship between and traffic
11 signs except the latest study by Babić et al. (2014, 2017). They analyzed traffic signs in terms of
12 traffic safety in Croatia, and found out that improper installation and maintenance of traffic signs
13 can affect the traffic safety, which is the base of this study.

14 So far there have been different applications of SUR model in different areas (Zellner, 1962;
15 Moon and Perron, 2006; Zhou and Kockelman, 2009; You and Zhou, 2014; Kim et al., 2015),
16 but the application in transportation safety area is very limited. Initially, Porter et al. (2007)
17 investigated relationships between speed behavior, roadway and roadside geometrics and traffic
18 control in work zones using seemingly unrelated regression (SUR) model. The SUR model
19 accounted for contemporaneous correlation of the disturbance terms in 85th percentile passenger
20 car speed and standard deviation of passenger car speed. Then Prozzi and Hong (2008) examined
21 the performance of a transportation infrastructure through seemingly unrelated regression models.
22 The results demonstrated international roughness index and rut depth were associated with
23 different mechanisms, and model estimation was efficient when accounting for the correlation
24 among the models.

25 Subsequently, SUR model was extended to various forms by integrating different models from
26 cross-sectional to panel data model. Wang and Kockelman (2007) firstly proposed a spatially and
27 temporally auto-correlated seemingly unrelated regression model and exemplified the accident
28 rates in China. The results showed that investment in transportation infrastructure was estimated
29 to have statistically significant effects on reducing severe accident rates, and the incorporation of
30 spatial and temporal effects improved the model. Then Zhou and Kockelman (2009) predicted
31 the distribution of households and employment with seemingly unrelated regression model with
32 two spatial processes. The results demonstrated that SUR estimation of land-use intensities from
33 parcel-level data with two types of spatial dependence was feasible and meaningful. Later Kim et
34 al. (2015) evaluated the impact of a household head's telecommuting on household travel while
35 controlling for the interdependence within a household and across travel purposes by applying
36 seemingly unrelated censored regression models in the Seoul Metropolitan Area. The analysis
37 showed that when vehicle travel budgets of a given household were limited, the compensatory
38 travel mechanism can make optimum use of limited resources, but offset the travel-substituting
39 effect of telecommuting. From the perspective of economics, the latest study by Mao (2016)
40 analyzed relationship between capital income, economic growth and income inequality. The
41 seemingly unrelated regression with fixed effects panel data model and instrumental variable
42 estimation technique were employed, which provides the foundation for this study theoretically,
43 and the findings suggested that although capital income taxes hindered economic growth, they
44 achieved the purpose of income redistribution.

Summarized by the abovementioned literatures, employing seemingly unrelated regression with panel data model to investigate different severity levels as functions of a series of traffic signs would be initial attempt in transportation area, which can address the correlations of severity levels and heterogeneity issue of different state roads.

Data Description

Our dataset integrated the data of Traffic Accident Database System from 2012 to 2016 with the traffic signs maintained by the Republic of Croatia Ministry of the Interior. Four main components from the Traffic Accident Database System were included: the accident consequence, the traffic signs profiles, the accident environment, and roadway characteristics.

About 135 state roads were elaborately selected from Croatia in five years. As shown in Figure 1, totally there were 665 state roads, but after removing some roads without data and processing, 410 valid state roads were determined for the analysis. In Croatia, the consequences of accidents are typically categorized as death, injured, or material damage. In our sample, the death cases only accounted for 2.1%. Given that the two adjacent accident categories were quite similar, merging the death and injured categories was not expected to substantially affect the inference. Consequently, the dependent variables in the proposed model were bivariate accident severity levels in which the response of interest referred to death and injured (DI), and material damage (MD) was treated as the contrast. As required by the SUR model, dependent variables should be continuous, thus accident rate (expressed as accidents per million vehicles miles travelled (MVMT)) is introduced to represent MD and DI. The way that accident rate is considered is because incorporates the effect of volume and road length, it is more adequate to measure the accident risk faced and perceived by individual drivers than accident frequency, which is highly related to the traffic volume (Schultz et al. 2008). Thus, the variables road length, traffic volume and the number of accidents cannot be included in the independent variables to avoid the estimation bias.

By aggregating the traffic signs and accident environment profiles, the explanatory variables reflecting the traffic signs including classification, types, the number of different traffic signs, and the proportion of all traffic signs, the environmental factors (i.e., visibility, weather conditions), the roadway and traffic characteristics (i.e., road type, speed limit, road length and annual average daily traffic (AADT)), and demographic characteristics of passengers (i.e., average, minimum and maximum age) were extracted.

Classification of traffic signs in terms of their function and retroreflective material is the major assortment. In Croatian regulations, traffic signs are, by their function, classified into the followings: A-warning signs, B-mandatory signs, C-information signs, D-directional signs, E-additional panels and K-traffic equipment; while in Europe retroreflective materials are divided into materials of Classes I, II and III. Materials of Classes I and II use spherical or prismatic retroreflective sheeting, while materials of Class III use exclusively prismatic retroreflective sheeting (Babic et al, 2014; 2017). Both classifications with valid and non-valid numbers and proportions were collected for the five years. Figure 2 shows one segment of state road D1 with valid and non-valid signs. The blue marks represent technically correct signs while the red marks are technically defective signs.

Figure 1 Selected State Road in Croatia

Figure 2 State Road D1 with Traffic Signs

Table 1 Summary of the parameters

Factor	Attribute	Count(proportion)	
State roads	2012	131(19.7%)	
	2013	131(19.7%)	
	2014	131(19.7%)	
	2015	131(19.7%)	
	2016	141(21.2%)	
Accident consequence	Death or injured	8613(40.7%)	
	Material damage	12562(59.3%)	
Traffic signs	Class I	381540(65.3%)	
	Class II	161079(27.5%)	
	Class III	41664(7.2%)	
Environmental factors		Accidents (proportion)	
Visibility	Day	13513(67.5%)	
	Low visibility for MD	4063(20.3%)	
	Low visibility for DI	2439(12.2%)	
Weather	Clear sky	12902(64.5%)	
	Cloudy	4043(20.2%)	
	Bad weather	3039(15.3%)	
	(rain/fog/snow/other)		
	Range	Mean	S.D.
Accident rate			
Material damage (MD)	Min:0.022; Max:19.56	0.78	1.22
Death or injured (DI)	Min:0.017; Max:7.95	0.56	0.80
Traffic Signs			
Number of warning signs	Min:6; Max:2342	262.47	32.86
Number of mandatory signs	Min:10; Max:4457	367.39	57.06
Number of information signs	Min:4; Max:4224	40.66	56.50
Number of directional signs	Min:2; Max:973	73.12	11.63
Number of additional panels	Min:0; Max:942	87.73	13.18
Number of traffic equipment	Min:0; Max:2339	230.83	30.57
Number of traffic signs in Class I	Min:27; Max:10436	932.86	133.95
Number of traffic signs in Class II	Min: 3; Max: 4030	393.84	55.93
Number of traffic signs in Class III	Min: 0; Max:890	101.87	12.73
Number of valid traffic signs per km	Min:5.39; Max:72.23	18.32	8.02
Number of invalid traffic signs per km	Min:0.28; Max:15.97	4.31	2.26
Traffic characteristics			
Speed limit, km/hr	Min:28; Max:98	62.54	10.38
Road length, km	Min:1.5; Max:644.6	66.73	8.92
AADT	Min:229; Max:14637	3546.61	227.99
Driver's characteristics			
Average age for MD	Min:20; Max: 68	42.3	6.12
Average age for DI	Min:19; Max: 65	40.7	6.69

3 The environmental factors collected include visibility and weather conditions. In terms of
4 visibility degree, the number of MD and DI was categorized for high visibility (i.e. daytime) and
5 low visibility (including night, twilight and dawn). Similarly, the atmospheric conditions were
6 collected according to the weather variation, and three types were classified, clear sky, cloudy
7 and bad weather conditions (including rain, fog, snow, salty road and other).

The variables used for model development are displayed in Table 1, with the proportions of the categorical variables before and the descriptive statistics of the continuous variables in the following.

Methodology

Pooled seemingly unrelated regression model

SUR model is a system of linear equations with errors that are correlated with across equations for a given individuals but are uncorrelated across individuals. The model consists of $j=1 \dots m$ linear regression equations for $i=1 \dots N$ individuals. Consider a set of separate regressions expressed as:

$$y_j = X_j \beta_j + u_j \quad (1)$$

where y_j and u_j are n -dimensional vectors, β_j is the p_j -dimensional parameter vector, and X_j is the $n \times p_j$ matrix of covariates. If the m equations are taken together, it can be described as:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & & X_m \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix} \quad (2)$$

Here the error terms are assumed to have zero mean and to be independent across individuals and homoskedastic, in which u_j satisfies the following assumptions:

- Mean of error terms: $E(u_j|X) = 0$;
- Variance of error term in equation j : $E(u_j u_j' | X) = \sigma_{jj} I_N$
- Covariance of error terms across equations j and j' : $E(u_j u_{j'}' | X) = \sigma_{jj'} I_N$ where $j \neq j'$
- Overall variance-covariance matrix: $\Omega = E(uu') = \sum \otimes I_N$

Generally, for single linear regression models, ordinary least square (OLS) estimation for each equation yields a consistent estimator of β , but the optimal estimator is generalized least square (GLS) estimator:

$$\hat{\beta}_{GLS} = \{X'(\Sigma^{-1} \otimes I_N)X\}^{-1} \{X'(\Sigma^{-1} \otimes I_N)y\} \quad (3)$$

with $Var(\hat{\beta}) = \{X'(\Sigma^{-1} \otimes I_N)X\}^{-1}$.

Thus, the estimation procedures include two steps:

- First step: each equation is estimated by OLS, and the residuals from the m equations are used to estimate Σ using $\hat{u}_j = y_j - X_j\hat{\beta}_j$ and $\hat{\sigma}_{jj} = \frac{\hat{u}_j\hat{u}_j'}{N}$

- Second step: substitute $\hat{\Sigma}$ for Σ of GLS estimator:

$$\hat{\beta}_{GLS} = \{X'(\hat{\Sigma}^{-1} \otimes I_N)X\}^{-1} \{X'(\hat{\Sigma}^{-1} \otimes I_N)y\} \quad (4)$$

As for the cross-equation testing, we can test whether the coefficients are jointly significantly different from zero $\beta_j = \beta_j = 0$ or whether the coefficients are significantly from each other $\beta_j = \beta_{j'}$. More details can be referred to Moon and Perron (2006).

Panel data seemingly unrelated regression model

The SUR model in the context of unbalanced panel data using random effect estimators is based on constructing a multistep (stepwise) algorithm using Generalized Least Squares (GLS) and the Maximum Likelihood (ML) procedures. The method was originally developed by Biørn (2004).

Consider the system of G equation models:

$$\begin{aligned} y_{git} &= X_{git}\beta_g + u_{git} & i=1...N, t=1...T_i, g=1...G \\ u_{git} &= v_{gi} + e_{git} \end{aligned} \quad (5)$$

where v_{gi} are unobserved individual-level effects in the g^{th} equation; e_{git} are the observation-specific errors in the g^{th} equation; X_{git} is a vector of strictly exogenous covariates in the g^{th} equation; β_g are vectors of parameters to be estimated in the g^{th} equation; Specifically, in this study y_{git} represents the vector of accident rate for MD and DI levels, X_{git} represents the vector of exogenous variables in the two equations, β_g are vectors of parameters to be estimated, v_{gi} are unobserved state-roads-level effects, and e_{git} are the observation-specific errors.

Observations in the unbalanced dataset are observed in at least one and at most P periods, and let N_p denotes the number of state roads observed in p periods, where $p = 1, \dots, P$. Hence, we can rearranged the unbalanced dataset in the way that the N_1 roads observed once come first, the N_2 roads observed twice come second, N_3 roads observed three times come third, and etc. In this

way, the set of N_I roads is considered as a cross section, and the sets of N_p roads where $p = 2, \dots, P$ as balanced panels.

For each g^{th} equation in the system, the model can be written as follows. Here, t is a sequence index, not a time index.

$$y_{git} = X_{git}\beta_g + v_{gi} + e_{git} \quad i=1\dots N, t=1\dots T_i, g=1\dots G \quad (6)$$

Observations in the unbalanced panel are rearranged into P balanced panels. In each balanced panel, all observations are observed with the same number of periods. Therefore, the model can be rewritten for each p in P :

$$y_{i(p)} = X_{i(p)}\beta + e_p \# v_i + e_{i(p)} \quad (7)$$

where e_p is vector of ones ($p \times 1$); $\#$ is the Kronecker direct product.

Using the overall within-individual (within-roads in this study) and between-individual (between-roads in this study) covariation matrices, Biørn (2004) derived the unbiased estimators of σ_v and σ_e for each balanced panel. By using those estimated covariance matrices, the GLS problem is considered by minimizing the usual sandwich form with respect to parameter estimates. Then the beta GLS estimator (β_{GLS-p}) is obtained for the state roads observed p times.

The overall GLS estimator can be shown to be the function of β_{GLS-p} and their variances $V(\beta_{GLS-p})$ for $p = 1, \dots, P$.

In order to get the efficient estimator of the SUR system, the multistep (stepwise) Maximum Likelihood estimation is implemented. The multistep is the problem of two sub-problems: (a) maximizing the log likelihood with respect to β parameters for given σ_v and σ_e , which is the same as the GLS part above, and (b) maximizing the log likelihood with respect to σ_v and σ_e for given β parameters. The multistep algorithm jointly solves (a) and (b), and it will stop until convergence of the overall estimates.

Measures of goodness of fit

Followed the McElroy (1977), R square is a measure of goodness of fit for SUR model, which gives the correlation of the residuals. This kind of R square is a monotonic transformation of the F statistic from the test that all the coefficients in the system are zero. It is bounded on $[0, 1]$, so it has the same kind of interpretation as the ordinary R square:

$$R^2 = 1 - \frac{U'W*U}{Y'W*Y} \quad (8)$$

where U is the stacked vector of residuals, Y is the stacked vector of dependent variables and W is the variance-covariance matrix of residuals. If the vector of residuals is close to zero, R square is equal to 1; if the model has no explanatory power, then $U=Y$, so R square is equal to 0.

Moreover, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are introduced for the number of predictors along with the likelihood values of the model. The lower the AIC and BIC values, the better the statistical fit of the model. The AIC and BIC can be estimated from

$$\begin{aligned} AIC &= -2\log(L) + m \\ BIC &= -2\log(L) + m\log n \end{aligned} \quad (9)$$

where L is the likelihood of the data given the proposed model, m is the number of parameters and n is the number of observations.

Results and Discussion

As stated in the modeling, SUR model is employed to evaluate the interactions between accident rate and traffic signs. Since each year the selected roadways are not the same as last year, the data included are not from the same observed locations from 2012 to 2016, thus unbalanced panel data is used to address the heterogeneity issue among the roadways. Table 2 gives the estimated results for pooled, fixed-effects and random-effects seemingly unrelated regression models, and 95% confidence intervals for statistically significant variables obtained from software STATA 15.

Before the proposed model estimation, the correlation test is conducted to avoid the multicollinearity issues among the independent variables. The test result shows that visibility in daytime is highly related to average speed limit, number of traffic signs in Class I, II and III, while the average speed limit is highly related to clear sky, cloudy and bad weather conditions, thus they are not adopted at the same time as the independent variables. Other variables, such as number of information signs, number of directional signs, number of additional panels, average age of drivers, etc., didn't show up in the results because they are not significant for the MD and DI accident rate.

Shown from Table 2, it can be found that the correlation for the residuals (R square) in the accident rate for pooled, fixed-effects and random-effects SUR models are 0.756, 0.765 and 0.771, respectively, and we can reject the hypothesis that this correlation is zero, meaning that there indeed exists correlation between two accident rate equations of MD and DI. Moreover, the R square in random-effects SUR model is larger than that in pooled and fixed-effects ones. Additionally, the AIC and BIC values in random-effects SUR model are smaller than those in pooled and fixed-effects ones. Therefore, it can be concluded that the performance of random-effects SUR model is better than that of pooled and fixed-effects ones.

1 **Table 2 Estimation Results for Pooled, Fixed-effects and Random-effects SUR Models**

Variable	Pooled				Fixed-effects				Random-effects			
	Coef.	Std. Err.	z	Elasticity	Coef.	Std. Err.	z	Elasticity	Coef.	Std. Err.	z	Elasticity
Accident Rate of MD												
Low visibility	0.021*	0.005	4.73*	0.014	0.022*	0.005	4.74*	0.016	0.014*	0.002	5.86*	0.017
Average speed limit	0.010*	0.005	2.01*	0.001	0.011*	0.005	2.02*	0.001	0.001	0.001	0.58	
No. of mandatory signs	-0.001*	0.001	-4.47*	0.001	-0.001*	0.001	-4.74*	0.001	-0.001*	0.001	-5.91*	0.001
No. of invalid traffic signs per km	0.025*	0.007	3.49*	0.027	0.030*	0.006	4.93*	0.026	0.036*	0.005	6.27*	0.025
Constant	1.112*	0.395	2.81*									
Accident Rate of DI												
Low visibility	0.018*	0.004	4.65*	0.020	0.017*	0.004	5.46*	0.021	0.015*	0.002	8.55*	0.022
Average speed limit	0.011*	0.003	3.30*	0.009	0.007*	0.003	4.03*	0.010	0.004*	0.001	4.84*	0.011
No. of warning signs	-0.001*	0.001	-4.92*	0.001	-0.001*	0.001	-9.24*	0.001	-0.001*	0.001	-10.16*	0.001
No. of invalid traffic signs per km	0.034*	0.004	7.60*	0.035	0.043*	0.004	9.06*	0.035	0.048*	0.003	16.01*	0.034
Constant	0.752*	0.242	3.11*									
Goodness-of-fit												
No. of observations			20,047									
Correlation of Residuals R²			0.756			0.765			0.771			
AIC			1982.42			1928.24			1829.02			
BIC			1986.06			1968.61			1869.16			

2 Note: * Significant at the 5% level.

1 Compared with visibility in the daytime, the low visibility (weak natural light) is positively
2 significant to both MD and DI accident rate, implying that the lower visibility the higher the
3 accident rate for both MD and DI, which makes sense. Generally, the visibility in the daytime is
4 much better than that in the low visibility, thus the chances of causing accidents are higher under
5 the ambiguous natural light conditions. On the other hand, the impact of DI on accident rate
6 (marginal effect or elasticity is 0.022) is a little higher than that of MD (marginal effect is 0.017),
7 which implies that the DI is more likely to happen in the low visibility than MD. The reason is
8 such that the perception-reaction time in the low visibility is slower than that in the daytime, thus
9 requiring longer maneuver time to adapt the light condition and easily leading to more severe
10 accidents. This finding is uniform with those by Anarkooli and Hosseinlou (2016), Uddin and
11 Huynh (2017), and Wei et al. (2017), which suggests that injury severity levels are contributed to
12 different lightning conditions in different ways.

13 Another significant variable, average speed limit is positively related to accident rate, but it is
14 significant for DI while not for MD. Higher speed limit lead to higher accident rate, which is
15 uniform with previous studies (Xu et al., 2013a; Xu et al., 2013b; Xu and Kouhpaejad, 2015),
16 whereas the difference is that the probability of leading to DI is larger at higher speed limit
17 because it's hard to control the maneuver, hence causing the severity levels to be more severe.
18 The results are consistent with those by Wu and Wang (2013), Uddin and Huynh (2017),
19 whereas Anarkooli and Hosseinlou (2016) found that speed limit does not have a significant
20 effect injury severity on two-lane rural roads.

21 Two types of traffic signs are negatively significant for the MD and DI accident rate,
22 respectively. As for MD, the number of mandatory signs is significantly associated with the
23 accident rate, indicating that people pay more attention to the mandatory signs, thus contributing
24 less severe collision, and the more this type of traffic signs, the fewer MD accidents. Specifically
25 if one more warning sign is installed, the MD accidents would be reduced 40%; as for DI, the
26 number of warning signs is highly related to the accident rate, implying that the more this type of
27 traffic signs, the fewer DI accidents. Usually the warning signs are installed at dangerous or
28 specific locations, and the severity level would be more severe if accidents occur, so if one more
29 this type of traffic signs is installed, the DI accidents would be decreased 50%. However, the
30 findings by Carson and Mannering (2001) showed that the placement of specific warning signs,
31 i.e. ice-warning signs, is ineffective on frequency or severity of accidents. This suggests that
32 although the total number of warning signs is significant, one specific type of warning signs may
33 not.

34 The total number of invalid traffic signs per km is positively related to both MD and DI accident
35 rate, which is in line with the intuition. For all types of traffic signs listed above, every year there
36 may be various reasons making the signs invalid, such as the sun light, rain washing, erosion,
37 etc., the more invalid traffic signs per km, the higher chances for drivers to cause both MD and
38 DI accidents because not all the drivers are familiar with the roadways in the whole country,

1 where appropriate instructions at certain locations are definitely necessary. Particularly, one
2 more invalid traffic signs per km occurs, the accident rate for MD and DI is increased 25% and
3 34%, respectively.

4 In practice, according to the analysis results of the accident rate for both MD and DI, from the
5 environmental aspect it is suggested to adjust the streetlight during certain time periods,
6 especially at dawn, twilight and night, because during those periods the natural light is weak, and
7 the drivers' eyesight is ambiguous, thus leading to the accidents easily. What's more, because the
8 range covered by each streetlamp is limited, the streetlamps should be installed at appropriate
9 distances so as to increase the visibility of the roadways all along and keep the light continuous,
10 hence decreasing the collision risk; As for the state roadways in Croatia, the speed limit plays an
11 important role in determining the accident severity levels, especially DI, so before the speed limit
12 signs are installed, it should be determined carefully after a variety of investigation according to
13 traffic volume, roadway characteristics and other influencing factors; Since the number of
14 mandatory signs and warning signs are significant to accident rate for MD and DI, respectively,
15 the two types of traffic signs should be installed more at appropriate intervals correspondingly,
16 compared to other types of traffic signs. Particularly, because severe accidents usually occur at
17 dangerous locations, warning signs should be set up at certain distances before the dangerous
18 locations so as to remind the drivers in advance, and at some specific places maybe double
19 warning; As for the total traffic signs, roadway management departments should maintain and
20 change them regularly in case of bad conditions so as to keep the signs valid all the time, and
21 avoid the collision risk as much as possible.

22 **Conclusions**

23 This study investigated the interactions between accident severity levels and traffic signs in state
24 roads located in Croatia through the analysis of detailed accident severity levels, a series of
25 traffic signs, traffic characteristics, environmental factors and driver's characteristics, and the
26 seemingly unrelated regression models in unbalanced panel data approach were developed.

27 Some key findings are achieved from this analysis. To our knowledge, it's the first attempt to
28 analyze the interactions between accident severity levels and traffic signs by employing a panel
29 data seemingly unrelated regression model, in which the seemingly unrelated model addresses
30 the correlation of residuals, while the panel data model accommodates the heterogeneity
31 attributed to unobserved factors; by comparing with the pooled, fixed-effects and random-effects
32 SUR models, the results reveals that random-effects SUR model performs better than the other
33 two, and the dataset collected from states roads, Croatia from 2012 to 2016 was employed to
34 illustrate the proposed model; by involving the visibility, the present study demonstrates that the
35 low visibility causes a relatively higher risk of MD and DI; It is noteworthy that average speed
36 limit corresponds with DI accident rate, whereas not with that of MD; The number of mandatory
37 signs is more likely to reduce the MD accident rate, while the number of warning signs is

significant for DI accident rate; The number of invalid traffic signs per km are important for both MD and DI accidents, thus regular maintenance should be kept for a safer roadway environment.

Some weakness still exists in this study. The data collected only include four main components, which may not cover all the aspects of influencing accidents, and if possible more data will be collected in the future. Although the proposed model concerns over the heterogeneity issue, the role of spatial correlation of severity levels in different state roads has not been strongly addressed. Thus, in the future the spatial correlation, such as Bayesian hierarchical approach, could be attempted. Furthermore, as the death cases only accounted for 2.1%, the death and injured were merged into one category as DI, so future efforts to consider the small proportion of death accidents as one separate category could be tried. Besides, since the results of the study were based on the dataset from Croatia, it is worthwhile to try out different data sources to confirm the findings and transferability of this study in future studies. Additionally, although the aggregated traffic signs are significant for accident rate, the effectiveness of each roadway sign may be different, so in the next step the impact of each type of traffic sign on accident will be examined.

Acknowledgements

The authors are grateful for the data collection from Republic of Croatia Ministry of the Interior. The study was jointly supported by Fundamental Research Fund for the Central Universities (HUST: 2018KFYYXJJ001), National Road Safety Program of the Republic of Croatia 2011-2020 and Singapore Ministry of Education Academic Research Fund Tier 1 MOE2017-T1-001-069.

References

- Abay, K.A., 2013. Examining pedestrian-injury severity using alternative disaggregate models. *Research in Transportation Economics*, 43(1), 123-136.
- Anarkooli, A.J., Hosseinlou, M.H., 2016. Analysis of the injury severity of crashes by considering different lighting conditions on two-lane rural roads. *Journal of Safety Research*, 56, 57-65.
- Anastasopoulos, P. C., 2016. Random parameters multivariate tobit and zero-inflated count data models: addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis. *Analytic methods in accident research*, 11, 17-32.
- Anastasopoulos, P. C., Mannering, F. L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. *Accident Analysis and Prevention*, 41(1), 153-159.
- Anastasopoulos, P. C., Mannering, F. L., 2011. An empirical assessment of fixed and random parameter logit model using crash-and non-crash-specific injury data. *Accident Analysis and Prevention* 43(3), 1140-1147.
- Anastasopoulos, P. C., Mannering, F. L., Shankar, V.N., Haddock, J. E., 2012. A study of factors affecting highway accident rates using the random-parameter tobit model. *Accident Analysis and Prevention* 45, 628-633.

- 1 Avery, R.B., 1977. Error components and seemingly unrelated regressions. *Econometrica*, 45,
- 2 199-209.
- 3 Aziz, H.A., Ukkusuri, S.V., Hasan, S., 2013. Exploring the determinants of pedestrian-vehicle
- 4 crash severity in New York City. *Accident Analysis and Prevention*, 50, 1298-1309.
- 5 Babić D, Babić D, Macura D., 2017. Model for predicting traffic signs functional service life-the
- 6 Republic of Croatia case study. *Promet-Traffic & Transportation*, 29(3), 343-349.
- 7 Babić, D., Ščukanec, A., Fiočić, M., 2014. Traffic sign analysis as a function of traffic safety on
- 8 Croatian state road D3. *ICTTE-International Conference on Traffic and Transportation*
- 9 *Engineering*, Belgrade.
- 10 Biørn, E., 2004. Regression systems for unbalanced panel data: a stepwise maximum likelihood
- 11 procedure. *Journal of Econometrics*, 122(2), 281-291.
- 12 Carson, J., Mannering, F.L., 2001. The effects of ice warning signs on accident frequencies and
- 13 severities. *Accident Analysis and Prevention*, 33(1), 99-109.
- 14 El-Basyouny, K., Sayed, T., 2009. Accident prediction models with random corridor parameters.
- 15 *Accident Analysis and Prevention*, 41(5), 1118-1123.
- 16 Fountas, G., Anastasopoulos, P.C., 2017. A random thresholds random parameters hierarchical
- 17 ordered probit analysis of highway accident injury-severities. *Analytic methods in accident*
- 18 *research*, 15, 1-16.
- 19 Jørgensen, F., Wentzel-Larsen, T., 1999. Optimal use of warning signs in traffic. *Accident*
- 20 *Analysis and Prevention*, 31, 729-738.
- 21 Kim, S.N., Choo, S., Mokhtarian, P.L., 2015. Home-based telecommuting and intra-household
- 22 interactions in work and non-work travel: A seemingly unrelated censored regression
- 23 approach. *Transportation Research Part A: Policy and Practice*, 80, 197-214.
- 24 Loo, R., 1979. Role of primary personality factors in the perception of traffic signs and driver
- 25 violations and accidents. *Accident Analysis and Prevention*, 11, 125-127.
- 26 Lord, D., Mannering, F. L., 2010. The statistical analysis of crash-frequency data: a review and
- 27 assessment of methodological alternatives. *Transportation Research Part A* 44(5), 291–305.
- 28 Mannering, F. L., Bhat, C. R., 2014. Analytic methods in accident research: Methodological
- 29 frontier and future directions. *Analytic Methods in Accident Research* 1: 1-22.
- 30 Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical
- 31 analysis of highway accident data. *Analytic methods in accident research*, 11, 1-16.
- 32 Mao, C., 2016. Growth, income inequality, and capital income taxes: evidence from a seemingly
- 33 unrelated regression model on panel data. *Economics Bulletin*, 36(3), 1463-1478.
- 34 McElroy, M.B., 1977. Goodness of fit for seemingly unrelated regressions, *Journal of*
- 35 *Econometrics*, 6(3), 381-387.
- 36 Moon, H.R., Perron, B., 2006. Seemingly unrelated regresstions. *New Palgrave Dictionary of*
- 37 *Economics*, 1-9.
- 38 Park, E. S., Lord, D., 2007. Multivariate Poisson-lognormal models for jointly modeling crash
- 39 frequency by severity. *Transportation Research Record* 2019: 1–6.
- 40 Porter, R.J., Mahoney, K.M., Mason, J.M., 2007. Seemingly unrelated regression model of car
- 41 speeds and speed deviations in freeway work zones. *Transportation Research Record*, 2023,
- 42 44-51.
- 43 Prozzi, J.A., Hong, F., 2008. Transportation infrastructure performance modeling through
- 44 seemingly unrelated regression systems. *Journal of Infrastructure Systems*, 14(2), 129-137.

- 1 Savolainen, P.T., Mannering, F. L., Lord, D., Quddus, M.A., 2011. The statistical analysis of
2 highway crash-injury severities: A review and assessment of methodological alternatives.
3 Accident Analysis and Prevention 43(5), 1666-1676.
- 4 Schultz, G.G., Braley, K.T., Boschert, T., 2008. Correlating access management to crash rate,
5 severity and collision type. 87th Annual Meeting of the Transportation Research Board,
6 January 13–17, Washington, DC.
- 7 Strawderman, L., Rahman, M.M., Huang, Y., Nandi, A., 2015. Driver behavior and accident
8 frequency in school zones: assessing the impact of sign saturation. Accident Analysis and
9 Prevention, 82, 118-125.
- 10 Uddin, M., Huynh, N., 2017. Truck-involved crashes injury severity analysis for different
11 lighting conditions on rural and urban roadways. Accident Analysis and Prevention, 108, 44-
12 55.
- 13 Vilchez, J.L., 2017. Representativity and univocity of traffic signs and their effect on trajectory
14 movement in a driving simulation task: warning signs. Traffic Injury Prevention, 18(5), 544-
15 549.
- 16 Villalón-Sepúlveda, G., Torres-Torriti, M., Flores-Calero, M., 2017. Traffic sign detection
17 system for locating road intersections and roundabouts: the Chilean case. Sensors,17(6), 1207.
18 DOI:10.3390/s17061207.
- 19 Wang, X., Kockelman, K.M., 2007. Specification and estimation of a spatially and temporally
20 autocorrelated seemingly unrelated regression model: application to crash rates in China.
21 Transportation, 34, 281-300.
- 22 Washington, S., Karlaftis, M.G., Mannering, F., 2010. Statistical and Econometric Methods for
23 Transportation Data Analysis, 2nd Edition, London: Chapman and Hall/CRC.
- 24 Wei, X., Shu, X., Huang, B., Taylor, E.L., Chen, H., 2017. Analyzing traffic crash severity in
25 work zones under different light conditions. Journal of Advanced Transportation, UNSP
26 5783696. DOI: 10.1155/2017/5783696.
- 27 Wu, Q., Wang, J.J., 2013. Exploration on speed-limit signs of traffic accident section on
28 freeways. Applied Mechanics and Materials, 361-363, 2349-2352.
- 29 Xu, X., Kouhpanejad, A., Željko, Š., 2013a. Analysis on influencing factors identification of
30 crash rate using Tobit model with endogenous variable. Promet-Traffic & Transportation,
31 25(3), 217-224.
- 32 Xu, X., Kwigizile, V., Teng, H., 2013b. Identifying access management factors associated with
33 safety of urban arterials mid-blocks: A panel data simultaneous equation models approach.
34 Traffic Injury Prevention, 14(7), 734-742.
- 35 Xu, X., Kouhpanejad, A., 2015. Simultaneous Equation Tobit Analysis of Arterial Crash Rates.
36 Proceedings of the Institution of Civil Engineers-Transport, 168(4), 346-355.
- 37 Xuan, Y., Kanafani, A., 2014. Evaluation of the effectiveness of accident information on freeway
38 changeable message signs: a comparison of empirical methodologies. Transportation
39 Research Part C, 48, 158-171.
- 40 You, J., Zhou, X., 2014. Asymptotic theory in fixed effects panel data seemingly unrelated
41 partially linear regression models. Econometric Theory, 30, 407-435.
- 42 Zellner, A., 1962. An efficient method of estimating seemingly unrelated regressions and tests
43 for aggregation bias. Journal of American Statistical Association, 57(298), 348-368.

- 1 Zeng, Q., Wen, H., Huang, H., Abdel-Aty, M., 2017. A Bayesian spatial random parameters
2 Tobit model for analyzing crash rates on roadway segments. *Accident Analysis and*
3 *Prevention*, 100, 37-43.
- 4 Zhou, B., Kockelman, K.M., 2009. Predicting the distribution of households and employment: a
5 seemingly unrelated regression model with two spatial processes. *Journal of Transport*
6 *Geography*, 17(5), 369-376.