Modelling Crash Spatial Heterogeneity using Semi-Parametric Geographically Weighted Poisson Regression

Richard Amoh-Gyimah [a], Meead Saberi[a], Majid Sarvi[b]

[a] *Institute of Transport Studies, Department of Civil Engineering, Monash University, Australia*

[b] *Department of Infrastructure Engineering, The University of Melbourne, Australia*

Abstract

Crash data are typically collected with reference to location dimension. Such data suffer from unobserved heterogeneity. The objective of this paper is twofold: (1) to develop zonal crash prediction models using the Semi-Parametric Geographically Weighted Poisson Regression (S-GWPR) to address the issue of unobserved heterogeneity and (2) compare the performance of the S-GWPR with a non-spatial negative binomial (NB) model. The result indicates that by accounting for unobserved heterogeneity, the S-GWPR models performed better than the NB models. It was also found that unlike the NB models that show fixed parameters, all the variables except three have spatially varying coefficients in the S-GWPR.

Background

Crash data are typically collected with reference to location dimension, measured as points in space. Such data leads to two main problems; spatial dependence between the observations and spatial heterogeneity among relationship that are modelled (LeSage, 1999). Traditional Generalized Linear Models (GLM) models such as Poisson and NB largely ignore this important issue of spatial correlation and unobserved heterogeneity. Recently, many researches have investigated the issue of spatial correlation, neglecting unobserved heterogeneity. To account for spatial correlation, Bayesian spatial models with conditional autoregressive prior has been widely employed in safety research (Huang, Abdel-Aty, & Darwiche, 2010; Siddiqui, Abdel-Aty, & Choi, 2012; Zeng & Huang, 2014). However, both the GLM models and the Bayesian models with conditional autoregressive prior are unable to address unobserved heterogeneity. These models are thought of as global or semi-local, as they have a set of fixed parameter estimates across the region of analysis (Xu & Huang, 2015). It is however believed that the impact of a variable in one part of the region might be different in other parts of the region. Therefore the possibility of accounting for this spatial heterogeneity by allowing some or all parameters to vary spatially has considerable potential (Xu & Huang, 2015). The objective of this paper is twofold: (1) develop zonal crash prediction models using the S-GWPR to address the issue of unobserved heterogeneity and (2) compare the performance of the S-GWPR with a NB.

Data and Methodology

Four sets of data were obtained for this study; crash data and network data from VicRoads, census and land use data from Australian Bureau of Statistics. Crash data for 2010-2012 were used in the study. The 2011 census data contains both socio-economic and demographic data, whereas the land use data contains various land use activities such as residential and commercial. The data were aggregated into the Australian statistical area level 2. In total, four models were developed; two total crash models and two serious crash models using both the NB and S-GWPR techniques. Performance comparison between the non-spatial NB models and the GWPR models were made using the mean absolute deviation (MAD) and Akaiki information criterion (AIC).

Results

The result indicates that parameters of the NB models fall into the range of its corresponding counterparts in S-GWPR. This confirms that the NB estimated parameters generally represent the averages effects of the explanatory variables on crashes. Mapping the local parameter estimates of the S-GWPR models show that though some variables such as percentage of young population may have a negative mean parameter, there are some part of Melbourne where the parameter rather has a positive effect on crashes. Lastly, the MAD and AIC measure confirm that the S-GWPR models perform better than the NB models largely because the S-GWPR models accounted for unobserved heterogeneity.

Conclusion

This study intended to investigate the various factors that affect road traffic crashes by using the S-GWPR technique that is able to address the issue of unobserved heterogeneity. The result clearly shows that by addressing unobserved heterogeneity, the S-GWPR models performed better than the non-spatial NB models.

References

Huang, H., Abdel-Aty, M., & Darwiche, A. (2010). County-Level Crash Risk Analysis in Florida. *Transportation Research Record: Journal of the Transportation Research Board, 2148*(-1), 27-37. doi: 10.3141/2148-04

LeSage, J. P. (1999). The Theory and Practice of Spatial Econometrics, working paper, Department of Economics: University of Toledo.

Siddiqui, C., Abdel-Aty, M., & Choi, K. (2012). Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accident Analysis & Prevention, 45*(0), 382-391.

Xu, P., & Huang, H. (2015). Modeling crash spatial heterogeneity: Random parameter versus geographically weighting. *Accident Analysis & Prevention, 75*, 16-25.

Zeng, Q., & Huang, H. (2014). Bayesian spatial joint modeling of traffic crashes on an urban road network. *Accident Analysis & Prevention, 67*, 105-112.