Analytic Methods in Accident Research 19 (2018) 33–48

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

journal homepage: www.elsevier.com/locate/amar

Benchmarking regions using a heteroskedastic grouped random parameters model with heterogeneity in mean and variance: Applications to grade crossing safety analysis

Shahram Heydari [a,*], Liping Fu [b], Lalita Thakali [b], Lawrence Joseph [c]

[a] Imperial College London, Centre for Environmental Policy, London, UK
[b] Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, Canada
[c] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada

ARTICLE INFO

Article history:
Received 11 April 2018
Received in revised form 15 June 2018
Accepted 18 June 2018
Available online xxxx

ABSTRACT

Comparing regions while adjusting for differences in characteristics of sites located in those regions is valuable since it identifies possible inter-regional dissimilarities in crash risk propensities according to specific safety performance measures (e.g., crash frequency of a specific type). This paper describes a framework to benchmark different regions (neighborhoods, provinces, etc.) in terms of a selected safety performance measure. To avoid issues relating to aggregated (macro-level) data, we use disaggregate (micro-level) data to draw inferences at a macro/region-level, which is often needed for developing large-scale transportation safety and planning programs and policies. To overcome unobserved heterogeneity, we employ a multilevel Bayesian heteroskedastic Poisson lognormal model with grouped random parameters allowing heterogeneity in both mean and variance parameters. The proposed approach is illustrated through a comprehensive study of highway railway grade crossings across Canada. The results indicate that the proposed model addresses unobserved heterogeneity more efficiently and provides more insight compared to conventional random parameters models. For example, we found that as traffic exposure increases, grade crossing safety deteriorates at a higher rate in the Canadian Prairies than in the other regions. Our benchmarking framework is also affected by different model specifications. The results indicate the need for further in-depth investigations, which could help to identify possible reasons for inter-region differences in terms of specific safety indicators. This study provides valuable guidelines to Canadian transportation authorities, revealing important underlying crash mechanisms at highway railway grade crossings in Canada.

1. Introduction

The presence of a vast railway network in Canada imparts some risk to road/rail users and to residents living around railway lines. As reported by the Transportation Safety Board of Canada, 11,736 rail accidents of various forms have been observed over a ten-year period from 2006 to 2015 (Transportation Safety Board of Canada, 2015). According to these data, around 17% of all rail accidents have occurred at highway-railway grade crossings. Despite improvements in the recent years,

* Corresponding author.
E-mail addresses: s.heydari@imperial.ac.uk (S. Heydari), lfu@uwaterloo.ca (L. Fu), lthakali@uwaterloo.ca (L. Thakali), lawrence.joseph@mcgill.ca (L. Joseph).

https://doi.org/10.1016/j.amar.2018.06.003
2213-6657/

the number of grade crossing crashes remains high so that grade crossing safety is still a significant concern for transportation authorities and Canadian society as a whole. For example, the Transportation Safety Board of Canada reports 1953 crossing accidents for the years 2006–2015, causing 219 fatalities and 267 serious injuries (Transportation Safety Board of Canada, 2015).

To address the safety concerns at grade crossings, Transport Canada initiated a funding program called "grade crossing improvement program" that provides contributes to the eligible costs of crossing improvements completed by railways and road authorities across Canada. More recently, another funding program, rail safety improvement program has started, which invests millions of dollars each year on various safety improvement programs across Canada. One of the emerging needs to support improved decision-making for this program is to develop knowledge and tools for fair distribution of funding among Canada's vastly dispersed and different regions. This need calls for research to better understand the complex crash mechanisms at highway-railway grade crossings and develop robust methodology for benchmarking regions in terms of grade crossing safety.

1.1. Unobserved heterogeneity

A key issue in modeling crash data is addressing unobserved heterogeneity – which is caused by unobserved variables (Mannering and Bhat; 2014; Mannering et al., 2016). That is, there are factors (e.g., driver behavior) that contribute to the safety at a site; however, data on these factors may not be available in crash data sets. Consequently, differences between sites, which are similar in their known characteristics, with different crash frequencies may go unexplained by the data. Essentially, an important question that arises here is: could the effects of explanatory variables on safety vary across observations or groups of observations? The answer to this question is affirmative; therefore, approaches such as random parameters modeling and latent class modeling that allow model parameters to vary across sites (or groups of sites) are frequently used in traffic safety studies (Anastasopoulos and Mannering, 2009; El-Basyouny and Sayed, 2009; Dinu and Veeraragavan, 2011; Anastasopoulos et al., 2012; Venkataraman et al., 2014; Chen and Tarko, 2014; Barua et al., 2016; Coruh et al., 2015; Park et al., 2016; Bhat et al., 2017; Heydari et al., 2017a).

A number of traffic safety studies accounted for group effects (instead of site effects) in random effects, random parameters, and latent class modeling to capture variations in unknown/unmeasured factors that vary across groups (Wu et al., 2013; Heydari et al., 2014a, 2016b; Sarwar et al., 2017; Fountas et al., 2018a, 2018b; Cai et al., 2018). A discussion on the importance of accounting for group (e.g., region) effects in traffic safety research is provided in Heydari et al. (2016b). For example, the latter study discusses how spatially and non-spatially related unobserved factors can be captured to some extent through accounting for group (e.g., region) specific effects. In this paper, we focus on random parameters (slopes) models. Interested readers are referred to Mannering et al. (2016) for a comprehensive discussion on the unobserved heterogeneity problem and other statistical approaches that help mitigate this problem.

Recently, a few traffic safety studies have attempted to address the unobserved heterogeneity issue by employing random parameters models placing covariates on varying means and/or variances (so-called heterogeneity in means and/or variances models) mostly in crash injury-severity analysis (Seraneeprakarn et al., 2017; Behnood and Mannering, 2017; Xin et al., 2017). Venkataraman et al. (2014) developed a heterogeneity in mean count model for evaluating the effects of interchange type on heterogeneous influences of interstate geometrics on crash frequencies. The above studies highlight the advantages of the latter approach over conventional random parameters models. Basically, the heterogeneity in mean/variance approach models the within covariate variability (in both mean and variance) as a function of explanatory variables available in data. This could add robustness to the random parameters modeling approach while allowing the analyst to explain variability in model parameters through existing explanatory variables, shedding more light on the underlying mechanisms of traffic safety.

Most traffic safety studies, including those adopting a heterogeneity in mean and/or variance approach, have assumed homoskedasticity in their analysis. One could instead allow for heteroskedasticity that considers a heterogeneous variance. For example, Hong et al. (2016) considered heteroskedasticity in pedestrian exposure modeling. To accommodate heteroskedasticity, a potentially robust approach is to model the variance of the observational level error term as a function of explanatory variables. This allows the analyst to discover the source of heteroskedasticity or dispersion in a data set. Note that Bayesian methods could be extremely useful in terms of the ease of implementing heterogeneity in mean and/or variance models. The use of Bayesian statistics however has been rare if nonexistent in developing heterogeneity in mean/variance models in the extent of traffic safety literature.

While a general discussion on Bayesian methods and their advantages can be found in Gelman et al. (2004), here we provide a brief discussion. The Bayesian approach is promising not only in extending standard models to include more complex components including random parameters models, multilevel models, and multivariate models, but also easily enables combining these models together. While in frequentist statistics, new (often simulation-based) algorithms are needed to estimate more complex models, in Bayesian statistics standard Markov chain Monte Carlo (MCMC) algorithms are mostly used, without any need for developing tailored algorithms. In particular, using freely available software such as WinBUGS (Lunn et al., 2000), complex extensions and their computation through standard MCMC methods is often straightforward. The Bayesian approach propagates uncertainty in all layers and across all parameters of a model better (more fully and more intuitively) than its classical counterpart. In fact, it provides full probability densities for model parameters, whereas the frequentist approach typically provides only point estimates and confidence intervals. Bayesian (credible) intervals directly