Accident Analysis and Prevention 120 (2018) 114–121

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Multivariate linear intervention models with random parameters to estimate the effectiveness of safety treatments: Case study of intersection device program

Emanuele Sacchi[a,*], Karim El-Basyouny[b]

[a] Department of Civil, Geological and Environmental Engineering, University of Saskatchewan, Saskatoon, SK, Canada
[b] Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB, Canada

ARTICLE INFO

Keywords:
Random parameters
Multivariate models
Intervention models
Before-after studies
Crash modification functions

ABSTRACT

A novel intervention model that analyzes time-series crash data was recently introduced in the road safety statistical field. The model allows the computation of components related to direct and indirect treatment effects using a linearized time-series intervention model. The isolation of a component corresponding to the direct treatment effects, known as the crash modification function (CMFunction), enables the assessment of safety countermeasures over time. To gain new insights into how crash counts are influenced by covariates and to account for the fact that many components affecting crash occurrence are not easily available (unobserved heterogeneity), the linear intervention models with random parameters are implemented to evaluate the safety impacts of a specific treatment. Both matched-pair and full random parameter models were applied. In addition, the analysis was carried out in a multivariate context to account for possible correlation between dependent variables. The safety treatment selected for this study was the Intersection Safety Device (ISD) program implemented in the City of Edmonton (Alberta, Canada). The safety impacts were estimated by assessing the change in crash severity (property-damage-only vs. fatal-plus-injury) over time. Overall, the results showed a lower deviance information criterion (better goodness of fit) of the multivariate linear intervention model with random parameters compared to the univariate form with fixed parameters. The difference of the indexes of treatment effectiveness between the proposed modeling framework and the univariate model with fixed parameters was estimated up to 2.7%, which indicates the importance of accounting for unobserved heterogeneity.

1. Introduction

A variety of analytical methods have been employed to analyze crash counts (Lord and Mannering, 2010; Mannering and Bhat, 2014). Recently, crash data modeling has utilized increasingly sophisticated statistical techniques to uncover the relationships between crash occurrence and road characteristics while accounting for the randomness and any unobserved heterogeneity in the data. The main outcome from the modeling process is typically a regression model that produces an estimate of the expected crash frequency for a location based on the traffic exposure (volume) and site-specific traffic and geometrical characteristics, i.e., a safety performance function (SPF). Traditionally, the techniques used to develop SPFs have accounted for Poisson variation (crashes are random, discrete, nonnegative, and sporadic events) and extra-Poisson variation due to potential population heterogeneity that leads to over-dispersion (Miaou, 1994; Hauer, 1997).

In addition to the Poisson error structure, other count data distributions have been proposed to deal with specific crash data issues (Lord and Mannering, 2010; Mannering and Bhat, 2014). Recently, important advancements in the field have been brought about by adopting new modeling techniques that are able to address issues other than the choice of the best error structure for crash counts. These techniques mainly focus on gaining new insights into how crash counts are influenced by a variety of factors captured by the model's covariates.

To this end, some researchers have focused their attention on novel modeling forms (Pawlovich et al., 2006; Li et al., 2008; Park et al., 2010; El-Basyouny and Sayed, 2011). In particular, a novel model form for panel data has been introduced in road safety statistical analysis (El-Basyouny and Sayed, 2011, 2012a,b). The model allows the computation of components related to direct and indirect treatment effects under a linear intervention SPF. The isolation of a component

* Corresponding author.
E-mail addresses: emanuele.sacchi@usask.ca (E. Sacchi), basyouny@ualberta.ca (K. El-Basyouny).

https://doi.org/10.1016/j.aap.2018.08.007
Received 13 November 2017; Received in revised form 3 August 2018; Accepted 6 August 2018
0001-4575/

corresponding to the direct treatment effects enables assessment of the effectiveness of a safety countermeasure in terms of reducing collisions while isolating the effects of local (site-related) environmental factors.

The form, which makes use of linear slopes, was developed to deal with both immediate and gradual treatment impacts while accounting for countermeasure implementation, time effects, traffic volumes, and the effects of other covariates representing various site characteristics (Li et al., 2008; El-Basyouny and Sayed, 2011). Moreover, Park et al. (2010) extended the intervention model to include multivariate dependent variables with multiple regression links and proposed an algorithm for the computation of the treatment effectiveness index to determine the efficacy of the countermeasure. Other recent studies have also focused on multivariate spatial analysis to model crash frequency (Huang et al., 2017; Cheng et al., 2018).

Another important model advancement is using random parameters models that can account for components affecting crash occurrence but not easily available to the analyst (Anastasopoulos and Mannering, 2009; El-Basyouny and Sayed, 2009; Erdong and Tarko, 2014; Barua et al., 2015; Emine et al., 2015). Random parameters models have the potential to overcome inherent deficiencies of current crash records and other data sources by explicitly accounting for unobserved heterogeneity among different road sites to provide more precise inference. Random parameters count models were first introduced to the road safety literature by Anastasopoulos and Mannering (2009), who used a simulation-based maximum likelihood method for parameter estimation. These models can be structured to account for heterogeneity among specific groups of locations that share common features (matched-pairs) or to allow each parameter to vary across all locations (full random parameters) (Mannering et al., 2016). For completeness, it is important to mention that some research has also explored possible causes of the heterogeneity using a particular case of random parameters models (i.e., random effects) which assumes that the true means being estimated at the different sites are not identical but follow the normal distribution around a linear predictor such as a SPF (residual heterogeneity). This approach can also be applied when a group of treatment sites is clustered into N pairs or corridors, similar for traffic, geometric and environmental conditions and, therefore, a common mean can be assumed among matched sites (Mannering et al., 2016).

Overall, linear intervention models have been investigated and applied mostly in the context of univariate studies to account for the unobserved heterogeneity. Hence, there is a need to estimate the effectiveness of safety countermeasures by incorporating the multivariate nature of crash counts while developing linear intervention models with random parameters. Spreading the application of random parameters within the framework of multivariate linear intervention models would represent an important advancement to road safety assessments, with the ultimate goal of improving the estimation of treatment effectiveness.

To this end, this study focuses on applying this advanced statistical framework to a specific treatment implemented in the City of Edmonton, Alberta, Canada. The safety treatment selected was the Intersection Safety Device (ISD) program. ISD program consists of cameras that combine red-light running enforcement and speed enforcement at signalized intersections. Although ISD cameras are similar to red-light cameras (RLCs), the addition of speed enforcement can have an influence on the intersection's safety performance as it target drivers who are speeding through the intersection and drivers who enter the intersection after the red-light. The effectiveness of RLCs has been extensively investigated in the literature (see for instance Council et al., 2005; Shin and Washington, 2007; Lord and Geedipally, 2014). However, the safety impacts of ISD cameras have not been as widely studied. Hence, in this study the safety impact of ISDs was estimated by assessing the change in property-damage-only and severe (fatal-plus-injury) crashes using an observational before-after (BA) study design.

2. Linear intervention model

Consider a BA study in which crash data are available for a reasonable period of time before and after an intervention at treatment and comparison sites. Let Y_{it} denote the crash count recorded at location i ($i = 1, 2, \ldots, n$) during year t ($t = 1, 2, \ldots, t_B, t_B+1, \ldots, t_B+t_A$), where t_B represents the last year before treatment and t_A represents the number of years after treatment. To introduce the linear intervention model, let T_i denote the treatment indicator (1 for treatment sites, 0 for comparison sites), $t_{B,i}+1$ the intervention year for the i^{th} treatment site and its matching comparison group, I_{it} the time indicator (equals 1 in the after period, 0 in the before period), and $V_{1,it}$, $V_{2,it}$ the annual average daily traffic (AADT) at the major and minor approaches (for intersections), respectively.

2.1. Univariate model

For a Poisson-lognormal intervention (PLNI) model, Y_{it} are assumed to be independently distributed (El-Basyouny and Sayed, 2012a, b):

$$Y_{it}|\lambda_{it} \sim Poisson(\lambda_{it}), \tag{1}$$

$$ln(\lambda_{it}) = ln(\mu_{it}) + \varepsilon_i, \tag{2}$$

$$ln(\mu_{it}) = \alpha_0 + \alpha_1 T_i + \alpha_2 t + \alpha_3[t- (t_{B,i} + 1)]I_{it} + \alpha_4 T_i t + \alpha_5 T_i[t - (t_{B,i} + 1)]\, I_{it} + \alpha_6 T_i I_{it} + \beta_1 ln(V_{1,it}/V_{2,it}), \tag{3}$$

with

$$\varepsilon_i \sim Normal(0, \sigma_\varepsilon^2), \tag{4}$$

where α_1 represents the difference in log crash count between treated and comparison sites, α_2 represents a linear time trend, α_3 represents the slope due to the intervention, α_4 and α_5 respectively allow for different time trends and intervention slopes across the treated and comparison sites, α_6 accounts for a possible sudden change (drop or increase) of crashes at treated sites in the post-intervention period (usually referred to as the "jump" term), $V_{1,it}$ and $V_{2,it}$ respectively denote the AADT at the major and minor approaches (for intersections) with β_1 the exposure coefficient of the ratio of AADTs between major and minor roadways (Wang et al., 2006), ε_i accounts for random effects for latent variables across the sites, and σ_ε represents the extra-Poisson variation.

2.2. Multivariate model

For a collision count of severity level k ($k = 1, 2, \ldots, K$), multivariate analysis can be conducted to understand the relationship between severity levels (Park et al., 2010). In this case, the PLNI model in Eq. (3) remains the same except the superscript k is added to the coefficients. However, to account for the correlation among crash counts of different severity levels at site i, it is assumed for random effects that:

$$\varepsilon_i = (\varepsilon_i^1, \varepsilon_i^2, \ldots, \varepsilon_i^K) \sim N_K\,(0, \Sigma), \tag{5}$$

where Σ is a covariance matrix in which the diagonal element σ_{kk} represents the variance of ε_i^k and the off-diagonal element σ_{jk} represents the covariance of ε_i^j and ε_i^k.

2.3. Random parameters model

Model coefficients can also be allowed to vary randomly from one group of sites to another (matched pairs). Hence, representing the variation due to comparison-treatment pairing is possible by allowing the model coefficients to vary randomly from one pair to another, such that:

$$\alpha_{p(i),j} \sim N(\alpha_j, \sigma_j^2),\; j = 0, 1, 2, 3, 4, 5, 6, \tag{6}$$

$$\beta_{\,p(i),1} \sim N(\beta_{\,1}, \sigma_1^2), \tag{7}$$