Accident Analysis and Prevention 64 (2014) 78–85

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Dynamic compositional modeling of pedestrian crash counts on urban roads in Connecticut

Volodymyr Serhiyenko[a,*], John N. Ivan[b,1], Nalini Ravishanker[a,2], Md Saidul Islam[b,3]

[a] Department of Statistics, University of Connecticut, 215 Glenbrook Road, Storrs, CT 06269, USA
[b] Department of Civil and Environmental Engineering, University of Connecticut, 261 Glenbrook Road, Storrs, CT 06269, USA

ARTICLE INFO

Article history:
Received 19 March 2013
Received in revised form
13 November 2013
Accepted 14 November 2013

Keywords:
Compositional time series
Box–Cox transformation
Dynamic Generalized Linear Models
Injury severity levels
Pedestrian safety
Vector autoregressive (VAR) models

ABSTRACT

Uncovering the temporal trend in crash counts provides a good understanding of the context for pedestrian safety. With a rareness of pedestrian crashes it is impossible to investigate monthly temporal effects with an individual segment/intersection level data, thus the time dependence should be derived from the aggregated level data. Most previous studies have used annual data to investigate the differences in pedestrian crashes between different regions or countries in a given year, and/or to look at time trends of fatal pedestrian injuries annually. Use of annual data unfortunately does not provide sufficient information on patterns in time trends or seasonal effects. This paper describes statistical methods uncovering patterns in *monthly* pedestrian crashes aggregated on urban roads in Connecticut from January 1995 to December 2009. We investigate the temporal behavior of injury severity levels, including fatal (*K*), severe injury (*A*), evident minor injury (*B*), and non-evident possible injury and property damage only (C and O), as *proportions* of all pedestrian crashes in each month, taking into consideration effects of time trend, seasonal variations and VMT (vehicle miles traveled). This type of dependent multivariate data is characterized by positive components which sum to one, and occurs in several applications in science and engineering. We describe a dynamic framework with vector autoregressions (VAR) for modeling and predicting compositional time series. Combining these predictions with predictions from a univariate statistical model for *total* crash counts will then enable us to predict pedestrian crash counts with the different injury severity levels. We compare these predictions with those obtained from fitting separate univariate models to time series of crash counts at each injury severity level. We also show that the dynamic models perform better than the corresponding static models. We implement the Integrated Nested Laplace Approximation (INLA) approach to enable fast Bayesian posterior computation.

Taking CO injury severity level as a baseline for the compositional analysis, we conclude that there was a noticeable shift in the proportion of pedestrian crashes from injury severity *A* to *B*, while the increase for injury severity *K* was extremely small over time. This shift to the less severe injury level (from *A* to *B*) suggests that the overall safety on urban roads in Connecticut is improving. In January and February, there was some increase in the proportions for levels *A* and *B* over the baseline, indicating a seasonal effect. We found evidence that an increase in VMT would result in a decrease of proportions over the baseline for all injury severity levels. Our dynamic model uncovered a decreasing trend in all pedestrian crash counts before April 2005, followed by a noticeable increase and a flattening out until the end of the fitting period. This appears to be largely due to the behavior of injury severity level *A* pedestrian crashes.

1. Introduction

Pedestrian safety is a serious problem in the United States. In 2010, the total number of pedestrian traffic fatalities was 4280, which is a four percent increase from the number reported in 2009 (National Highway Traffic Safety Administration, 2012). Various studies have been performed to identify factors which affect pedestrian crashes and severity. Many factors contribute to the frequency and severity of pedestrian crashes and conflicts (Pasanen and Salmivaara, 1993; Garber and Lineau, 1996; Jensen, 1999; Klop

* Corresponding author. Tel.: +1 860 486 5455; fax: +1 860 486 3414.
E-mail addresses: volodymyr.serhiyenko@uconn.edu, volodymyr.serhiyenko@gmail.com (V. Serhiyenko), john.ivan@uconn.edu (J.N. Ivan), nalini.ravishanker@uconn.edu (N. Ravishanker), md.islam@uconn.edu (M.S. Islam).
[1] Tel.: +1 860 486 0352; fax: +1 860 486 2298.
[2] Tel.: +1 860 486 4760; fax: +1 860 486 4113.
[3] Tel.: +1 860 486 0586; fax: +1 860 486 2298.

0001-4575/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aap.2013.11.006

and Khattak, 1999; LaScala et al., 2000; Retting et al., 2003; Lee and Abdel-Aty, 2005). For example, Garber and Lineau (1996) found that the age of the pedestrian, location of the crash, the type of facility, the use of alcohol, and the type of traffic control at the site are associated with pedestrian conflicts and the likelihood of severe injury in motor vehicle crashes. This same study also found that pedestrian involvement rates are significantly higher at locations within 150 feet of an intersection stop line. Zajac and Ivan (2003) found similar results for roadway features and pedestrian characteristics having significant correlation with pedestrian injury severity from their study on rural Connecticut state-maintained highways. In addition, they also studied influence of area features on pedestrian injury severity and found that villages, downtown fringe, and low-density residential areas tend to experience higher pedestrian injury severity than downtown, compact residential, and medium- and low-density commercial areas. As one would expect, vehicle speed is seen as a significant contributor to crash severity. According to a mathematical model, a speed of 50 km/h increases the risk of death almost eight-fold compared to a speed of 30 km/h (Pasanen and Salmivaara, 1993). Crash environment also affects crash severity as Klop and Khattak (1999) found that rain, fog, or snow as well as dark environment increases injury severity.

Along with the study on factors affecting pedestrian crash and its severity, also understanding the crash trend provides a good insight into the magnitude of the pedestrian crash problem. In a study of pedestrian crash trends from around the world, Zegeer and Bushell (2012) collected pedestrian safety statistics at the global, regional, and national levels, and studied driver factors, roadway factors, vehicle factors, demographic factors, and pedestrian factors which affect the risk and/or severity of a pedestrian crash. They presented lessons for improving pedestrian safety learned from several countries, especially in Europe, and from USA. A few of the pedestrian safety strategies that they mentioned were to provide pedestrian-friendly geometric guidelines, implement effective traffic control and other pedestrian safety treatments, expand funding for safety education programs, and develop pedestrian friendly ITS vehicle features.

Spainhour et al. (2006) studied pedestrian crash trends and causative factors in Florida. The paper focused on finding primary contributing factors for pedestrian crashes, and concluded that pedestrians were at fault in 78 percent of the cases reviewed. However, they did not study pedestrian crashes by severity of injuries. Rather they categorized pedestrian crashes as follows: crossing not in a crosswalk, crossing at intersection, in road, walking along roadway (with traffic), walking along roadway (against traffic), exit vehicle, vehicle turn or merge, unique (crashes with some unusual circumstances which are not likely to happen again) and other (crashes with unknown circumstances). Recently, Hu et al. (2012) introduced dynamic time series modeling in a Bayesian framework to uncover temporal patterns in the safety of senior and non-senior drivers in Connecticut.

The objective of our study is to discover temporal changes in pedestrian crashes with a particular injury severity level as proportions of total crashes of all injury levels. In other words, our goal is to investigate whether over the given time period, there is an increase in crashes of one or more injury severity levels with attendant decreases in other levels. We collected records for all crashes on state-maintained roads in Connecticut from January 1995 to December 2009 from the Connecticut Department of Transportation (ConnDOT). Crashes were classified into the following severity groups: K = fatal injury, A = severe injury, B = evident minor injury, C = non-evident minor injury and O = property damage only. The study design is described in Section 2.

The compositional time series modeling described in Section 3 enables us to model transformed crash proportions of different injury severity levels in order to discuss changes and connections among them. Section 4 describes a dynamic modeling framework for the time series of total pedestrian crash counts and compares it with a static model. Section 5 describes predictions of proportions and total counts which enables us to obtain predictions of the crash counts by injury severity levels. Section 6 provides a summary and discussion..

2. Study design

Crash records of State-maintained roads are recorded and preserved by Connecticut Department of Transportation (ConnDOT). Crash data from January 1995 to December 2009 from ConnDOT was used. Property damage only (PDO) crashes were not reported in the database before 2007 for local roads. PDOs were reported starting in 2007. Crash data were aggregated by each month at five different severity level e.g. fatal (K), severe injury (A), evident minor injury (B), non-evident possible injury (C) and property damage only (O). The C and O severity levels were combined into one response variable for analysis because the O level is rare for pedestrian crashes, while the other severity levels were each defined as individual response variables.

Vehicle-miles-traveled (VMT) was used as a predictor variable in compositional modeling. For this purpose we needed monthly VMTs during the analysis period. ConnDOT has average daily VMT for each year for various definitions of facility type based on urban or rural location and functional classification. Also ConnDOT has pneumatic tube and induction loop counters from which these annual average daily VMT estimates are derived. To obtain monthly VMTs, monthly expansion factors obtained from ConnDOT were used. Descriptive statistics of the response variables and VMT used in the analysis are given in Table 1. Please note that throughout the paper we use VMT divided by 1000 in the data analyses.

3. Dynamic compositional time series modeling

The main characteristic of compositional data, which occurs frequently in various areas such as chemistry, demography, geology, survey analysis, consumer demand analysis, etc., is that at each time point, all components are positive and sum to one. There is a need to model different proportions or compositions that are observed over time, i.e., to model temporal changes of such compositional time series using suitable models. A compositional time series is defined as a G-variate vector of positive components $x_t = (X_{t1}, \ldots, X_{tG})$, for $t = 1, \ldots, T$, where the structure is completely defined by $g = G - 1$ components, so that x_t lies in a g-dimensional simplex:

$$S^g = \{(X_{t1}, \ldots, X_{tg}) : X_{t1} > 0, \ldots, X_{tg} > 0; X_{t1} + \cdots + X_{tg} = 1\}$$

Statistical analysis follows via a suitable transformation of the data from the g-dimensional simplex S^g into the Euclidean space R^g. An excellent approach for compositional data analysis is given by Aitchison (1982, 1986), who introduced the Additive Log Ratio (ALR) transformation and the Centered Log Ratio (CLR) transformation, by Rayens and Srinivasan (1991) who discussed the more general Box–Cox (Box and Cox, 1964) transformation, and by Egozcue et al. (2003) who proposed the Isometric Log Ratio transformation. Aitchison (1986) along with Brunsdon (1987), Smith and Brunsdon (1989), Brunsdon and Smith (1998), and Ravishanker et al. (2001) discussed compositional time series analysis. In these papers, compositional time series were first transformed via the *ALR* (or more general Box–Cox) transformation, and were then analyzed with standard time series model techniques, such as Vector AutoRegression (*VAR*), Vector AutoRegressive Moving Average (*VARMA*), or Dynamic Linear Modeling via the Kalman Filter, or in a Bayesian framework.