Accident Analysis and Prevention 88 (2016) 124–137

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

A Bayesian ridge regression analysis of congestion's impact on urban expressway safety

Qi Shi*, Mohamed Abdel-Aty, Jaeyoung Lee

Department of Civil, Environmental and Construction Engineering, University of Central Florida, Engineering II-215, Orlando, FL 32816, United States

ARTICLE INFO

Article history:
Received 15 August 2014
Received in revised form 23 November 2015
Accepted 1 December 2015

Keywords:
Urban expressway safety
Congestion
Bayesian inference
Ridge regression

ABSTRACT

With the rapid growth of traffic in urban areas, concerns about congestion and traffic safety have been heightened. This study leveraged both Automatic Vehicle Identification (AVI) system and Microwave Vehicle Detection System (MVDS) installed on an expressway in Central Florida to explore how congestion impacts the crash occurrence in urban areas. Multiple congestion measures from the two systems were developed. To ensure more precise estimates of the congestion's effects, the traffic data were aggregated into peak and non-peak hours. Multicollinearity among traffic parameters was examined. The results showed the presence of multicollinearity especially during peak hours. As a response, ridge regression was introduced to cope with this issue. Poisson models with uncorrelated random effects, correlated random effects, and both correlated random effects and random parameters were constructed within the Bayesian framework. It was proven that correlated random effects could significantly enhance model performance. The random parameters model has similar goodness-of-fit compared with the model with only correlated random effects. However, by accounting for the unobserved heterogeneity, more variables were found to be significantly related to crash frequency. The models indicated that congestion increased crash frequency during peak hours while during non-peak hours it was not a major crash contributing factor. Using the random parameter model, the three congestion measures were compared. It was found that all congestion indicators had similar effects while Congestion Index (CI) derived from MVDS data was a better congestion indicator for safety analysis. Also, analyses showed that the segments with higher congestion intensity could not only increase property damage only (PDO) crashes, but also more severe crashes. In addition, the issues regarding the necessity to incorporate specific congestion indicator for congestion's effects on safety and to take care of the multicollinearity between explanatory variables were also discussed. By including a specific congestion indicator, the model performance significantly improved. When comparing models with and without ridge regression, the magnitude of the coefficients was altered in the existence of multicollinearity. These conclusions suggest that the use of appropriate congestion measure and consideration of multicolilnearity among the variables would improve the models and our understanding about the effects of congestion on traffic safety.

1. Introduction

Providing motorists with efficient and safe traffic system has long been considered a priority of traffic professionals. With the growth in traffic demand outpacing construction of road infrastructure, congestion and safety concerns arise. In urban areas, many traffic authorities have turned to toll/turnpike facilities and efficient use of Intelligent Transportation Systems (ITS) techniques as remedies for congestion and to improve safety. Whether ameliorating congestion or safety would have positive effects on the other depends on the relationship between congestion and safety. It is widely acknowledged that crashes lead to non-recurrent congestion while the effects of congestion on crashes seem inconclusive.

This paper aims at identifying the relationship between congestion and crashes for urban expressway State Road (SR) 408 in Central Florida based on the existing ITS facilities. The 21.4-mile expressway of interest is managed by Central Florida Expressway Authority (CFX). The toll expressway accommodates large commuter traffic and travels through the downtown Orlando area. Consequently, recurrent congestion during peak hours is observed on the expressway. For more efficient operation, CFX has installed Automatic Vehicle Identification (AVI) system for Electronic Toll

* Corresponding author.
E-mail addresses: shiqi@knights.ucf.edu (Q. Shi), m.aty@ucf.edu (M. Abdel-Aty), jaeyoung@knights.ucf.edu (J. Lee).

http://dx.doi.org/10.1016/j.aap.2015.12.001
0001-4575/

Collection (ETC) and travel time estimation. In 2013, Microwave Vehicle Detection System (MVDS) was introduced for traffic monitoring on the expressways. Based on the real-time data derived from the detection systems, CFX provides motorists with travel time and other information via the Dynamic Message Signs (DMS). The multiple ITS instruments deployed on the expressway give the authority a favorable margin for better understanding of the traffic conditions and potential for improvement. In this study, traffic parameters from the MVDS and AVI systems along with roadway geometric characteristics were collected to identify their relationship with crash occurrence. Compared with previous crash frequency studies exploring congestion's effects on safety, the current work applied ridge regression within Bayesian framework to avoid the problem of multicollinearity. Through this method, it is expected that more accurate parameter estimates will be achieved.

2. Background

Traffic congestion especially in urban areas has been receiving increasing attention during the past few decades. Congestion is not only costly in terms of time losses and fuel waste, but also causes unstable traffic flow because of the stop-and-go traffic pattern. Whether congestion has an impact on traffic safety has been researched in many previous studies. One early study by Shefer and Rietveld (1997) stated that congestion leads to lower numbers of fatalities because of lower speed in congestion. Baruya (1998) investigated crash frequencies in four European countries and declared that congestion is positively associated with crash frequency. Golob and Recker (2003) in their study of Southern California freeways found rear-end crashes are generally associated with high variations in relatively low speeds which are commonly observed in congested traffic. Kononov et al. (2008) developed safety performance functions (SPFs) using Colorado, California, and Texas data. Their conclusion suggested that safety deteriorates with the degradation of congestion expressed through level of service but segments with the highest volumes experienced lower crash rates. Wang et al. (2009) explored the relationship between crash and congestion for an orbital motorway around London using a spatial analysis approach. They found that congestion imposes little or no impact on the frequency of road accidents. In a recent review of previous works about the congestion's effect on road safety, Wang et al. (2013) concluded that existing research has not yet reached an agreement on the impact of traffic congestion. To understand the complex phenomenon, several factors should be taken into account; (1) how is congestion measured. With the development of traffic detection systems, new congestion measurements become available. Whether these detection systems reflect the congestion accurately is crucial in the congestion-safety studies. (2) Congestion could be time and location specific in urban area. Freeways/expressways in urban areas carry large commuter traffic in morning and evening peak hours. The necessity to consider peak and non-peak traffic hours is worth investigation. (3) The complex nature of crashes often requires multiple explanatory variables included in safety analysis. The traffic speed, volume, and density could be correlated which would introduce multicollinearity. Multicollinearity could affect accurate interpretation of the variables' effects (Berry and Feldman, 1985). Yet existing studies haven't addressed the problem adequately. The conflicting results in previous congestion-safety studies might be resulting from this issue.

2.1. Congestion performance measures

Congestion is easy to be observed and interpreted. Nevertheless, how to measure congestion could be a challenging task. Whether congestion estimates are accurate and useful largely depends on the available data. For a long time, Volume-to-Capacity Ratio (V/C Ratio) and the Level of Service (LOS) have been the primary congestion indicators since continuous data collection was not feasible (National Research Council, 2010). However, they lack the flexibility to capture the dynamics of congestion. With the development of ITS equipment, more direct measurement of speed, travel time, and volume becomes available. New approaches to measure congestion are proposed (Lomax et al., 1997; Schrank and Lomax, 2007). Current congestion measures can be broken down into three general categories: density-based, travel-time-based and travel-speed-based. Traffic occupancy has become more available to reflect the traffic density on roadways thanks to the deployment of ITS facilities. The numerous types of ITS traffic detectors also promote the creation of other congestion measures. Travel Time Index (TTI) is a widely used travel-time-based congestion measure. It is defined as the ratio of the travel time during the peak period to the time required to make the same trip at free-flow speeds (Schrank and Lomax, 2007). TTI can be measured by the AVI system which identifies vehicles at different detecting locations and calculates the travel time. In the meantime, speed-based congestion measures are also applied. Christidis and Rivas (2012) in their report of European road congestion defined the intensity of congestion as the ratio of the average to free flow speeds. Hossain and Muromachi (2012) used the rate of reduction in speed caused by congestion from the free-flow speed condition as congestion index. In this study, the expressway has both AVI and MVDS systems. Multiple congestion measures were tested to investigate their feasibility to be used in traffic safety studies.

2.2. Time-specific nature of congestion

Another issue related to the congestion in crash frequency studies is the problem of averaging. Mensah and Hauer (1998) discussed the potential problems resulting from averaging traffic flows over a period of time. They argued that the cause-effect relationship is between crashes and the flows prevailing near the time of crash occurrence and there are two (daytime and nighttime) and perhaps many more cause-effect functions in the course of the study period. Wang et al. (2009) in their study suggested that in reality, the level of congestion varies over time and averaging may have an impact on the effect of congestion on crashes. In this paper, as it is still a crash frequency study, aggregating the traffic flow data is unavoidable. Nevertheless, we divided the traffic into peak and non-peak hours to test if there is significant difference of the effects of traffic variables during these two time periods.

2.3. Multicollinearity

Multicollinearity occurs when the explanatory variables are correlated among each other. The inherent collinearity won't affect the overall model fitting (Burnham and Anderson, 2002). However, some consequences could result from multicollinearity as pointed out by Washington et al. (2010) such as high standard errors of the estimated parameters of the collinear variables, large sampling variability which means estimated parameters based on different sample could vary widely, and inaccurate interpretation of the effects of explanatory variables because the change of one variable would inevitably lead to the change of another variable. In crash frequency studies, the traffic data are observational data in most cases and not controlled. Correlations exist among speed, density and volume. Nevertheless, existing literature shows that this issue is not adequately addressed. This might be a reason why researchers haven't reached a conclusive statement about the congestion's effect on traffic safety. Solutions to the multicollinearity problem were generalized by Congdon (2007) which include: (1) the introduction of extra information; (2) the multivariate