Accident Analysis and Prevention 118 (2018) 29–37

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Are estimates of crash modification factors mis-specified?

Robert B. Noland*, Yemi Adediji

Alan M. Voorhees Transportation Center, Edward J. Bloustein School of Planning Public Policy, Rutgers University, New Brunswick, NJ, United States

ARTICLE INFO

Keywords:
Crash modification factors
Spatial analysis
Geometric design

ABSTRACT

Transportation planners and traffic engineers are using crash modification factors to evaluate how changes in road geometry and design features can reduce crashes. Crash modification factors are typically estimated based on segmenting links on a highway and associating with geometric features. This allows statistical methods to be applied to the data. Concurrently there is a stream of research that relies on spatial units of analysis to examine crashes; these typically use broad features of the road network combined with socio-economic and demographic factors that are associated with crashes. In this paper, we examine whether omission of these spatial factors in a link-based model results in mis-specified models, in particular, omitted variable bias. Our results suggest that there is no change in coefficient signs, but that there is a reduction in the magnitude of estimates, suggesting that omitted variable bias exists. The sign of spatial variables differ substantially when combined with a link-based model. We also find substantial variability in coefficient estimates, and discuss the implications of these results for the use of crash modification factors in cost-benefit analysis of road safety projects.

1. Introduction

Crash modification factors for the evaluation of safety interventions are being developed to provide guidance to highway engineers and planners on which infrastructure interventions can best reduce vehicle crashes. The *Highway Safety Manual* (AASHTO, 2010) was developed to provide guidance on specific crash modification factors for a large range of possible interventions. Crash modification factors are developed using statistical models that link a variety of geometric design variables to police-reported crashes on road links (AASHTO, 2010). As such, the unit of analysis tends to be links or segments of a highway, or intersection zones. This allows one to match the highway characteristics with the crashes that occur along the specific link in the road network. These models focus on the geometric design of the highway including features such as turning radius, road curvature, access points, lane widths, and number of lanes, among others. The output of these models is a parameter estimate associated with a specific design element that can then be used in a cost-benefit analysis to determine whether a design change should be implemented. The objective of this paper is to examine whether these models are properly specified and in particular, whether they suffer from omitted variable bias; and also to point towards a possible solution to this problem by incorporating spatial variables into these models.

Both Mannering and Bhat (2014) and Mannering et al. (2016) note that omitted variable bias is a common problem in crash analysis. The former study points out that parsimonious models (common in practice) will suffer from omitted variable bias, while the latter study considers methods to control for unobserved heterogeneity while noting that these models do not identify its cause. Many models developed in the *Highway Safety Manual* (AASHTO, 2010) tend to be univariate and would clearly suffer from some omitted variable bias. One critique of adding additional control variables is that some of these may be correlated. This is a trade-off the analyst must consider, but multi-collinearity simply means that a model may not provide the correct inference if the correlated variables are both included and can be easily dealt with to determine the best model fit.

One critique of crash modification factors and the link-based models on which these are based, is that the coefficient estimates are not transferable (Hauer et al., 2012) and often are highly variable (Elvik, 2015). The latter may be due to variation in the traffic environment, or in other words, the omission of important variables associated with crashes but correlated with variables in the model. This means that one may find unreliable or diminished parameter estimates, including statistically significant effects from factors that are not associated with the measured crash outcome, or non-significant effects for factors that are associated with crashes (Wu et al., 2015). One source of omission is the spatial context where the crashes occur. There is a large literature that examines spatial associations with crashes such as median income levels, population density, and other census-level data. These include studies of pedestrians in New Jersey (Noland et al., 2013), child and adult pedestrians in London (Graham et al., 2005), motor-vehicle crashes in England (Noland and Quddus, 2004a) and Pennsylvania

* Corresponding author.
E-mail address: rnoland@rutgers.edu (R.B. Noland).

https://doi.org/10.1016/j.aap.2018.05.017
Received 21 July 2017; Received in revised form 10 May 2018; Accepted 19 May 2018

(Aguero-valverde and Jovanis, 2006), and a number of studies conducted in Florida, including a spatial analysis of counties (Huang et al., 2010), analysis of pedestrians and cyclist crashes at a small spatial scale (Siddiqui et al., 2012) and analysis of the most desirable spatial unit to use (Abdel-aty et al., 2013; Lee et al., 2014), suggesting that traffic-analysis zones (TAZ) while convenient to use may not provide the best model fit.

Combining the spatial context with a link-based analysis may shed light on whether omitted variable bias is occurring in link-based models. Basically, we know that some of these spatial factors are associated with crashes, therefore not including them in a link-based model is problematic. We are aware of one empirical study that identified issues associated with omitted variable bias when intersection-based models do not control for spatial attributes that may also affect crashes (Mitra and Washington, 2012). Another study examined how parameter estimates vary using artificially generated data (Wu et al., 2015) concluding that parameter estimates are unreliable when important variables are omitted. Similar studies using artificially generated data have identified other potential mis-specifications in estimates of crash modification factors (Wu and Lord, 2017, 2016). A recent study using data from the state of Virginia found differences in local crash modification factors, arguing that this was largely due to spatial variation (Liu et al., 2017); however, this study did not include sufficient variables that might control for the localized variation found in their estimates. To examine these questions, we use data from New Jersey to estimate a link-based model combined with spatial variables to determine whether there is omitted variable bias. We also examine the variability of coefficient estimates between and within models.

In the following sections, we first review some of the results from link-based studies, primarily to highlight the variation in results found in the literature, recognizing that a difficulty with comparing results is the multitude of variables and modeling approaches used, as well as different data sources. We then discuss our own data and the challenges of combining the link-based and spatial approaches. This is followed by our analysis, including estimates for a spatial analysis and for five highways in New Jersey. Results and discussion follow with implications for crash analysis and the use of crash modification factors in cost-benefit analysis.

2. Link-based studies

The Poisson and Negative Binomial models have emerged as the standard in crash analysis as these best match the distributional characteristics of crash data. Though the Poisson model is often the starting point due to its suitability for analyzing count data, the Negative Binomial model is often chosen because of the Poisson model's assumption of equi-dispersion, i.e., the mean and variance are equal. Real data usually violates this assumption, including most crash datasets.

Most of the studies we reviewed were estimated using the Negative Binomial model, however a handful employed other models, including Chiou and Fu (2013) who used a multinomial-generalized Poisson model with error components, and Garnowski and Manner (2011) who used a random parameters negative binomial model in addition to a fixed parameter model. Our review focuses on the variables included in each model and the results of the estimates, and we make no judgement about the appropriateness of various modeling techniques. Our review is by no means comprehensive, but we sought out studies that are repeatedly cited in the literature (except some that are more recent).

In reviewing link-based studies that provide estimates of crash modification factors, we found a wide variety of different variables included in the models presented. Different measures were sometimes used for the same variables, complicating efforts to compare coefficient values. Our comparison focuses on results for the following geometric design elements included in some of the studies: number of traffic lanes, lane width, average annual daily traffic (AADT), horizontal curvature, shoulder width, and median width. These are variables that we use in our analysis based on the availability of data in New Jersey. We examine each of the geometric variables in turn.

2.1. Horizontal curvature

An early study used data for an interstate highway in Washington state and the focus was on horizontal curvature for different design speeds, while controlling for weather conditions (Shankar et al., 1995). Lane and shoulder widths were virtually constant over the stretch of road analyzed given that interstate highways follow standard design guidelines. Estimates ranged from 0.046 (lower design speed) to 0.117 (higher design speed) measured using the number of horizontal curves at the two design speeds based on segmenting the interstate into ten sections. Further work in Washington state, based on data for principal arterials, measured curvature as the horizontal curve radius (Milton and Mannering, 1998), that is wider longer curves have a larger radius; coefficient values for two different data sets varied between −0.0021 and −0.000221, and both were statistically significant.

An analysis of an arterial roadway in Florida also controlled for horizontal curvature, but used a different measure of "degrees/100 m arc" (Abdel-aty and Radwan, 2000). Their coefficient estimate is positive and significant with a value of 0.124, which cannot be directly compared to the results of Shankar et al. (1995). A study of a 4-lane motorway in Italy used a measure of 1/radius (km^{-1}), and found positive coefficients, for all crashes, of about 0.26 (Caliendo et al., 2007). Using data from Indiana, another study used a measure of 18,000/($\pi \cdot r$), with r (radius) defined in feet (Malyshkina and Mannering, 2010). The coefficient estimate is −0.0562, negative and opposite that of other studies; that is, wider curves increased risk.

Another study using data from Indiana focused on rural two-lane roads (Labi, 2011). Horizontal curvature was based on "average horizontal curve radius", coefficients were estimated for different crash severity levels and for different functional road classes of rural two-lane roads; for fatal plus injury crashes, this varied from 0.0262 to 0.0580.

Returning to data from Washington state, Bauer and Harwood, (2014) define horizontal curvature as $1/\ln(2 \times 5730/r)$. The coefficient estimated is statistically significant at a 95% level and is 0.19. A variable was also included in the estimated model that assessed the interaction between horizontal curves and vertical grades, which was also statistically significant. This report, conducted for the US Federal Highway Administration, is of note partly because it was conducted to develop crash modification factors for the *Highway Safety Manual* and presents precise figures for both fatal/injury crashes and property-damage only crashes.

Curvier roads are assumed to increase the probability of crashes, all else equal, and these results largely support that notion. However, curvier roads may also lead to reduced speeds if they are perceived as riskier (Noland, 2013). Only one study notes that the curviest stretches of roads seem to have fewer crashes and that curves tend to be riskiest following a long tangent section, i.e., a straight road leading into a curve (Milton and Mannering, 1998). While all these studies controlled for different variables (some of which we discuss below), none included spatial variables that might provide a better context for the driving population and the local area in which the crash occurred.

2.2. Shoulder width

Another commonly included geometric design variable is the shoulder width. Larger shoulders are assumed to decrease the crash rate. A larger right-hand shoulder provides greater space for a driver to recover if a loss of control occurs. Some of the models we reviewed include parameters for shoulder width, which can be easily measured.

Variation in the results is large. One study that analyzed two-lane rural roads in four different states had coefficient estimates ranging from −0.1230 to −0.4541 (Council, Stewart 1999). Two of the models included surface width as an additional control variable; this might be