Transportation Research Part A 116 (2018) 220–231

Contents lists available at ScienceDirect

Transportation Research Part A

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

Modeling the endogeneity of lane-mean speeds and lane-speed deviations using a Bayesian structural equations approach with spatial correlation

Wen Cheng*, Gurdiljot Singh Gill, Taha Sakrani, Dennis Ralls, Xudong Jia

Department of Civil Engineering, California State Polytechnic University, Pomona, 3801 W. Temple Ave., Pomona, CA 91768, United States

ARTICLE INFO

Keywords:
Random effects
Spatial correlation
Multivariate
Multilane
Endogeneity
Speed prediction

ABSTRACT

This study focused on the development of speed prediction models for a multilane highway which incorporate the potential endogenous relationship between adjacent lane speeds and speed deviations while considering geometric design, traffic flow, and other variables in the model specification and accounting for the correlation structures due to multilevel nature of data. The Full Bayesian framework was employed to build the hierarchical models which accounted for three correlation structures at multiple levels: the correlation between speeds of adjoining lanes due to multivariate nature; spatially structured correlations between the adjacent segments, and spatially unstructured correlations among segments.

The model estimates which influence the lane-mean speed indicated the directional variation of exogenous factors. For the westbound traffic, the afternoon and night hours were observed to be influential while eastbound traffic was more sensitive to the morning period. The segment length revealed a lane-varying correlation where longer segments influenced a speed reduction for the three lanes closer to the median while the speed in outermost lane exhibited a reverse trend as it increased with longer segments. Both models, with mean speed and speed deviations, demonstrated the significant presence of endogeneity due to mean speeds and speed deviations of adjacent lanes, respectively. This study also assessed the accuracy of predicted mean speed and speed deviations by calculating the measures of discrepancy between the observed and model predicted speeds. The Bayesian residuals, which incorporated the normal, multivariate, and spatial correlation structures, exhibited significant superiority at the prediction accuracy than the Normal ones. This discrepancy in prediction performance reflected the impact of consideration or exclusion of random effects.

1. Introduction

In the field of transportation engineering, the geometric design and level of service of roadways are greatly influenced by the operational speed. Speed plays a vital role in traffic management and control as it is fundamental to the performance measures of traffic. Due to the general tendency of unavailability of speed distribution information of road entities, it has been best estimated in terms of its basic relationship with other fundamental variables of the road entity such as density or flow. The basis of this approach was laid by the seminal research studies from several decades ago (Greenshields, 1935; Lighthill and Whitham, 1955) which established the relationships between flow, density, and space mean speed. To construct a superior function of speed-flow relationship,

* Corresponding author.
E-mail addresses: wcheng@cpp.edu (W. Cheng), gurdiljotg@cpp.edu (G.S. Gill), tsakrani@cpp.edu (T. Sakrani), dbralls@cpp.edu (D. Ralls), xjia@cpp.edu (X. Jia).

https://doi.org/10.1016/j.tra.2018.06.014
Received 15 May 2017; Received in revised form 4 March 2018; Accepted 20 June 2018
0965-8564/

some studies incorporated different explanatory variables which could potentially influence the vehicular speed such as an array of environment conditions and traffic characteristics. The study by Fitzpatrick et al. (2005) on two-lane tangent sections developed the operational speed model and considered roadway and roadside characteristics such as number of access points, roadside development type, lane width, and others. The study by Einbeck and Dwyer (2011) proposed the use of local principal curves to model the speed-flow relationship. Following the variable selection recommended by previous studies focused on flow theories (Shankar et al. 1995; Li et al. 2009; Golob and Recker, 2003), this study included variables pertaining to roadway conditions such as speed limit, ramp and signal control, and speed distribution. Pei et al. (2012) employed Full Bayesian method to establish a prediction model for speed distribution, in terms of average travel speed and speed deviation, by considering the effects of traffic flow, road geometry, and weather conditions. A study by Silvano and Bang (2015) focused on the impact of posted speed limit (PSL) on the free-flow speed on urban roads by incorporating the road characteristics, such as carriageway width, road environments, and the presence of on-street parking and sidewalks. As noted by Bassani et al. (2016), the recent research on speed prediction models has focused on improving the model prediction capability, unlike the previous studies which aimed to investigate the influential factors.

Apart from the aforementioned studies which were centered on the development of speed prediction models with different exogenous variables, several studies related with other transportation modeling context have utilized endogeneity models to address the simultaneity problems common to transportation field (Derrig et al., 2000; Eisenberg, 2003; Cheng et al., 2013; Washington et al., 2010). The exogenous and endogenous variables differ on the basis of dependency between the independent and dependent variables in a model. Majority of models assume that the causal relationship between such variables is unidirectional, which means that only the independent variables could impact the dependent ones and not the other way around. However, the transportation field poses different simultaneity issues where the dependent variable could also impact the independent ones. For example, in case of safety analysis of intersections crashes, different covariates may be considered which may impact the dependent variable of crash occurrence. One such factor is the presence of left-turn lanes at the intersection. However, the introduction of left-turn lanes on an intersection is also influenced by the perceived safety concern. This builds up an endogeneity problem where the left-turn lane could not be assumed to be influenced by crash occurrence (dependent variable) (Kim and Washington, 2006). Different studies have attempted to circumvent the issue of endogeneity bias in different aspects of transportation research (Dane et al., 2014). The study by Bhat and Koppelman (1993) used the Simultaneous Equations Model (SEM) approach to develop an integrated model of employment, income, and household car ownership. They regarded employment and income as endogenous variables in disaggregate travel demand frameworks and estimated a joint model of employment, income, and car ownership. Abu-Eisheh (2001) developed a vehicle demand and driver population models by employing simultaneous equation estimation techniques as a function of socioeconomic and political variables and observed simultaneity among the dependent variables of automobile ownership and driver population. Kim and Washington (2006) noted that inclusion of endogeneity revealed the unbiased effect of left-turn lanes on crash frequency as left-turn lanes were observed to reduce angle crash frequencies as expected by engineering judgment. Medina and Tarko (2007) explored the relationship between roadway characteristics, driver behavior as influenced by the perceived risk, and safety by formulating an SEM that consists of two structural equations involving two endogenous variables, the objective risk and the speed selected by the drivers.

Some studies incorporated the SEM formulation to address the endogeneity issue for speed prediction models. Shankar and Mannering (1998) investigated the role of endogeneity at a macroscopic level by developing models for lane-mean speeds and speed deviations on a three-lane highway. Due to the linear nature of functional form, three-stage least squares (3SLS) estimation was employed for model estimation. They found the presence of endogeneity in the form of variables pertaining to lean-mean speed and speed deviations of the adjacent lanes while exogenous variables related to the environment, traffic flow, and temporal factors were noted to be influential for lane speeds. Boyle and Mannering (2004) utilized the same 3SLS approach to explore the effect of driver behavior for in- and out-of-vehicle travel advisory systems. The results showed the negative endogenous relationship between mean speed and speed deviation as various driver-related variables and advisory message indicators were found to be associated with speed and speed deviation. The study by Himes and Donnell (2010) focused on the investigation of geometric design and traffic flow factors to model the mean operating speed and speed deviation on four-lane highways by using 3SLS estimator while exploring the endogeneity and incorporating the contemporaneous correlation between the disturbances across the equations. This study recommended the exploration of more geometric variables (shoulder width, lane width) and focus on areas with more substantial traffic flow environments.

Similar to the above studies which controlled the endogeneity to reduce bias and inconsistency of estimated parameters, some studies aimed to develop more precise models by incorporating different types of correlation structures found in transportation data. Some studies employed the multivariate approach for simultaneously modeling different crash outcomes based on accident type or severity (Aguero-Valverde and Jovanis, 2009; Cheng et al., 2017b; Cheng et al., 2018b) while other studies expanded to explore the spatial correlations (Jonathan et al., 2016; Cheng et al., 2017a; Cheng et al., 2018c). Such correlations are typical in studies of crash prediction models but rare in lane speed models. Recently, a study by Bassani et al. (2016) utilized random effects to account for the variability in the time and space of data for the estimation of free-flow speed on two-lane rural highways. The adopted model structure separated the estimate of the central tendency of speeds from the typical deviations of individual speeds. Unlike the aforementioned crash prediction models, the random effects in this study were normally distributed and hence it did not consider any multivariate or spatial correlations.

In summary, few multilane highway speed prediction models have been estimated. Of those, some have included traffic flow, seasonal, and time-of-day variables in the model specification using a simultaneous equations framework. Other multilane speed prediction models considered only inside and outside lane mean speeds, or estimated a model of mean speed and speed deviation without considering interaction of speed across travel lanes. There is a need to estimate a multilane highway speed prediction model