An Improved Structural Equations System for Lane-mean Speeds Prediction and Modeling the Endogeneity of Adjacent Lane Speeds (Proposal)

1. Problem Statement

Like other transportation data, lane-mean speeds are also best modeled by a system of structural equations. Several studies omit the interrelation between adjacent lane speeds, which may produce biased and inconsistent results if models are solved by ordinary least squares (OLS). The uncorrelatedness of regressors and disturbances assumption of OLS is violated since one or more independent variables are endogenous in the system. This study attempts to propose a structural equations approach to model the lane-mean speeds in multi-lane traffic with the endogeneity of adjacent lane speeds and also the downstream speeds being considered. Additionally, the equations system can serve as a prediction model for lane-mean speeds.

2. Research Objectives

Lane-mean speeds modeling is very important in the field of multi-lane traffic researching. Some transportation data are best modeled by a system of structural equations, including lane-mean speeds (Washington et al., 2010). In Shankar and Mannering (1998), a structural equations approach was used to model the lane-mean speeds and lane-speed deviations where three-stage least squares (3SLS) was applied to handle the endogeneity problem, i.e. the correlation between regressors and disturbances that means changing the value of dependent variables may lead to the changing of independent variables. The researchers conducted an empirical study based on this approach, combining the relevant data of a six-lane road (three lanes in each direction) with environmental data, temporal data and traffic flow factors. However, although the methodology solved the endogeneity problem, obtaining a best linear unbiased estimator (BLUE), its predicting accuracy was not assessed, the authors focusing on the interpretation of significant variables instead. In addition, Cheng et al. (2018) proposed hierarchical models based on a Bayesian framework in order to address the same problem. However, in their empirical study, the model did not take into account the effect of trucks in traffic flow, which may result in unrealistic and unreliable results. Most importantly, neither of the two studies mentioned above accounted for the impact of downstream speed on upstream speed.

In this stufy, an improved structural equations system is proposed for lane-mean speed prediction. In relation to previous literature, the underlying methodology solves the endogeneity problem by utilizing a 3SLS model. One of the main contributions of the proposed system is the fact that it considers the effect of downstream speed on upstream speed. The results of this approach are compared to those obtained by Shankar and

Mannering (1998). Several tests are conducted in order to assess the feasibility and reliability of the approach.

3. Methodology

The methodology used in this study consists of several steps. First, a road segment is divided into shorter cells (see Fig. 1). Data on important factors (i.e. traffic data, environmental data and temporal data, etc.) that will serve as inputs in the estimation are collected in each cell for constructing an integrated and reliable model. Finally, the structural equations system will be employed on each cell to estimate the average vehicles speed in corresponding space.

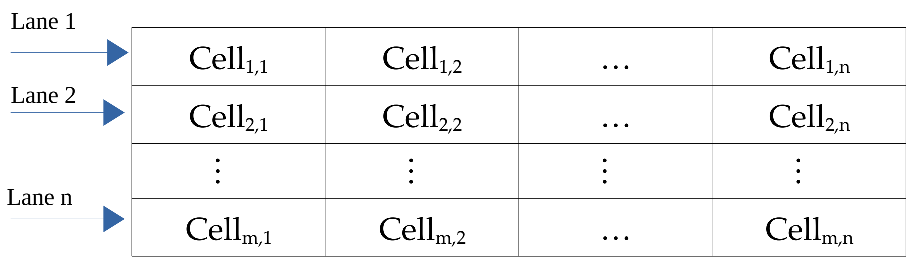

Fig. 1. Cells structure in a road segment

As previously described in Shankar and Mannering (1998), when using a structural equations system, a correlation between adjacent lane speeds was observed. This means that the lane-mean speed of a cell is determined not only by its own basic determinants (i.e. traffic data, environmental data and temporal data, etc.), but also by the speed of vehicles in adjacent lanes. This is because the adjacent lanes belonging to a particular segment are not isolated in space and the vehicle speeds in such lanes are interdependent (Cheng et al., 2018). As cells in the same segment are considered in a same structural equations system and their lane-mean speeds are estimated simultaneously, the lane-mean speed of a cell, which is the dependent variable in one equation, may be a regressor in other equations. Those lane-mean speeds that act as dependent variables and regressors at the same time are endogenous variables of the system.

Furthermore, it has been proven that upstream traffic flow is correlated with the downstream traffic condition (Lighthill and Whitham, 1955). Thus, it is necessary to consider the impact of downstream speed in the same lane when estimating the lane-mean speed of a cell. In this study, the lane-mean speed of a cell is one of the exogenous variables in the equation with lane-mean speed in its backward cell as the dependent variable. For instance, based on Fig. 1, the lane-mean speed of $\text{Cell}_{1,2}$ is an exogenous independent variable when estimating the lane-mean speed of $\text{Cell}_{1,1}$.

Hence, for lane-mean speeds, the system of equations can be written as follows (focused on segment j),

$$u_{1,j}=\alpha_{1,j}+\boldsymbol{\beta}_{\mathbf{1},\boldsymbol{j}}\cdot\boldsymbol{X}_{\mathbf{1},\boldsymbol{j}}+\boldsymbol{\lambda}_{\mathbf{1},\boldsymbol{j}}\cdot\boldsymbol{Z}_{\mathbf{1},\boldsymbol{j}}+\theta_{1,j}\cdot v_{1,j}+\eta_{1,j}\cdot u_{1,j+1}+\varepsilon_{1,j}$$

$$u_{2,j}=\alpha_{2,j}+\boldsymbol{\beta}_{2,j}\cdot\boldsymbol{X}_{2,j}+\boldsymbol{\lambda}_{2,j}\cdot\boldsymbol{Z}_{2,j}+\theta_{2,j}\cdot v_{2,j}+\eta_{2,j}\cdot u_{2,j+1}+\varepsilon_{2,j}$$
$$\dots$$
$$u_{i,j}=\alpha_{i,j}+\boldsymbol{\beta}_{i,j}\cdot\boldsymbol{X}_{i,j}+\boldsymbol{\lambda}_{i,j}\cdot\boldsymbol{Z}_{i,j}+\theta_{i,j}\cdot v_{i,j}+\eta_{i,j}\cdot u_{i,j+1}+\varepsilon_{i,j} \tag{1}$$
$$\dots$$
$$u_{m,j}=\alpha_{m,j}+\boldsymbol{\beta}_{m,j}\cdot\boldsymbol{X}_{m,j}+\boldsymbol{\lambda}_{m,j}\cdot\boldsymbol{Z}_{m,j}+\theta_{m,j}\cdot v_{m,j}+\eta_{m,j}\cdot u_{m,j+1}+\varepsilon_{m,j}$$

where i is the lane number, j is the segment number. $u_{i,j}$ is the lane-mean speed. $\boldsymbol{X}_{i,j}$ is the vector of exogenous variables that affect the speed of $\text{Cell}_{i,j}$ except for the average vehicles speed of next segment in the same lane. $\boldsymbol{Z}_{i,j}$ is the vector of endogenous variables that varies as the dependent variable $u_{i,j}$ varies except for the mean speeds of adjacent lanes. $v_{i,j}$ is the mean speed in the crucial lane adjacent to lane i in the same segment. $\alpha_{i,j},\boldsymbol{\beta}_{i,j},\boldsymbol{\lambda}_{i,j},\theta_{i,j},\eta_{i,j}$ are estimable coefficients. ε_i is disturbance term.

The crucial lane adjacent to lane i , is defined as the lane that has a slower average speed (i.e. more congested) between the two lanes adjacent to lane i . And $v_{i,j}$ is the average vehicles speed in this lane,

$$v_{i,j}=\begin{cases} min\{u_{i-1,j},u_{i+1,j}\}, if\ i\neq 1\wedge i\neq m \\ u_{2,j}, if\ i=1 \\ u_{m-1,j}, if\ i=m \end{cases} \tag{2}$$

In order to obtain unbiased and consistent results in estimating equations (1) which contain endogenous independent variables, 3SLS is appropriate.

4. Empirical Studies

We will analyze and assess the improved structural equations system by four scenarios with empirical traffic data:

- The first scenario is to illustrate the improvement and superiority of the presented structural equations system by comparing its predicting accuracy with the approach proposed in Shankar and Mannering (1998).
- The second scenario is to compare the predicting accuracy of mean speeds of different lanes that belong to same segments in order to analyze if it is reasonable to design a crucial adjacent lane for the lanes caught in the middle and also to assess the model's results in different lanes.
- The third scenario is to reveal in what transverse condition (number of lanes) can the system receive the best forecasting result, while the longitudinal condition (length of segment) is controlled for.
- The fourth scenario is the reverse of the third scenario: in what longitudinal condition can the system perform best with the transverse condition being controlled for.

References

Cheng, W., Gill, G. S., Sakrani, T., Ralls, D., & Jia, X. (2018). Modeling the endogeneity of lane-mean speeds and lane-speed deviations using a Bayesian structural equations approach with spatial correlation. *Transportation Research Part A: Policy and Practice*, *116*(March), 220–231.

Lighthill, M. J., & Whitham, G. B. (1955). On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*.

Shankar, V., & Mannering, F. (1998). Modeling the endogeneity of lane-mean speeds and lane-speed deviations: A structural equations approach. *Transportation Research Part A: Policy and Practice*.

Washington S. P., Karlaftis M. G., & Mannering F. (2010). Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC, 145-160.