ELSEVIER

Accident Analysis and Prevention 38 (2006) 1094–1100

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

The significance of endogeneity problems in crash models: An examination of left-turn lanes in intersection crash models

Do-Gyeong Kim[a,*], Simon Washington[b,1]

[a] *Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, AZ 85712-0072, United States*
[b] *Department of Civil and Environmental Engineering, Arizona State University, Tempe, AZ 85287-5306, United States*

Received 10 January 2006; received in revised form 6 April 2006; accepted 15 April 2006

Abstract

Crash prediction models are used for a variety of purposes including forecasting the expected future performance of various transportation system segments with similar traits. The influence of intersection features on safety have been examined extensively because intersections experience a relatively large proportion of motor vehicle conflicts and crashes compared to other segments in the transportation system.

The effects of left-turn lanes at intersections in particular have seen mixed results in the literature. Some researchers have found that left-turn lanes are beneficial to safety while others have reported detrimental effects on safety. This inconsistency is not surprising given that the installation of left-turn lanes is often endogenous, that is, influenced by crash counts and/or traffic volumes. Endogeneity creates problems in econometric and statistical models and is likely to account for the inconsistencies reported in the literature.

This paper reports on a limited-information maximum likelihood (LIML) estimation approach to compensate for endogeneity between left-turn lane presence and angle crashes. The effects of endogeneity are mitigated using the approach, revealing the unbiased effect of left-turn lanes on crash frequency for a dataset of Georgia intersections. The research shows that without accounting for endogeneity, left-turn lanes 'appear' to contribute to crashes; however, when endogeneity is accounted for in the model, left-turn lanes reduce angle crash frequencies as expected by engineering judgment. Other endogenous variables may lurk in crash models as well, suggesting that the method may be used to correct simultaneity problems with other variables and in other transportation modeling contexts.

Keywords: Endogeneity; Angle crashes; Left-turn lane indicator; Limited-information maximum likelihood estimation

1. Introduction

Crash prediction models are used for a variety of purposes; most frequently to estimate the expected crash frequencies from various roadway entities (highways, intersections, interstates, etc.) and also to identify geometric, environmental, and operations factors that are associated with crashes. Much research related to crash prediction relies heavily upon the use of econometric modeling techniques. As examples, numerous studies have modeled the safety impacts of installation of left-turn lanes on accident rates (Foody and Richardson, 1973; Gluck et al., 1999), passing-related accidents (Parker et al., 1983), specific types of accidents (McCoy and Malone, 1989; Poch and Mannering, 1996; Kim et al., 2006), and accident frequencies (Lacy, 1972; Dale, 1973; Bauer and Harwood, 1996).

Empirical examination of the safety effects of left-turn lanes is of interest for a variety of reasons. First, left-turn lanes are often justified in part for operational reasons; by providing additional capacity, reducing delay, and providing refuge for left-turning vehicles. Second, removing turning vehicles from through lanes is thought to reduce rear-end crashes and improve safety. In many cases left-turn lanes are warranted as a result of turning movement volumes, turning-related accident frequencies, or both. Finally, public monies are often invested to install left-turn lanes and so the safety 'return' on investment should be examined.

The prior empirical evidence on the relationship between the presence of left-turn lanes and accidents is mixed. Some of the previous research has revealed that left-turn lanes have been effective in decreasing the potential for accidents (Foody and

* Corresponding author. Tel.: +1 480 727 9806; fax: +1 480 965 0557.
E-mail addresses: dokkang@u.arizona.edu (D. Kim), simon.washington@asu.edu (S. Washington).
[1] Tel.: +1 480 965 3589; fax: +1 480 965 0557.

0001-4575/$ – see front matter
doi:10.1016/j.aap.2006.04.017

Richardson, 1973; Gluck et al., 1999; Parker et al., 1983; Lacy, 1972; Dale, 1973), while other studies have revealed that left-turn lanes are associated with higher frequencies of accidents (McCoy and Malone, 1989; Poch and Mannering, 1996; Kim et al., 2006). The detailed descriptions for the safety effects of left-turn lanes on different types and/or severity of accidents are well documented in Harwood et al. (2002). The mixed results in the literature tend to confuse the issue and are difficult to explain. One possibility is that left-turn lanes improve safety in some cases and worsen safety in others; however, it is hard to imagine a set of physical circumstances where left-turn lanes influence the operation of intersections in seemingly systematic and opposite ways. Perhaps a more likely explanation for the mixed results is that prior researchers have not controlled for the potential endogeneity of left-turn lanes, which will bias the results obtained from crash models. In this paper we aim to examine whether this explanation may account for the mixed results for left-turn lanes, and whether correcting for endogeneity offers promise for other improvements in safety models.

Many prior studies have treated the presence of left-turn lanes as exogenous and have not controlled for potential endogeneity. A study by Kim et al. (2006) illustrated that the presence of left-turn lanes is potentially endogenous because crash warrants (left-turn or total crashes exceeding a threshold) may often be used to warrant the installation of left-turn lanes. For example, the 2003 Oregon State Highway Design Manual states that installation of a left-turn lane is based upon vehicular volume, crash experience, and special cases such as railroad crossings, passing lanes, geometric/safety concerns, non-traversable medians, and signalized intersections. This finding suggests that in cases when crash warrants are met there exists a bi-directional relationship between the presence of left-turn lanes and accidents (i.e., the endogenous independent and dependent variable are interchangeable).

The simultaneous relationship between the presence of left-turn lanes and accidents requires that model parameters be simultaneously estimated. Without controlling for endogeneity (i.e., single equation modeling techniques), estimated parameters are biased and inconsistent (Ramanathan, 2002; Washington et al., 2003). Since biased estimators might over- or under-estimate the true effect of the variable of interest, endogeneity should be properly accounted for in order to obtain unbiased and consistent estimates. One method for obtaining unbiased estimators is to estimate parameters simultaneously within a set of structural equations, called simultaneous equation models (SEM). Simultaneous equations techniques can also be applied to effectively deal with endogeneity.

The paper describes the continuation of work conducted by the authors in Kim et al. (2006). The primary objective of this paper is to examine the endogeneity of left-turn lanes and angle crashes in crash models. For the analysis, two different models are estimated using the limited-information maximum likelihood (LIML) approach: a negative binomial model for angle crashes and a logit model for left-turn lanes (lane present or not-present). The objective of this research is to identify the true relationship between left-turn lanes and angle crashes by accounting for endogeneity, to assess the magnitude of the bias, and to shed light on the problem raised by endogeneity. The remainder of the paper is organized as follows: The literature review is presented in the next section, followed by a description of the data used in the study. Then, the LIML model estimation technique is described. Finally, a discussion of the model estimation results and conclusions are presented.

2. Literature review

Previous research has shown that endogeneity problems persist in crash data. Researchers have attempted to account for endogenous relationships to better understand the true effects on crashes of various kinds of variables such as seat belt usage rate on fatalities (Derrig et al., 2000), state-level policies related to drunk driving on fatal crash reductions (Eisenberg, 2003), and no-fault automobile insurance on fatal accident rates (Cummins et al., 2001). However, the vast majority of prior studies related to roadway crash prediction models have not addressed potential endogeneity problems. A Canadian study by Gaudry and Vernier (2002) dealt with the endogeneity between speed and road safety such as crash frequency and severity on Canadian roads. They proposed a simultaneous structure controlling for endogeneity consisting of three equations, two of which are logit models and one of which is a linear regression model.

Endogeneity problems, however, have been identified and discussed in other transportation topic areas. Shankar and Mannering (1998) developed structural models of mean speeds and speed deviations across lanes of a multilane highway to provide an improved understanding of highway operations. They found that the mean speed in each lane is endogenously related to the mean speeds in adjacent lanes, and in a similar fashion speed deviations in each lane have endogenous relationships with speed deviations in adjacent lanes. For the model estimation, three-stage least squares (3SLS) estimation was used because the model functional forms are linear.

Bhat and Koppelman (1993) used the SEM approach to develop an integrated model of employment, income, and household car ownership. They proposed a fundamental change in the traditional view of employment and income as exogenous variables in travel demand models. They regarded these variables as endogenous variables in disaggregate travel demand frameworks and estimated a joint model of employment, income, and car ownership, which takes account of the interdependencies among these variables and their structural relationships with relevant exogenous variables.

Abu-Eisheh (2001) also employed simultaneous equation estimation techniques to estimate automobile demand and driver population in the Palestinian Territories as a function of socioeconomic and political variables. In this study, Abu-Eisheh considered the interaction and simultaneity between automobile ownership and driver population, and thus formulated two linear regression models as a set of equations. These models were simultaneously estimated by the three-stage least squares (3SLS) methods. He also developed a dynamic automobile demand simulation model by utilizing a simultaneous-estimation system, which considers the interaction between supply and demand and the resulting equilibrium (Abu-Eisheh and Mannering, 2002).

Table 1
Variables used in the study

Variables	Definition	Mean	Min	Max	S.D.
ANGLE	Number of angle crashes	2.05	0	13	2.70
Log of AADTMAJ	Logarithm of AADT on major road	7.98	6.04	9.63	0.93
Log of AADTMIN	Logarithm of AADT on minor road	6.49	4.38	9.25	0.95
SIGNAL	Intersection type (0 if non-signalized intersection, 1 if signalized intersection)	0.27	0	1	0.45
RTLMAJ	Right-turn lane indicator (1 if at least one right-turn lane on the major road, 0 otherwise)	0.12	0	1	0.32
LTLMAJ	Left-turn lane indicator (1 if at least one left-turn lane on the major road, 0 otherwise)	0.20	0	1	0.40
RTLMIN	Right-turn lane indicator (1 if at least one right-turn lane on the minor road, 0 otherwise)	0.10	0	1	0.30
LTLMIN	Left-turn lane indicator (1 if at least one left-turn lane on the minor road, 0 otherwise)	0.12	0	1	0.33
DRWYMAJ	Number of driveways on major road within 250 ft of the intersection center	1.68	0	9	1.88
DRWYMIN	Number of driveways on minor road within 250 ft of the intersection center	1.81	0	11	1.70
LIGHTMAJ	Lighting indicator (1 if lighting exists on the major road, 0 otherwise)	0.18	0	1	0.38
LIGHTMIN	Lighting indicator (1 if lighting exists on the minor road, 0 otherwise)	0.16	0	1	0.39
TERNMAJ	Terrain on major road (0 = flat, 1 = rolling, 2 = mountainous)	0.57	0	2	0.56
TERNMIN	Terrain on minor road (0 = flat, 1 = rolling, 2 = mountainous)	0.71	0	2	0.55
SPDLIMAJ	Speed limit on major road in mph	46.73	25	55	8.05
SPDLIMIN	Speed limit on minor road in mph	38.77	20	55	6.76
HAU	Intersection angle variable in degree	0.53	−57.5	50.0	20.50

They also used the 3SLS methods to model the quantity and price of automobiles within a supply-demand framework.

3. Data descriptions

The data used to test for endogeneity in crash models are derived from 38 counties within the state of Georgia, which were also used in the previous study (Kim et al., 2006). The reason for using the same dataset is to explore how the estimated safety effect of left-turn lanes on angle crashes changes when controlling for endogeneity as compared to the prior study. A total of 155 rural intersections were included in the data: 113 non-signalized intersections and 42 signalized intersections of 2-lane 4-legged roads, with a total of 317 angle crashes (147 occurring at non-signalized intersections and 170 occurring at signalized intersections). Angle crashes occurring within the intersection or within 76 m (250 ft) from the intersection center along the major and minor road are included in the analysis. While this classification scheme may omit some intersection crashes and include some non-intersection crashes, it is commonly used in the US because it is a non-arbitrary criterion that is easily repeatable and generalizable across jurisdictions. For the analysis, 16 variables based on their theoretical role in safety were extracted from road characteristic files and geographic information system (GIS) roadmaps using the GIS technique. Descriptions of the variables used in the analysis and basic statistics are provided in Table 1.

4. Methodological approach

In this section, the methodological approach taken to account for endogeneity in the crash data is described. It is important to note that standard or 'canned' software could not be used to deal with the endogeneity due to complexities in the system of models.

4.1. Limited information maximum likelihood (LIML) estimation

The estimation of linear models with endogeneity is somewhat straightforward, but a situation in which a count dependent variable depends on a binary endogenous variable and vice versa is more complex because a simple reduced form does not exist.

Consider the following two equations:

$$y_{1i} = \exp(\alpha_1 y_{2i} + x_i'\beta_1 + \varepsilon_i) = \exp(\alpha_1 y_{2i} + x_i'\beta_1)\exp(\varepsilon_i) = \exp(\alpha_1 y_{2i} + x_i'\beta_1)u_i \qquad (1)$$

$$y_{2i}^* = \alpha_2 y_{1i} + z_i'\beta_2 + \nu_i \qquad (2)$$

where y_{1i} is a count endogenous variable, y_{2i} is a binary endogenous variable, x_i and z_i are exogenous variables, u_i and ν_i are disturbance terms that follow a gamma distribution and a logistic distribution, respectively, and y_{2i}^* is an unobservable variable. What we observe is a binary variable y_{2i} defined as

$$y_{2i} = \begin{cases} 1 & \text{if } y_{2i}^* > 0 \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

If both α_1 and α_2 are not equal to zero (i.e., dependent variables appear as a regressor in different equations), endogeneity exists and arises from simultaneity. As such the endogeneity cannot be adequately explained in a single equation estimation. Instead, model parameters must be simultaneously estimated through a system of equations. From the viewpoint of model estimation, however, the model described above cannot be estimated using fully simultaneous estimation due to logical consistency (Winkelmann, 2003; Windmeijer and Santos Silva, 1997; Cameron and Trivedi, 1998), which arises when a count dependent variable depends on a binary endogenous variable (as one case).

This model is logically consistent when the following condition is satisfied:

$$Pr(y_{2i}=1)+Pr(y_{2i}=0)=1 \tag{4}$$

where $Pr(y_{2i}=1)$ is the probability of y_{2i} being equal to 1 and $Pr(y_{2i}=0)$ is the probability of y_{2i} being equal to 0. The probability of y_{2i} being equal to 1 is obtained from the relations (2) and (3):

$$\begin{aligned}Pr(y_{2i}=1)&=Pr[\nu_i>-(\alpha_2 y_{1i}+z_i'\beta_2)]\\&=1-F[-(\alpha_2 y_{1i}+z_i'\beta_2)]=F(\alpha_2 y_{1i}+z_i'\beta_2)\end{aligned} \tag{5}$$

where F is the cumulative distribution function for ν_i, assuming symmetry of the density of ν_i. Assuming that ν_i is logistic distributed, Eq. (5) is rewritten as

$$F(\alpha_2 y_{1i}+z_i'\beta_2)=\frac{\exp(\alpha_2 y_{1i}+z_i'\beta_2)}{1+\exp(\alpha_2 y_{1i}+z_i'\beta_2)} \tag{6}$$

Substituting Eq. (1) for y_{1i} in Eq. (5), Eq. (4) is rewritten as

$$\begin{aligned}&F[\alpha_2\exp(\alpha_1+x_i'\beta_1+\varepsilon_i)+z_i'\beta_2]\\&\quad+\{1-F[\alpha_2\exp(x_i'\beta_1+\varepsilon_i)+z_i'\beta_2]\}=1\end{aligned} \tag{7}$$

or

$$\begin{aligned}&F[\alpha_2\exp(\alpha_1+x_i'\beta_1+\varepsilon_i)+z_i'\beta_2]\\&\quad=F[\alpha_2\exp(x_i'\beta_1+\varepsilon_i)+z_i'\beta_2]\end{aligned} \tag{8}$$

This condition implies that either $\alpha_1=0$ or $\alpha_2=0$ and therefore it is concluded that logical consistency prohibits fully simultaneous estimation (Winkelmann, 2003). By assuming that $\alpha_2=0$, Eqs. (1) and (2) are rewritten as

$$y_{1i}=\exp(\alpha_1 y_{2i}+x_i'\beta_1+\varepsilon_i) \tag{9}$$

$$y_{2i}^*=z_i'\beta_2+\nu_i \tag{10}$$

and Eq. (9) is estimated by replacing y_{2i} with its conditional mean, $E(y_{2i}|z_i)=F(z_i{}'\beta_2)$. However, estimation of Eq. (9) with the conditional mean function, $\exp[\alpha_1 F(z_i{}'\beta_2)+x_i{}'\beta_1]$ does not yield consistent estimates of the parameters. Consistent estimators for α_1 and β_1 are obtained by replacing y_{2i} by $F(z_i'\hat\beta_2)$, instead of $F(z_i{}'\beta_2)$, where $\hat\beta_2$ is the logit or probit estimator of β_2 (Windmeijer and Santos Silva, 1997). In this current study $\hat\beta_2$ is estimated using a logit model.

For the binary outcome model, consistent estimators for α_2 and β_2 are obtained in similar fashion. Assuming $\alpha_1=0$, the binary outcome model is estimated by substituting for y_{1i} in Eq. (5), $\exp(x_i'\hat\beta_1)$ where $\hat\beta_1$ is obtained using a negative binomial model. As a result, all of the model parameters are estimated using the following equations:

$$y_{1i}=\exp(\alpha_1\hat y_{2i}+x_i'\beta_1+\varepsilon_i) \tag{11}$$

$$y_{2i}^*=\alpha_2\hat y_{1i}+z_i'\beta_2+\nu_i \tag{12}$$

where $\hat y_{2i}=F(z_i'\hat\beta_2)$ and $\hat y_{1i}=\exp(x_i'\hat\beta_1)$. This approach mimics the logic of instrumental variable estimation in linear simultaneous equations models (Cameron and Trivedi, 1998).

Since Eqs. (11) and (12) do not reduce to simple forms for y_{1i} and y_{2i} as described previously, the parameters are estimated by maximizing the log-likelihood function using limited-information maximum likelihood (LIML) estimation (for the GMM estimation, see Windmeijer and Santos Silva, 1997). As a result contemporaneous correlation effects are not examined (see Washington et al., 2003).

All estimation and computation is completed using the GAUSS programming language. The standard errors of the parameters are obtained from the inverse of the negative of the Hessian matrix of the log-likelihood function.

4.2. Endogeneity tests

A general approach used to test for endogeneity is the Durbin–Wu–Hausman (DWH) test. However, LaFrance (1993) stated that a problem with the DWH test in small samples occurs if the difference between the covariance matrices for the two sets of parameter estimates is not positive semi-definite, and suggested an alternative approach based on the DWH test. Thus, the endogeneity test used in this study follows the approach proposed by LaFrance (1993), which is called the modified DWH test in this paper.

The test is based on the difference between parameter estimates with and without controlling for potential endogeneity. The null hypothesis is that parameters estimated without controlling for endogeneity are consistent. The alternative hypothesis is that parameters estimated without controlling for endogeneity are inconsistent, which implies endogeneity of the explanatory variables.

The modified DWH test statistic is given by

$$H=(\hat\theta_{\mathrm{SEE}}-\hat\theta_{\mathrm{LIML}})'[\mathrm{var}(\hat\theta_{\mathrm{SEE}})-\mathrm{var}(\hat\theta_{\mathrm{LIML}})]^{-1}(\hat\theta_{\mathrm{SEE}}-\hat\theta_{\mathrm{LIML}})$$

where $\hat\theta_{\mathrm{SEE}}$ is the vector of estimated parameters obtained from single equation estimation (without controlling for endogeneity) and $\hat\theta_{\mathrm{LIML}}$ is the vector of estimated parameters using LIML estimation (with controlling for endogeneity). Under the null hypothesis, the test statistic is asymptotically distributed as $\chi^2(k)$, where k is the number of positive elements in the diagonal of $\mathrm{var}(\hat\theta_{\mathrm{SEE}})-\mathrm{var}(\hat\theta_{\mathrm{LIML}})$.

5. Modeling results

First, two individual models are estimated separately to compare estimated parameters with and without controlling for the endogeneity of a left-turn lane indicator variable. One is a negative binomial model for angle crashes and the other is a logit model for left-turn lane indicator. A total of 16 variables were used at the beginning of the estimation procedure for both of models, and were then re-estimated with only significant variables. Because the models are used primarily for forecasting purposes a 90% level of confidence is used throughout.

The estimation results not accounting for endogeneity are summarized in Tables 2 and 3. Five variables were found to be statistically significant in the angle crash model, while four variables were statistically significant in the left-turn lane indicator

Table 2
Estimation results for angle crashes (negative binomial regression model)

Variables	Estimated coefficient	*t*-Statistic	*p*-Value
Constant	−4.2040	−3.98	0.000
Log of AADT on the major road	0.3227	2.57	0.010
Log of AADT on the minor road	0.2868	2.80	0.005
Left-turn lane indicator (1 if at least one left-turn lane on the major road, 0 otherwise)	0.6897	2.87	0.004
Number of driveways on the major road within 250 ft of the intersection center	0.1142	2.23	0.026
Lighting indicator (1 if light on the major road, 0 otherwise)	−0.5925	−2.23	0.026
α (Dispersion parameter)	0.5346	3.72	0.000
Number of observations	155		
Log-likelihood at convergence	−269.64		

Table 3
Estimation results for left-turn lanes (logit model)

Variables	Estimated coefficient	*t*-Statistic	*p*-Value
Constant	−16.359	−3.46	0.001
Log of AADT on the major road	1.6469	2.91	0.004
Number of angle crashes	0.2134	1.75	0.080
Signal indicator (1 if signalized intersections, 0 otherwise)	3.1564	4.20	0.000
Number of driveways on the major road within 250 ft of the intersection center	−0.4476	−2.52	0.012
Number of observations	155		
Log-likelihood at convergence	−35.36		

model. For the angle crash model, four variables have a positive relationship with the number of angle crashes, while one variable reveals a negative relationship. As described in prior studies, AADT and the number of driveways are found to be positively associated with angle crashes; the former capturing roadway exposure and the latter representing increased opportunities for angle crashes. Also, the presence of left-turn lanes (the focus of this study) is positively associated with angle crashes. As discussed previously, it is typically assumed that installation of a left-turn lane can effectively reduce angle crashes; however the estimation result is counter-intuitive. The reason, as previously discussed in detail, might be that the left-turn lane indicator is endogenous. Finally, a lighting indicator variable has a negative relationship with angle crashes, which suggests that the presence of lighting effectively reduces angle crashes (although this variable too could potentially be endogenous).

For the left-turn lane indicator model, AADT on the major road, number of angle crashes, signal indicator, and number of driveways on the major road were found to be statistically significant as expected. AADT on the minor road was found to be

Table 4
Estimation results with controlling for endogeneity (LIML approach)

Variables	Estimated coefficient	*t*-statistic	*p*-value
Eq. (1): angle crashes (dependent variable)			
Constant	−2.0147	−1.87	0.063
Log of AADT on the major road	0.1289	2.12	0.036
Log of AADT on the minor road	0.2756	2.47	0.015
Left-turn lane indicator (1 if at least one left-turn lane on the major road, 0 otherwise)	−0.1653	−2.06	0.041
Number of driveways on the major road within 250 ft of the intersection center	0.1583	2.51	0.013
Lighting indicator (1 if light on the major road, 0 otherwise)	−0.6520	−2.24	0.027
α (Dispersion parameter)	0.6073	4.08	0.000
Number of observations	155		
Log-likelihood at convergence	−271.95		
Eq. (2): left-turn lane indicator (dependent variable)			
Constant	−18.255	−3.45	0.000
Log of AADT on the major road	1.9397	2.91	0.004
Number of angle crashes	0.0750	1.99	0.048
Signal indicator (1 if signalized intersections, 0 otherwise)	3.6051	4.22	0.000
Number of driveways on the major road within 250 ft of the intersection center	−0.4774	−2.77	0.006
Number of observations	155		
Log-likelihood at convergence	−37.01		

insignificant in this model. This is likely because the dependent variable of interest is the presence of left-turn lanes on the major routes. The significant variables associated with installation of left-turn lanes are consistent with the left-turn warrant criteria presented in the 2003 Oregon State Highway Design Manual. Increased AADT, higher angle crashes, and signalized intersections are more likely to be associated with the installation of left-turn lanes. In contrast, a relatively high number of driveways tends to control installation of left-turn lanes.

Next, a two-model system was estimated simultaneously using the LIML approach described previously. For the simultaneous estimation only significant variables were used: five variables for the angle crash model and four variables for the left-turn indicator model. The estimation results when controlling for endogeneity are presented in Table 4.

The results show that there are no major differences between the estimated safety effects of estimated parameters with and without controlling for endogeneity except for the effect on angle crashes of the left-turn lane indicator. Single equation estimation results in a positive relationship between the left-turn lane indicator and angle crashes (0.6897), while the presence of left-turn lanes was found to be negatively associated with angle crashes (−0.1653) when the LIML approaches are applied, as engineering logic and theory would suggest. In addition, the number of angle crashes was found to be statistically significant in the left-turn lane indicator model when the LIML approaches are applied. This finding suggests that angle crashes are endogenously associated with a left-turn lane indicator and vice versa, and the true effect of the presence of left-turn lanes for these data is obtained by accounting for endogeneity. It also suggests that ignoring the endogeneity will lead to erroneous, counter-intuitive conclusions.

Finally, the modified DWH test was applied to examine the endogeneity of angle crashes and the left-turn lane indicator. For the modified DWH test, the null hypothesis is that parameters estimated without controlling for endogeneity are consistent. Table 5 shows the results for endogeneity test. For the angle crash model, the null hypothesis is rejected at the 95% significance level because the modified DWH test statistic (110.09) is greater than the critical value (14.067, d.f. = 7, $p = 0.05$). Therefore, it is concluded that parameters estimated without controlling for endogeneity are inconsistent, implying that independent variables are endogenous. The test result for the left-turn lane indicator model also shows that independent variables are endogenous since the test statistic (30.045) is greater than the critical value (11.071, d.f. = 5, $p = 0.05$).

Table 5
Results for endogeneity test

	DWH test statistic	Critical value
Angle crash model	110.09	14.067 (d.f. = 7, $p = 0.05$)
Left-turn lane model	30.045	11.071 (d.f. = 5, $p = 0.05$)

Note: Angle crash model is used to test for the endogeneity of a left-turn lane indicator, while left-turn lane indicator model is used to test for the endogeneity of angle crashes. H_0: Parameters estimated without controlling for endogeneity are consistent. H_1: Parameters estimated without controlling for endogeneity are inconsistent.

6. Conclusions and recommendations

Past research has revealed inconsistent relationships between the presence of left-turn lanes and different types and/or severity of accidents. Some prior studies have found that installation of left-turn lanes is effective in reducing crashes and improving safety, while others have shown that the presence of left-turn lanes is associated with higher frequencies of accidents, contrary to engineering expectation and theory. We believe that the inconsistency observed in the literature is not an artifact of the underlying phenomenon but instead a problem in numerical estimation arising from simultaneity. This simultaneity means that in some cases (studies, sites, etc.) or for a fair portion of intersections the presence of left-turn lanes are installed (are affected by) as a result of the crash history of the intersection, whereas the left-turn lane also directly impacts safety through the channeling of movements and impact on conflicts and crashes. Thus, to sort out the ‘real’ effect of left-turn lanes on safety simultaneity must be addressed. Although prior studies have recognized this potential effect, none to the author’s knowledge have controlled for endogeneity between left-turn lanes and safety—the prime focus of this study.

The results show that there are no substantive differences among model parameters estimated using either single equation estimation or simultaneous equation estimation techniques, except for the effect of left-turn lanes (the suspected endogenous variable). The presence of left-turn lanes is found to be positively associated with angle crashes when the angle crash model and left-turn indicator models are estimated separately (ignoring simultaneity), while the LIML approach reveals that the presence of left-turn lanes reduces angle crashes, in agreement with engineering expectation. Furthermore, the endogenous relationship between a left-turn lane indicator and angle crashes is found to be statistically significant using an endogeneity test. This finding suggests that the endogenous relationship must be accounted for in the estimation of crash models to better understand the safety effect of the presence of left-turn lanes on crashes.

The model developed in this study has some limitations. For example, the predictive ability of the angle crash model may be improved by including turning movements at intersection approaches rather than using overall intersection AADT. In addition, installation of left-turn lanes is more likely to be affected by turning movements as well as opposite through traffic volumes. However, directional movements were unavailable for this study. Therefore, it is necessary to improve the ability of model prediction in future studies using turning movements. It is also possible that other endogenous variable exist, such as the presence of intersection lighting. These other possibilities need to be explored.

The approach described in this paper is generalizable to other safety modeling applications as well as to other transportation disciplines. It is possible and likely that many features of intersections are installed in response to traffic crashes, raising the strong likelihood of endogeneity related problems. Ignoring these relationships yields at times, as shown here, counterintuitive findings. The interpretation of model outputs, as a result, can lead conclusions that are opposite to expectation.

Further research needs to be conducted to deal with more than one endogenous variable in these models, and to seek out other possible endogeneity issues. The available software and estimation of such models remains relatively difficult, and so improvements in this area are also welcomed.

References

Abu-Eisheh, S.A., 2001. Modeling Automobile Demand and Driver Population in Palestinian Territories. Transportation Research Record 1752. National Academy of Sciences, Washington, DC, pp. 108–116.

Abu-Eisheh, S.A., Mannering, F.L., 2002. Forecasting automobile demand for economies in transition: a dynamic simultaneous-equation system approach. Transport. Plann. Technol. 25 (4), 311–331.

Bauer, K.M., Harwood, D.W., 1996. Statistical Models of At-Grade Intersection Accidents, FHWA-RD-96-125, Federal Highway Administration.

Bhat, C.R., Koppelman, F.S., 1993. An endogenous switching simultaneous equation system of employment, income, and car ownership. Transport. Res. Part A 27 (6), 447–459.

Cameron, A.C., Trivedi, P.K., 1998. Regression Analysis of Count Data. Econometric Society Monographs 30, Cambridge University Press.

Cummins, J.D., Philips, R.D., Weiss, M.A., 2001. The incentive effects of no-fault automobile insurance. J. Law Econ. XLIV, 427–464.

Dale, C.W., 1973. Cost-Effectiveness of Safety Improvement Project. Federal Highway Administration, Washington, DC.

Derrig, R.A., Segui-Gomez, M., Abtani, A., 2000. The effect of seat belt usage rates on the number of motor vehicle-related fatalities. In: Proceedings of the Risk Theory Society, Baltimore, MD.

Eisenberg, D., 2003. Evaluating the effectiveness of policies related to drunk driving. J. Policy Anal. Manage. 22 (2), 249–274.

Foody, T.J., Richardson, W.C., 1973. Evaluation of Left Turn Lanes as a Traffic Control Device. Ohio Department Transportation.

Gaudry, M., Vernier, K., 2002. Effects of Road Geometry and Surface on Speed and Safety. University of Montreal, Publication AJD-18.

Gluck, J., Levinson, H.S., Stover, V., 1999. Impacts of Access Management Techniques. NCHRP Report 420, Transportation Research Board.

Harwood, D.W., Bauer, K.M., Potts, I.B., Torbic, D.J., Richard, K.R., Kohlman Rabbani, E.R., Hauer, E., Elefteriadou, L., 2002. Safety Effectiveness of Intersection Left- and Right-Turn Lanes. FHWA-RD-02-089, Federal Highway Administration.

Highway Design Manual: Appendix F, 2003. Oregon Department of Transportation.

Kim, D., Washington, S.P., Oh, J., 2006. Modeling Crash Types: New Insights into the Effects of Covariates on Crashes at Rural Intersections. J. Transport. Eng., Am. Soc. Civil Eng. 132 (4), 282–292.

Lacy, J.D., 1972. Traffic operations program to increase capacity and safety. Traffic Q 26 (3), 327–340.

LaFrance, J.T., 1993. Weak separability in applied welfare analysis. Am. J. Agricult. Econ. 75 (3), 770–775.

McCoy, P.T., Malone, M.S., 1989. Safety Effects of Left-Turn Lanes on Urban Four-Lane Roadways. Transportation Research Record 1239, TRB. National Research Council.

Parker, M.R., Flak, M.A., Tsuchiyama, K.H., Wadenstorer, S.C., Hutcherson, F., 1983. Geometric Treatments for Reducing Passing Accidents at Rural Intersections on Two-Lane Rural Highways: vol. I—Final Report, Publication FHWA-RD-83-074, Federal Highway Administration.

Poch, M., Mannering, F.L., 1996. Negative binomial analysis of intersection accident frequencies. In: Presented at the 75th Annual Meeting of the Transportation Research Board.

Ramanathan, R., 2002. Introductory Econometrics with Applications, 5th ed. Harcourt College Publisher.

Shankar, V., Mannering, F.L., 1998. Modeling the endogeneity of lane-mean speeds and lane-speed deviations: a structural equations approach. Transport. Res. Part A 32 (5), 311–322.

Washington, S., Karlaftis, M., Mannering, F., 2003. Statistical and Econometric Methods for Transportation Data Analysis. Chapman and Hall, Boca Raton.

Windmeijer, F.A.G., Santos Silva, J.M.C., 1997. Endogeneity in count data models: an application to demand for health care. J. Appl. Econometr. 12, 281–294.

Winkelmann, R., 2003. Econometric Analysis of Count Data. Springer, Germany.