Multivariate Analysis of Accident Related Outcomes with Respect to Contemporaneous Correlation and Endogeneity: Application of Simultaneous Estimation Techniques

MULTIVARIATE ANALYSIS OF ACCIDENT RELATED OUTCOMES WITH RESPECT TO CONTEMPORANEOUS CORRELATION AND ENDOGENEITY: APPLICATION OF SIMULTANEOUS ESTIMATION APPROACHES

By

Do-Gyeong Kim

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF CIVIL ENGINEERING AND ENGINEERING MECHANICS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY
WITH A MAJOR IN CIVIL ENGINEERING

In the Graduate College

THE UNIVERSITY OF ARIZONA

2006

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation

prepared by Do-Gyeong Kim

entitled Multivariate Analysis of Crash Related Outcomes with Respect to Contemporaneous Correlation and Endogeneity: Application of Simultaneous Estimation Approaches

and recommend that it be accepted as fulfilling the dissertation requirement for the

Degree of Doctor of Philosophy

_____ Date: 04/12/2006
Simon Washington

_____ Date: 04/12/2006
Mark Hickman

_____ Date: 04/12/2006
Wei Lin

_____ Date: 04/12/2006
Satheesh Aradhyula

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____ Date: 04/12/2006
Dissertation Director: Simon Washington

_____ Date: 04/12/2006
Dissertation Director: Mark Hickman

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED:_____

Do-Gyeong Kim

ACKNOWLEDGEMENTS

My most sincere thanks and gratitude go to my advisor, Dr. Simon Washington, for his continued support, guidance and assistance, and encouragement and belief in my potential throughout the entire doctorate program. Without his support and guidance, this research would have been impossible. I would also like to express my appreciation to other member of my committee, Dr. Mark Hickman, Dr. Wei Hua Lin, and Dr. Satheesh Aradhyula, for their valuable suggestions and comments to improve the quality of my dissertation research.

I am also thankful to all of my colleagues in the Transportation Program of the Department of Civil Engineering and Engineering Mechanics, Aichong, Sudeshna, Wen, Felipe, Alejandro, Kangwon, and many others, for their friendship, support, and valuable challenges.

Finally, I would like to express my deepest appreciation and love to my wife, Yuhwa, and my daughter, Yeeun. They provided me with a great deal of encouragement and support throughout the duration of this research. They also sacrificed a great deal of time away from me. There is no doubt in my mind that without their support, sacrifice, patience and understanding, I could not have completed this dissertation work. My special thanks and gratitude go to my parents and brother for their encouragement and support to enable me to pursue my higher education.

DEDICATION

To my loving wife and daughter,

Yuhwa and Yeeun,

for their endless love, support, and sacrifice.

And also to my parents

for encouraging me to realize my dream.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS - Continued

Page

TABLE OF CONTENTS - Continued

Page

LIST OF ILLUSTRATIONS

Page

LIST OF TABLES

LIST OF TABLES – Continued

Page

ABSTRACT

Motor vehicle crashes have increasingly become a serious concern for highway safety engineers and transportation agencies over the past few decades. This serious concern has led to a great deal of research activities. One of these activities is to develop safety analysis tools, specifically crash prediction models, for the purpose of reducing crashes and enhancing highway safety.

Crash prediction models based on statistical or econometric modeling techniques are used for a variety of purposes; most commonly 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. A comprehensive review of prior literature indicates that many researchers have mainly focused on the development of aggregate crash prediction models based on single equation estimation techniques to identify the influences of geometric, environmental, and traffic variables on a single counted outcome. In some cases, however, more than one dependent variable might be of interest and hence several equations are formulated at the same time. Such a multiple equation structure may require simultaneous (or joint) estimation techniques under some situations.

This dissertation research develops simultaneous estimation approaches to account for contemporaneous correlation and endogeneity problems in crash data. Specifically, seemingly unrelated negative binomial models and simultaneous equation models are developed to account for contemporaneous correlation between the disturbance terms across crash type models and to control for the endogenous relationship between the

presence of left-turn lanes and angle crashes.

Modeling crash types may provide certain advantages to gain insights as to 1) identification of high-risk sites with respect to specific types of crashes, which is not revealed through crash totals, and 2) the differences between conditions that lead to various crash types, but the disturbance terms across crash types might be contemporaneously correlated due to the unobserved common characteristics. Therefore, individual and simultaneous crash type models were estimated and the results of both models were compared. The results showed that a simultaneous estimation approach provides more efficient estimators relative to a single equation estimation technique.

The presence of left-turn lanes has been treated as exogenous in crash prediction models, but in fact they are affecting each other. The bi-directional relationship between left-turn lanes and crashes results in endogeneity. This research investigated the endogenous relationship between left-turn lanes and crashes and developed simultaneous equation models to control for the endogeneity. The findings indicated that the presence of left-turn lanes is endogenously associated with crashes and the real effect of left-turn lanes on crashes can be obtained by controlling for endogeneity.

CHAPTER 1
INTRODUCTION

1.1 Research Background

Every year a number of drivers and road users are killed and injured in motor vehicle crashes. According to the 2003 Traffic Safety Facts, 42,643 people lost their lives and 2,889,900 people were injured in motor vehicle crashes in 2003. In addition, deaths and injuries resulting from motor vehicle crashes were the leading cause of death for persons of every age from 2 through 33 years old. To do this, motor vehicle crashes have increasingly become a serious concern for highway safety engineers and transportation agencies over the past few decades. This serious concern has led to a great deal of research activities. One of these activities is to develop safety analysis tools, specifically crash prediction models, for the purpose of reducing crashes and enhancing highway safety.

Crash prediction models based on statistical or econometric modeling techniques are used for a variety of purposes; most commonly 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. Being able both to predict the total expected frequencies of crashes and to identify the safety effects of various variables on crashes enables highway safety engineers and transportation agencies to:

1. Identify high risk sites with respect to total accidents, fatalities, and serious injuries based on the predicted frequencies; therefore, an efficient project can be

prioritized to enhance roadway safety with regard to the specific type of crashes;

2. Establish highway safety plans to reduce each type of crash (total crashes, fatalities, injuries, run-off-road crashes, etc); for example, more appropriate and effective countermeasures can be established to reduce run-off-road crashes based on the factors that are significantly associated with run-off-road crashes, and;
3. Make roads safer by applying crash prediction models to road design.

Econometric modeling techniques have been extensively used for estimating crash prediction models on highway segments and at intersections as a function of geometric characteristics, environmental factors, and traffic control elements. Although much research has dealt with a variety of crash related problems and identified the safety effects of various explanatory variables on crashes, there still remain a number of issues that have rarely been discussed in prior research, but should be identified and explained to better understand the complicated relationships between crashes and various factors.

Many researchers have mainly focused on the development of aggregate crash prediction models, whereby concentration is on the influences of geometric, environmental, and traffic variables on a single counted outcome variable such as the total expected number of crash frequencies, fatal crashes, injury crashes, and so on. Consequently, much research related to crash prediction models relies upon single equation modeling techniques. In some cases, however, more than one dependent variable might be of interest, and hence several equations are formulated at the same time as a set of equations. Although a set of equations can be estimated by conventional modeling

techniques such as single equation estimation (i.e., equation-by-equation estimation), such a multiple equation structure may require simultaneous (or joint) estimation in some situations. The popular examples that require simultaneous estimation include contemporaneous correlation and endogeneity, which are issues addressed in this research.

It is well known that contemporaneous correlation results in consistent but inefficient estimators (Greene, 2003; Washington et al., 2003) and endogeneity leads to biased and inconsistent estimators (Ramanathan, 2002; Washington et al., 2003). Of course these issues have been discussed in prior transportation-related studies, but most of previous research has concentrated on dealing with contemporaneous correlation and endogeneity problems related to continuous outcome variables. In other words, contemporaneous correlation and endogeneity problems in count data models have rarely been discussed. A possible reason for this might be the fact that it is not easy to develop simultaneous estimation techniques for count data models relative to linear regression models.

Therefore, this research aims to investigate whether or not contemporaneous correlation and endogeneity problems exist in crash data and develop simultaneous estimation techniques to control for contemporaneous correlation and endogeneity for crash prediction models.

1.2 Problem Statement

Due to the complexity of traffic movements and the potential number of conflicts between motor vehicles, pedestrians, and bicyclists, there is relatively large potential for crashes at intersections (compared to segments). From the desire to understand factors

that are associated with crashes and for estimating expected crash frequencies, researchers routinely estimate crash prediction models of intersections using roadway geometric variables, environmental factors, and traffic conditions as predictors. However, many researchers have mainly focused on the development of aggregate crash prediction models to predict the total expected number of crashes at intersections, but crash prediction models focused on predicting different crash types have rarely been developed.

According to the results of a previous study (Shankar et al., 1995), separate regression models for specific types of crashes would have the potential for providing greater explanatory power relative to a single overall frequency model. In other words, specific types of crashes are associated with roadway geometric variables and environmental factors in different ways, and thus separate models allow coefficient estimates to vary by the type of crash. This indicates that it is necessary to estimate individual crash models for different crash types because modeling crash types separately enables researcher and highway safety engineers to better understand the different effects of a variety of independent variables on specific types of crashes.

Moreover, there are at least three important and defensible reasons for estimating separate crash type models:

1. The first is the need to identify sites that are high-risk with respect to specific crash types but that are not revealed through crash totals. For example, an urban intersection may produce fairly typical numbers of total crashes but possess outlying numbers of rear-end or angle crashes with respect to similar sites. Being able to determine the ‘expected’ number of crashes by crash types enables a

researcher to identify these outlying types of crashes, which in turn may indicate a specific deficiency at the intersection;

2. A second use of these models is to gain an understanding of the differing effects of geometric, traffic, and environmental factors on crash type, so that countermeasure effects may be better understood. For example, it may be learned that pavement resurfacing increases roll over crashes but reduces rear-end crashes, and;
3. Finally, there is *a priori* reason to believe that certain types of crashes are associated with road geometry, the environment, and traffic variables in different ways. For example, dark conditions at intersections may influence angle crashes more than run-off-road crashes—suggesting that these crash types may be better explained in crash models through different coefficients.

From a methodological point of view, separate crash type models can be estimated by equation-by-equation count data models such as Poisson and negative binomial regression models. However, crash type models might be simultaneously estimated in a system of seemingly unrelated negative binomial regression equations due to potential contemporaneous correlation between the disturbance terms across crash type models.

Contemporaneous correlation arises when the disturbance terms in the system of equations are correlated. Neglecting contemporaneous correlation (i.e., use of single equation estimation) may yield consistent but inefficient estimators (Greene, 2003; Washington et al., 2003) because the assumption that the disturbance terms are

independent across equations is violated; that is, the covariance of the disturbance terms in different equations is not equal to zero. In such a case, all of the parameters in different equations should be estimated simultaneously to obtain more efficient estimators. The common way to solve a contemporaneous correlation problem is a Seemingly Unrelated Regression (SURE) approach, which was proposed by Zellner (1962).

To explicitly illustrate contemporaneous correlation problems in crash type models, suppose that one may want to estimate crash prediction models for angle and rear-end crashes for urban intersections. Since the number of angle and rear-end crashes are collected from same intersections, angle and rear-end crash models are estimated using identical independent variables as follows:

E(Angle) = f(roadway geometry, environmental and traffic factors, and error term)

E(Rear-end)= f(roadway geometry, environmental and traffic factors, and error term)

In such a case, it is reasonable to believe that angle and rear-end crashes might be correlated because they share unobserved characteristics as well as observed independent variables. As a result, contemporaneous correlation problems may exist in this case due to the common omitted variables. Besides this example, there are a number of crash-related applications in which contemporaneous correlation may exist. Examples include the estimation of crash severities (i.e., fatal, injury, and property damage only) and roadway and roadside crashes. They also share unobserved common characteristics, and thus contemporaneous correlation might be expected to exist across models.

A few efforts to account for contemporaneous correlation can be found in prior

literature. Ladron de Guevara et al. (2004) developed planning level crash prediction models for fatal and injury crashes by accounting for contemporaneous correlation, and Chayanan et al. (2003) explored contemporaneous correlation between roadway and roadside accident rates. However, contemporaneous correlation across crash type models has never been addressed in prior research, suggesting the need for estimating crash type models simultaneously to accommodate contemporaneous correlation.

Next, endogeneity problems need to be addressed. The majority of previous studies have estimated crash prediction models as a function of geometric characteristics, environmental factors, and traffic control elements, and in most cases, these kinds of independent variables have been mainly treated as exogenous variables and used to explain crash phenomenon. However, not all of these independent variables might be exogenous. A representative example of potentially endogenous variables is a left-turn lane indicator variable, which indicates the presence of left-turn lanes, because there is an interaction between left-turn lane indicator variables and crashes.

Regarding left-turn lane indicator variables as exogenous implies that installation of left-turn lanes may affect accident frequencies or accident severities, but cannot be affected by accident frequencies or accident severities. The fact that crash frequencies have been frequently used to justify installation of left-turn lanes, however, indicates that crash frequencies also influence installation of left-turn lanes. According to the 2003 Highway Design Manual of the state of Oregon, installation of a left-turn lane at intersections is determined based on vehicular volume, crash experience, and special

cases such as railroad crossings, passing lane, geometric/safety concerns, non-traversable median, and signalized intersection. Since installing left-turn lanes is affecting crash frequencies and vice versa, there exists a bi-directional relationship between left-turn lanes and crashes, and endogeneity arises from the bi-directional relationship (i.e., independent and dependent variables are affecting each other and thus dependent variables appear as a regressor in different equations).

Another evidence for the potential endogeneity of left-turn lanes is that the safety effects of left-turn lanes on crashes have been mixed. Some of the previous research has revealed that left-turn lanes have been effective in decreasing the potential for accidents (Foody and 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 mixed results in the literature tend to confuse the issue. A possible explanation for the mixed results is that prior researchers have not controlled for the potential endogeneity of left-turn lanes.

In addition to the left-turn lane indicator variable, there are a number of examples for endogeneity arising from a bi-directional relationship. Another example of endogenous variables might be a traffic signal indicator since traffic signals are warranted as a result of crash frequencies, implying that they are affecting each other.

Compared to contemporaneous correlation, endogeneity creates serious estimation problems. Without controlling for endogeneity that arises from a bi-directional relationship, the parameters estimated will be biased and inconsistent (Ramanathan,

2002; Washington et al., 2003) due to the violation of the assumption that the conditional expectation of ε_i given X_i must be equal to zero. Since biased estimators might over- or underestimate the true effects of the variable of interest, endogeneity should be properly explained to obtain unbiased and consistent estimators. One effective way to obtain unbiased estimators is to estimate parameters simultaneously within a set of equations, called a Simultaneous Equation Model (SEM).

Numerous studies have been conducted to account for endogenous relationships to better understand the real 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, all of these studies have controlled for endogeneity associated with safety policy related variables. The potential endogeneity of geometric variables have never been controlled for in previous studies. Furthermore, these studies estimated crash prediction models by the use of single equation estimation techniques, whereas simultaneous estimation techniques that control for a bi-directional endogenous relationship have never been developed in crash-related studies. The reason for this is likely that the potential endogeneity of geometric characteristics have been neglected. Consequently, the potential endogeneity of geometric characteristics needs to be examined and should be properly accounted for to better understand the relationships between crashes and geometric design features if they are endogenously related.

1.3 Research Objectives

This research aims to develop multiple equation estimation techniques, which jointly estimate a set of equations, for predicting more reliable crash prediction models and for better understanding the effects of the variables of interest by accounting for contemporaneous correlation and endogeneity. The following research objectives have been identified through this research to achieve the main goal:

1. This research develops crash prediction models for different crash types such as angle, rear-end, head-on, sideswipe (both same direction and opposite direction), and pedestrian-involved crashes to 1) demonstrate that different crash types are associated to predictor variables in different ways, 2) illustrate that high-risk sites with respect to specific types of crashes are not revealed through crash totals, and 3) show that estimation of crash type models may lead to greater insights regarding crash occurrence and countermeasure effectiveness.
2. This research investigates whether or not the disturbance terms in different crash type models are contemporaneously correlated. Also, a simultaneous estimation technique, a Seemingly Unrelated Negative Binomial (SUNB) model, to accommodate contemporaneous correlation is developed. Then, the parameters estimated by two approaches are compared to show the gain in efficiency.
3. The potential endogeneity of left-turn lanes and angle crashes is also investigated. In order to account for endogenous relationship between the presence of left-turn lanes (a binary endogenous variable) and angle crashes (a count endogenous variable), this research develops a simultaneous negative binomial-logit model.

1.4 Dissertation Organization

This dissertation is organized into six chapters. Chapter 1 is the introduction of the research in which the research background, problem statement, and research objectives are detailed. Chapter 2 presents a detailed literature review of prior studies regarding crash prediction models, contemporaneous correlation, endogeneity, and econometric analysis. Chapter 3 introduces statistical methodologies used in this research for estimating crash prediction models. Specifically, this chapter describes multiple equation estimation techniques that deal with contemporaneous correlation and control for the potential endogeneity. Chapter 4 presents data sources and data collection efforts. In Chapter 5 the estimation results and findings are discussed. Chapter 6 provides the conclusions that highlights the main results, and gives recommendations and suggestions for future research.

CHAPTER 2

LITERATURE RIVEW

2.1 Introduction

This chapter introduces the literature on crash type models, the effect of left-turn lanes, and econometric analysis concerning contemporaneous correlation and endogeneity. The literature review begins with an overview of previous research related to crash type models. Then, significant studies that employed simultaneous equation estimation techniques to solve contemporaneous correlation and endogeneity problems are presented.

2.2 Literature on Crash Type Models

Considerable past research has concentrated on the development of crash prediction models at intersections. Much of this research (Harwood et al., 2000; Vogt, 1999; Vogt and Bared, 1998; Oh et al., 2003; Lyon et al., 2003) has focused on modeling the relationship(s) between total, fatal, and injuries crashes with intersection geometric characteristics, environmental factors, and traffic-related explanatory variables. While this research is extremely useful for understanding and forecasting total crashes and for understanding the general impacts of countermeasures and/or road features, it does not reveal a disaggregate “picture” of crash events at intersections. In contrast, research conducted to investigate the safety effects of roadway geometric, traffic, and environmental factors on the different crash types is scant.

An early study by Hauer et al. (1988) developed crash type prediction models for 15 different accident patterns at urban and suburban signalized intersections in Toronto,

Canada. They defined each accident pattern based on the maneuvers of the two vehicles before the collision. These models required turning movements at intersections as inputs—which in fact were shown to significantly improve predictive ability of the models. Another study conducted on Canadian roads (Persaud and Nguyen, 1998) developed two levels of models based on data inputs for three- and four-legged signalized intersections. The Level 2 models were similar to the model developed by Hauer et al. (1988), while the Level 1 models developed 'aggregate' prediction models for crash types such as rear-end, right-angle, and turning movement crashes. For the Level 2 models, a total of 25 accident patterns were defined by the movements of involved vehicles prior to collision.

Another study by Shankar et al. (1995) resulted in the estimation of separate crash models to evaluate the effects of roadway geometric variables and environmental factors on specific crash types such as sideswipes, rear-end, parked vehicles, fixed objects, overturns, and same direction, and mainly focused on identifying the safety effects of environmental variables (i.e., weather information) on crash types rather than identifying the safety effects of roadway geometric variables. Shankar et al. concluded that separate regression models for specific types of crashes would have the potential for providing greater explanatory power relative to a single overall frequency model. Specifically, separate models allow coefficient estimates to vary by the type of crash, which agrees with intuition as described previously. Stutts et al. (1996) identified the effects of pedestrian characteristics, location, environment, and roadway on pedestrian-vehicle crashes. They divided pedestrian crash types into 15 subgroups and presented the

distribution of pedestrian-vehicle crashes by pedestrian age, pedestrian injury severity, hour of day, light condition, roadway system, and detailed pedestrian location for each of 15 pedestrian crash subgroups. In addition to these studies, Greibe (2003) and Ryan et al. (1998) attempted to identify the safety effects on different crash types of signal control and different age groups, respectively, through the use of statistical analysis.

As described in these studies, therefore, it can be concluded that specific crash type models are more likely to provide greater power to understand and analyze accident characteristics compared to overall crash models, but no comprehensive studies have been conducted to identify the impacts of various explanatory factors on specific types of crashes at rural intersections.

2.3 Literature on Contemporaneous Correlation

While many econometric applications have dealt with contemporaneous correlation problems, prior research applying seemingly unrelated regression modeling techniques to estimate crash prediction models is relatively scant. Only a few researches have employed the SURE approach for count data to accommodate contemporaneous correlation between the disturbance terms.

A recent study by Clarke and Loeb (2005) developed crash models to investigate the determinants of three types of railroad-related fatalities such as trespasser fatalities, fatalities at grade crossings, and fatalities involving passengers and employees. Unlike traditional crash models that have been developed in past studies, they estimated a number of single equation models for each type of fatalities using a log linear functional

form. Among the numerous models, three independent models were selected as final models for three types of fatalities by examining specification errors such as omission of variables, misspecification of functional form, and so on. Then, they re-estimated final models using seemingly unrelated regression analysis to gain more efficient estimators.

Chayanan et al. (2003) also employed seemingly unrelated regression analysis to estimate two linear regression models to explore the relationship between roadway and roadside accident rates for Washington State highways. They believed that roadway and roadside accident rates for a given roadway section are affected by geometric, traffic, and environmental factors in different ways as well as may be correlated because of unobserved common effects across the roadway and roadside. Thus, they modeled the roadway and roadside accident rates simultaneously, but concluded that the SURE approach was found to produce minimal gains in efficiency over single-equation regression models.

However, these studies focused on the development of simultaneous models with log linear and linear models. Considering the fact that crash counts are discrete and non-negative integers and the variance of the crash frequency is greater than the mean, a negative binomial model is regarded as the preferred modeling approach, but a seemingly unrelated negative binomial (SUNB) model has never been developed for estimating crash prediction models.

Research conducted by Ladron de Guevara et al. (2004) employed seemingly unrelated regression analysis to estimate planning level crash prediction models. Three negative binomial models were estimated for fatal, injury, and property damage only

(PDO) crashes. Only fatal and injury crashes, however, were re-modeled simultaneously by a system of negative binomial models because estimation problems prohibited the simultaneous estimation of injury and PDO crashes. For the estimation of a system of negative binomial models, however, they used a commercial software called aML. The limitation of using a commercial software is that a researcher is unable to deal with more complex model functional forms that are not provided by the software.

2.4 Literature on the Effect of Left-Turn Lanes

Many research 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; Chin and Quddus, 2003). In addition to these studies, numerous researchers have explored the safety effectiveness of left-turn lanes (Greiwe, 1986; Agent, 1983; Ben-Yakov and Craus, 1980; Craus and Mahalel, 1980; McFarland, 1979; Tamburri and Hammer, 1968; Wilson et al., 1967; Maze et al., 1994), but it is virtually impossible to cite all of them. So, several representative references relevant to this research are described here and the detailed descriptions for the safety effects of left-turn lanes on different types and/or severity of accidents, which are not described here, are summarized in Table 2.1.

Table 2.1. Summary of the safety effectiveness of installing left-turn lanes

Source	Reported LTL effectiveness (Percent change in accident frequency)		Conditions/comments
	Total intersection accidents	Left-turn accidents	
Greiwe (1986)	-58	-62	eight LTLs added by restriping
Agent (1983)	-77	–	unsignalized intersection
	-54	–	signalized intersections
Ben-Yakov and Craus (1980) / Craus and Mahalel (1980)	-38	–	LTL installation
McFarlane (1979)	-70	–	LTL with curbed median; urban
	-65	–	LTL with curbed median; suburban
	-60	–	LTL with curbed median; rural
	-15	–	LTL with painted median; urban
	-30	–	LTL with painted median; suburban
	-50	–	LTL with painted median; rural
	-36	–	signalized intersection with LTL and exclusive phase
	-15	–	signalized intersection with LTL but no exclusive phase
Tamburri and Hammer (1968) / Wilson et al. (1967)	-18	–	unsignalized intersection
Maze et al. (1994)	-6	–	signalized intersection; LTL with permitted phasing
	-35	–	signalized intersection; LTL with protected/permitted phasing

Source: Harwood et al. (2002)

Foody and Richardson (1973) found that accident rates decreased by 38 percent with the addition of a left-turn lane at signalized intersections and by 76 percent at unsignalized intersections. Parker et al. (1983) showed that the addition of left-turn lanes at rural intersections along two-lane highways can reduce the potential for passing-related accidents. Similarly, Gluck et al. (1999) reported that accident rates were reduced by ranging from 18 to 77 percent due to the installation of left-turn lanes. While these studies analyzed the safety effects of left-turn lanes, when installed alone, on crashes, earlier studies conducted by Lacy (1972) and Dale (1973) presented that installation of left-turn lanes was very effective in increasing safety when the left-turn treatment was implemented with additional safety measures.

Lacy (1972) found that a left-turn lane, when coupled with several other safety improvements, reduced accident frequency by 35 percent and accident severity by 80 percent. Dale (1973) found that installation of a traffic signal and left-turn channelization at intersections along rural two-lane highways reduced the total number of accidents by 20 percent, while the installation of a traffic signal without any channelization reduced the total number of accidents by only 6 percent.

Not all studies, however, have shown that left-turn lanes reduce accidents. McCoy and Malone (1989) found that installation of left-turn lanes reduced rear-end, sideswipe, and left-turn accidents, but significantly increased right-angle accidents at unsignalized intersections. Poch and Mannering (1996) and Kim et al. (2006) also found some situations in which accidents of specific types increased with installation of left-turn lanes. Bauer and Harwood (1996) revealed that left-turn lanes were associated with

higher frequencies of both total multiple-vehicle accidents and fatal and injury multiple-vehicle accidents. Another study by Chin and Quddus (2003) showed that having an uncontrolled left-turn lane increases the likelihood of accidents, perhaps of the sideswipe and head-to-side types. However, they concluded that total annual accidents decreases by 24 percent when a longer section on the left-turn lane for acceleration prior to merge into cross traffic stream is provided.

Although numerous research has attempted to identify the safety effects of left-turn lanes on various crashes, the potential endogeneity of left-turn lanes has never been discussed in the prior research. Moreover, the potential endogeneity of other geometric characteristics such as lighting has also never been discussed. On the other hand, some studies have shown that endogeneity problems persist in crash data.

2.5 Literature on Endogeneity

Researchers have attempted to account for endogenous relationships to better understand the safety effects on crashes of various kinds of variables such as seat belt usage rate on motor vehicle-related 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).

Derrig et al. (2000) evaluated the effect of seat belt usage rates on the number of motor vehicle-related fatalities. In this research, they treated seat belt usage rates as an endogenous variable because safety incentive grants sponsored by the National Highway Traffic Safety Administration (NHTSA) may affect the seat belt usage rates. For

example, the NHTSA provided a total of $500 million safety incentive grants to states with higher than average observed usage rates during 1999-2003 under the TEA-21 funding. This indicates that the incentive grants resulted in increased seat belt usage rates for states by establishing mandatory seat belt laws. Consequently, they developed crash prediction models to estimate motor vehicle-related fatalities by the use of Instrumental Variable (IV) methods.

Similarly, Cummins et al. (2001) conducted an empirical analysis of automobile accident fatality rates in all U.S. states over the period 1968-94, controlling for the potential endogeneity of no-fault laws. They proposed that the presence of no-fault in states is likely to be endogenous due to the fact that states tend to adopt no-fault in response to high auto insurance costs. In order to control for endogeneity, they employed two estimation techniques such as instrumental variables (IV) and inverse Mill's ratios (IM).

Eisenberg (2003) presented new findings on the effectiveness of state-level public policies related to drunk driving such as 0.08 and 0.10 percent blood alcohol content (BAC) standards. Although he addressed that there may exist a bi-directional relationship between fatal crash rates and the enactment of stricter drunk driving policies, he did not develop simultaneous equation models for the analysis. Instead, he tried to address the potential endogeneity of state-level policies related to drunk driving by estimating the timing of the policy effects with respect to the date of adoption.

However, these studies have estimated crash prediction models using single equation modeling techniques such as instrumental variables (IV) methods and weighted least

squares. The SEM approaches have rarely been used for the estimation of crash prediction models. In contrast, only a few researchers have applied SEM techniques to other transportation applications to control for endogeneity.

Shankar and Mannering (1998) developed structural models of mean speeds and speed deviations for each lane of a multilane highway to provide a better understanding of highway safety and the effects that lane-mean and lane speed deviations have on highway safety. They found that the mean speed in each lane is endogenously related to the mean speeds in the 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 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, he 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 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). They also used the 3SLS methods to model the quantity and price of automobiles within a supply-demand framework.

2.6 Results of Literature Review

Comprehensive literature review was conducted with respect to crash prediction models related to crash types, the safety effects of left-turn lanes, and econometric modeling techniques that are used in prior research to control for contemporaneous correlation and endogeneity. Thorough literature review has led to the following conclusions:

1. As described in prior research, modeling crash types would provide more insights into the understandings of the safety effects of various factors on crash occurrence and have the potential for providing greater explanatory power relative to a single overall frequency model. However, crash prediction models focused on predicting different crash types have rarely been developed. Therefore, comprehensive and in-depth studies need to be done for estimating crash prediction models for specific types of crashes.
2. Different crash types might be expected to be correlated because they share unobserved common characteristics, but prior studies did not consider

contemporaneous correlation between the disturbance terms across crash types when they estimated crash type models. Therefore, contemporaneous correlation should be examined and it is necessary to develop simultaneous estimation techniques to deal with contemporaneous correlation.

3. Although the safety effects of left-turn lanes on crashes have been extensively investigated, the potential endogeneity of left-turn lanes has never been discussed in crash prediction models. This might lead to the mixed results for the safety effects of the presence of left-turn lanes on crash occurrence. To better understand the influence of left-turn lanes on crashes, it is necessary to control for the potential endogeneity of left-turn lanes.
4. Simultaneous estimation techniques have been mainly employed to estimate several linear regression models in transportation applications, but never been developed to account for simultaneity between a count variable and a binary variable.

CHAPTER 3

METHODOLOGY

This chapter presents econometric modeling techniques developed to accommodate contemporaneous correlation and control for endogeneity. Since the methodological approaches proposed in this research are the extension of traditional modeling techniques such as Poisson and negative binomial models, this chapter first describes the model specifications of Poisson and negative binomial models. Then, a seemingly unrelated negative binomial model and a simultaneous negative binomial-logit model, which are the proposed methods in this research, are presented.

3.1 Conventional Modeling Techniques

Two different approaches are generally used to estimate crash prediction models, Poisson regression and negative binomial regression (Jovanis and Chang, 1986; Washington et al. 2003), although various other modeling approaches are possible (see for example Lord et al., 2005). Crash counts are approximated well by a Poisson process (Joshua and Garber, 1990), since crash counts are discrete and positive integers. The Poisson regression model requires that the variance of the crash frequency is approximately equal to its mean. In much crash data, however, the variance of the crash frequency is greater than the mean and overdispersion occurs (Miaou et al., 1992). Miaou et al. introduced the negative binomial distribution for modeling traffic safety which accommodates greater variance in the data than allowed by the Poisson distribution. The overdispersion typically arises from variation in crash means across sites. As a result, the

negative binomial regression model is the preferred modeling approach when overdispersion is present (Washington et al., 2003). Therefore, the Poisson and negative binomial model specifications are briefly described in this section. A comprehensive and detailed review of alternative modeling approaches is provided in Lord et al. (2005).

3.1.1 The Poisson Model

Let $y_i, i = 1, \ldots, n$, be the observations of a discrete and non-negative integer variable, which is assumed to be independently Poisson distributed, with the conditional mean specified as:

$$E(y_i \mid x_i) = \lambda_i = \exp(x_i'\beta) \tag{3.1}$$

where x_i is a *k*-vector of explanatory variables associated with the *i*-th observation and β is a *k*-vector of unknown parameters. Equation (3.1) is referred to as mean function, conditional expectation function, or regression function (Winkelmann, 2003).

The Poisson distribution has only one parameter that simultaneously determines (conditional) mean and variance; that is, the mean is assumed to be equal to the variance. Therefore,

$$E(Y) = Var(Y) = \lambda . \tag{3.2}$$

The density function of y_i given x_i is

$$f(y_i \mid x_i) = \frac{e^{-\lambda_i} \lambda_i^{\,y_i}}{y_i!}. \qquad y_i = 0, 1, 2, \ldots \tag{3.3}$$

Therefore, the likelihood function can be obtained by multiplying the density function of y_i across all observations as follows:

$$L(\beta \mid y) = \prod_{i=1}^{n} f(y_i \mid \beta) = \prod_{i=1}^{n} \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \tag{3.4}$$

and the log-likelihood function is

$$\begin{aligned} \ln L(\beta \mid y) &= \sum_{i=1}^{n} \left[-\lambda_i + y_i \ln \lambda_i - \ln(y_i!) \right] \\ &= \sum_{i=1}^{n} \left[-\exp(x_i'\beta) + y_i x_i'\beta - \ln(y_i!) \right]. \end{aligned} \tag{3.5}$$

The unknown parameters, β, can be estimated by maximizing the log-likelihood function. The maximizing value for β, denoted as $\hat{\beta}_{ML}$, is derived by computing the first derivatives of the log-likelihood function:

$$\frac{\partial \ln L(\beta \mid y)}{\partial \beta} = \sum_{i=1}^{n} \left[-x_i' \exp(x_i'\beta) + y_i x_i' \right] \tag{3.6}$$

and then solving the first order conditions for a maximum

$$\sum_{i=1}^{n} \left[-x_i' \exp(x_i'\beta) + y_i x_i' \right] = 0 . \tag{3.7}$$

The standard errors of the unknown parameters are obtained from the inverse of negative of the Hessian matrix of the log-likelihood function. The Hessian matrix is obtained from the second derivatives of the log-likelihood function with respect to β.

$$H(\beta;\, y,\, x) = \frac{\partial^2 \ln L(\beta \mid y)}{\partial \beta\, \partial \beta'} = \sum_{i=1}^{n} \left[- x_i'\, x_i \exp(x_i'\beta) \right] \tag{3.8}$$

and then the variance of $\hat{\beta}_{ML}$ is given by

$$Var(\hat{\beta}_{ML}) = \left(- E\left(\frac{\partial^2 \ln L(\beta \mid y)}{\partial \beta\, \partial \beta'} \right) \right)^{-1} = \left(\sum_{i=1}^{n} - x_i'\, x_i \exp(x_i'\beta) \right)^{-1} \tag{3.9}$$

Since the Hessian matrix is negative definite, the log-likelihood function of the Poisson regression model is globally concave.

3.1.2 Negative Binomial Model

The Poisson distribution has a single (one) parameter distribution; that is, the Poisson model requires that the conditional mean and conditional variance are equal, and hence the Poisson model is unable to account for overdispersion in the data. The alternative to the Poisson model is a negative binomial model which allows for overdispersion by introducing a disturbance in the definition of the parameter of the Poisson distribution. Therefore, the parameter of traditional negative binomial models for an univariate counted outcome can be rewritten as

$$\lambda_i^{\,*} = \exp(x_i'\beta + \varepsilon_i) = \exp(x_i'\beta) \exp(\varepsilon_i) = \lambda_i u_i \tag{3.10}$$

where ε_i is a specification error due to omitted explanatory variables.

Then the unconditional distribution of y_i given x_i is

$$f(y_i \mid x_i) = \int \frac{e^{-\lambda_i u_i}(\lambda_i u_i)^{y_i}}{y_i!} g(u_i) du_i \tag{3.11}$$

where $g(u_i)$ is a probability density function of u_i and it is usually assumed that $u_i = \exp(\varepsilon_i)$ is distributed as a $\Gamma(\alpha, \beta)$. The mean and variance of a gamma distribution are $E(u_i) = \alpha\beta$ and $Var(u_i) = \alpha\beta^2$, respectively. For mathematical convenience, $E(u_i)$ is assumed to be one (as a result, $Var(u_i)$ is equal to α^{-1}).

Therefore, the probability density function of u_i is

$$g(u_i) = \frac{\theta^\theta}{\Gamma(\theta)} u_i^{\theta-1} e^{-\theta u_i} \tag{3.12}$$

where θ is a gamma's parameter and is equal to $\beta(=1/\alpha)$.

The integration of a Poisson-gamma mixture (eqn. (3.11)) leads to the negative binomial distribution for y_i and the density function of y_i is given by

$$\begin{aligned} f(y_i \mid x_i) &= \int \frac{e^{-\lambda_i u_i}(\lambda_i u_i)^{y_i}}{y_i!} \frac{\theta^\theta}{\Gamma(\theta)} u_i^{\theta-1} e^{-\theta u_i} du_i \\ &= \frac{\Gamma(\theta + y_i)}{\Gamma(y_i + 1)\Gamma(\theta)} \left(\frac{\theta}{\lambda_i + \theta}\right)^\theta \left(\frac{\lambda_i}{\lambda_i + \theta}\right)^{y_i}. \end{aligned} \tag{3.13}$$

which has conditional mean λ_i and conditional variance $\lambda_i + \alpha^{-1}\lambda_i^2$.

The log-likelihood function is

$$\begin{aligned} \ln L(\theta \mid y) = \sum_{i=1}^{n} \big[&\ln\Gamma(\theta + y_i) - \ln\Gamma(\theta) - \ln\Gamma(y_i + 1) \\ &+ \theta \ln\theta + y_i x_i'\beta - (\theta + y_i)\ln(\lambda_i + \theta)\big]. \end{aligned} \tag{3.14}$$

and by replacing θ by α^{-1} equation (3.14) can be written as:

$$\ln L(\theta \mid y) = \sum_{i=1}^{n}\left[\ln\Gamma\left(\frac{1}{\alpha}+y_i\right) - \ln\Gamma\left(\frac{1}{\alpha}\right) - \ln\Gamma(y_i+1) + y_i \ln\alpha + y_i x_i'\beta - \left(\frac{1}{\alpha}+y_i\right)\ln(1+\alpha\lambda_i)\right]. \tag{3.15}$$

The unknown parameters, β and α (overdispersion parameter), can be estimated by maximizing the log-likelihood function and derived by computing the first derivatives of the log-likelihood function with respect to β and α:

$$\frac{\partial \ln L(\beta \mid y)}{\partial \beta} = \sum_{i=1}^{n}\left[\frac{(y_i-\lambda_i)}{1+\alpha\lambda_i}x_i'\right] \tag{3.16}$$

$$\frac{\partial \ln L(\beta \mid y)}{\partial \alpha} = \sum_{i=1}^{n}\left[-\frac{1}{\alpha^2}\Psi\left(\frac{1}{\alpha}+y_i\right) + \frac{1}{\alpha^2}\Psi\left(\frac{1}{\alpha}\right) + \frac{y_i}{\alpha} + \frac{1}{\alpha^2}\ln(1+\alpha\lambda_i) - \left(\frac{1}{\alpha}+y_i\right)\frac{\lambda_i}{1+\alpha\lambda_i}\right] \tag{3.17}$$

where $\lambda_i = \exp(x_i'\beta)$ and $\Psi(x)$ is a digamma function that equals:

$$\Psi(x) = \frac{d\ln\Gamma(x)}{dx} = \frac{\Gamma'(x)}{\Gamma(x)}.$$

The standard errors of the parameters, $\hat{\beta}_{ML}$ and $\hat{\alpha}_{ML}$, are obtained from the inverse of the Hessian matrix. The Hessian matrix is obtained from the second derivatives of the log-likelihood function with respect to β and α. The detailed information on the second derivatives of the log-likelihood function is presented in Lawless (1987). The (2×2) Hessian matrix is given by

$$H(\beta, \alpha; y, x) = \begin{bmatrix} \dfrac{\partial^2 \ln L(\beta \mid y)}{\partial \beta \, \partial \beta'} & \dfrac{\partial^2 \ln L(\beta \mid y)}{\partial \beta \, \partial \alpha} \\ \dfrac{\partial^2 \ln L(\beta \mid y)}{\partial \alpha \, \partial \beta} & \dfrac{\partial^2 \ln L(\beta \mid y)}{\partial \alpha \, \partial \alpha'} \end{bmatrix} \tag{3.18}$$

and then the variances of $\hat{\beta}_{ML}$ and $\hat{\alpha}_{ML}$ can be obtained from:

$$\begin{aligned} Var(\hat{\beta}_{ML}) &= \left[- E\left(\frac{\partial^2 \ln L(\beta \mid y)}{\partial \beta \, \partial \beta'} \right) \right]^{-1} \\ Var(\hat{\alpha}_{ML}) &= \left[- E\left(\frac{\partial^2 \ln L(\beta \mid y)}{\partial \alpha \, \partial \alpha'} \right) \right]^{-1} \end{aligned} . \tag{3.19}$$

3.2 Seemingly Unrelated Regression (SURE) Model

3.2.1 Seemingly Unrelated Negative Binomial (SUNB) Model

A modeling technique for dealing with contemporaneous correlation between the disturbance terms in linear regressions was proposed by Zellner (1962), whereas the approach to account for correlation amongst count data, namely Seemingly Unrelated Negative Binomial (SUNB) model, has been pursued by Hausman, Hall and Griliches (1984). This is also referred to as a Multivariate Poisson-Gamma Mixture model (Winkelmann, 2003).

The SUNB models can estimate a set of negative binomial models simultaneously. For J outcomes, the unconditional distribution of $y_i = (y_{i1}, \cdots, y_{iJ})'$ given x_i can be obtained by integrating the product of the distribution for each outcome over u_i (Winkelmann, 2003).

$$
\begin{aligned}
f(y_i \mid x_i) &= \int \left[f(y_{i1} \mid x_{i1}) \times f(y_{i2} \mid x_{i2}) \times \cdots \times f(y_{iJ} \mid x_{iJ}) \right] g(u_i) du_i \\
&= \int \left[\prod_{j=1}^{J} f(y_{ij} \mid x_{ij}) \right] g(u_i) du_i \qquad (3.20) \\
&= \int \left[\prod_{j=1}^{J} \frac{\exp(-\lambda_{ij} u_i)(\lambda_{ij} u_i)^{y_{ij}}}{y_{ij}!} \right] g(u_i) du_i \, .
\end{aligned}
$$

If u_i is gamma distributed with $E(u_i) = 1$ and $Var(u_i) = \alpha^{-1}$ as same with univariate negative binomial models, it can be shown that the joint distribution function of y_i is

$$
\begin{aligned}
f(y_i \mid x_i) &= \int \left[\prod_{j=1}^{J} \frac{\exp(-\lambda_{ij} u_i)(\lambda_{ij} u_i)^{y_{ij}}}{y_{ij}!} \right] \frac{\theta^{\theta}}{\Gamma(\theta)} u_i^{\theta-1} e^{-\theta u_i} du_i \\
&= \left(\prod_{j=1}^{J} \frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!} \right) \frac{\theta^{\theta}}{\Gamma(\theta)} \int e^{-(\lambda_{i.}+\theta)u_i} u_i^{y_{i.}+\theta-1} du_i \qquad (3.21) \\
&= \left(\prod_{j=1}^{J} \frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!} \right) \frac{\Gamma(\theta + y_{i.})}{\Gamma(\theta)} \theta^{\theta} \left(\lambda_{i.} + \theta \right)^{-\theta + y_{i.}}
\end{aligned}
$$

where $y_{i.} = \sum_{j=1}^{J} y_{ij}$, $\lambda_{i.} = \sum_{j=1}^{J} \lambda_{ij}$ and $\lambda_{ij} = \exp(x_{ij}^{'} \beta_j)$.

The parameters of the seemingly unrelated negative binomial model can be estimated by maximum likelihood estimation that maximizes the log-likelihood function. The log-likelihood function for the SUNB model is

$$
\begin{aligned}
\ln L(\theta \mid y) = \sum_{i=1}^{n} \Bigg[& \sum_{j=1}^{J} y_{ij} x_{ij}^{'} \beta_j - \sum_{j=1}^{J} \ln(y_{ij}!) + \ln \Gamma(\theta + y_{i.}) \\
& - \ln \Gamma(\theta) + \theta \ln \theta - (y_{i.} + \theta) \ln(\lambda_{i.} + \theta) \Bigg]
\end{aligned} \qquad (3.22)
$$

where $y_{i.} = \sum_{j=1}^{J} y_{ij}$, $\lambda_{i.} = \sum_{j=1}^{J} \lambda_{ij}$ and $x_{ij}^{'}$ is the *i*-th row of X_j .

The first derivatives of equation (3.22) with respect to β and α are given by:

$$\frac{\partial \ln L(\beta \mid y)}{\partial \beta_j} = \sum_{i=1}^{n} \left[y_{ij} x'_{ij} - \left(y_{i.} + \frac{1}{\alpha} \right) \left(\frac{\alpha\, x'_{ij} \exp(x'_{ij} \beta_j)}{1 + \alpha \lambda_{i.}} \right) \right] \tag{3.23}$$

$$\begin{aligned} \frac{\partial \ln L(\beta \mid y)}{\partial \alpha} = \sum_{i=1}^{n} \Bigg[& -\frac{1}{\alpha^2} \Psi\left(\frac{1}{\alpha} + y_i \right) + \frac{1}{\alpha^2} \Psi\left(\frac{1}{\alpha} \right) - \frac{1}{\alpha^2} \ln\left(\frac{1}{\alpha} \right) - \frac{1}{\alpha^2} \\ & + \frac{1}{\alpha^2} \ln\left(\frac{1}{\alpha} + \lambda_{i.} \right) + \frac{1}{\alpha^2} \left(\frac{1 + \alpha y_{i.}}{1 + \alpha \lambda_i} \right) \Bigg] \end{aligned} \tag{3.24}$$

and the standard errors of the parameters are obtained from the inverse of the Hessian matrix of the log-likelihood function.

3.2.2 Test for Contemporaneous Correlation between Disturbances

As stated in Chapter 1, the SURE approach yields more efficient estimators when the disturbance terms across equations are contemporaneously correlated. If there is no correlation between the disturbance terms, equation-by-equation estimation can yield efficient estimators.

In general, two common approaches, the Likelihood Ratio (LR) test and Lagrange Multiplier (LM) test, are used for testing contemporaneous correlation (Greene, 2003). The Likelihood Ratio (LR) test is based on the statistic:

$$\lambda_{LR} = T \left(\sum_{i=1}^{n} \ln \hat{\sigma}_i^2 - \ln \left| \hat{\Sigma} \right| \right) \tag{3.25}$$

where $\hat{\sigma}_i^2$ is $e_i' e_i / T$ from the individual least squares regressions and $\hat{\sum}$ is the maximum likelihood estimator of $\sum$, which is calculated by $e_i' e_j / T$. The null hypothesis of the LR test is that the off-diagonal elements of $\sum$ are zero, that is, that there is no correlation across equations, and this statistic has a limiting chi-squared distribution with $M(M-1)/2$ degrees of freedom under the hypothesis.

Another approach proposed by Breusch and Pagan (1980) is the Lagrange Multiplier (LM) test. The LM statistic is

$$\lambda_{LM} = T \sum_{i=2}^{M} \sum_{j=1}^{i-1} r_{ij}^2 \tag{3.26}$$

where r_{ij} is the estimated correlation, $\hat{\sigma}_{ij} / \left[\left(\hat{\sigma}_{ii} \right) \left(\hat{\sigma}_{jj} \right) \right]^{1/2}$, T is the number of observations and M is the number of equations. This statistic has a limiting chi-squared distribution with $M(M-1)/2$ degrees of freedom.

From a computational perspective, the LM test has the advantage that it does not require computation of the maximum likelihood estimator of a diagonal matrix, $\sum$ (Greene, 2003). Hence, a test for contemporaneous correlation between the disturbance terms was conducted using the LM test.

The null hypothesis of the LM test is that the disturbance terms are independent across equations. If the LM statistic is greater than a critical value with $M(M-1)/2$ degrees of freedom, then the null hypothesis is rejected and it can be concluded that the disturbance terms are contemporaneously correlated across equations.

3.3 Simultaneous Equation Models

3.3.1 Simultaneous Negative Binomial-Logit Model

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:

$$\begin{aligned} y_{1i} &= \exp(\alpha_1 y_{2i} + x_i' \beta_1 + \varepsilon_i) = \exp(\alpha_1 y_{2i} + x_i' \beta_1)\exp(\varepsilon_i) \\ &= \exp(\alpha_1 y_{2i} + x_i' \beta_1) u_i \end{aligned} \tag{3.27}$$

$$y_{2i}^* = \alpha_2 y_{1i} + z_i' \beta_2 + \nu_i \tag{3.28}$$

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

$$y_{2i} = \begin{cases} 1 & \text{if } y_{2i}^* > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{3.29}$$

If both α_1 and α_2 are not equal to zero (i.e., dependent variables appear as a regressor in different equations), there exists endogeneity that arises from simultaneity and the endogeneity cannot be adequately explained in single equation estimation set. In such cases, therefore, all the parameters should be simultaneously estimated as a system of equations. From the viewpoint of model estimation, however, the model described above cannot be estimated by fully simultaneous estimation due to logical consistency

(Winkelmann, 2003; Windmeijer and Santos Silva, 1997; Cameron and Trivedi, 1998), which is an issue in a situation where a count dependent variable depends on a binary endogenous variable.

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

$$\Pr(y_{2i} = 1) + \Pr(y_{2i} = 0) = 1 \tag{3.30}$$

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 can be obtained from the relations (3.28) and (3.29):

$$\begin{aligned} \Pr(y_{2i} = 1) &= \Pr[\nu_i > -(\alpha_2 y_{1i} + z_i' \beta_2)] \\ &= 1 - F[-(\alpha_2 y_{1i} + z_i' \beta_2)] \\ &= F(\alpha_2 y_{1i} + z_i' \beta_2) \end{aligned} \tag{3.31}$$

where F is the cumulative distribution function for ν_i, assuming symmetry of the density of ν_i. Since ν_i is assumed to be a logistic distribution, equation (3.31) can be rewritten as

$$F(\alpha_2 y_{1i} + z_i' \beta_2) = \frac{\exp(\alpha_2 y_{1i} + z_i' \beta_2)}{1 + \exp(\alpha_2 y_{1i} + z_i' \beta_2)}. \tag{3.32}$$

Substituting equation (3.27) for y_{1i} in equation (3.31), equation (3.30) can be rewritten as

$$F[\alpha_2 \exp(\alpha_1 + x_i' \beta_1 + \varepsilon_i) + z_i' \beta_2] + \{1 - F[\alpha_2 \exp(x_i' \beta_1 + \varepsilon_i) + z_i' \beta_2]\} = 1 \tag{3.33}$$

or

$$F[\alpha_2 \exp(\alpha_1 + x_i' \beta_1 + \varepsilon_i) + z_i' \beta_2] = F[\alpha_2 \exp(x_i' \beta_1 + \varepsilon_i) + z_i' \beta_2]. \tag{3.34}$$

This condition implies that either $\alpha_1 = 0$ or $\alpha_2 = 0$ and therefore it can be concluded that logical consistency prohibits fully simultaneous estimation (Winkelmann, 2003). By assuming that $\alpha_2 = 0$, equations (3.27) and (3.28) are rewritten as

$$y_{1i} = \exp(\alpha_1 y_{2i} + x_i' \beta_1 + \varepsilon_i) \tag{3.35}$$

$$y_{2i}^* = z_i' \beta_2 + \nu_i \tag{3.36}$$

and equation (3.35) can be estimated by replacing y_{2i} by its conditional mean, $E(y_{2i} \mid z_i) = F(z_i' \beta_2)$. However, estimation of equation (3.35) 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 can be 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 research, $\hat{\beta}_2$ is estimated using a logit model.

For the binary outcome model, consistent estimators for α_2 and β_2 can be obtained in similar fashion. Assuming $\alpha_1 = 0$, the binary outcome model is estimated by substituting $\exp(x_i' \hat{\beta}_1)$ for y_{1i} in equation (3.31), where $\hat{\beta}_1$ is obtained from a negative binomial model. As a result, all of the parameters are estimated by the following equations:

$$y_{1i} = \exp(\alpha_1 \hat{y}_{2i} + x_i' \beta_1 + \varepsilon_i) \tag{3.37}$$

$$y_{2i}^* = \alpha_2 \hat{y}_{1i} + z_i' \beta_2 + \nu_i \tag{3.38}$$

where $\hat{y}_{2i} = F(z_i' \hat{\beta}_2)$ and $\hat{y}_{1i} = \exp(x_i' \hat{\beta}_1)$. This mimics the logic of instrumental variable estimation in linear simultaneous equations models (Cameron and Trivedi, 1998).

The SEM model can be generally estimated by Indirect Least Squares (ILS), Instrumental Variable (IV), and Two-Stage Least Squares (TSLS) methods (Ramanathan, 2002), but it is not possible to apply these methods to special cases when a simple reduced form does not exist (Winkelmann, 2003). For such cases, the Maximum Likelihood Estimation (MLE) and the Generalized Method of Moments (GMM) are considered as the alternative ways to estimate unbiased estimators. Since equations (3.37) and (3.38) do not reduce to simple reduced forms for y_{1i} and y_{2i}, the MLE method is employed for the estimation of structural parameters in the analysis (for the GMM estimation, see Windmeijer and Santos Silva, 1997).

All of the parameters are estimated by maximizing the log-likelihood function using limited-information maximum likelihood (LIML) estimation, and the standard errors of the parameters are obtained from the inverse of the Hessian matrix of the log-likelihood function.

3.3.2 Test for Endogeneity

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 semidefinite, 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 research..

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

$$H = (\hat{\theta}_{SEE} - \hat{\theta}_{LIML})' \left[\text{var}(\hat{\theta}_{SEE}) - \text{var}(\hat{\theta}_{LIML})\right]^{-1} (\hat{\theta}_{SEE} - \hat{\theta}_{LIML}) \quad (3.29)$$

where $\hat{\theta}_{SEE}$ is the vector of estimated parameters obtained from single equation estimation (without controlling for endogeneity) and $\hat{\theta}_{LIML}$ 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 $\text{var}(\hat{\theta}_{SEE}) - \text{var}(\hat{\theta}_{LIML})$.

CHAPTER 4

DATA COLLECTION AND REDUCTION

This chapter discusses the data sources used in this research and presents all of the data collection activities associated with this research. The data used in this research came from the Georgia Department of Transportation. The data was comprised and collected from four data sources: crash files, road characteristic files, aerial photographs, and geographic information system (GIS) roadmaps. Crash data and road characteristic files are stored in MS ACCESS format and two years of crash data, 1996 and 1997, were used to develop the statistical models described in the previous chapter. This chapter first describes the data sources, and then detailed procedures on data collection are presented.

4.1 Crash Files

Georgia covers 59,441 square miles, making it the 24th largest of the 50 states. Georgia comes close to being an average size state, but has the second highest number of counties (159 counties) of any state in the United States. Figure 4.1 depicts the geographical locations of counties in the state of Georgia. Despite of the higher number of counties, crash files from the Georgia Department of Transportation contained detailed information on crashes occurring within 36 counties. Therefore, crash data used in this research were obtained from 36 counties out of 159 counties.

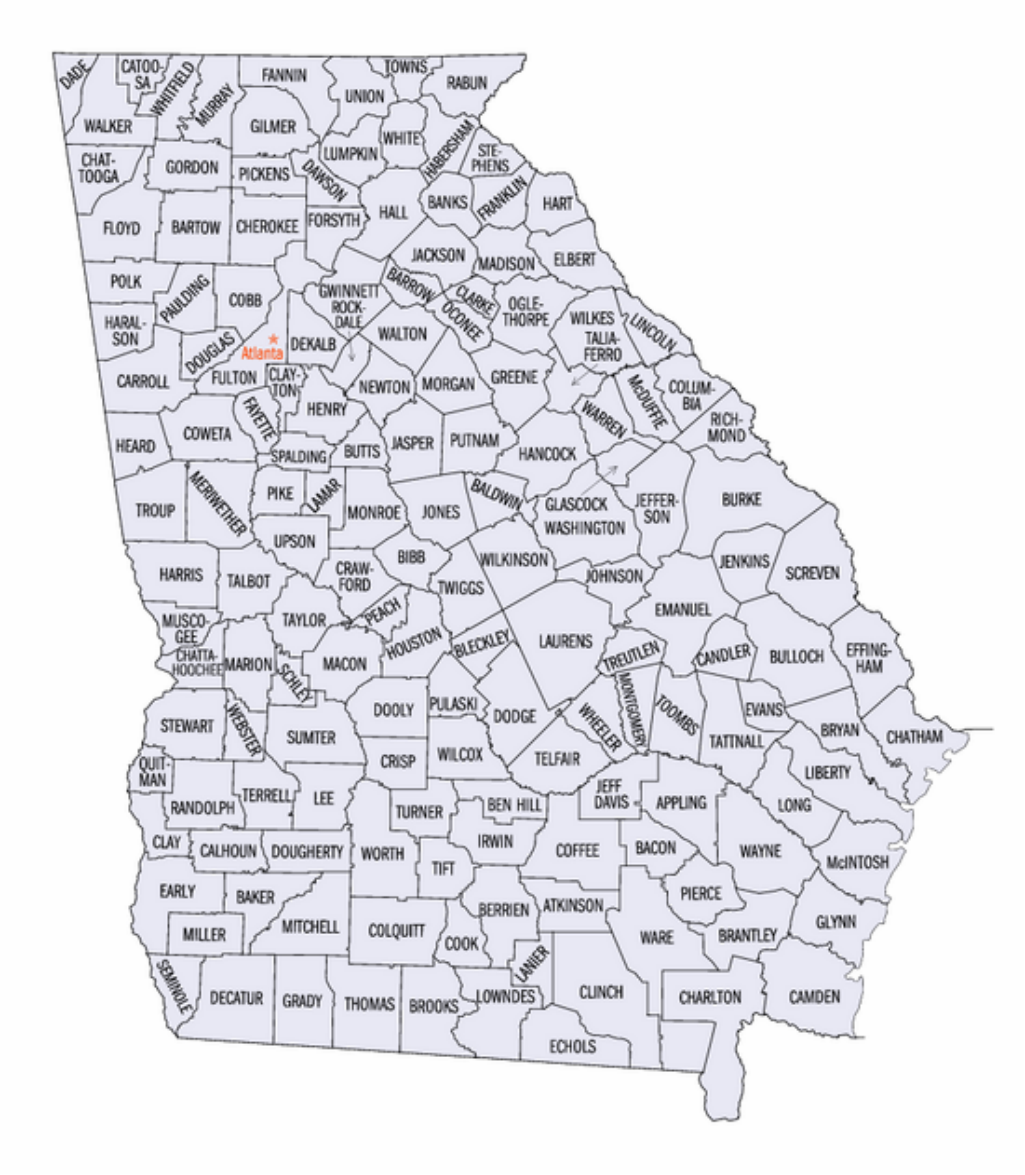

Figure 4.1. Geographical location of counties in the state of Georgia

Crash data including detailed information related to traffic accidents were available for two years, 1996 – 1997. The crash files contained basic accident, vehicle, and occupant information as well as location information such as site ID and county name where the crash occurred on a case-by-case basis. The following information is included

in the crash files:

- Accident information
  - Case number
  - Accident date and location (site ID and county name)
  - Number of vehicles involved
  - Accident type: angle, head-on, rear-end, sideswipe (both same and opposite direction), and pedestrian related crashes
- Occupant information
  - Number of occupants
  - Number of fatalities and injuries
- Day of week: weekday and weekend
- Weather condition: clear, rainy, snowy, and fog
- Surface condition: dry and wet
- Characteristics of roadway at which crashes occurred: functional classification, curve, and grade

4.2 Road Characteristic Files

Road characteristic files for years 1996 and 1997 provide detailed information on roadway characteristics and traffic related attributes. Road characteristic files also contained site ID and county name where crashes occurred. The variables included in the road characteristic files are listed below:

- Average Annual Daily Traffic (AADT)

- Median type and width
- Shoulder type and width
- Number of lanes and lane width
- Traffic control type: signalized and non-signalized intersections
- The presence of left- and right-turn lanes
- Number of commercial/ residential driveways
- The presence of lighting
- Terrain and roadside hazard rating
- Posted speed limits
- Sight distance
- Vertical curvature

4.3 Aerial Photographs

Digital Orthophotography Quarter-Quadrangles (DOQQs) aerial photos were used from 1994 and 2000 to extract information regarding intersection angle and degree of horizontal curvature of selected intersections by overlapping with GIS roadmaps. The DOQQs aerial photos were obtained from the GIS center of Georgia Institute of Technology.

4.4 Data Collection

According to the Crash Analysis, Statistics & Information (CASI) Notebook developed by the Department of Motor Vehicle Safety, Georgia Department of

Transportation, a total of 298,288 and 301,822 crashes occurred at roadway segments and intersections in the state of Georgia in 1996 and 1997, respectively, and a total of 85,104 and 87,078 crashes occurred within 36 counties in 1996 and 1997, respectively. Table 4.1 shows the number of motor vehicle crashes and crash rate per 10,000 licensed drivers by 36 counties.

Since this research focused on the development of crash prediction models for two-lane four-legged rural intersections (both signalized and stop-controlled intersections), only crashes occurring at two-lane four-legged rural intersections were extracted from the data with the aid of GIS technology. More specifically, crash data used in this research were obtained as follows:

First, both crash and road characteristic files were transferred from MS ACCESS format into "DBF" GIS format in order to utilize GIS technology for data collection. This enables us to locate accidents on GIS roadmaps. Then, crash files and road characteristic files were overlapped on GIS roadmaps in ArcView GIS software. By doing so, crash files were able to be integrated with road characteristic files by the use of location information (site ID and county name). Figure 4.2 illustrates locations of 1996 and 1997 crashes in Appling County, Georgia. The crashes were connected with road characteristic files so that it was possible to obtain detailed information on road characteristics around accident locations.

Table 4.1. Total number of crashes and rates per 10,000 licensed drivers by 36 counties

County	1996		1997	
	Number	Rate	Number	Rate
BARTOW	2,650	502.9	2,697	502.8
BIBB	7,288	652.0	7,210	651.1
BULLOCH	1,676	534.4	1,792	568.3
BUTTS	532	446.5	432	358.0
CARROLL	2,530	435.1	2,770	475.2
CATOOSA	1,917	571.0	1,837	540.3
COLUMBIA	2,427	393.0	2,644	418.5
COOK	434	432.3	433	426.6
COWETA	2,434	437.4	2,334	410.9
DOUGLAS	3,444	531.6	3,525	537.4
EFFINGHAM	702	314.9	662	289.2
EVANS	177	278.6	197	309.8
FANNIN	427	290.5	424	285.8
FAYETTE	2,555	409.1	2,596	399.2
FLOYD	3,811	595.3	4,067	640.3
GORDON	1,693	529.5	1,647	512.7
GWINNETT	19,807	539.4	21,113	553.4
HABERSHAM	1,056	428.5	1,050	421.0
HALL	5,120	587.1	5,120	581.6
HANCOCK	67	127.7	48	91.4
HARALSON	596	346.4	643	373.9
HEARD	53	82.8	64	99.9
JEFFERSON	325	267.5	288	239.5
LUMPKIN	589	536.1	634	564.9
MADISON	543	296.4	571	310.4
MERIWETHER	454	306.8	430	292.5
MONTGOMERY	72	159.4	74	165.4
PIKE	148	168.5	156	176.2
RABUN	367	346.6	409	383.9
RICHMOND	10,140	768.7	10,261	802.2
ROCKDALE	2,841	571.4	2,768	549.3
SPALDING	2,547	624.3	2,506	617.4
STEPHENS	624	310.6	721	359.5
TATTNALL	190	144.9	220	166.9
WALTON	1,267	363.3	1,179	332.0
WHITFIELD	3,601	582.7	3,556	574.8
TOTAL	85,104	414.3	87,078	416.2

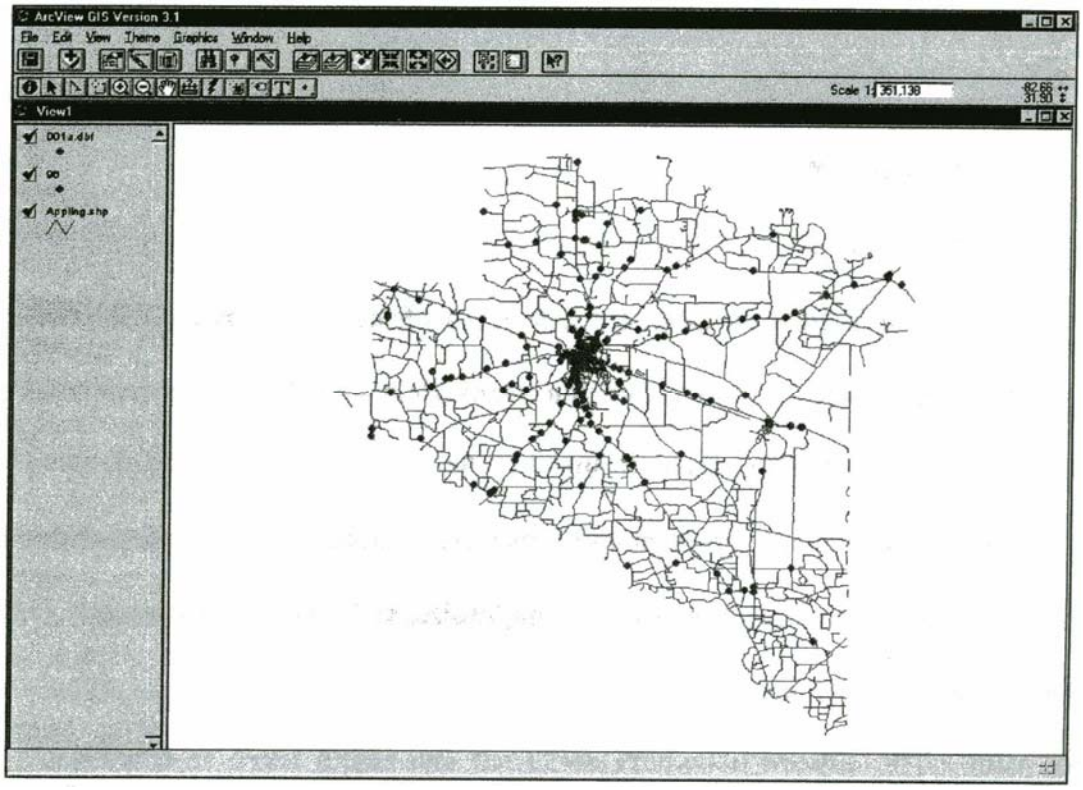

Figure 4.2. 1996 and 1997 crashes in Appling County, Georgia

After integrating crash data with road characteristic files, rural intersections were selected and intersection-related crashes for corresponding intersections were extracted from the integrated crash data. In most cases, crashes occurring at the intersection or within 250 ft (76.2 m) from the intersection center along the major and minor road were regarded as intersection-related crashes, and this criterion was also used in this study.

A total of 390 rural intersections were selected as target intersections. These rural intersections can be divided into five types based on the number of legs, traffic control type, and the number of lanes of approaches:

- Three-legged stop-controlled intersections of two-lane rural roads
- Four-legged stop-controlled intersections of two-lane rural roads
- Three-legged stop-controlled intersections with two lanes on minor and four lanes on major road
- Four-legged stop-controlled intersections with two lanes on minor and four lanes on major road
- Four-legged signalized intersections of two-lane rural roads

The number of each type of intersection is presented in Table 4.2. As described previously, this research focuses on the development of crash prediction models for two-lane four-legged rural intersections, and as a result, a total of 165 rural intersections were included in the data: 114 stop-controlled intersections and 51 signalized intersections.

Table 4.2. Number of intersections by types

Intersection Type	No. of intersections
3-legged stop-controlled intersections of two-lane roads	121
4-legged stop-controlled intersections of two-lane roads	114
3-legged stop-controlled intersections with 2 lanes on minor and 4 lanes on major road	52
4-legged stop-controlled intersections with 2 lanes on minor and 4 lanes on major road	52
4-legged signalized intersections of two-lane roads	51
Total	390

4.4.1 Dependent Variables

This research developed crash prediction models for different types of crashes, and thus total crashes were divided into six different crash types: angle crashes, head-on crashes, rear-end crashes, sideswipe crashes (both same and opposite direction), and pedestrian-involved crashes. It should be noted that total number of crashes does not include single vehicle crashes such as run-off-road and fixed object crashes because single vehicle crashes were unavailable in the dataset. Table 4.3 presents the summary of the crash data used in this study.

Table 4.3. Summary of crash data

	Stop-Controlled		Signalized		Total	
	Number	%	Number	%	Number	%
No. of intersections	114	69.1	51	30.9	165	100
Total crashes	345	41.2	492	58.8	837	100
Angle	152	18.2	211	25.2	363	43.4
Head-on	9	1.0	13	1.6	22	2.6
Rear-end	62	7.4	175	20.9	237	28.3
Sideswipe – same direction	14	1.7	28	3.3	42	5.0
Sideswipe – opposite direction	12	1.4	9	1.1	21	2.5
Pedestrian-involved	96	11.5	56	6.7	152	18.2

A total of 837 crashes were included in the data: 345 occurring at stop-controlled intersections and 492 occurring at signalized intersections. On average, five crashes per each intersection occurred for two years (837 crashes / 165 intersections = 5.07). This is a small number of crashes compared to typical numbers of intersection-related crashes. One possible reason for this might be the fact that these crashes were collected from rural

intersections.

The newly collected data indicate that more than half of the crashes (58.8%) occurred at signalized intersections. With respect to crash types, more than 40% of the crashes were angle crashes, and only 2.6% and 2.5% of the total crashes were head-on crashes and opposite direction sideswipe crashes, respectively.

4.4.2 Independent Variables

For this research, only intersection-level characteristics were used to develop crash prediction models. Crash-level characteristics such as day of week, weather information, surface condition (wet or dry), and so on were not included in the data because the fundamental unit of analysis is an intersection.

A total of 26 variables that were available in the road characteristic files were selected as independent variables: 5 traffic related attributes and 21 roadway related characteristics. The basic statistics of variables are summarized in Table 4.4.

Table 4.4. Dependent and independent variables used in the research

Variables	Definition	Mean	Min	Max	S.D.
Dependent Variable					
TOTACC	Number of total crashes	5.07	0	53	6.64
ANGLE	Number of angle crashes	2.20	0	33	3.58
HDON	Number of head-on crashes	0.13	0	2	0.38
REND	Number of rear-end crashes	1.44	0	15	2.48
SSSAME	Number of sideswipe same direction crashes	0.25	0	6	0.72
SSOPPO	Number of sideswipe opposite direction crashes	0.13	0	3	0.40
PEDINV	Number of pedestrian-involved crashes	0.92	0	4	0.98

Table 4.4. Dependent and independent variables used in the research – continued

Variables	Definition	Mean	Min	Max	S.D.
Independent Variable					
Log of AADTMAJ	Logarithm of AADT on major road	8.04	6.04	9.63	0.94
Log of AADTMIN	Logarithm of AADT on minor road	6.57	4.38	9.25	0.99
MEDWDMAJ	Median width on major road in feet	0.75	0	14	2.95
MEDWDMIN	Median width on minor road in feet	0.21	0	12	1.56
SHLWDMAJ	Shoulder width on major road in feet	1.43	0	10	1.42
SHLWDMIN	Shoulder width on minor road in feet	0.64	0	10	1.12
SIGNAL	Intersection type (0 if non-signalized intersection, 1 if signalized intersection)	0.31	0	1	0.46
RTLMAJ	Right-turn lane indicator (1 if at least one right-turn lane on the major road, 0 otherwise)	0.15	0	1	0.35
LTLMAJ	Left-turn lane indicator (1 if at least one left-turn lane on the major road, 0 otherwise)	0.25	0	1	0.43
RTLMIN	Right-turn lane indicator (1 if at least one right-turn lane on the minor road, 0 otherwise)	0.12	0	1	0.33
LTLMIN	Left-turn lane indicator (1 if at least one left-turn lane on the minor road, 0 otherwise)	0.14	0	1	0.35
HZRATMAJ	Roadside hazard rating on major road (from 1, least hazardous case, to 7, most hazardous case)	3.47	1	6	1.16
HZRATMIN	Roadside hazard rating on minor road	3.64	1	7	1.12
DRWYMAJ	Number of driveways on major road within 250ft of the intersection center	1.82	0	9	2.03
DRWYMIN	Number of driveways on minor road within 250ft of the intersection center	1.90	0	11	1.77
LIGHTMAJ	Lighting indicator (1 if lighting exists on the major road, 0 otherwise)	0.20	0	1	0.40
LIGHTMIN	Lighting indicator (1 if lighting exists on the minor road, 0 otherwise)	0.19	0	1	0.39
TERNMAJ	Terrain on major road (0=flat, 1=rolling, 2=mountainous)	0.57	0	2	0.55
TERNMIN	Terrain on minor road (0=flat, 1=rolling, 2=mountainous)	0.69	0	2	0.55
SPDLIMAJ	Speed limit on major road in mph	46.35	25	55	8.10
SPDLIMIN	Speed limit on minor road in mph	38.70	20	55	6.77
SDMAJ	Sight distance on major road in feet	1068	235	2000	425.9
SDMIN	Sight distance on minor road in feet	825	224	2000	355.1
VIMAJ	Sum of absolute change of grade in percent per hundred feet for each curve on major road or	1.74	0	5.63	1.13
VIMIN	minor road within 250ft of the intersection center, divided by the number of such curves	2.51	0	8.93	1.69
HAU	Intersection angle variable in degree	0.44	-57.5	50.0	20.61

CHAPTER 5

MODEL ESTIMATION AND RESULTS

In this chapter the estimation results, along with the statistical tests for the significance of contemporaneous correlation and endogeneity, are discussed. First, this chapter provides the estimation results of crash type models using single equation estimation techniques (i.e., conventional modeling techniques such as Poisson and negative binomial models) for two-lane four-legged rural intersections. Then, the estimation results of crash type models in a system of seemingly unrelated regression equations, along with the statistical test for the significance of contemporaneous correlation, are presented. Finally, the results of simultaneous equation models that control for endogeneity are discussed.

5.1 Single Equation Estimation for Crash Type Models

First, six individual models were estimated separately for crash types defined previously, along with a model of total expected number of crashes. Total number of crashes was estimated to illustrate that high-risk sites with respect to specific crash types cannot be revealed through crash totals. Note that it is assumed in this section that there is no contemporaneous correlation between the disturbance terms across crash type models.

An array of models were estimated and compared, various collections of independent variables were fitted, and ‘best’ models were selected using standard goodness of fit criteria and logical defensibility. All resultant models were of the negative binomial type except head-on and pedestrian-involved crash models, which were fitted using Poisson

regression (overdispersion parameter for these crash type models was not significantly different from zero). It proved difficult to obtain statistically significant explanatory variables for the opposite direction sideswipe model. The primary reason is likely that crash frequencies of this type were too small to lead to meaningful results. Although these crash types were omitted from the analysis, an alternative approach might be to include these crash types with head-on crashes (they arise under similar pre-crash conditions) since these two crash types may respond similarly to specific countermeasures and intersection features. As a result nothing can be concluded about opposite direction sideswipe crashes in this analysis.

5.1.1 Total Crash Model

Estimation results for the total crash model are presented in Table 5.1. Six explanatory variables are included in this model: AADT on the major road, AADT on the minor road, median width on the major road, right-turn lane on the major road, the number of driveways on the major road, and lighting on the major road. All of these variables are statistically significant at the 95% significance level. Of the six variables, four variables have a positive relationship with total number of crashes, while two variables reveal a negative relationship.

Table 5.1. Estimation results for total crashes (Negative binomial regression model)

Variables	Estimated coefficient	*t*-statistic	*p*-value
Constant	-4.4552	-6.101	0.0000
Log of AADT on the major road	0.4356	5.116	0.0000
Log of AADT on the minor road	0.3196	4.103	0.0000
Major road median width in feet	-0.0757	-3.210	0.0016
Right-turn lane indicator (1 if at least one right-turn lane on the major road, 0 otherwise)	0.7409	3.803	0.0002
Number of driveways on the major road within 250ft of the intersection center	0.1168	2.951	0.0036
Lighting indicator (1 if light on the major road, 0 otherwise)	-0.4785	-2.470	0.0145
α (dispersion parameter)	0.4139	5.260	0.0000
Number of observations	165		
Log-likelihood at zero	-464.54		
Log-likelihood at convergence	-394.52		
ρ^2	0.15		

Similar to prior studies, AADT is positively associated with total crashes. Increased traffic volumes (AADT) or exposure to risk on both the major and minor roads are associated with increased crashes at intersections. Median width for the major road is negatively associated with crashes also in agreement with prior research (Oh et al., 2003). The explanation for this finding is that physical separation of travel directions on average improves safety. Surprising, the presence of exclusive left-turn lanes, which typically is associated with a reduction in crashes, was not found to be statistically significant. This is likely because angle crashes, which are affected by left-turn lanes, represent a subset of total crashes, and so the effect of left-turn lanes is muted by the remaining crash types. In contrast, the presence of a right-turn lane on a major road is associated with a higher

number of crashes at intersections. There is a plausible explanation for this finding. Typically, right-turn lanes are installed when right turning movements are high and also when the opportunity for angle crashes is also high—so it may merely indicate the presence of a traffic movement that is not significant at similar sites without right-turn lanes. The number of commercial driveways close to the intersection was also associated with higher total crash frequencies at intersections. This is not surprising; in general traffic movements are more complex at access points, thus traffic crashes are more likely to occur at access points, especially when access points are close to intersections (prohibition of access points near intersections has become a popular access management strategy). Finally, a positive relationship was revealed between lighting and safety, which suggests that the presence of lighting effectively reduces traffic crashes—providing greater visibility to drivers at night so they can perceive and react quicker to potential hazardous situations.

5.1.2 Angle Crash Model

For the negative binomial angle crash model five independent variables were found to be statistically significant, as shown in Table 5.2. The variables include AADT on the major road, AADT on the minor road, left-turn lane on the major road, the number of driveways on the major road, and lighting on the major road. In contrast with the total crash model, the presence of left-turn lanes on major roads was found to be statistically significant at the 95% significance level (positively associated with angle crashes). From

the standpoint of the safety effects of significant variables on angle crashes, it is worthwhile to discuss the effect of left-turn lanes on angle crashes.

Table 5.2. Estimation results for angle crashes (Negative binomial regression model)

Variables	Estimated coefficient	*t*-statistic	*p*-value
Constant	-4.6298	-4.265	0.0000
Log of AADT on the major road	0.3692	2.821	0.0054
Log of AADT on the minor road	0.3015	3.010	0.0030
Left-turn lane indicator (1 if at least one left-turn lane on the major road, 0 otherwise)	0.5433	2.231	0.0270
Number of driveways on the major road within 250ft of the intersection center	0.1050	2.092	0.0380
Lighting indicator (1 if light on the major road, 0 otherwise)	-0.7573	-2.917	0.0040
α (dispersion parameter)	0.6372	4.275	0.0000
Number of observations	165		
Log-likelihood at zero	-318.67		
Log-likelihood at convergence	-293.86		
ρ^2	0.09		

As described in Chapters 1 and 2, the safety effect of left-turn lanes on crashes has been mixed due to the potential endogeneity of a left-turn lane indicator variable. In the angle crash model the presence of the left-turn lane indicates installation of the left-turn lane for handling large turning movements or in reaction to higher than average crash counts at a location. In other words, crash warrants (left-turn or total crashes exceeding a threshold) may have been used to justify installation of the left-turn lanes, and in fact crash experience is being used as one of the criteria for determining installation of left-

turn lanes (2003 Highway Design Manual, Oregon). This may indicate that there exits a bi-directional relationship between left-turn lanes and crashes (i.e., independent and dependent variables are affecting each other and thus dependent variables appear as a regressor in different equations), and the endogeneity arises from the bi-directional relationship, which leads to biased parameter estimates in a regression (Greene, 2003). It is also possible that some left-turn lanes were not installed due to crash warrants but exposure warrants—in these cases a right-hand side variable (left-turn lane indicator) would be influenced by the exposure variables (AADT). This condition is also generally undesirable in a regression. A promising solution to this problem is to adopt an instrumental variables approach, more specifically simultaneous equation estimation techniques, whereby a system of equations is such that,

$$E(\mu_i) = aAADT_{Major}^{\alpha 1} AADT_{Minor}^{\alpha 2} \exp\sum \beta_j x_{ij} \tag{5.1}$$

$$E(LTMAJ) = f\left(LT_{Maj}, LT_{Min}, AADT_{Maj}, AADT_{Min}, Angle\ Crashes\right), \tag{5.2}$$

and where $\exp\sum \beta_j x_{ij}$ includes $E(LTMAJ)$—the predicted value (instrument) for the presence of left-turn lanes, and all other predictors as shown in Table 5.2. In Equation (5.2), $E(LTMAJ)$ is a function (most likely a logistic regression) of left-turn volumes (LT) and AADT on the major and minor roads and angle crashes. Unfortunately, without left-turning traffic volumes at the sampled sites (unlike Hauer et al., 1988), which are often a primary determinant in left-turn lane installation warrants, such a model is not possible with the available data. The simultaneous estimation of equations (5.1) and (5.2) is discussed in more details in Section 5.3.

The presence of right-turn lanes and major road median width was not statistically significant. Because median width is mildly correlated with presence of left-turn lane (0.45), the positive role that median width typically plays—providing turning vehicles a refuge while turning—is thought to be subsumed by the presence of left-turn lane effect.

5.1.3 Head-on Crash Model

For the Poisson head-on crash model shown in Table 5.3, only traffic volume variables for the major and minor roads are statistically significant. Generally, median width is expected to be negatively associated with head-on crashes, that is, a narrower median width prevents fewer head-on crashes by providing less recovery distance (and time), by providing less of a vehicle refuge, etc. Contrary to expectations, this variable was not statistically significant in this model. The reason might be due to the small number of observations—of the 837 total crashes, only 22 crashes (2.6%) were head-on crashes.

Table 5.3. Estimation results for head-on crashes (Poisson regression model)

Variables	Estimated coefficient	*t*-statistic	*p*-value
Constant	-11.6926	-4.329	0.0000
Log of AADT on the major road	0.6585	1.899	0.0576
Log of AADT on the minor road	0.5864	2.786	0.0053
Number of observations	165		
Log-likelihood at zero	-67.71		
Log-likelihood at convergence	-56.46		
ρ^2	0.17		

5.1.4 Rear-end Crash Model

The rear-end negative binomial crash model presented in Table 5.4 shares the same set of explanatory variables with the total crash model, albeit with different coefficient estimates (this is an argument, as discussed previously, for not using simultaneous model estimation), except for a lighting indicator variable. The results showed that lighting is not associated with rear-end crashes. The directions of the associations with safety of the model variables are identical to the safety effects on total crashes: AADT, the presence of a right-turn lane, and density of driveways are positively associated with rear-end crashes.

Table 5.4. Estimation results for rear-end crashes (Negative binomial regression model)

Variables	Estimated coefficient	*t*-statistic	*p*-value
Constant	-11.3326	-8.075	0.0000
Log of AADT on the major road	0.9571	6.135	0.0000
Log of AADT on the minor road	0.4610	4.490	0.0000
Major road median width in feet	-0.0838	-2.805	0.0056
Right-turn lane indicator (1 if at least one right-turn lane on the major road, 0 otherwise)	0.7425	3.123	0.0021
Number of driveways on the major road within 250ft of the intersection center	0.1389	2.480	0.0141
α (dispersion parameter)	0.4119	2.464	0.0148
Number of observations	165		
Log-likelihood at zero	-214.29		
Log-likelihood at convergence	-206.12		
ρ^2	0.04		

The sign associated with the presence of a right-turn lane is again likely to be a cross-sectional data artifact and endogeneity problem: sites with right-turn lanes had experienced high numbers of rear-end crashes to justify their installation. As before, an instrumental variables approach making use of right-turning traffic volumes could address the bias introduced by the endogeneity. The relationship between median width and rear-end crashes is a bit puzzling and suggests that median width is serving as a proxy for an unobserved variable—most likely degree of urbanization, with greater widths positively associated with more urbanized and complex driving environments. Similar to the angle crash model, the effect of left-turn lane and median width are confounded—and only one variable enters the model, unlike previous results found by Mitra et al. (2002).

5.1.5 Sideswipe (Same Direction) Crash Model

A surprising finding for the sideswipe same direction crash model is that a statistically significant relationship between traffic volumes and sideswipe crashes could not be found. This is a bit troubling and is expected to represent an anomalous finding—it is anticipated that replicated studies would find a significant relationship. Only three variables were associated with sideswipe crashes as presented in Table 5.5.

Table 5.5. Estimation results for sideswipe (same direction) crashes (Negative binomial)

Variables	Estimated coefficient	*t*-statistic	*p*-value
Constant	-2.6925	-7.669	0.0000
Major road median width in feet	-0.2368	-2.436	0.0159
Left-turn lane indicator (1 if at least one left-turn lane on the major road, 0 otherwise)	1.8042	4.782	0.0000
Number of driveways on the major road within 250ft of the intersection center	0.2698	3.235	0.0015
α (dispersion parameter)	0.7364	2.013	0.0457
Number of observations	165		
Log-likelihood at zero	-87.39		
Log-likelihood at convergence	-85.07		
ρ^2	0.03		

According to the modeling results, median width on major roads is negatively associated with sideswipe crashes, whereas the presence of a left-turn lane and the density of nearby driveways are associated with higher crash frequencies; vehicles slowing to turn are struck by through moving vehicles. A similar result was observed in a previous study (Chin and Quddus, 2003)—it appears that when left-turn lanes are present, motor vehicles in the approach spend proportionately more time side-by-side (e.g. a queue formed in the left-turn lane is passed by all adjacent lane through traffic) and proportionately more time weaving into and out of the left-turn lane—when combined lead to greater opportunities for same direction sideswipe crashes.

The number of driveways is also associated with an increase in sideswipe crashes. This may be explained by conflicts between through vehicles and merging vehicles from driveways—which in the presence of an already complicated driving task (intersection)

presents increased risk. This finding supports access management policies near intersections.

5.1.6 Pedestrian-involved Crash Model

A Poisson model was developed to estimate pedestrian-involved crashes, and is shown in Table 5.6. It should be noted at the outset that pedestrian exposure is entirely missing (for the obvious reason that pedestrian crossing counts are not routinely recorded), which complicates model estimation and subsequent interpretation. Thus, the variables that are available require a modeler to provide educated guesses as to what surrogate 'pedestrian-involved' variables are being captured by standard traffic related variables, a difficult task at best. Three variables were found to be statistically associated with pedestrian-involved crashes: AADT on minor roads, shoulder width on major roads, and lighting on major roads.

Table 5.6. Estimation results for pedestrian-involved crashes (Poisson regression model)

Variables	Estimated coefficient	*t*-statistic	*p*-value
Constant	-2.2038	-4.053	0.0000
Log of AADT on the minor road	0.3069	3.805	0.0002
Major road shoulder width in feet	0.1055	2.039	0.0430
Lighting indicator (1 if light on the major road, 0 otherwise)	-0.5810	-2.487	0.0139
Number of observations	165		
Log-likelihood at zero	-209.66		
Log-likelihood at convergence	-198.80		
ρ^2	0.05		

Unlike the other models, AADT on major roads was not significant, but AADT on minor roads is positively associated with pedestrian-involved crashes. The reason for this finding is not entirely clear; however, it is suggested that AADT on the minor road is suggestive of a relatively busy intersection, with greater numbers of pedestrians.

Shoulder width is positively associated with pedestrian-involved crashes. This finding is somewhat intuitive and suggests that pedestrians may feel a greater sense of security (compared with narrow shoulders where they may refuse to walk) walking on a road with wide shoulders—increasing their exposure to potential conflict with motor vehicles. In addition, crossing a roadway with wider shoulders increases the exposure time to motor vehicles. The relationship of safety with AADT on minor roads may be indicative of higher road standards associated with a greater local built environment and increased pedestrian activity (again in the absence of pedestrian exposure). The presence of lighting is found to be negatively associated with pedestrian safety. Contrary to the other two variables, lighting appears not to be a surrogate for pedestrian activity, and suggests that lighting is beneficial for reducing pedestrian and motor vehicle conflicts.

5.1.7 Comparison of Coefficients and Variables across Crash Type Models

Table 5.7 provides a summary of model estimation results across all models estimated in this study. As stated previously, one of the justifications for modeling crash types is to identify which variables contribute to certain types of crashes and to compare how different significant variables affect safety for different crash types. The results show that a handful of the available roadway geometric and traffic volume variables

affect the safety of two-lane rural intersections: AADT on major and minor roads, median width of major roads, shoulder width of major roads, the presence of turning lanes, the number of driveways on major roads, and the presence of lighting for major roads are found in various groupings in the models.

Among these variables, AADT on the major and minor roads in general are the most influential variables associated with crash risk. As AADT increases, exposure to risk (at the site) increases. Past research has shown that some safety performance functions with respect to AADT are non-linear (e.g. exponential)—whereby increasing AADT eventually will lead to improvement in safety; non-linear relationships between safety and AADT were also found in this study. AADT is generally not viewed as a controllable factor but instead an important predictor of crashes, since controlling AADT is generally not an option to engineers or planners.

With respect to influential roadway geometric variables, median widths of major roads are negatively associated with safety on total crashes, rear-end crashes, and sideswipe crashes.

Shoulder width is significant only for pedestrian-involved crashes. The number of driveways near an intersection (driveway density) is associated with higher crash frequencies for all crash types except head-on and pedestrian-involved crashes.

Table 5.7. Comparison of coefficients and variables across crash type models

Variables	Estimated Coefficients					
	TC[2]	AC[2]	HD[2]	RE[2]	SS[2]	PR[2]
Constant	-4.4552	-4.6298	-11.6926	-11.3326	-2.6925	-2.2038
Log of AADT on the major road	0.4356	0.3692	0.6585	0.9571	n/s	n/s
Log of AADT on the minor road	0.3196	0.3015	0.5864	0.4610	n/s	0.3069
Major road median width in feet	-0.0757	n/s	n/s	-0.0838	-0.2368	n/s
Major road shoulder width in feet	n/s[1]	n/s	n/s	n/s	n/s	0.1055
Right-turn lane indicator (1 if at least one right-turn lane on the major road, 0 otherwise)	0.7409	n/s	n/s	0.7425	n/s	n/s
Left-turn lane indicator (1 if at least one left-turn lane on the major road, 0 otherwise)	n/s	0.5433	n/s	n/s	1.8042	n/s
Number of driveways on the major road within 250ft of the intersection center	0.1168	0.1050	n/s	0.1389	0.2698	n/s
Lighting indicator (1 if light on the major road, 0 otherwise)	-0.4785	-0.7573	n/s	n/s	n/s	-0.5810
Estimated Models	NB	NB	P	NB	NB	P

Note: All of the variables are statistically significant at 5% significance level except a AADT on the major road for the head-on crash model and a lighting indicator variable for the rear-end crash model, which are significant at 10% significance level.

[1] "n/s" indicates that a zero-effect coefficient could not be rejected at 95% level of confidence

[2] TC: Total crash model, AC: Angle crash model, HD: Head-on crash model, RE: Rear-end crash model, SS: Sideswipe (same direction) crash model, PR: Pedestrian-involved crash model

[3] NB: Negative Binomial model, P: Poisson model

The turning-lane indicator variables revealed a variety of statistically significant effects. The presence of right-turn lanes on major roads was significant in the rear-end and total crash models, whereas the presence of left-turn lanes was significant in angle and same direction sideswipe crash models. The coefficient estimates for the right-turn indicator were effectively the same; whereas the coefficient estimates for the left-turn indicator variables differed by a factor of two.

Finally, the lighting indicator variable revealed a fairly consistent negative association with safety across models, although a zero effect could not be rejected for the head-on and same direction sideswipe crash type models, probably because the observed number of these two crash types during the observation period is small, representing 22 and 42 out of 837 crashes, respectively.

Examination of the coefficient estimates across crash type models clearly demonstrates, at least statistically, that crash types are associated with a different set of predictors. This result is not surprising, since pre-crash conditions will largely determine the crash type—and this result is in agreement with engineering expectations. For example, a relatively short yellow clearance interval at a signalized intersection may affect rear-end crashes differently than same direction sideswipe crashes. The insights afforded by estimating crash type models may lead to insights as to the relative effectiveness of various countermeasures and/or predictive variables.

The second justification claimed in this research for estimating crash type models is that ‘hazardous’ sites may reveal themselves through elevated crash type frequencies and not total crash frequencies. Although this research is not focused on ‘hot spot

identification' methodology, an inspection of a simple ranking of 'highest risk' sites using the models estimated in this study can reveal this phenomenon. Table 5.8 lists the top 5 'high risk' sites identified using each of the models. To identify the sites listed in the table, the total and crash type models were used to predict crash frequencies for each of the sites. Predicted frequencies were compared to observed frequencies, and the top 5 'high-risk' sites (observed – predicted) were identified.

The table shows that for the total crash model, Sites 129, 134, 124, 157, and 136 were ranked as 'most hazardous'. That is, the model of total crashes predicts the expected crash frequency for these sites; however these sites recorded significantly more than predicted (and based on similar sites). In comparison, the head-on crash model ranked only two of these sites as 'highest risk', site 129 and site 124, and identified the highest risk site as site 14. Similarly, the rear-end crash model identified site 149 as most hazardous, because it recorded 6 more sideswipe crashes than expected compared to similar sites. Site 50, again not identified as top priority by the total crash model, recorded 3 more pedestrian crashes than expected compared to similar sites.

Table 5.8. Top 5 outlying observations with respect to expected crash counts (in order of decreasing rank)

Model		Rank				
		1st	2nd	3rd	4th	5th
Total Crash	*Site ID*	*129*	*134*	*124*	*157*	*136*
	Observed	53	23	27	23	25
	Predicted	19	10	18	16	20
	Residual [1]	34	13	9	7	5
Angle Crash	*Site ID*	*129*	*134*	*124*	*136*	*159*
	Observed	33	11	12	13	10
	Predicted	15	5	7	9	9
	Residual	18	6	5	4	1
Head-on Crash	*Site ID*	*14*	*129*	*124*	*158*	
	Observed	2	2	1	1	
	Predicted	0	1	0	0	
	Residual	2	1	1	1	
Rear-end Crash	*Site ID*	*149*	*129*	*124*	*136*	*159*
	Observed	8	15	10	10	8
	Predicted	2	10	6	8	6
	Residual	6	5	4	2	2
Sideswipe Crash (same direction)	*Site ID*	*157*	*131*	*149*	*130*	*158*
	Observed	6	3	2	3	3
	Predicted	1	1	0	2	2
	Residual	5	2	2	1	1
Pedestrian-involved Crash	*Site ID*	*50*	*107*	*118*	*124*	*83*
	Observed	4	3	3	3	3
	Predicted	1	1	1	2	2
	Residual	3	2	2	1	1

[1] Residual = Observed crashes – Predicted crashes

While it is acknowledged that sample sizes are small and that this analysis is preliminary (we do not assess 'how outlying' observations are with respect to expected safety), the results do lend credence to the claim that crash type models may reveal 'problems' at intersections (road segments, interchanges, etc.) that will not be revealed through total crash models. A more thorough hot spot identification analysis would consider site selection bias (regression to the mean) and might apply Bayesian methods to correct for this phenomenon.

This section provided the estimation results of separate crash prediction models for crash types, assuming that the disturbance terms across crash type models are not correlated. In fact, however, these crash types are expected to be contemporaneously correlated due to the unobserved common variables, and thus the estimation results of crash type models with controlling for contemporaneous correlation are discussed in the next section.

5.2 Seemingly Unrelated Regression Estimation for Crash Type Models

Previous work has shown that estimation of crashes originating from the same entity (e.g. intersection, road segment, city, traffic analysis zone) may require simultaneous estimation techniques (Ladron de Guevara et al., 2004) due to contemporaneous correlation of disturbances—representing a Seemingly Unrelated REgression (SURE) system (Zellner, 1962). Simultaneity in this case may arise from common omitted variables across models of crash types as a result of common important omitted variables. In such a case, regression models applied equation-by-equation yield consistent but

inefficient coefficient estimates. By accounting for this correlation across errors using simultaneous equation models, more efficient parameter estimates can be obtained. However, parameter estimates in either singular or simultaneous estimation methods are unbiased (Washington et al., 2003).

If simultaneity is present across models (i.e. in various models of crash type) then the error terms (fitted minus observed values) will be highly correlated across models (Washington et al., 2003; Greene, 2002). Therefore, this research first explored whether or not the error terms are contemporaneously correlated across models through the Lagrange Multiplier (LM) test.

Based on the results from section 5.1, the predicted values of three negative binomial and two Poisson models were calculated, and then residuals (observed minus predicted values) were generated for five models. Using equation (3.26), the LM test statistic was calculated. The LM test statistic was 110.15, which is greater than the critical value (18.31 with $d.f.$=10 and p=0.05), and thus the null hypothesis that the off-diagonal elements of $\sum$ are zero was rejected, implying that there exists contemporaneous correlation between the error terms across models and all of the models should be estimated simultaneously.

Simultaneous estimation of five crash models is required to joint the two different distributions of error terms in this case: negative binomial and Poisson distribution. However, this case has not been solved so far due to severe computational problems (Jung and Winkelmann, 1993). For this research, therefore, all the models were assumed to follow a negative binomial distribution and estimated simultaneously although head-on

and pedestrian-involved crashes were found to be better fitted using Poisson regression model. Simultaneous estimation results using seemingly unrelated negative binomial regression models are presented in Tables 5.9 through 5.13.

Table 5.9. Simultaneous estimation results (Angle crashes)

Variables	Estimated coefficient	Standard error	*p*-value
Constant	5.4550	0.9799	0.0000
Log of AADT on the major road	-0.1308	0.1166	0.2637
Log of AADT on the minor road	-0.5835	0.0722	0.0000
Left-turn lane indicator (1 if at least one left-turn lane on the major road, 0 otherwise)	1.2577	0.1944	0.0000
Number of driveways on the major road within 250ft of the intersection center	0.1395	0.0417	0.0010
Lighting indicator (1 if light on the major road, 0 otherwise)	0.1201	0.2024	0.5538

Table 5.10. Simultaneous estimation results (Head-on crashes)

Variables	Estimated coefficient	Standard error	*p*-value
Constant	-16.162	4.2248	0.0001
Log of AADT on the major road	1.8669	0.4760	0.0001
Log of AADT on the minor road	-0.2804	0.1874	0.1367

Table 5.11. Simultaneous estimation results (Rear-end crashes)

Variables	Estimated coefficient	Standard error	*p*-value
Constant	-3.4688	1.2101	0.0047
Log of AADT on the major road	0.6416	0.1355	0.0000
Log of AADT on the minor road	-0.2752	0.0843	0.0013
Major road median width in feet	-0.0431	0.0254	0.0916
Right-turn lane indicator (1 if at least one right-turn lane on the major road, 0 otherwise)	0.5946	0.1926	0.0024
Number of driveways on the major road within 250ft of the intersection center	0.1301	0.0474	0.0068

Table 5.12. Simultaneous estimation results (Sideswipe same direction crashes)

Variables	Estimated coefficient	Standard error	*p*-value
Constant	-2.7498	0.3868	0.0000
Major road median width in feet	-0.2568	0.0906	0.0052
Left-turn lane indicator (1 if at least one left-turn lane on the major road, 0 otherwise)	1.5419	0.3559	0.0000
Number of driveways on the major road within 250ft of the intersection center	0.3747	0.0794	0.0000

Table 5.13. Simultaneous estimation results (Pedestrian-involved crashes)

Variables	Estimated coefficient	Standard error	*p*-value
Constant	3.2682	0.6144	0.0000
Log of AADT on the minor road	-0.4537	0.0898	0.0000
Major road shoulder width in feet	0.0041	0.0563	0.9415
Lighting indicator (1 if light on the major road, 0 otherwise)	0.5104	0.2384	0.0338

Despite the significance of the LM test statistic, some of the estimation results seem to be questionable; for example, AADT is usually expected to be positively associated with all types of crashes, but AADT on minor roads was found to have a negative relationship with angle, head-on, rear-end, and pedestrian-involved crashes. In addition, AADT variables were found to be insignificant for angle and head-on crashes, and lighting and major road shoulder width were not found to be significant for angle and pedestrian-involved crashes, respectively. This is likely because head-on and pedestrian-involved crashes were estimated by negative binomial models although they were better fitted using Poisson models. Another possible reason is that not all the disturbance terms of crash type models might be contemporaneously correlated.

In order to explore whether all of the disturbance terms are contemporaneously correlated across models, the crash models were paired with two models and the LM test was conducted for a total of 10 pairs of models. Table 5.14 presents the results of the contemporaneous correlation test. It should be noted that the LM statistics shown in Table 5.14 were computed with negative binomial models for angle, rear-end, and sideswipe crashes and Poisson models for head-on and pedestrian-involved crashes.

As shown in Table 5.14, the LM statistics between pedestrian-involved crashes and all other types of crashes were smaller than the critical value, indicating that this type of crashes are not correlated with other types of crashes and thus the pedestrian-involved crash model can be estimated separately. In contrast, angle crashes were found to be correlated with rear-end and sideswipe crashes, but contemporaneous correlation between head-on and sideswipe crashes was found to be not significant.

Table 5.14. Test results for contemporaneous correlation between two disturbance terms

	AC	HD	RE	SS	PR
AC	-	3.6002	48.724	14.890	0.0843
HD	3.6002	-	11.322	0.8288	1.4412
RE	48.724	11.322	-	28.788	0.4504
SS	14.890	0.8288	28.788	-	0.0205
PR	0.0843	1.4412	0.4504	0.0205	-

Note: 1) Angle, rear-end, and sideswipe crashes were of the negative binomial type, whereas head-on and pedestrian-involved crashes were fitted with Poisson regression.
2) Critical value = 3.84 (*d.f.*=1, *p*=0.05)
3) AC: Angle crash model, HD: Head-on crash model, RE: Rear-end crash model, SS: Sideswipe (same direction) crash model, PR: Pedestrian-involved crash model

Through the statistical test for contemporaneous correlation, it was decided to estimate only three crash models simultaneously: angle, rear-end, and sideswipe crashes. All of these models were estimated in a system of seemingly unrelated negative binomial regression equations. Simultaneous estimation results for angle, rear-end, and sideswipe crashes are summarized in Table 5.15. All of the parameters estimated in angle, rear-end, and sideswipe crash models were also found to be statistically significant at 95% significance level and the directions of the associations with the model variables are identical to the safety effects obtained from single equation estimation.

Table 5.15. Simultaneous estimation results for angle, rear-end, and sideswipe crashes

Variables	Estimated coefficient	*t*-statistic	*p*-value
Angle Crash Model			
Constant	-3.1421	-3.417	0.0008
Log of AADT on the major road	0.2364	2.114	0.0361
Log of AADT on the minor road	0.2289	3.204	0.0001
Left-turn lane indicator (1 if at least one left-turn lane on the major road, 0 otherwise)	0.7412	3.906	0.0020
Number of driveways on the major road within 250ft of the intersection center	0.1268	3.141	0.0010
Lighting indicator (1 if light on the major road, 0 otherwise)	-0.7517	-3.793	0.0002
Rear-end Crash Model			
Constant	-11.072	-8.631	0.0000
Log of AADT on the major road	0.9636	6.788	0.0000
Log of AADT on the minor road	0.4167	4.867	0.0000
Major road median width in feet	-0.0638	-2.484	0.0140
Right-turn lane indicator (1 if at least one right-turn lane on the major road, 0 otherwise)	0.4349	2.246	0.0261
Number of driveways on the major road within 250ft of the intersection center	0.1225	2.566	0.0112
Sideswipe (Same Direction) Crash Model			
Constant	-2.6654	-8.048	0.0000
Major road median width in feet	-0.1875	-2.038	0.0432
Left-turn lane indicator (1 if at least one left-turn lane on the major road, 0 otherwise)	1.7164	5.030	0.0000
Number of driveways on the major road within 250ft of the intersection center	0.2481	3.396	0.0008
α (dispersion parameter)	0.7067	5.358	0.0000
Number of observations	165		
Log-likelihood at convergence	-943.92		

Compared to single estimation results, it was found that simultaneous estimation approaches provided nonidentical but close estimates for the coefficients in angle, rear-end, and sideswipe crash models as shown in Tables 5.16 through 5.18. Furthermore, more efficient estimators were obtained when the models were jointly estimated in a system of seemingly unrelated regression equations (i.e., the standard errors of the coefficients were smaller when the models were jointly estimated).

Table 5.16. Comparison of coefficients and standard errors (Angle crash model)

Variables	Single Equation Estimation	Simultaneous Equation Estimation
Constant	-4.6298 (1.0855)	-3.1421 (0.9195)
Log of AADT on the major road	0.3692 (0.1309)	0.2364 (0.1118)
Log of AADT on the minor road	0.3015 (0.1003)	0.2289 (0.0714)
Left-turn lane indicator (1 if at least one left-turn lane on the major road, 0 otherwise)	0.5433 (0.2435)	0.7412 (0.1898)
Number of driveways on the major road within 250ft of the intersection center	0.1050 (0.0501)	0.1268 (0.0404)
Lighting indicator (1 if light on the major road, 0 otherwise)	-0.7573 (0.2596)	-0.7517 (0.1982)

Note: Standard errors appear in parentheses.

Table 5.17. Comparison of coefficients and standard errors (Rear-end crash model)

Variables	Single Equation Estimation	Simultaneous Equation Estimation
Constant	-11.3326 (1.4034)	-11.072 (1.2828)
Log of AADT on the major road	0.9571 (0.1560)	0.9636 (0.1419)
Log of AADT on the minor road	0.4610 (0.1027)	0.4167 (0.0856)
Major road median width in feet	-0.0838 (0.0299)	-0.0638 (0.0256)
Right-turn lane indicator (1 if at least one right-turn lane on the major road, 0 otherwise)	0.7425 (0.2377)	0.4349 (0.1937)
Number of driveways on the major road within 250ft of the intersection center	0.1389 (0.0560)	0.1225 (0.0478)

Note: Standard errors appear in parentheses.

Table 5.18. Comparison of coefficients and standard errors (Sideswipe crash model)

Variables	Single Equation Estimation	Simultaneous Equation Estimation
Constant	-2.6925 (0.3511)	-2.6654 (0.3312)
Major road median width in feet	-0.2368 (0.0972)	-0.1875 (0.0920)
Left-turn lane indicator (1 if at least one left-turn lane on the major road, 0 otherwise)	1.8042 (0.3773)	1.7164 (0.3413)
Number of driveways on the major road within 250ft of the intersection center	0.2698 (0.0834)	0.2481 (0.0731)

Note: Standard errors appear in parentheses.

5.3 Simultaneous Equation Estimation to Control for Endogeneity

In section 5.1.2, the potential endogeneity of a left-turn lane indicator variable in crash prediction models was proposed due to simultaneity between left-turn lanes and crashes. Moreover, since endogeneity leads to biased and inconsistent estimators, simultaneous equation estimation techniques were suggested as an efficient method to capture endogeneity. Thus, this section describes estimation results of simultaneous equation models to compensate for endogeneity between left-turn lane presence and angle crashes, along with a statistical test for endogeneity.

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. Because a negative binomial model for angle crashes was already estimated in section 5.1.2, a logit model for left-turn lane indicator was estimated with the same set of explanatory variables, including the number of angle crashes. As stated in section 5.1.2, five variables were found to be significant at the 95% significance level for angle crashes, while four variables were found to be statistically significant at the 90% significant level for left-turn lane indicator. The reason for using the 90% significance level for left-turn lane indicator is to include the number of angle crashes into the left-turn lane indicator model. The number of crashes is usually expected to be highly associated with installation of left-turn lanes, and in fact crash experience has been used to justify installation of left-turn lanes as described previously. However, this variable was found to be non-significant at the 95% significance level, but significant

at the 90% significance level. This might be due to the fact that the number of angle crashes is endogenously associated with installation of left-turn lanes. Therefore, this variable was included in the final model in order to test the potential endogeneity. The estimation results without controlling for endogeneity (i.e., single equation estimation results) are summarized in Table 5.19.

Table 5.19. Estimation results without controlling for endogeneity

Variables	Estimated coefficient	*t*-statistic	*p*-value
Equation 1: Angle crashes (dependent variable)			
Constant	-4.6298	-4.265	0.0000
Log of AADT on the major road	0.3692	2.821	0.0054
Log of AADT on the minor road	0.3015	3.010	0.0030
Left-turn lane indicator (1 if at least one left-turn lane on the major road, 0 otherwise)	0.5433	2.231	0.0270
Number of driveways on the major road within 250ft of the intersection center	0.1050	2.092	0.0380
Lighting indicator (1 if light on the major road, 0 otherwise)	-0.7573	-2.917	0.0040
α (dispersion parameter)	0.6372	4.275	0.0000
Equation 2: Left-turn lane indicator (dependent variable)			
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 intersection, 0 otherwise)	3.1564	4.20	0.000
Number of driveways on the major road within 250ft of the intersection center	-0.4476	-2.52	0.012

Note: The results of angle crash model are the same as Table 5.2.

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 similarly described in prior studies, AADT and the number of driveways are found to be positively associated with angle crashes. Also, the presence of left-turn lane, which is the focus of this research, has an influence on increase in angle crashes. It is typically regarded that installation of left-turn lane can effectively reduce angle crashes, but the estimation result is opposite to our expectations. The reason for this might be the fact that the left-turn lane indicator may be an endogenous variable as described previously. Finally, a lighting indicator variable has a negative relationship with angle crash, which suggests that the presence of lighting effectively reduces angle crashes.

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. In particular, AADT on the minor road was found to be insignificant for the model. This is likely because the dependent variable of interest is left-turn lanes on major roads. These significant variables that may affect installation of left-turn lanes are consistent with the left-turn lane criteria presented in the 2003 Highway Design Manual of Oregon. Increased AADT, higher angle crashes, and signalized intersections are more likely to be associated with installation of left-turn lanes. On the other hand, a higher number of driveways tends to mitigate against installation of left-turn lanes.

Next, a two model system was estimated simultaneously by applying the LIML approaches (equations (3.37) and (3.38)). For the simultaneous estimation of models,

only significant variables were used: five variables for the angle crash model and four variables for the left-turn indicator model. The estimation results are presented in Table 5.20.

Table 5.20. Estimation results with controlling for endogeneity (LIML approach)

Variables	Estimated coefficient	*t*-statistic	*p*-value
Equation 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 250ft 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
Equation 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 250ft of the intersection center	-0.4774	-2.77	0.006

The results show that there is no difference between the safety effects of estimated parameters with and without controlling for endogeneity, except for the effect on angle crashes of a left-turn lane indicator. Single equation estimation results in a positive relationship between the left-turn lane indicator and angle crashes (0.5433), 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 understanding would predict. This suggests that angle crashes are endogenously associated with a left-turn lane indicator, and ignoring the endogeneity will lead to erroneous conclusions.

Finally, the modified DWH test was applied to examine the endogeneity of angle crashes and a left-turn lane indicator. For the modified DWH test, the null hypothesis is that parameters estimated without controlling for endogeneity are consistent. Table 5.21 shows the results for the 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 5.21. Results for endogeneity test

Endogeneity Test		
H_0: Parameters estimated without controlling for endogeneity are consistent. H_1: Parameters estimated without controlling for endogeneity are inconsistent.		
	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.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

This research focused on the development of crash prediction models for different types of crashes such as angle, head-on, rear-end, same direction sideswipe, and pedestrian-involved crashes for two-lane and four-legged rural intersections, and developed simultaneous equation estimation methods, with the intent to accommodate contemporaneous correlation and control for endogeneity in crash models. The results of this research showed that certain types of crashes are associated with roadway characteristics and traffic attributes in different ways, and crash type models enable researchers to identify sites that are high-risk with respect to specific crash types. Statistical tests that were introduced to examine contemporaneous correlation and endogeneity suggested the needs for more efficient and advanced analytical methods. Further and deeper understanding into the safety effects of traffic and geometric features on highway safety has been revealed through advanced analytical techniques, which are seemingly unrelated negative binomial models and simultaneous equation models, by accounting for contemporaneous correlation and endogeneity.

The more detailed discussions associated with these research activities are provided in this chapter, followed by the recommendations for potential research that could be undertaken in the future to advance the concepts and initiatives that have been presented in this dissertation.

6.1 Discussion of Crash Type Models

The justification for this research rested upon the claim that modeling crash types provides certain advantages and is complementary to total crash modeling (totals, fatals, injuries, etc.). Much research has focused on estimating total crash models, while relatively little has focused on modeling crash types. Through analysis of rural intersections in Georgia, total crashes and crash type models were estimated for total, angle, head-on, rear-end, same direction sideswipe, and pedestrian involved crashes. The data were fit with negative binomial or Poisson regression models, resulting in satisfactory predictive models that are shown to be generally consistent with past research findings and with engineering knowledge of crash occurrence (with a few justifiable exceptions).

The results support the claim that crash types are associated with different conditions at rural intersections in Georgia—although the expectation is that analysis results would be similar elsewhere and for different roadway elements. Coefficient estimates varied considerably in some cases, and models were fit with different subsets of explanatory variables. One advantage of modeling crash type, therefore, is to gain insight as to the differences between conditions that lead to various crash types. Armed with this knowledge it may be possible to better understand the impact of various countermeasures on crash type.

It was also argued, secondly, that crash type models might also provide insight for ranking sites (intersections in this research), with total crash models failing to identify specific problems related to specific crash types. Although unsophisticated, it was shown

through comparison of expected to actual counts that crash type models may in fact identify different sites for further engineering scrutiny. This finding again bolsters the argument for the development of crash type models as complements to total crash models.

A larger and more comprehensive data set spanning several states would offer improved analysis potential. The inclusion of single-vehicle accidents is necessary to provide a complete 'picture' of intersection crashes, but these crash types were unavailable. Turning movements at intersection approaches would likely lead to improvements in explanatory ability of the models; however, turning movement counts are rarely known in practice and may limit practicality of the models. In addition, environmental variables and in particular inclement weather related data (i.e. wet pavement days, icy days, snow days, etc.) would be useful to include in predictive models but were unavailable for this research. Finally, a more sophisticated analysis of hot spot identification would reveal further insights into differences between total crash and crash type models and should be conducted. It is recommended that crash type models be estimated more routinely in conjunction and as complements to total crash models. It should be emphasized that crash type models are intended as a complement to models of crash severity, which are used to answer different but equally important safety questions.

6.2 Discussion of Seemingly Unrelated Regression Estimation

It is reasonable to believe that different types of crashes originating from the same entity (e.g., intersection, roadway segment, traffic analysis zone) contemporaneously correlated across the disturbance terms because they share unobserved common

characteristics (i.e., omitted common variables). This research estimated crash type models in a system of seemingly unrelated negative binomial equations using geometric characteristics and traffic attributes. Although environmental factors (weather conditions, visibility, wind, etc) and driver's characteristics (attention, mood, eyesight, reaction times, driving skills, etc.) are also highly associated with crashes, these variables were not included in the data due to unavailability. These factors can be regarded as omitted variables in this research, and lead to contemporaneous correlation across the disturbance terms. This correlation can be also found in other crash-related applications. If contemporaneous correlation exists, several equations should be estimated simultaneously in a system of seemingly unrelated regression equations because neglecting contemporaneous correlation leads to consistent but inefficient estimators.

This research investigated the potential contemporaneous correlation between the disturbance terms across crash type models through a statistical test, and the result indicated that the disturbance terms are contemporaneously correlated across all of the models. As a result, a simultaneous estimation technique was applied to estimate the parameters. It should be noted that although individual crash type models were fitted using Poisson and negative binomial models, all of the models were simultaneously estimated, assuming that all types of crashes followed a negative binomial distribution due to severe computational problems. However, some of the estimation results seem to be doubtful. The reason for this might be the fact that all of the models were simultaneously estimated using negative binomial regression models. Another possible reason is that not all the disturbance terms of crash type models might be

contemporaneously correlated. Thus, statistical tests were conducted again by making a pair of crash type models and it was shown that contemporaneous correlation exists among only three types of crashes such as angle, rear-end, and sideswipe crashes. Head-on and pedestrian-involved crashes were found to be uncorrelated with each other as well as other types of crashes.

Based on the results of statistical tests, angle, rear-end, and sideswipe crash models were estimated simultaneously as a set of equations. Estimation results presented that, compared to single equation estimation, simultaneous estimation techniques yield nonidentical but close estimates for the coefficients in angle, rear-end, and sideswipe crash models. The results also showed that more efficient estimators can be obtained when the crash models are jointly estimated, as expected.

6.3 Discussion of Simultaneous Equation Estimation

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 significantly associated with higher frequencies of accidents, contrary to expectation. The inconsistency observed in the literature is not an artifact of the underlying phenomenon but instead a problem in numerical estimation resulting 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 research.

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 presence of left-turn lanes are 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 indicates 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 research 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 research. Therefore, it is necessary to improve the ability of model prediction in future studies using turning movements.

The approach described in this dissertation is generalizeable 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. For example, a traffic signal indicator might be expected to have endogenous relationship with crashes since installation of traffic signals are in general warranted as a result of crash frequencies. Another example of potential endogeneity is the presence of lighting. Lighting facilities are installed at sites where the likelihood of crashes occurring during the nighttime is relative high or the number of crashes occurred at night is high. These evidences suggest that a traffic signal indicator and the presence of lighting are endogenously associated with crashes. 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 expectations.

6.4 Recommendations for Future Research

Based on the results of the present research, the following issues can be addressed in the future:

1. This research utilized overall intersection AADT for estimating an angle crash model. Since angle crashes are highly associated with turning movements at intersections, however, including turning movements at intersection approaches

may improve the predictive ability of the angle crash model relative to using overall intersection AADT. In addition, including turning movements into the angle crash model may provide greater explanatory power for the endogenous relationship between left-turn lanes and angle crashes since installation of left-turn lanes is more likely to be affected by turning movements as well as opposite through traffic volumes. Therefore, it is necessary to improve the ability of model prediction in future studies using turning movements.

2. Despite the advantage of a seemingly unrelated regression model, the simultaneous estimation of negative binomial and Poisson models has not been developed due to complicated mathematical derivations. Thus, a seemingly unrelated negative binomial-Poisson model needs to be developed for better investigating contemporaneous correlation between the disturbance terms of negative binomial and Poisson models.
3. This research has dealt with endogeneity problems for one independent variable. In some cases, however, more than one independent variable might be endogenously correlated with dependent variables. Therefore, further research needs to be conducted to deal with more than one endogenous variable in crash prediction models, and to seek out other possible endogeneity issues. The available software and estimation of such models remains relatively difficult, and so it is also necessary to concentrate more efforts on the development of estimation techniques.

APPENDIX A: DATA STATISTICS

Table A.1. The number of crash frequencies by intersections

Int. ID	Total crashes	Angle crashes	Head-on crashes	Rear-end crashes	Sideswipe same direction crashes	Sideswipe opposite direction crashes	Pedestrian-involved crashes
1	2	0	0	2	0	0	0
2	1	1	0	0	0	0	0
3	0	0	0	0	0	0	0
4	1	0	0	0	0	0	1
5	8	7	0	0	0	0	1
6	2	1	0	0	0	0	1
7	2	1	0	0	0	0	1
8	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0
10	3	2	0	0	1	0	0
11	2	0	0	0	0	0	2
12	2	1	0	0	0	0	1
13	1	0	0	0	0	0	1
14	5	1	2	0	0	0	2
15	1	1	0	0	0	0	0
16	5	3	0	2	0	0	0
17	3	0	0	0	0	0	3
18	6	4	0	2	0	0	0
19	4	3	0	0	0	0	1
20	4	1	0	0	0	0	3
21	2	0	0	0	0	0	2
22	5	3	1	0	0	1	0
23	3	2	0	0	0	0	1
24	1	0	0	0	0	0	1
25	1	0	0	0	1	0	0
26	1	0	0	0	0	0	1
27	1	0	0	0	1	0	0
28	6	3	0	2	0	0	1
29	3	1	0	0	0	0	2
30	0	0	0	0	0	0	0
31	4	1	0	1	0	1	1
32	9	2	1	3	1	1	1
33	5	0	1	0	0	1	3
34	0	0	0	0	0	0	0
35	3	0	0	0	0	1	2

Int. ID	Total crashes	Angle crashes	Head-on crashes	Rear-end crashes	Sideswipe same direction crashes	Sideswipe opposite direction crashes	Pedestrian-involved crashes
36	2	1	0	0	0	0	1
37	0	0	0	0	0	0	0
38	2	0	0	0	0	0	2
39	0	0	0	0	0	0	0
40	1	1	0	0	0	0	0
41	3	3	0	0	0	0	0
42	2	1	0	1	0	0	0
43	2	1	0	0	0	0	1
44	2	2	0	0	0	0	0
45	0	0	0	0	0	0	0
46	0	0	0	0	0	0	0
47	1	1	0	0	0	0	0
48	4	1	0	1	0	0	2
49	4	1	0	2	1	0	0
50	6	0	0	2	0	0	4
51	6	1	0	3	0	1	1
52	11	5	0	3	0	0	3
53	1	0	0	0	0	0	1
54	2	2	0	0	0	0	0
55	2	1	0	0	0	0	1
56	8	3	0	3	0	0	2
57	1	0	0	0	0	0	1
58	3	1	0	0	0	0	2
59	2	1	0	0	0	0	1
60	5	3	0	0	0	1	1
61	3	1	0	0	1	0	1
62	1	0	0	1	0	0	0
63	1	1	0	0	0	0	0
64	1	0	0	0	0	0	1
65	0	0	0	0	0	0	0
66	1	1	0	0	0	0	0
67	2	0	0	0	0	1	1
68	3	1	0	0	0	0	2
69	1	1	0	0	0	0	0
70	9	7	1	1	0	0	0
71	0	0	0	0	0	0	0
72	1	0	0	1	0	0	0
73	10	7	0	2	0	1	0

Int. ID	Total crashes	Angle crashes	Head-on crashes	Rear-end crashes	Sideswipe same direction crashes	Sideswipe opposite direction crashes	Pedestrian-involved crashes
74	4	3	0	1	0	0	0
75	5	4	0	0	0	0	1
76	4	3	0	1	0	0	0
77	1	0	0	0	0	0	1
78	4	3	0	0	0	0	1
79	6	4	0	2	0	0	0
80	3	2	0	0	0	0	1
81	2	0	0	0	0	1	1
82	1	0	0	0	0	0	1
83	10	2	1	3	1	0	3
84	7	3	0	2	1	0	1
85	1	0	0	1	0	0	0
86	9	2	1	4	0	0	2
87	2	1	0	0	0	0	1
88	2	0	0	0	0	0	2
89	2	1	0	0	0	0	1
90	5	3	0	1	0	1	0
91	2	0	0	0	0	0	2
92	4	1	0	0	1	0	2
93	12	7	0	5	0	0	0
94	0	0	0	0	0	0	0
95	7	2	0	3	1	0	1
96	10	4	0	4	1	0	1
97	4	1	1	1	0	1	0
98	0	0	0	0	0	0	0
99	1	0	0	0	1	0	0
100	1	1	0	0	0	0	0
101	0	0	0	0	0	0	0
102	3	1	0	0	0	0	2
103	2	1	0	0	0	0	1
104	0	0	0	0	0	0	0
105	2	1	0	0	0	0	1
106	0	0	0	0	0	0	0
107	10	5	0	1	1	0	3
108	4	1	0	0	0	0	3
109	0	0	0	0	0	0	0
110	1	0	0	0	0	0	1
111	11	8	0	1	1	0	1

Int. ID	Total crashes	Angle crashes	Head-on crashes	Rear-end crashes	Sideswipe same direction crashes	Sideswipe opposite direction crashes	Pedestrian-involved crashes
112	1	1	0	0	0	0	0
113	1	0	0	0	0	0	1
114	2	1	0	0	0	0	1
115	12	6	0	5	0	0	1
116	11	1	1	8	0	0	1
117	4	3	0	0	0	0	1
118	13	4	0	3	2	1	3
119	6	0	1	2	0	0	3
120	1	1	0	0	0	0	0
121	1	0	0	1	0	0	0
122	0	0	0	0	0	0	0
123	18	8	0	8	0	0	2
124	27	12	1	10	1	0	3
125	10	2	1	6	0	0	1
126	17	8	0	7	0	0	2
127	1	1	0	0	0	0	0
128	5	1	0	1	0	0	3
129	53	33	2	15	1	0	2
130	14	5	0	4	3	0	2
131	12	5	0	4	3	0	0
132	6	3	0	3	0	0	0
133	7	3	0	3	0	0	1
134	23	11	0	7	2	1	2
135	10	6	0	2	1	0	1
136	25	13	0	10	0	0	2
137	20	10	0	7	0	0	3
138	0	0	0	0	0	0	0
139	5	1	0	4	0	0	0
140	10	6	0	1	0	0	3
141	11	4	1	3	1	0	2
142	1	0	0	1	0	0	0
143	0	0	0	0	0	0	0
144	6	1	0	2	1	1	1
145	10	6	0	3	0	0	1
146	2	1	0	0	0	0	1
147	2	1	0	0	0	0	1
148	12	3	1	6	0	0	2
149	19	6	1	8	2	0	2

Int. ID	Total crashes	Angle crashes	Head-on crashes	Rear-end crashes	Sideswipe same direction crashes	Sideswipe opposite direction crashes	Pedestrian-involved crashes
150	3	0	0	1	0	0	2
151	2	1	0	1	0	0	0
152	5	2	0	2	1	0	0
153	0	0	0	0	0	0	0
154	2	1	0	0	0	0	1
155	0	0	0	0	0	0	0
156	0	0	0	0	0	0	0
157	23	9	0	8	6	0	0
158	20	7	1	7	3	2	0
159	21	10	1	8	0	0	2
160	5	2	1	2	0	0	0
161	7	4	0	2	0	1	0
162	5	2	0	2	0	0	1
163	11	3	0	2	1	3	2
164	5	2	0	2	0	0	1
165	9	3	1	4	0	0	1
Total	837	363	22	237	42	21	152

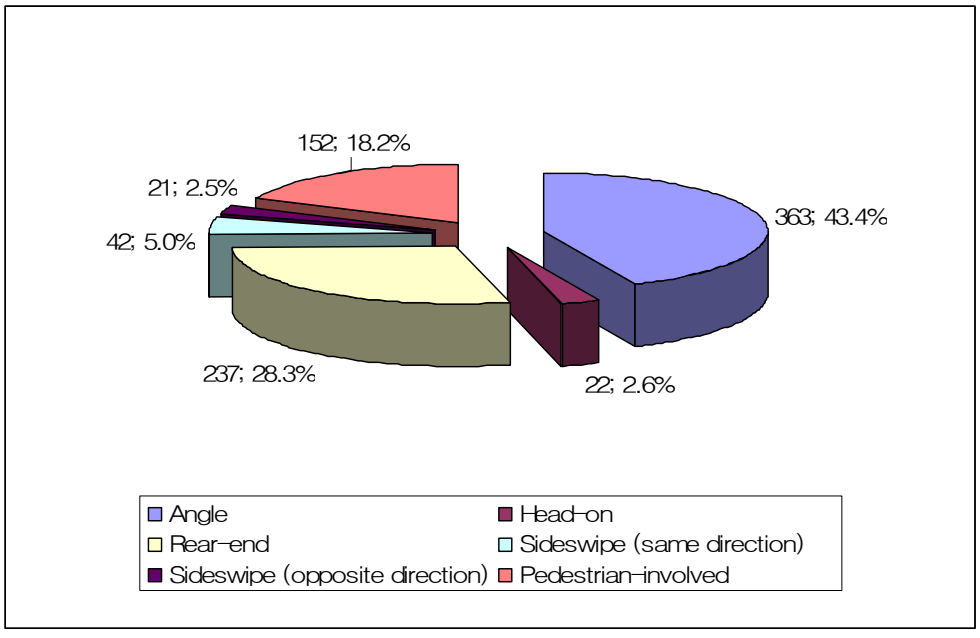

Figure A.1. Constitution of crash data

Table A.2. The number of intersections according to the number of total crashes

Number of crashes	Number of intersections		
	Non-signal	Signal	Total
0	17 (10.30)	6 (3.64)	23 (13.94)
1	26 (15.76)	4 (2.42)	30 (18.18)
2	23 (13.94)	4 (2.42)	27 (16.36)
3	11 (6.67)	1 (0.61)	12 (7.27)
4	11 (6.67)	1 (0.61)	12 (7.27)
5	7 (4.24)	6 (3.64)	13 (7.88)
6	5 (3.03)	3 (1.82)	8 (4.85)
7	2 (1.21)	2 (1.21)	4 (2.42)
8	2 (1.21)	0 (0.00)	2 (1.21)
9	3 (1.82)	1 (0.61)	4 (2.42)
10	4 (2.42)	4 (2.42)	8 (4.85)
11	2 (1.21)	3 (1.82)	5 (3.03)
12	1 (0.61)	3 (1.82)	4 (2.42)
13	0 (0.00)	1 (0.61)	1 (0.61)
14	0 (0.00)	1 (0.61)	1 (0.61)
17	0 (0.00)	1 (0.61)	1 (0.61)
18	0 (0.00)	1 (0.61)	1 (0.61)
19	0 (0.00)	1 (0.61)	1 (0.61)
20	0 (0.00)	2 (1.21)	2 (1.21)
21	0 (0.00)	1 (0.61)	1 (0.61)
23	0 (0.00)	2 (1.21)	2 (1.21)
25	0 (0.00)	1 (0.61)	1 (0.61)
27	0 (0.00)	1 (0.61)	1 (0.61)
53	0 (0.00)	1 (0.61)	1 (0.61)
Total	114	51	165

Note: The percentages are in parentheses.

Table A.3. The number of intersections according to the number of angle crashes

Number of crashes	Number of intersections		
	Non-signal	Signal	Total
0	46 (27.88)	10 (6.06)	56 (33.94)
1	36 (21.82)	10 (6.06)	46 (27.88)
2	9 (5.45)	5 (3.03)	14 (8.48)
3	12 (7.27)	6 (3.64)	18 (10.91)
4	4 (2.42)	3 (1.82)	7 (4.24)
5	2 (1.21)	2 (1.21)	4 (2.42)
6	0 (0.00)	5 (3.03)	5 (3.03)
7	4 (2.42)	1 (0.61)	5 (3.03)
8	1 (0.61)	2 (1.21)	3 (1.82)
9	0 (0.00)	1 (0.61)	1 (0.61)
10	0 (0.00)	2 (1.21)	2 (1.21)
11	0 (0.00)	1 (0.61)	1 (0.61)
12	0 (0.00)	1 (0.61)	1 (0.61)
13	0 (0.00)	1 (0.61)	1 (0.61)
33	0 (0.00)	1 (0.61)	1 (0.61)
Total	114	51	165

Note: The percentages are in parentheses.

Table A.4. The number of intersections according to the number of head-on crashes

Number of crashes	Number of intersections		
	Non-signal	Signal	Total
0	106 (64.24)	39 (23.64)	145 (87.88)
1	7 (4.24)	11 (6.67)	18 (10.91)
2	1 (0.61)	1 (0.61)	1 (1.21)
Total	114	51	165

Note: The percentages are in parentheses.

Table A.5. The number of intersections according to the number of rear-end crashes

Number of crashes	Number of intersections		
	Non-signal	Signal	Total
0	83 (50.30)	12 (7.27)	95 (57.58)
1	13 (7.88)	6 (3.64)	19 (11.52)
2	9 (5.45)	9 (5.45)	18 (10.90)
3	6 (3.64)	5 (3.03)	11 (6.67)
4	2 (1.21)	4 (2.42)	6 (3.64)
5	1 (0.61)	1 (0.61)	2 (1.21)
6	0 (0.00)	2 (1.21)	2 (1.21)
7	0 (0.00)	4 (2.42)	4 (2.42)
8	0 (0.00)	5 (3.03)	5 (3.03)
10	0 (0.00)	2 (1.21)	2 (1.21)
15	0 (0.00)	1 (0.61)	1 (0.61)
Total	114	51	165

Note: The percentages are in parentheses.

Table A.6. The number of intersections according to the number of sideswipe same direction crashes

Number of crashes	Number of intersections		
	Non-signal	Signal	Total
0	100 (60.61)	37 (22.42)	137 (83.03)
1	14 (8.48)	7 (4.24)	21 (12.73)
2	0 (0.00)	3 (1.82)	3 (1.82)
3	0 (0.00)	3 (1.82)	3 (1.82)
6	0 (0.00)	1 (0.61)	1 (0.61)
Total	114	51	165

Note: The percentages are in parentheses.

Table A.7. The number of intersections according to the number of sideswipe opposite direction crashes

Number of crashes	Number of intersections		
	Non-signal	Signal	Total
0	102 (61.82)	45 (27.27)	147 (89.09)
1	12 (7.27)	4 (2.42)	16 (9.70)
2	0 (0.00)	1 (0.61)	1 (0.61)
3	0 (0.00)	1 (0.61)	1 (0.61)
Total	114	51	165

Note: The percentages are in parentheses.

Table A.8. The number of intersections according to the number of pedestrian-involved crashes

Number of crashes	Number of intersections		
	Non-signal	Signal	Total
0	50 (30.30)	19 (11.52)	69 (41.82)
1	41 (24.85)	14 (8.48)	55 (33.33)
2	15 (9.09)	12 (7.27)	27 (16.36)
3	7 (4.24)	6 (3.64)	13 (7.88)
4	1 (0.61)	0 (0.00)	1 (0.61)
Total	114	51	165

Note: The percentages are in parentheses.

APPENDIX B: COMPUTER CODES (GAUSS)

B.1 Poisson Regression Model

```
new;

load dat[165,10]=c:\ml\poisson.txt;
y=dat[.,1]; x=ones(rows(dat),1)~dat[.,2:cols(dat)];

theta={1,1,1,1,1,1,1,1,1,1};
add=ones(rows(theta),1);
simp=theta~(theta+add.*eye(rows(theta)));

{theta,min}=amoeba(simp,&loglik,100000);

proc likeli(theta);
      local likeln, loglik, i, yhat;

      yhat=x*theta;
      likeln=zeros(rows(y),1);
      i=1;
      do while i <= rows(y);
            likeln[i]=-exp(yhat[i])+(yhat[i]*y[i])-ln(y[i]!);
            i=i+1;
      endo;
      retp(likeln);
endp;

proc loglik(theta);
retp(-sumc(likeli(theta)));
endp;

print "likel=" loglik(theta);
print "theta's:" theta;

/* Calculation of standard errors of betas */

lamda=exp(x*theta);
vartheta=zeros(cols(dat),cols(dat));
z=zeros(cols(dat),cols(dat));
k=1;
do while k <= rows(dat);
      z=-(lamda[k]*x[k,.]'*x[k,.]);          /* Hessian metrix */
      vartheta=vartheta+z;
```

```
      k=k+1;
endo;

se=diag(sqrt(inv(-vartheta)));                /* var(beta)=inv([-H]) */

print "standard errors of betas" se;
```

B.2 Negative Binomial Regression Model

```
new;

load dat[165,7]=c:\dokkang\dissertation\mlcodes\nbtest.txt;
y=dat[.,1];
x=ones(rows(dat),1)~dat[.,3:7];

theta={1,1,1,1,1,1,1};
add=ones(rows(theta),1);
simp=theta~(theta+add.*eye(rows(theta)));

{theta,min}=amoeba(simp,&loglik,100000);

proc likeli(theta);
      local loglik, i, yhat;

      yhat=x*theta[1:6];
      theta[7]=abs(theta[7]);
      loglik=zeros(rows(y),1);
      i=1;
      do while i <= rows(y);
            loglik[i]=ln(gamma((1/theta[7])+y[i]))-ln(gamma(1/theta[7]))-ln(y[i]!)
                  +(y[i]*yhat[i])-
                  ((1/theta[7])+y[i])*ln(1+theta[7]*exp(yhat[i]))+y[i]*ln(theta[7]);
            i=i+1;
      endo;
      retp(loglik);
endp;

proc loglik(theta);
retp(-sumc(likeli(theta)));
endp;

print "likel=" loglik(theta);
print "theta's:" theta;

/* Calculation of standard errors of betas */

lamda=exp(x*theta[1:6]);
vartheta=zeros(cols(x),cols(x));
z=zeros(cols(x),cols(x));
j=1;
do while j <= rows(dat);
```

```
      z=-(((1+theta[7]*y[j])*lamda[j])/(1+(theta[7]*lamda[j]))^2)*(x[j,.]'*x[j,.]);
          /* Hessian metrix */
      vartheta=vartheta+z;
      j=j+1;
endo;

se=diag(sqrt(inv(-vartheta)));                  /* var(beta)=inv([-H]) */
print "standard errors of betas" se;

/* Calculation of standard errors of alpha (overdispersion parameter) */

varalpha=zeros(rows(dat),1);
i=1;
do while i <= rows(dat);
      varalpha[i]=(2/theta[7]^3)*digamma(y[i]+(1/theta[7]))
                   +(1/theta[7]^4)*trigamma(y[i]+(1/theta[7]))
                    -(2/theta[7]^3)*digamma(1/theta[7])-
                   (1/theta[7]^4)*trigamma(1/theta[7])
                   -(y[i]/theta[7]^2)-(2/theta[7]^3)*ln(1+theta[7]*lamda[i])
                   +((2*lamda[i])/(theta[7]^2*(1+theta[7]*lamda[i])))
                   +(((y[i]+(1/theta[7]))*lamda[i]^2)/(1+theta[7]*lamda[i])^2);
      i=i+1;
endo;

varalpa=inv(-sumc(varalpha));
sealpha=sqrt(varalpa);

print "standard errors of alpha" sealpha;
```

B.3 Binary Logistic Regression Model

```
new;

load dat[155,5]=c:\ml\logit.txt;
y=dat[.,1]; x=ones(rows(dat),1)~dat[.,2:5];

theta={1,1,1,1,1};
add=ones(rows(theta),1);
simp=theta~(theta+add.*eye(rows(theta)));

{theta,min}=amoeba(simp,&loglik,100000);

proc likeli(theta);
      local loglik, i, yhat;

      yhat=x*theta[1:5];
      loglik=zeros(rows(y),1);
      i=1;
      do while i <= rows(y);
            loglik[i]=y[i]*yhat[i]-ln(1+exp(yhat[i]));
            i=i+1;
      endo;
      retp(loglik);
endp;

proc loglik(theta);
retp(-sumc(likeli(theta)));
endp;

print "likel=" loglik(theta);
print "theta's:" theta;

/* Calculation of standard errors of betas */

lamda=exp(x*theta[1:5]);
vartheta=zeros(cols(x),cols(x));
z=zeros(cols(x),cols(x));
j=1;
do while j <= rows(dat);
      z=-(lamda[j]/(1+lamda[j])^2)*(x[j,.]'*x[j,.]);          /* Hessian metrix */
      vartheta=vartheta+z;
      j=j+1;
endo;
```

```
se=diag(sqrt(inv(-vartheta)));                  /* var(beta)=inv([-H]) */
print "standard errors of betas" se;

/* Calculation of standard errors of betas using a Hessian command */

proc g(theta);
      local vartheta, i, lamda1, lamda2;

      lamda1=x*theta[1:5];
      lamda2=exp(x*theta[1:5]);
      vartheta=zeros(rows(dat),1);
      i=1;
      do while i <= rows(dat);
            vartheta[i]=y[i]*lamda1[i]-ln(1+lamda2[i]);
            i=i+1;
      endo;
      retp(vartheta);
endp;

b0=theta[1:5];
h = hessp(&g,b0);

var=inv(-h);
print "Variance-Covariance Matirx" var;

se=diag(sqrt(inv(-h)));                  /* var(beta)=inv([-H]) */

print "standard errors of betas" se;
```

B.4 Seemingly Unrelated Negative Binomial Model (3 equations)

```
new;

load dat[165,14]=c:\dokkang\dissertation\mlcodes\finalsunb.txt;
y1=dat[.,2];       /* Anlge accidents */
y2=dat[.,3];       /* Head-on accidents */
y3=dat[.,4];       /* Rear-end accidents */
y4=dat[.,5];       /* Sideswipe same direction accidents */
y5=dat[.,6];       /* Pedestrian-involved accidents */
x1=ones(rows(dat),1)~dat[.,7 8 12:14];          /* ind. variables for angle crashes */
x2=ones(rows(dat),1)~dat[.,7 8];                /* ind. variables for head-on crashes */
x3=ones(rows(dat),1)~dat[.,7:9 11 13 14];       /* ind. variables for rear-end crashes */
x4=ones(rows(dat),1)~dat[.,9 12 13];            /* ind. variables for sideswipe crashes */
x5=ones(rows(dat),1)~dat[.,8 10 14];            /* ind. variables for pedestrian-involved */
x=x1~x2~x3~x4~x5;

theta={1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1};
add=ones(rows(theta),1);
simp=theta~(theta+add.*eye(rows(theta)));

{theta,min}=amoeba(simp,&loglik,100000);

proc likeli(theta);
      local loglik, i, y1hat, y3hat, y4hat;

      y1hat=x1*theta[1:6];
      y3hat=x3*theta[7:13];
      y4hat=x4*theta[14:17];
      theta[18]=abs(theta[18]);
      loglik=zeros(rows(dat),1);
      i=1;
      do while i <= rows(dat);
            loglik[i]=(y1[i]*y1hat[i]+y3[i]*y3hat[i]+y4[i]*y4hat[i])-ln(y1[i]!)-ln(y3[i]!)-
                  ln(y4[i]!)+ln(gamma((1/theta[18])+y1[i]+y3[i]+y4[i]))-
                  ln(gamma(1/theta[18]))+((1/theta[18])*ln(1/theta[18]))-
                  ((y1[i]+y3[i]+y4[i]+(1/theta[18]))*
                  ln(exp(y1hat[i])+exp(y3hat[i])+exp(y4hat[i])+(1/theta[18])));
            i=i+1;
      endo;
      retp(loglik);
endp;
```

```
proc loglik(theta);
retp(-sumc(likeli(theta)));
endp;

print "likel=" loglik(theta);
print "theta's:" theta;

/* Calculation of standard errors of betas */

lamda1=exp(x1*theta[1:6]);
lamda3=exp(x3*theta[7:13]);
lamda4=exp(x4*theta[14:17]);
lamda=lamda1+lamda3+lamda4;
w=theta[18];
denom=lamda+(1/w);
y=y1+y3+y4;
a=y+(1/w);

vartheta1=zeros(cols(x1),cols(x1));
vartheta3=zeros(cols(x3),cols(x3));
vartheta4=zeros(cols(x4),cols(x4));
z1=zeros(cols(x1),cols(x1));
z3=zeros(cols(x3),cols(x3));
z4=zeros(cols(x4),cols(x4));

j=1;
do while j <= rows(dat);
      z1=-a[j]*(((lamda1[j]/denom[j])-(lamda1[j]/denom[j])^2)*(x1[j,.]'*x1[j,.]));
      z3=-a[j]*(((lamda3[j]/denom[j])-(lamda3[j]/denom[j])^2)*(x3[j,.]'*x3[j,.]));
      z4=-a[j]*(((lamda4[j]/denom[j])-(lamda4[j]/denom[j])^2)*(x4[j,.]'*x4[j,.]));
      vartheta1=vartheta1+z1;
      vartheta3=vartheta3+z3;
      vartheta4=vartheta4+z4;
      j=j+1;
endo;

se1=diag(sqrt(inv(-vartheta1)));          /* var(beta)=inv([-H]) */
se3=diag(sqrt(inv(-vartheta3)));          /* var(beta)=inv([-H]) */
se4=diag(sqrt(inv(-vartheta4)));          /* var(beta)=inv([-H]) */

print "standard errors of betas for x1" se1;
print "standard errors of betas for x3" se3;
print "standard errors of betas for x4" se4;
```

```
/* Calculation of standard errors of alpha (overdispersion parameter) */

varalpha=zeros(rows(dat),1);
i=1;
do while i <= rows(dat);
      varalpha[i]=(2/w^3)*digamma(y[i]+(1/w))+(1/w^4)*trigamma(y[i]+(1/w))
                        -(2/w^3)*digamma(1/w)-(1/w^4)*trigamma(1/w)
                        +(2/w^3)*ln(1/w)+(3/w^3)-(2/w^3)*ln(lamda[i]+(1/w))
                        -(2/w^4)*(1/(lamda[i]+(1/w)))
                        -(y[i]+(1/w))*((2/(w^3*(lamda[i]+(1/w))))-
(1/(w^4*(lamda[i]+(1/w))^2)));
      i=i+1;
endo;

varalpa=inv(-sumc(varalpha));
sealpha=sqrt(varalpa);

print "standard errors of alpha" sealpha;

t1=theta[1:6]./se1;
t3=theta[7:13]./se3;
t4=theta[14:17]./se4;
talpha=theta[18]/sealpha;

print "t-statistics for x1" t1;
print "t-statistics for x3" t3;
print "t-statistics for x4" t4;
print "t-statistics for alpha" talpha;

/* Calculation of p-value */

pvalue1=2*cdftc(abs(t1),rows(dat)-rows(theta[1:6]));
pvalue3=2*cdftc(abs(t3),rows(dat)-rows(theta[7:13]));
pvalue4=2*cdftc(abs(t4),rows(dat)-rows(theta[14:17]));

print "p-value for x1" pvalue1;
print "p-value for x3" pvalue3;
print "p-value for x4" pvalue4;

/* Variance-Covariance matrix */

u1=y1-lamda1;
u2=y3-lamda3;
u3=y4-lamda4;
```

```
sigma11=u1'u1;
sigma22=u2'u2;
sigma33=u3'u3;
sigma21=u2'u1;
sigma31=u3'u1;
sigma32=u3'u2;

/* Lagrange Multiplier Statistics */

r21=sigma21/(sigma22*sigma11)^0.5;
r31=sigma31/(sigma33*sigma11)^0.5;
r32=sigma32/(sigma33*sigma22)^0.5;

LM=rows(dat)*(r21^2+r31^2+r32^2);

print "LM Statistics=" LM;
```

REFERENCES

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

Abu-Eisheh, S. A. and F. L. Mannering (2002), "Forecasting Automobile Demand for Economies in Transition: A Dynamic Simultaneous-Equation System Approach", *Transportation Planning and Technology*, Vol. 25, No. 4, pp. 311 – 331.

Agent, K. R. (1983), "Warrants for Left-Turn Lanes", *Transportation Quarterly*, Vol. 37, No. 1, pp. 99 – 114.

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

Ben-Yakov, Y. and J. Craus (1980), "Warrants for Left-Turn Lanes on Two-Lane Roads", *Publication No. 80-28*, Transportation Research Institute, Technion-Israel Institute of Technology.

Bhat, C. R. and F. S. Koppelman (1993), "An Endogenous Switching Simultaneous Equation System of Employment, Income, and Car Ownership", *Transportation Research Part A*, Vol. 27A, No, 6, pp. 447 – 459.

Breusch, T. and A. Pagan (1980), "The LM Test and Its Applications to Model Specification in Econometrics", *Review of Economic Studies* 47, pp. 239 – 254.

Cameron, A. C. and P. K. Trivedi (1998), "*Regression Analysis of Count Data*", Econometric Society Monograph No. 30, Cambridge University Press.

Chayanan, S., M. Nebergall, V. Shankar, N. Juvva, and Y. Ouyang (2003), "Interaction between the roadway and roadside–an econometric analysis of design and environmental factors affecting segment accident rates", *WA-RD 562.1*, Seattle, Washington

Chin, H. C. and M. A. Quddus (2003), "Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections", *Accident Analysis and Prevention*, Vol. 35, No. 2, pp. 253 – 259.

Clarke, W. and P. Loeb (2005), "The determinants of train fatalities: keeping the model on track", *Transportation Research Part E*, Vol. 41, No. 2, pp. 145 – 158.

Crash Analysis, Statistics & Information Notebook 1996 – 2003 (2004), Department of Motor Vehicle Safety, Georgia Department of Transportation.

Craus, J. and D. Mahalel (1986), "Analysis of Operations and Safety Characteristics of Left-Turn Lanes", *ITE Journal*, Vol. 56, No. 7, pp. 34 – 39.

Cummins, J. D., R. D. Philips, and M. A. Weiss (2001), "The Incentive Effects of No-Fault Automobile Insurance", *Journal of Law and Economics*, Vol. XLIV, pp. 427 – 464.

Dale, C. W. (1973), "Cost-Effectiveness of Safety Improvement Project", Federal Highway Administration, Washington, D.C.

Derrig, R. A., M. Segui-Gomez, and A. Abtani (2000), "The Effect of Seat Belt Usage Rates on the Number of Motor Vehicle-Related Fatalities", *Proceedings of the Risk Theory Society*, Baltimore, MD.

Eisenberg, D. (2003), "Evaluating the Effectiveness of Policies Related to Drunk Driving", *Journal of Policy Analysis and Management*, Vol. 22, No. 2, pp. 249 – 274.

Foody, T. J. and W. C. Richardson (1973), "Evaulation of Left Turn Lanes as a Traffic Control Device", Ohio Department of Transportation.

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

Greene, W. H. (2003), "*Econometric Analysis: 5th Edition*", Prentice Hall.

Greibe, P. (2003), "Accident Prediction Models for Urban Roads", *Accident Analysis and Prevention*, Vol. 35, No. 2, pp. 273 – 285.

Greiwe, R. R. (1986), "Intersection Management Techniques for the Left-Turning Vehicle", *ITE Journal*, Vol. 56, No. 6, pp. 23 – 28.

Harwood, D.W., F. M. Council, E. Hauer, W. E. Hughes, and A. Vogt (2000), "Prediction of the Expected Safety Performance of Rural Two-Lane Highways", *FHWA-RD-99-207*, Federal Highway Administration.

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

Hauer, E., JCN Ng, and J. Lovell (1988), "Estimation of Safety at Signalized Intersections (with discussions and closure)", *Transportation Research Record* 1185, National Academy of Sciences, Washington, D.C., pp 48 – 61.

Hausman, J. A, B. H. Hall, and Z. Griliches (1984), "Econometric Models for Count Data with An Application to the Patents-R&D Relationship", *Econometrica* 52, pp. 909 – 938.

Highway Design Manual: Appedix F. (2003), Oregon Department of Transportation.

Joshua, S. and N. Garber (1990), "Estimating Truck Accident Rate and Involvement Using Linear and Poisson Regression Models", *Transportation Planning and Technology 15*, pp. 41 – 58.

Jovanis, P., and H. Chang (1986), "Modeling the Relationship of Accidents to Miles Traveled", *Transportation Research Record 1068*, National Academy of Sciences, Washington, D.C., pp. 42 – 51.

Jung, R. C. and R. Winkelmann (1993), "Two Aspects of Labor Mobility: A Bivariate Poisson Regression Approach", *Empirical Economics 18*, pp. 543 – 556.

Kim, D., S. P. Washington, and J. Oh (2006), "Modeling Crash Types: New Insights into the Effects of Covariates on Crashes at Rural Intersections", *Journal of Transportation Engineering*, Vol. 132, No. 4, pp. 282 – 292.

Lacy, J. D. (1972), "Traffic Operations Program to Increase Capacity and Safety", *Traffic Quarterly*, Vol. 26, No. 3, pp. 327 – 340.

Ladron de Guevara, F., S. P. Washington, and J. Oh (2004), "Forecasting Crashes at the Planning Level: A Simultaneous Negative Binomial Crash Model Applied in Tucson, Arizona", *Transportation Research Record 1897*, National Academy of Sciences, Washington, D.C., pp 191 – 199.

LaFrance, J. T. (1993), "Weak Separability in Applied Welfare Analysis", *American Journal of Agricultural Economics*, Vol. 75, No. 3, pp. 770 – 775.

Lawless, J. F. (1987), "Negative Binomial and Mixed Poisson Regression", *The Canadian Journal of Statistics*, Vol. 15, No. 3, pp. 209 – 225.

Lord, D., S. Washington, and J. Ivan (2005), "Poisson, Poisson-Gamma, and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory", *Accident Analysis and Prevention*, Vol. 37, No. 1, pp. 35 – 46.

Lyon, C., J. Oh, B. Persaud, S. P. Washington, and J. Bared (2003), "Empirical Investigation of Interactive Highway Safety Design Model Accident Prediction Algorithm: Rural Intersections", *Transportation Research Record 1840*, National Academy of Sciences, Washington, D.C., pp 78 – 86.

Maze, T. H., J. L. Henderson, and S. Sankar (1994), "Impacts of Safety of Left-Turn Treatment at High-Speed Signalized Intersections", *Iowa Highway Research Board Project HR-347*, Iowa State University.

McCoy, P. T. and M. S. Malone (1989), "Safety Effects of Left-Turn Lanes on Urban Four-Lane Roadways", *Transportation Research Record 1239*, National Academy of Sciences, Washington, D.C., pp. 17 – 22.

McFarland, W. F., L. I. Griffin, J. B. Rollins, W. R. Stockton, D. T. Phillips, and C. L. Dudek (1979), "Assessment of Techniques for Cost-Effectiveness of Highway Accident Countermeasures", *FHWA-RD-79-53*, Federal Highway Association.

Miaou, S., P. Hu, T. Wright, A. Rathi, and S. Davis (1992), "Relationship Between Truck Accidents and Highway Geometric Design: A Poisson Regression Approach", *Transportation Research Record 1376*, National Academy of Sciences, Washington, D.C., pp. 10 – 18.

Mitra, S., H. C. Chin, and M. A. Quddus (2002), "Study of Intersection Accidents by Maneuver Type", *Transportation Research Record 1784*, National Academy of Sciences, Washington, D.C., pp. 43 – 50.

Oh, J., C. Lyon, S. P. Washington, B. Persaud, and J. Bared (2003), "Validation of the FHWA Crash Models for Rural Intersections: Lessons Learned", *Transportation Research Record 1840*, National Academy of Sciences, Washington, D.C., pp 41 – 49.

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

Persaud, B. and T. Nguyen (1998), "Disaggregate Safety Performance Models for Signalized Intersection on Ontario Provincial Roads", *Transportation Research Record 1635*, National Academy of Sciences, Washington D. C., pp. 113 – 120.

Poch, M. and F. L. Mannering (1996), "Negative Binomial Analysis of Intersection Accident Frequencies", *Presented at the 75th annual meeting of the Transportation Research Board*.

Ramanathan, R. (2002), "*Introductory Econometrics with Applications: 5th Edition*", Harcourt College Publisher.

Ryan, G. A, M. Legge, and D. Rosman (1998), "Age Related Changes in Drivers' Crash Risk and Crash Type", *Accident Analysis and Prevention*, Vol. 30, No. 3, pp. 379 – 387.

Shankar, V. and F. L. Mannering (1998), "Modeling the Endogeneity of Lane-Mean Speeds and Lane-Speed Deviations: A Structural Equations Approach", *Transportation Research Part A*, Vol. 32, No, 5, pp. 311 – 322.

Shankar, V., F. M. Mannering, and W. Barfield (1995), "Effect of Roadway Geometric and Environmental Factors on Rural Freeway Accident Frequencies", *Accident Analysis and Prevention*, Vol. 27, No. 3, pp. 371 – 389.

Stutts, J., W. Hunter, and W. Pein (1996), "Pedestrian-Vehicle Crash Types: An Update", *Transportation Research Record 1538*, National Academy of Sciences, Washington D. C., pp. 68 – 74.

Tamburri, T. N. and C. G. Hammer (1968), "Evaluation of Minor Improvements: (Part 5) Left-Turn Channelization", State of California, Division of Highways, Traffic Department.

Traffic Safety Facts (2003), National Highway Traffic Safety Administration, U.S Department of Transportation.

Vogt, A. and J. Bared (1998), "Accident Prediction Models for Two-Lane Rural Roads: Segments and Intersections", *FHWA-RD-98-133*, Washington, D.C.

Vogt, A. (1999), "Crash Models for Rural Intersections: Four-Lane by Two-Lane Stop-Controlled and Two-Lane by Two-Lane Signalized", *FHWA-RD-99-128*, Washington, D.C.

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

Wilson, J. C. et al. (1967), "Improved Street Utilization Through Traffic Engineering", *Highway Research Board Special Report No. 93*, Highway Research Board.

Windmeijer, F. A. G. and J. M. C. Santos Silva (1997), "Endogeneity in Count Data Models: An Application to Demand for Health Care", *Journal of Applied Econometrics*, Vol. 12, No. 3, pp. 281 – 294.

Winkelmann, R. (2003), "*Econometric Analysis of Count Data*", Springer, Germany.

Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, Vol. 57, No. 298, pp. 348 – 368.