5-1-2014

Simultaneous Equation Modeling for Crash Rate of Freeway Segments

Anthony Ramos
University of Nevada, Las Vegas, ramosa27@unlv.nevada.edu

SIMULTANEOUS EQUATION MODELING FOR CRASH RATE OF FREEWAY SEGMENTS

By

Anthony Ramos

Bachelor of Science in Engineering

University of Nevada, Las Vegas

2012

A thesis submitted for in partial fulfillment

of the requirements for the

Master of Science in Engineering – Civil and Environmental Engineering

Department of Civil and Environmental Engineering and Construction

Howard R. Hughes College of Engineering

The Graduate College

University of Nevada, Las Vegas

May 2014

THE GRADUATE COLLEGE

We recommend the thesis prepared under our supervision by

Anthony Ramos

entitled

Simultaneous Equation Modeling for Crash Rate of Freeway Segments

is approved in partial fulfillment of the requirements for the degree of

Master of Science in Engineering -- Civil and Environmental Engineering

Department of Civil and Environmental Engineering and Construction

Hualiang Teng, Ph.D., Committee Chair

Mohamed Kaseko, Ph.D., Committee Member

Moses Karakouzian, Ph.D., Committee Member

Kaushik Ghosh, Ph.D., Graduate College Representative

Kathryn Hausbeck Korgan, Ph.D., Interim Dean of the Graduate College

May 2014

Abstract

In 2010, the total number of reported traffic crashes in the state of Nevada was 51,664 of which 235 (0.6%) resulted in one or more deaths (Nevada Department of Transportation, 2012). The state's "Zero Fatalities" traffic safety campaign aims to reduce the rate even further. Out of the total number of crashes, Clark County (includes Las Vegas) accounted for 78.89%. This study examines safety improvement by developing advanced crash prediction models. The system of crash prediction equations consider geometric conditions and traffic volume using simultaneous equation modeling (SEM). The models are based on geometric characteristics and traffic volume data collected from Las Vegas freeway systems related to crash data provided by Nevada Department of Transportation (NDOT). All data characterizes the year 2010, chosen for the least amount of observed roadway construction zones when compared to other years.

The system of crash rate prediction equations represents connected freeway segment types. The types, defined by entrance (EN) and exit (EX) ramp-pair combinations, are estimated simultaneously instead of developing separate linear regression models. By modeling EX-EN segments connected to EN-EX using SEM, the relationship of crash rate in the EN-EX effects crash rate in EX-EN. Most EN-EX segments are considered weaving sections (lengths shorter than 2,500 feet) contributing to congestion.

The increase of significant model parameters is apparent when comparing SEM to single equation multiple linear regression. The additional information obtained confirms the correlation between crash rate prediction residuals exists between connecting EX-EN and EN-EX segment types and supports the existence of unobserved variables. SEM

method of “three-stage least squares” modeling allows for freeway segments containing different characteristics to be modeled together, i.e., presence of auxiliary lane in EN-EX can be modeled with EX-EN segments. Instrumental variables replace the missing auxiliary lane variable in EX-EN segments resulting in a system of regression equations for crash prediction. The models can be used for connecting paired segments of EX-EN and downstream EN-EX or a connecting three-segment semi-corridor of EN-EX, downstream EX-EN and downstream EN-EX.

Acknowledgments

I express many thanks to Dr. Hualiang (Harry) Teng for the opportunity of conducting research under his direction. As my academic advisor, Dr. Teng introduced me to research while providing a work study which allowed for the continuation of my educational experience. As my Advisory Committee Chair, he guided me through this thesis and other research projects with the intent of inspiring innovation and developing future topics. I am so very thankful for all his dedication to my academics, his understanding and patience in being my advisor.

I would also like to thank my thesis committee members, Dr. Kuashik Ghosh, Dr. Mohamed Kaseko, and Dr. Moses Karakouzian. The feedback and support I have received for my thesis was more than generous and has allowed me to gain the confidence in moving forward with my thesis topic.

Of course, none of this would be possible without the funding of the Nevada Department of Transportation. Their support for the previous research project for which this study is based on as well as other projects I have had the privilege to be a part of, has opened many doors of opportunity. I am also very grateful to the University of Las Vegas, Nevada for their finical support fueling my graduate education.

I am pleased to also thank my research colleague Eneliko Mulokozi, who aided in my understanding of research. His dedication to research projects is unmatched, even with a meddling partner such as myself.

Dedication

To my loving wife, Sandra, who believes in me even when I don't.

To my adorable daughter, Abigail, who even though she does not know it yet (just turned one) motives me to aspire for more.

And to my mother, who supported me the entire way.

To my entire extended family, I love you and thanks for putting up with my absence during midterms and finals week (that's pretty much the whole year).

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

In 2010, the total number of reported traffic crashes in the state of Nevada was 51,664 of which 235 (0.6%) resulted in one or more deaths (Nevada Department of Transportation, 2012). The fatality crash rate of 1.06 per 100 mvmt (1.06 fatal crashes occur for every 100 million vehicle miles traveled on the state's roadways) is down from previous years and on a par with the national rate. The state's "Zero Fatalities" traffic safety campaign aims to reduce the rate even further. Out of the total number of crashes, Clark County (includes Las Vegas) accounted for 78.89%.

1.1 – Geometric Design Issues

For roadway planners, designers and traffic operation engineers, safety improvements for urban freeways pose many challenges including spatial limitations. This study examines safety improvement by developing advanced crash rate prediction models. The models consider geometric conditions and traffic volume to predict crash rate. The variables used are based on geometric characteristics and traffic volume data collected from Las Vegas freeway systems related to crash data provided by Nevada Department of Transportation (NDOT).

In a previous study, Teng et al. (2013) observed the spatial correlations of Las Vegas freeway crash data and geometric conditions with the use of ArcGIS. The study considered individual freeway segments defined by American Association of State Highway and Transportation Officials (AASHTO) design standards of ramp-pair combinations seen in Figure 1. The individual segments, examples seen in Figure 2, were analyzed for their homogeneous characteristics such as number of lanes, etc. The segments were then digitized with the use of ArcGIS to geographically enclose the number of 2010 crashes in each segment. All observations were modeled for crash rate prediction using Multiple Linear Regression (MLR). As the study progressed, researchers observed possible correlations between successive freeway segments. Also, segments containing the most crashes were located near the various interchanges of the study area. The same study area (I-15, I-215 and US95 within the Las Vegas Valley) is examined in this thesis. The aforementioned possibility of correlations between successive segment types of exit-entrance (EX-EN) and entrance-exit (EN-EX) is explored by using Simultaneous Equation Modeling (SEM) to predict crashes based on varying geometric conditions and traffic volumes.

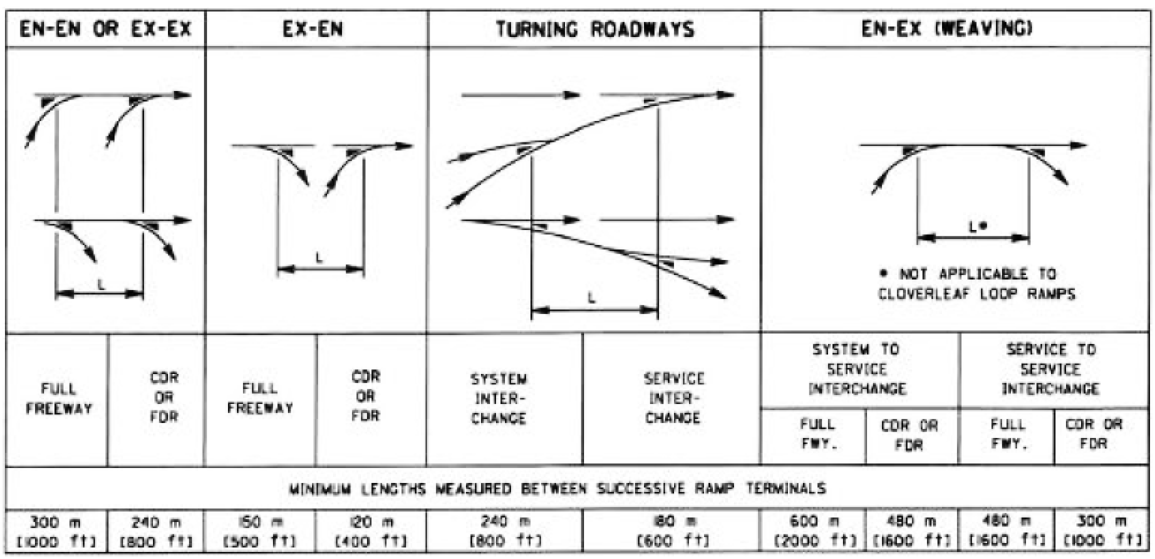

Figure 1 - Ramp-pair combinations define studied freeway segment types, EX-EN and EN-EX (AASHTO, 2001)

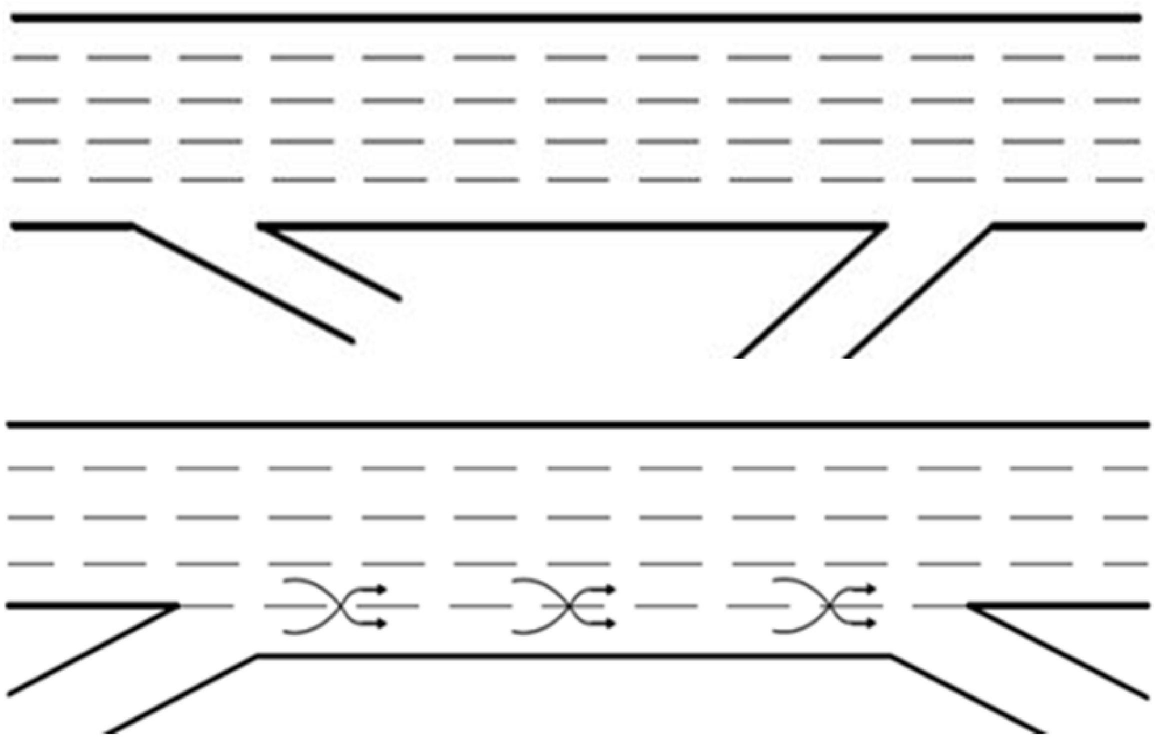

Figure 2 - EX-EN segment type (above). EN-EX segment type showing weaving movements (below)

1.2 – Weaving Sections

Identifying problematic weaving segments to assess risk factors in the Las Vegas freeway system presents the issue of differences among interchange characteristics. A generalized approach to analyzing crash data related to geometric design features is often practiced, especially when operating and designing interchanges and ramp terminals of various distinct interchanges. In the design stages, only the safety concerns of the interchange improvement is considered and the connecting freeway segments are overlooked. Also, Annual Average Daily Traffic (AADT), which is a measure of traffic volume, is non-specific when analyzing the unique geometric layout of each interchange.

Correlations of crash frequency to geometric features differ over peak and non-peak volume.

Unintentional freeway safety hazards of the Las Vegas freeways are seen in interchange spacing for the following reasons:

1. The Las Vegas Valley is contained within Clark County boundaries
2. Clark County follows standard township and sectioning guidelines
3. The townships consist of many one square-mile sections (usually 36 total)
4. Major arterials make up the boarders of these sections
5. Therefore, the arterials interchanges are about one mile apart

Ramp spacing must be considered along with interchange spacing. The design purpose of ramp type selection also dictates the length. The resulting ramp spacing is much shorter than interchange spacing. Short interchange spacing causes weaving sections in EN-EX segment types. Weaving contributes to traffic delay due to the short maneuvering lengths drivers must negotiate along with other competing drivers.

Any EN-EX ramp-pair freeway section over 2,500 feet is not considered a weaving section according to the Highway Capacity Manual 2010 (Roess et al. 2011). The AASHTO's "Green Book" (A Policy on Geometric Design of Highways and Streets) recommends, in Exhibit 10-68 (see Figure 1), minimum ramp spacing for five different terminal combinations including EN-EX (weaving segment). The recommend minimum distance is 2,000 feet between service (smaller features) interchanges or arterial interchanges. Many EN-EX segments are classified under the criteria of length less than 2,500 feet (45% out of the total EN-EX contained in the study area). Furthermore, the distance was established in the 1970s and was based on operational experience (Ray et al., 2011).

The short distance limiting driver maneuvering is known to cause turbulence in traffic flow (Roess et al., 2011). The turbulence effect stems from lane changing of drivers entering occurring at the same time drivers exit. Recent studies would suggest weaving sections are problematic in terms of crash frequency. Pulugurtha and Bhatt (2010) found crash frequency decreased when weaving section length increased. As well, they suggested the increase probably of crashes occurring in weaving section compared to all other freeway sections in Las Vegas.

In this thesis, crash frequency occurs more in the preceding EX-EN segment than in weaving sections when considering 2010 crash data. Traffic flow turbulence causes momentary speed reduction sending a shockwave effect upstream. As the shockwave moves past the influence of the weaving section, unsuspecting drivers fail to respond. The weaving segment is interrelated to crash frequency increases in the EX-EN segment upstream. Also, primary crashes which occur in weaving section produce shockwave speed reduction resulting in secondary crashes. These secondary crashes might not occur in the same weaving section. This further strengthens the basis to examine the adjoining segment types together.

1.3 – Simultaneous Equation Modeling

SEM allows for the analysis of an interrelated system of equations with independent (exogenous) variables, i.e., freeway segment characteristics, which differ over the two different segment types. In Teng et al. (2013), each segment type was modeled separately due to changing variables across successive freeway segments, e.g., EN-EX segments have presences of auxiliary lanes (dummy variable coded 0 or 1) and EX-EN do not (no 0 nor 1).

The endogenous variables (correlated to the error term or dependent) of crash rate for EX-EN and EN-EX segments can be used for regression in SEM. By the “three-stage least squares” (3SLS) method of SEM, the error structure of each equation, i.e., crash prediction residuals for EX-EN and EN-EX, must be correlated for both segment type equations. This procedure can be considered when crashes occur within the same study time period (1 year, 2010) even though each segment type presents different variables (Henningsen & Hamann, 2007). The resulting crash rate estimates are then regressed simultaneously to model crash rate of EX-EN and the downstream segment EN-EX, which for short lengths (< 2,500 feet) are known for undesirable vehicle weaving movements.

In MLR, the error terms from one observation to the next is assumed to be random. Therefore, comparing individual regression equations should result in zero covariance with the error terms proving they are independent of each other, i.e., they have no correlation. The contemporaneous correlation of the error terms across each equation needs to be determined for use of SEM. The contemporaneous correlation explains the relationship between the consecutive segment types, and how crash rate of EN-EX will affect the other in terms of safety.

Above all else, identification of the equations is required. The one-to-one relationship between the number of explanatory variables and parameters to be estimated ensures a system can be solved. More details of the identification process are in Chapter 3 Methodology.

1.4 – Study Objectives

This thesis includes extensive literature review done to ensure the proper use of 3SLS method of SEM. The results include analysis done with the data collection previously submitted to NDOT in Teng et al. (2013). However, further insight of the data is conveyed from the use of SEM. A system of crash prediction models includes one equation for each segment type. The correlation coefficients will be determined for the error terms of individual MLR equations of crash rate for each segment type. The results validate the use of SEM as a method of handling the heterogeneous, connecting freeway segment types. The connecting segment pairs and three-segment semi-corridors make up part of the urban freeway system influenced by interchanges. The resulting models incorporating geometric features and traffic volumes will aid in providing effective countermeasures to improve freeway safety.

Chapter 2 Literature Review

The literature review contained in this thesis focuses on traffic safety studies related to the following topics: freeway geometric features, econometric modeling techniques and simultaneous equation modeling in traffic studies. The chapter will lead to the methodology used for this study termed "three-stage least squares" (3SLS) and how the method is currently used in crash prediction. The significant relationship of geometric features and crash frequency of nearby freeway segments is emphasized. The importance of how the correlations validate the use of simultaneous equation modeling (SEM) is also emphasized.

The amount of studies focusing on urban freeway safety increased in recent years. The Highway Safety Manual (HSM), which is formulated from empirical data, is the preferred source for evaluating safety of roadway facilities. However, the HSM does not cover safety effects of freeway design functions. Space for freeway facilities decreases as any city grows. Lack of space stresses the consideration of all possible designs options to improve mobility. All of the aforementioned statements are cause for alarm when cogitating safety in freeway design.

Geometric design standards being used countrywide are congenital of empirical observation related to safety of design speeds. As those standards age, the need for reevaluation of design standards must keep up with concerns such as population grow and urban sprawl as well as innovative safety features established by the auto industry. Many studies were completed since the recent edition release of AASHTO's "Green Book" which is formulated from empirical results. Techniques in model estimation have branched into the econometric realm for more complex traffic prediction estimation resulting in more vigorous output on which to formulate design standards. Correspondingly, the use of Intelligent Transportation System (ITS) allows for more detailed database collection. A literature review was performed for this study to encompass these emerging techniques for crash frequency prediction for urban freeways.

2.1 – Highway Safety Manual

In order to circumvent intricate human behavioral concepts in experimental design, geometric features are commonly used for freeway safety studies. Identifying the key characteristics of freeway design attributed to crash frequency can produce significant prediction results without incorporating human factors. As seen in Figure 3, human factors alone account for 57% of all contributing factors of crashes according to Treat et al. (1979). Although roadway factors only covers 34% when combined with all other considered factors, the roadway is far more controllable than human behavior.

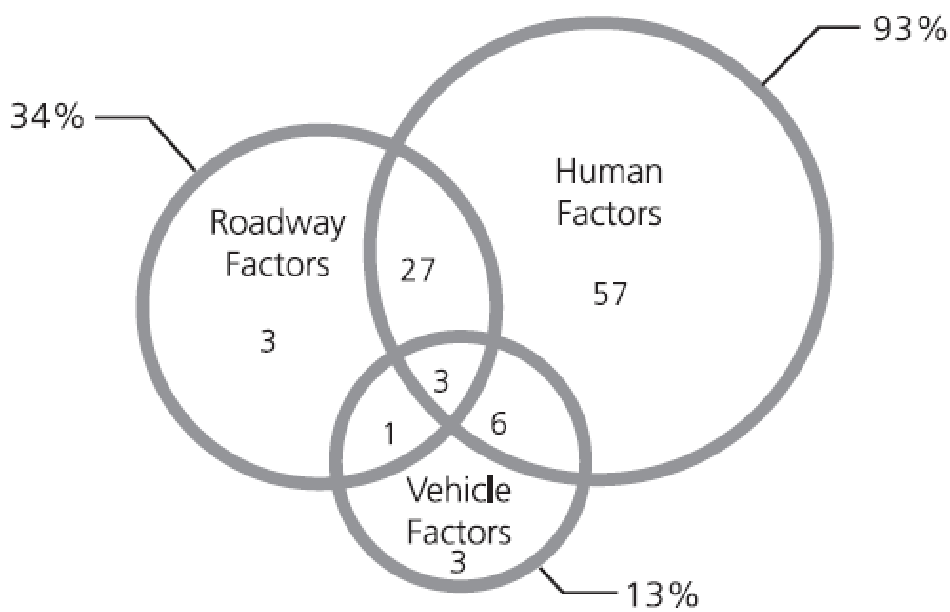

Figure 3 - Crash factors (AASHTO, 2010), (Treat et al. 1979)

HSM is a recently developed tool for engineers to assess safety issues of roadway design. The HSM offers step procedures on resolving issues through countermeasures. An outline of the process for selecting countermeasures for any specific site (except freeway facilities) is as follows:

1) Identify contributing factors at the site in question
2) Identify countermeasures to address the crash contributor
3) Conduct cost-benefit analysis for selected countermeasures

Most countermeasures are described in PART D of the manual along with the respective Crash Modification Factors (CMF). The CMF is derived from before-and-after studies. The results provided are used to understand the effects on crash frequency if the CMF is selected for that specific site. Some CMFs have a value over 1.0 which would indicate a rise in crash frequency and should not be considered for implementation of the site analyzed. For example, the installation of a signalized intersection where a stop was installed previously may increase rear-end crashes over preventing left-hand turn crashes.

The CMFs are then used with the crash predictive method describe in the HSM. The projected crash frequency of the analyzed facility is determined with historical data. Several years' worth of crash data is suggested in order to determine the "expected average crash frequency" per year. The results are used for alternative selection of countermeasures contained in the manual.

Utilization of the HSM to assess various facilities can be accompanied with the software, SafetyAnalyst, to preform simulation based modeling. However, the HSM does not include many CMFs to be used with various freeway facility types. Most CMFs listed for freeways are operational countermeasures such as crash and speed warning signs as well as providing sufficient lighting on freeways. As for geometric planning using CMFs, the manual is incomplete for now with the promise of CMFs for freeway geometric features in the next edition. This lack of factors related to freeway safety is due to the lack of studies done for freeway facilities.

2.2 – Geometric Features related to Crash Frequency

The HSM is comprised of the most relevant studies focused on relating geometric conditions and traffic safety. With the emergence of statistical software, simulation packages and ITS data, more studies have recently emerged dealing with freeways. Some of the studies considering freeway crash frequency and severity are Park et al. (2009), Golob et al. (2004), Ray et al. (2011), and Pilko et al. (2007). The researchers used geometric variables including vertical and horizontal alignment, number of through lanes and ramp spacing.

Park et al. (2009) defined freeway sections by curve radius and classified the sections by presence of ramps. Variables include number of lanes and median type which were related to crash frequency. Golob et al. (2004) derived severity prediction equations for weaving section types. Type A through C designates each probability equation in a multivariate probit model. Weaving type was the only geometric variable and crash frequency was found insignificant in their models. Ray et al. (2011) developed frequency models using ramp and mainline traffic volume along with interchange and ramp spacing to examine the tradeoff between freeway safety and adding a new interchange among existing interchanges. Pilko et al. (2007) modeled crash frequency by severity of freeway segments with variables such as interchange spacing, shoulder width and number of lanes.

Geedipally et al. (2012) suggested a reduction in high severity crashes on freeway segments with the increase of barrier presences portioned to either side of a freeway segment. Median barrier and shoulder barrier lengths are measured and divided by twice the length. Freeway segments with both median and shoulder barrier running alongside the entire segment would result in a ratio of 1.0. If a segment has median barrier along the entire length but no shoulder barrier, it would be given a ratio of 0.5. Ratios from 0.5 to 1.0 had shown a reduction of fatality crashes from 6.5% to 5.7%, respectively.

Haleem et al. 2013 demonstrated a two feet increase of shoulder width reduces fatal and injury crashes by 10%. However, this reduction was only for ramp influence areas of 0.3 miles upstream and downstream from painted gores. These studies are further evidence that urban freeway CMFs for the HSM are still in development.

Teng, et al. (2013) related shoulder width to crash frequency. The effects of narrow shoulder width had shown to increase crash frequency for Las Vegas freeways. Their crash prediction model also included the minimum number of weaving lanes. The presence of auxiliary lane insures at least two lanes are used in weaving movements. Weaving movements have long been a source of accidents due to the nature of competing lane changing.

Results from Zhang, et al. (2011) prove that designing for a two lane off-ramp without the lane change option can reduce fatal crashes by 0.2% (all other crashes by 3.6%) when compared to traditional parallel off-ramp design. By eliminating the option to exit at the painted gore, a two-lane off-ramp (seen in Figure 4) decreases the need for weaving movements just before the off-ramp. Figure 5 shows two-lane off-ramp with lane change option for comparison with Figure 4.

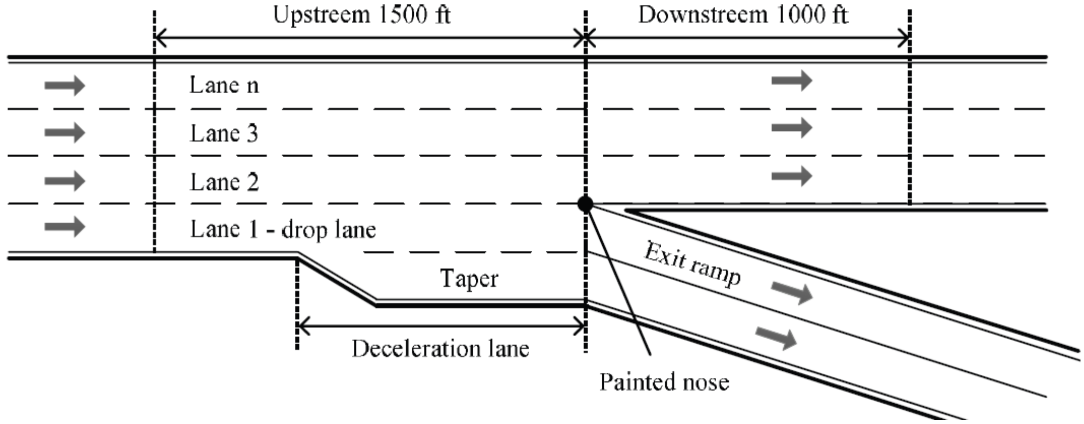

Figure 4 - Two-lane exit ramp without optional lane (Zhang, et al. 2011)

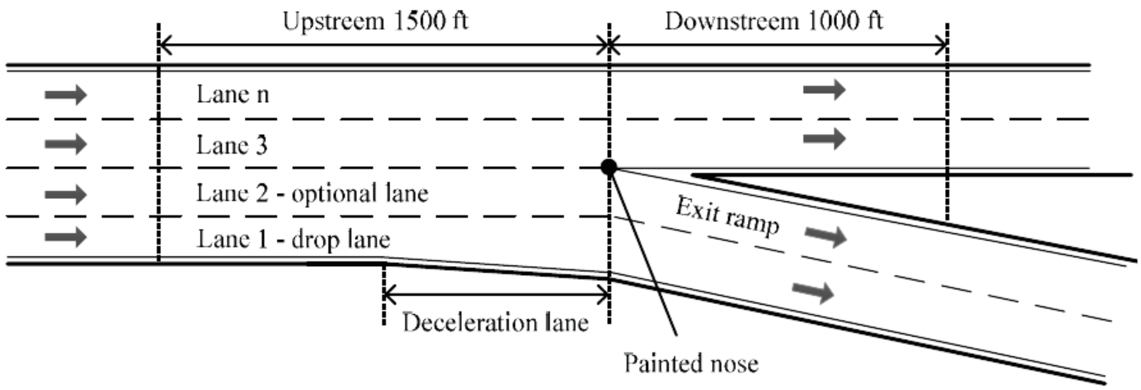

Figure 5 – Two-lane exit ramp with optional lane (Zhang, et al. 2011)

Park et al. (2009) related geometric design and safety to include ramp density and horizontal curves. The horizontal curves were recorded attributes such as number of lanes, median type, etc., found in rural and urban freeway segments to analyze with crash counts. Much like most recent studies, a negative binomial regression model was performed. The results clarified risk involved with the presences of exit and entrance ramps. The study was not able to capture the influence of weaving sections on crashes by considering ramp density in freeway sections.

Many geometric characteristics within a freeway segment can be considered when relating to crash occurrence. When pinpointing key regressors, focus should be placed on exactly the features exhibiting causality on crash locaion. Also, having too many or not enough variables can prove to be disadvantageous to the study. O'Cinneide (1998) pointed out that lane width, median width and shoulder width influence driver comfort which has been known to affect crash rate. Grade changes on freeway segments at 4% increased crash rate by 20% when compared to lower gradients. Las Vegas freeways contain only a few areas where grades are steep enough to influence vehicle acceleration and overall traffic flow.

When considering the installation of new interchanges between two existing interchanges, Pilko et al. (2007) conducted an experiment to factor geometric design in risk assessment models. Characteristics such as interchange spacing, shoulder widths and number of lanes in the freeway segment were just a few of the variables considered. The models showed that volume had a large sensitivity in freeway and ramp Annual Average Daily Traffic (AADT) when predicting fatal and injury crashes. Using AADT does not account for factors such as peak hour traffic flows, seasonal weather conditions and secondary crashes caused by initial crashes.

2.3 – Econometric Techniques in Traffic Studies

As any modeler will proclaim, not all data collected will result in inferential statistics especially when using classical regression techniques. Attending to assumption violations can be a daunting task when formulating significant results. Assumption violations include exogeneity, i.e., no correlation demonstrated between the explanatory variables and residuals, and heteroscedasticity, i.e., non-constant variance of error terms.

Other pitfalls include model selection when using mixtures of discrete and continuous data. Improper knowledge to detect these issues will lead to improper model results. Researches mustn't report results hastily if not all assumptions of classical regression are addressed.

The use of more complex modeling techniques in traffic engineering has increased as researchers embrace the use of econometrics. Golob (2003) lists the many studies already done using advance techniques of econometrics. Most of the studies deal with travel behavior but fundamentally the methods of structural equation modeling can be used for most other transportation related issues. Simultaneous equation modeling (SEM) is a form of structural equation modeling in which the interrelationship of dependent variables can be used to form a system of equations. SEM developed for more advanced econometric modeling can take many forms dealing with estimation issues expressed earlier in this section. Also, SEM is used for probability outcome variables used in probit regression.

Xu, et al. (2013) used structural equation modeling with crash rate and travel speed as endogenous variables. Tobit modeling was used to account for zero valued observations for crash rate in arterial mid-blocks. The use of average travel speed as an endogenous variable in the two-stage process resulted in more efficient estimators. Without the use of endogenous variables, the resulting models found four explanatory (exogenous, not effective by system) variables to be significant. With the two-stage Tobit model, eight explanatory variables were significant.

The increase in information picked up from the two-stage Tobit proves there is correlation between crash rate and average travel speed. The censored data issue was addressed through Tobit regression by accounting for the zero-valued, non-crash found many times in arterial segments. Left-censored data indicates crash rate is unobserved but those segments still contain information about crash rate occurrence. It is the difference between knowing what cause crashes versus what does not cause crashes.

Golob, et al. (2004) used multivariate probit model to evaluate crash occurrence of the conventional weaving section types. The three types are defined by weaving lane. Each observation (vehicle crash) was recorded for weaving section characteristics to understand the correlations between multiple characteristics resulting in crashes. Characteristic variables included roadway conditions crash severity and crash type, i.e., rear-end and side-swipe. The indication that crash severity diminished for weaving section with only one movement per merging vehicle had resulted. The researchers found no significant effect between weaving section type and crash frequency.

Multivariate probit model is a form of structural equation modeling in which endogenous variables (weaving types in the last example) show some correlation in error terms when regressed. The predictive estimates are found by maximum likelihood method against the exogenous variables (crash characteristics) in order to compare correlations. The probit model was chosen for the discrete case in which the endogenous dependent variables are jointly estimated.

2.4 – Simultaneous Equation Modeling

Few studies are available demonstrating use of SEM to develop freeway safety prediction models. Abdel-Aty et al. (2006) related various types of freeway traffic volume to crash frequency using SEM. Ye et al. (2009) and Medina et al. (2006) used SEM in highway safety prediction. Ye et al. (2009) modeled crash type frequencies, e.g., property damage and injury crashes, for rural intersections. They use traffic volumes and shoulder width to name a few variables. Medina et al. (2006) took into account the relationship between crash rate and mean speed as endogenous regressors.

In a study conducted by Ye, et al. (2009), SEM was used to predicted crash type frequencies. In rural areas of the state of Georgia, some 837 crashes spanned the period of two years at selected intersections. To understand the correlation between crash types, an equation for each of the following was formulated using multivariate Poisson regression: head-on, rear-end, sideswipe, sideswipe of opposing direction and pedestrian-involved. The endogenous variables of crash type then become independent variables in the estimating process along with traffic volumes, shoulder width, lighting, number of left and right turn lanes, grade and terrain of major and minor roads. The researcher did show the correlations in error terms which support that this type of modeling captured the unobserved factors of crash types related between intersections. Although the goodness-of-fit of the model and efficiency of the parameter estimates were improved, no additional information was obtained when modeling univariate approach compared to the simultaneous multivariate model.

Abdel-Aty, et al. (2006) used SEM to relate crash frequency of Florida freeways to traffic flow and geometric characteristics. Sticking with a count model, Seemingly Unrelated Negative Binomial regression was used for the discrete observations. The models related crash frequency of various scenarios, i.e., scenarios equal the number of equations in system to be modeled, with traffic and freeway geometric characteristics. One of the pair of scenarios had single and multi-vehicle crash frequencies simultaneous model against indicator variables such as radius category defined by threshold of 3,000 feet and pavement type. Also, continuous variables used include volume and speed. Other scenarios modeled are morning and afternoon peak hour crashes with the same explanatory variables. All of which resulted in more efficient parameter estimates and improved model fit when compared to the single equation modeling done in the same study.

Medina, et al. (2006) modeled endogenous variables mean speed and crash rate for Indiana highways using "three-stage least squares" method of SEM. A regression equation for crash rate was developed with mean speed as an endogenous variable along with posted speed limit, horizontal curve and cross-sectional width as well as behavioral indicator variables. The same explanatory variables were used to predict mean speed with crash rate as endogenous variable. All variables used for estimation were found significant, albeit, at the 10% significant level. Crash data for three years were converted to crash rate weighted by exposure as million vehicle miles traveled.

In this thesis, "three-stage least squares" method of simultaneous equation modeling was used based on the literature reviewed. The results show the interrelationship of endogenous variables of crash rate of EX-EN and EN-EX freeway segments used for estimation of more efficient regression parameters. The results also indicate the correlation of adjoining freeway segments which have heterogeneous geometric characteristics across adjoining segments and unobserved variables when relating crash occurrence to geometric variables such as median and shoulder widths, number of lanes, horizontal and vertical alignment as well as segment length with traffic volumes.

Chapter 3 Methodology

3.1 – Research Methodology

This research is a continuation of Teng, et al. (2013). The flow of research begins much earlier than the formulation of the problem statement for this thesis. To better understand the flow, Figure 5 begins with data collection from the previous study. Although the previous research project has a similar problem statement of crash reduction for the Las Vegas freeway system, the problem statement for this thesis was not observed and imagined until the previous project was drawing to a conclusion.

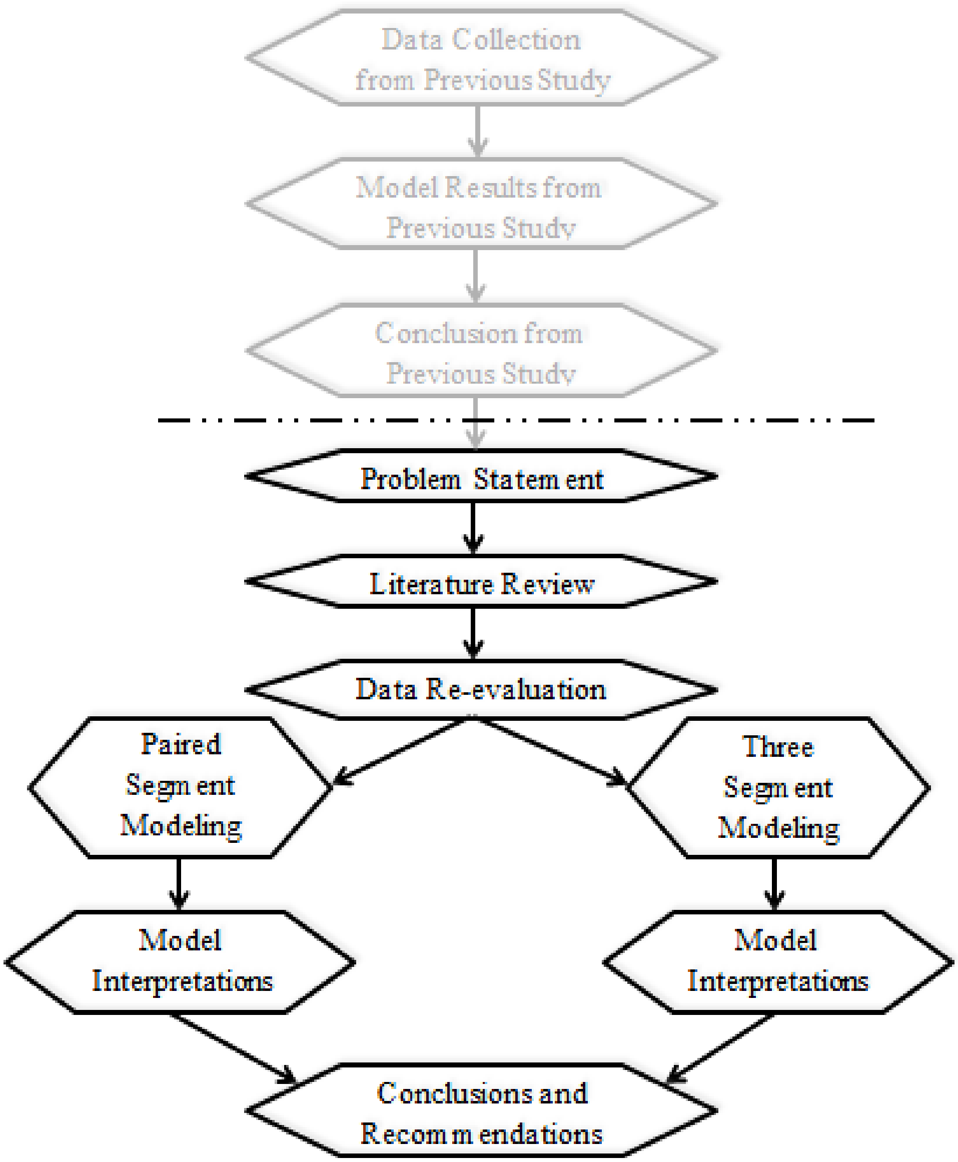

Figure 6 – Research methodology where dotted line represents beginning

After the problem statement was formulated, a literature review was completed to validate a modeling approach and to ensure the topic was not repeated regarding simultaneous equation modeling (SEM) for freeway segment crash prediction. All too often in freeway crash prediction studies, the geometric variables provide little information for predicting crash count or severity. Due to the limiting segmentation, the corridor geometry may change from segment to segment making the study area heterogeneous and variables difficult to capture empirically. Conversely, basic freeway segment crash prediction does not explain the influence of connecting segments.

The previous data set required rearrangement in order to model connecting freeway segments defined by ramp pairs. Most EX-EN segments are located at an interchange and have shown crash concentrations influenced by the connecting EN-EX. To better understand the causes of the concentration at or near the interchange, the upstream EN-EX was then added in another model to complete a semi-corridor without compromising exogenous geometric variables.

The interrelationship of connecting freeway segments in order to find latent variables was important to establish before SEM methods can be used. Also, identification of equations needs to be established. Modeling using "three-stage least squares" allows for crash rate correlations across freeway segments to be distinguishable. Through the utilization of instrumental variables (described in the next section), resulting model estimates are more efficient compared to estimates found using multiple linear regression.

The interpretation of the results assisted with recommendations for crash reduction. In the conclusion, caveats are suggested for future use of SEM for improved prediction modeling.

3.2 – Instrumental Variables

When least squares estimators are inconsistent, they do not converge on the probability of being an unbiased estimator as the number of observations increases (Wackerly et al. 2008). Instrumental variables allow for improvement for regression estimates as long as the assumptions are met. Instrumental variables must be correlated with the other set of explanatory (or exogenous) variables and have no correlation with the error term. The instrumental variable estimates can be found using "two-stage least squares" (2SLS) regression. Otherwise, instruments would be impossible to establish for regression use. The number of instrumental variables may equal or be less than the other set which allows for the use of data sets with different explanatory variables, e.g., EX-EN segments contain no auxiliary lanes but still can be regressed with EN-EX segments with presence of auxiliary lanes variables (Greene, 2008). Instrumental variables should then be able to fill the missing data columns of the uneven variable vectors using 2SLS estimation.

The endogenous variables are made up of data affected by the system. There must be the same number of endogenous variable as there are equations in the system in order to be considered a "complete system". The system of equations must be identified in which the number of unknown parameters equal the number of estimates coefficients. Each equation is identified by the Order Condition demonstrated by Green (2008) in equation (1). In equation (2), the difference between the total of exogenous variables in the system and the number of exogenous variables in the equation must be more than or equal to the difference between the total number of system endogenous variables and the number of endogenous variables in the equation minus one. Only then can the system of equations be determined through simultaneous equation modeling. The crash rate prediction equation for EX-EN segment type lacks the presence of auxiliary lane variable. To complete the system and address identification, instrumental variables are only included in the EX-EN regression equations.

$$K_j^* \geq M_j \tag{1}$$

$$K - K_j \geq M - M_j^* - 1 \tag{2}$$

where,

K = total number of system exogenous variables

Kj = number of equation exogenous variables

K_k^* = number of equation exogenous variables excluded

M = total number of system endogenous variables = total number of equations

Mj = number of equation endogenous variables

M_k^* = number of equation endogenous variables excluded

A system of equations for crash rate prediction for each segment type (G) included in the models can be seen in equation (3)

$$y_i = Y_i\alpha_i + X_i\beta_i + u_i, i = 1, 2, \dots G \tag{3}$$

where y is the vector of endogenous observed crash rate variables, Y is the vector of fitted endogenous crash rate variables with parameter α estimated with the use of instrumental variables, X is a matrix of exogenous geometric characteristics and traffic volume variables with β estimated parameter coefficients and u the error term. The EX-EN is when G=1 and EN-EX is when G=2. The assumption is that there is no correlation of error terms from observation to observation as seen in (4).

$$E[u_{it}u_{js}] = 0 \quad \forall \quad t \neq s \tag{4}$$

However, the contemporaneous correlation across equations for the combined observations in (5) is use to obtain the covariance matrix of contemporaneous correlation error terms used in SEM estimation techniques. The covariance matrix is seen in (6).

$$E[u_{it}u_{jt}] = \sigma_{ij} \tag{5}$$

$$E[uu^T] = \Omega = \sum \otimes I_T \tag{6}$$

where $\sum$ is the covariance matrix taken with the Kronecker product with *I*, the identity matrix for the *T* number of observations in each equation. The Kronecker product is not to be confused with matrix multiplication but results in a block matrix.

The instrumental variables, Z for the *i*th observation, are estimated in two-stage least square and are calculated in the following manner:

$$Z_i = \widehat{Y_i} = X[(X^T X)^{-1} X^T Y_i] \tag{7}$$

The next section explains how Z is used in the first stage in the modeling technique used in this thesis.

3.3 – Three-Stage Least Squares

The instrumental variables described in the last section take the place for latent variables and thus can account for measurement errors in data collection. The omitted variables in ordinary least squares estimation (OLS) cause biasness in single equation MLR. Acting as the unobserved variables, instrumental variable are estimated in the method of three-stage least squares (3SLS). The process of 3SLS simply explained by Washington, et al. (2011) is as follows:

1. Preform OLS regression for each equation to obtain predicted variables which are then used as instrumental variables, Z*i* in (7)
2. Find the error terms across equations for covariance matrix, Ω in (6)
3. Covariance matrix is then used for generalized least square (GLS) estimation of regression parameters, α and β in (7 and 8)

The first stage in the 3SLS method is to regress all exogenous variables using OLS on Y_i. The exogenous variables, X_i, are the geometric characteristics and traffic volume. The predicted endogenous variables of crash rate from each segment type become instrumental variables. The instrumental variables are used in the opposing equation for residual analysis. The residuals must be contemporaneously correlated.

The second stage is calculating the covariance matrix using the residuals from the included instruments. The third stage uses the covariance matrix to estimate parameters using GLS for the equation system seen in (7) and (8).

$$Y_1 = Y_2\alpha_2 + X_1\beta + \varepsilon_1 \tag{7}$$

$$Y_2 = Y_1\alpha_1 + X_2\beta + \varepsilon_2 \quad (8)$$

The results from 3SLS, as long as the disturbances are contemporaneously correlated, are consistent estimators which are asymptotically more efficient (Zellner & Theil, 1962) (Henningsen & Hamann, 2007). The method of 3SLS is better suited for systems of equations to aid in the understanding of the correlation between connecting freeway segments when predicting crash rates using geometric and volume variables.

Chapter 4 Data

The pooled data set used for modeling was collected from the previous study. A reevaluation of the data was required to combine sequential segments from Las Vegas freeways, I-15, I-215 and US 95. The previous study involved data collection including geometric characteristics, traffic volume and historical crash counts described in the following sections. Then, the explanation of how connected freeway segments were joined for SEM is found at the end of the chapter.

4.1 – Description of Variables

The following section contains simple explanations of the variables used for SEM. Table 1 includes the variable designations along with their units for quick reference. The crash rate was calculated by normalizing exposure of crash counts (recorded for a one year time period) taken over the traffic volume and length of the individual segment. All variables excluding AADT are characteristics of geometric design. Only crash rates are endogenous variables.

Table 1 - Description of Variables Used in Modeling

Variable	Type	Units	Description
CRASHRATE (EX-EN)	Endogenous	Million Vehicle Miles Traveled	Crash count per EX-EN segment taken over AADT and length in miles then multiplied by a million
CRASHRATE (EN-EX)		Million Vehicle Miles Traveled	Crash count per EN-EX segment taken over AADT and length in miles then multiplied by a million
TLANES	Exogenous	Number of lanes	Number of through lanes
SHOULDER		Feet	Measurement taken from outside lane to edge of pavement
MEDIAN		Feet	Measurement taken from inside lane to halfway of opposite direct inside laneedge of pavement
AADT		Vehicles per Day	Average Annual Daily Traffic
GRADE		Percent	Average grade change for segment
RADIUS		Feet	Curve radius
LENGTH		Feet	Measurement taken from painted gore of ramp terminal to following painted gore
AUX		No Units - Indicator Variable	Presents of auxiliary lane

4.2 – Segmentation

All measurements were taken from Google Earth image dated 5/28/2010 in order to evaluate the proper geometric features representative of 2010. The dated image also allows for the appropriate assessment of construction work zones which alter the driving conditions. If the image is not dated, improper geometric condition maybe recorded for the study year causing error in the dataset. The Las Vegas freeway system, as well as most urban freeways, necessitates improvement projects including freeway widening which can change geometric variables observed. The recorded observation changes introduce difficulties in yearly studies.

As-Built designs are ideal for recording geometric characteristic variables in any roadway study resulting in more accurate measurements for variables such as median and shoulder width. However, examining each plan set is very time consuming and would constitute a research project unto itself. The tradeoff of accuracy issues using Google Earth image is unequal to the amount of time spent recording data from As-Built plan sets.

The segmentation was done using the AASHTO ramp-pair combinations of EX-EN and EN-EX (as seen in Chapter 1). The length was considered from the painted gore of the first ramp terminal to the painted gore of the next ramp. The segment length in this study is considered base length in the Highway Capacity Manual 2010 for basic freeway and weaving segments.

The study area is highlighted in Figure 7. Not all freeway areas were used. Freeway segments with ramp pair combinations of EX-EX and EN-EN were excluded. Also, any segment with construction zone observed for 2010 was excluded.

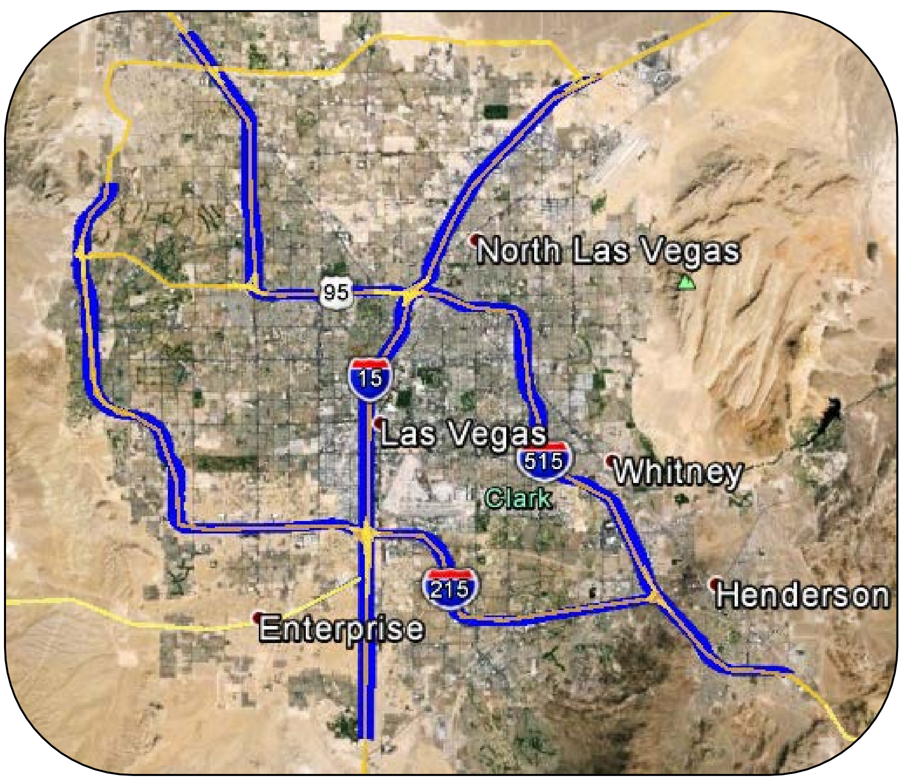

Figure 7 – Las Vegas studied freeways highlighted in blue

4.3 – Endogenous Variables

The crash rate variables were calculated using the following equation:

$$\text{CRASHRATE} = \frac{N}{V*365*L} * 10^6 \qquad (10)$$

where,

N = number of crashes in 2010 per segment

V = vehicles per day taken from AADT

L = segment length in miles

Crash data was supplied by the Nevada Department of Transportation and was previously geocoded. The accuracy of crash location is questionable; however, relative spatial correlations across segment types are still significant when making inferences related to geometric factors. Measurement errors in the recording of the crash data are prevalent. Mention of this fact is in the Highway Safety Manual 2010. Consistency in reporting techniques allow the observed crash data to still be compared over the years.

4.4 – Exogenous Variables

The geometric characteristics which include number of lanes, shoulder width, median shoulder width, average grade change, curve radius and segment lengths were recorded with the use of Google Earth. To relax accuracy issues when taking measurements, multiple measurements were taken with the use of Google Earth distance ruler and an average was recorded. Distance was averaged for shoulders, medians, radii and lengths for all segments. Taking an average for grade change is described by Roess (2011) in which freeway segment lengths less than 4,000 feet and grade change less than

5% can be done for composite grades. In this study, elevation was recorded at each end of the segment using Google Earth as well. Then, the average percent grade was recorded by taking the elevation difference over the segment length.

The shoulder width was measured according to the AASHTO "Green Book" definition. The freeway shoulder is the "usable" width intended for disabled vehicle parking and emergency vehicle passageway. The shoulder should be paved and continuous. The suggested width should be 10 to 12 feet depending on truck volume. The suggested width agrees with most of the measurements recorded for the studied freeway segments. The median is taken as the Green Book definition of median (the total distance between opposing interior lanes) and recorded as half for both directions. This measurement usually considered median shoulder width and should be 4 to 8 feet for each direction with 2 feet median barrier which agrees with the majority of recorded measurements in this thesis.

To overcome for the lack of state geometric data base and to simplify measurement, each horizontal curve observed was treated as a simple curve. Arc and chord length were recorded in Google Earth. The use of ArcGIS Curve Calculator under the COGO toolbar provided curve radii. When the same freeway segment encountered multiple curves, the shorter radius was recorded due to the stronger effect on driver comfort. The same reasoning was considered for segments containing both curve radius and tangent sections. Some freeway segments shared curve radius. In this case, every segment was designated with the same curve radius measurement.

Average annual daily traffic (AADT) volume data was taken from the Nevada Department of Transportation. When spot volumes were not included in their traffic report, a balance approach was considered so that all freeway segments in the study had AADT values in the data set. The ramp AADT along with nearby volumes assisted with the balance calculations. An example of balance calculation can be seen in Figure 8. The given nearby volume is added with each ramp volume resulting in the missing volume sum. A map showing which segments with missing volume from the published traffic report is shown in Figure 8.

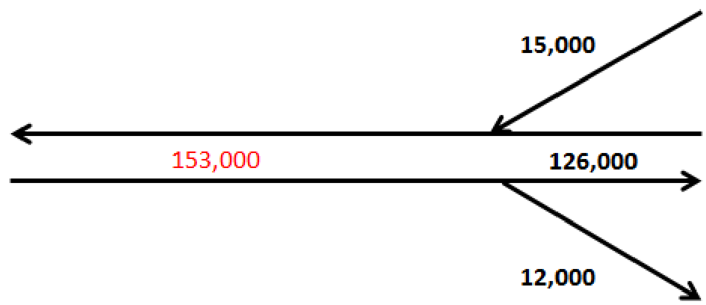

Figure 8 – The sum of given volumes (black) results in the missing volume (red)

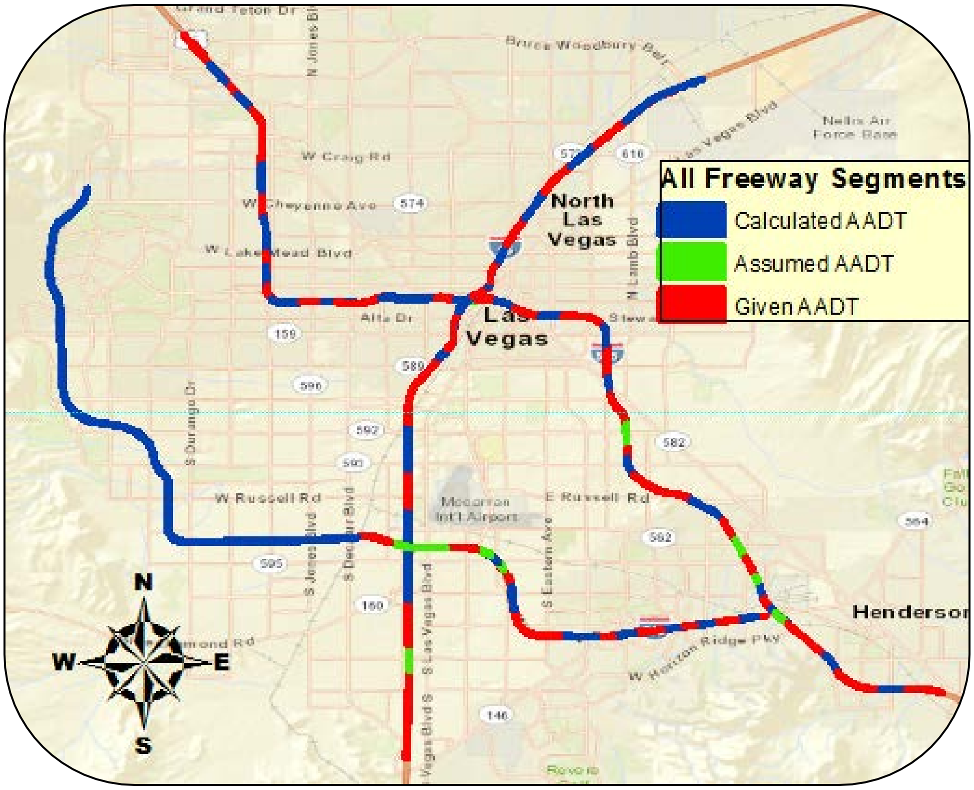

Figure 9 – Las Vegas freeway segments with given and calculated AADT

4.5 – Combining Freeway Segments

Freeway segments were paired initially by the EX-EN segment and the connecting (or downstream) EN-EX. Later, the upstream EN-EX was added for additional analysis of a combined three-segment semi-corridor. In Figure 10, the included EX-EN segment in the paired segment data set to be modeled must be followed by an EN-EX segment. The correlation of crash frequency between these two segment types was observed. The results chapter demonstrates the interrelationship when the two segments are modeled together.

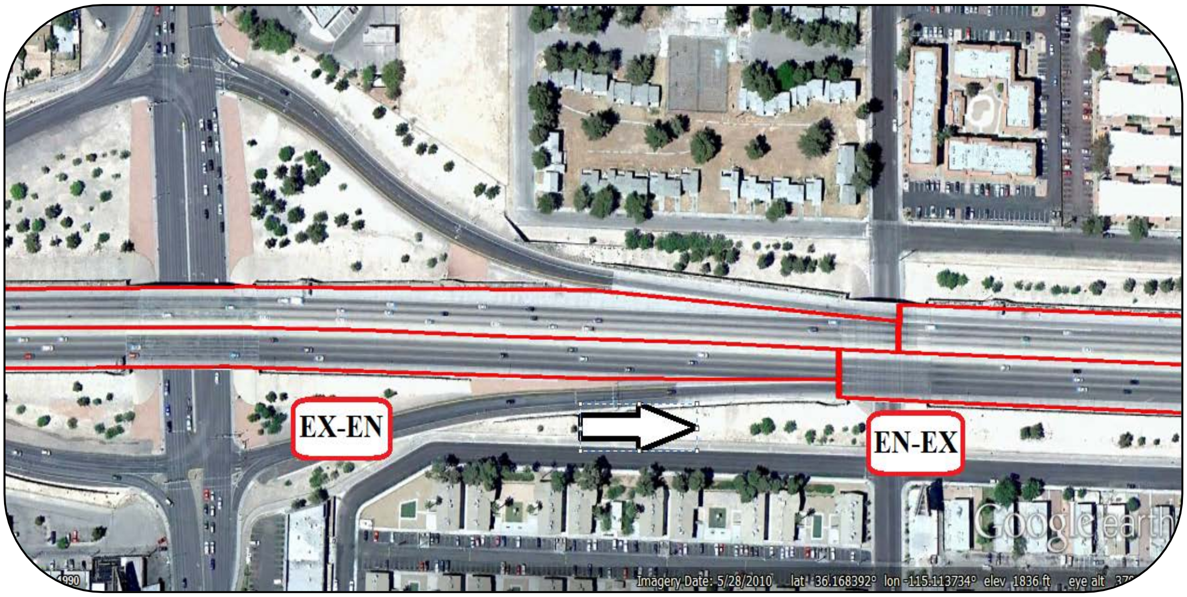

Figure 10 – Segment pairing of EX-EN to sequential EN-EX of US 95 at Eastern Ave.

The upstream EN-EX was then added to the paired segment observations resulting in an additional data set to be modeled. The three segments together form a heterogeneous semi-corridor in which explanatory variables differ across segments. Both examples of combining paired segments and combining the three-segment observations for data sets are seen in Figure 11. The resulting data sets to be modeled are pair-segment and three-segment, respectively. The effects of the interchange are better understood when including the upstream and downstream basic/weaving freeway segments (EN-EX). Each connected segment observation was modeled as one case in the SEM.

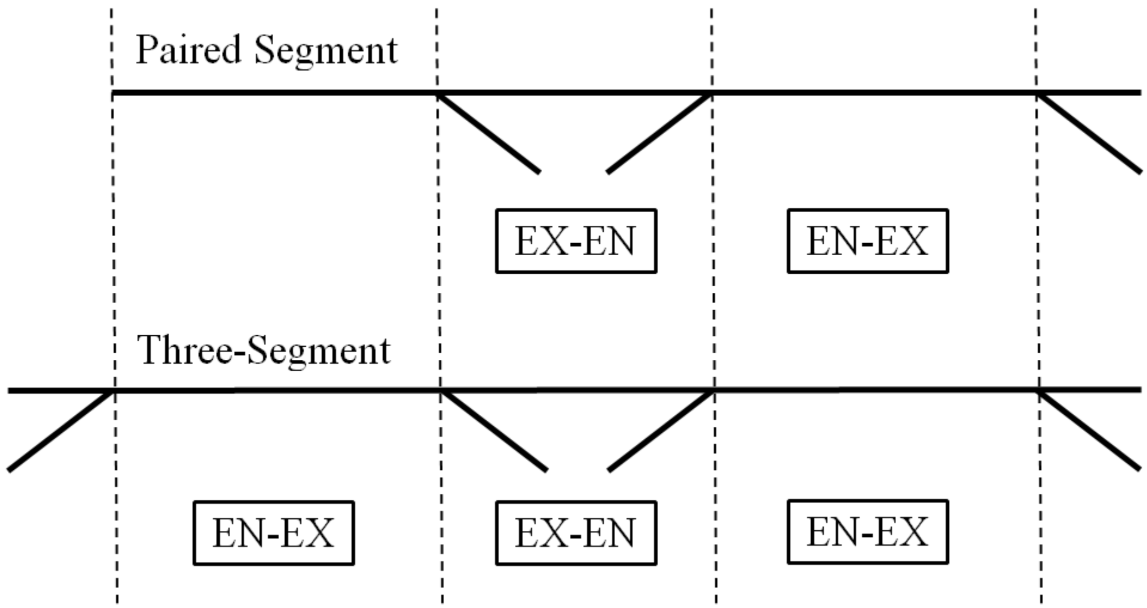

Figure 11 – Demonstration of combining successive freeway segments for data sets

Chapter 5 Results

The following chapter discusses the statistics of crash frequency analysis by segment type for pooled, paired segment and three-segment data sets. The tables and figures confirm intuitive reasoning for use of simultaneous equation modeling in order to show the unobserved relationship between segment type crash predictions.

Overall descriptive statistics including all endogenous and exogenous variables for each data set follows. The distribution of all variables dictates which model to use according to Empirical Bayes method descried in the Highway Safety Manual.

Lastly, the simultaneous equation modeling (SEM) results for the pair-segment and the three-segment models are shown. The interpretation of significant variables for each model is explained.

5.1 – Segment Type Crash Statistics

In past studies, weaving segments were thought to cause the majority of crashes on congested urban freeways. The influence area of weaving sections for density is defined by the Highway Capacity Manual. Influence area exceeds the ramp painted gore by 500 feet on either side of the segment. This can be seen in Figure 12 and can be compared to the crash frequency illustrated in Figure 13. The GIS exhibit (Figure 13) shows that the influence areas of weaving sections for safety, as opposed to density, might extend more than 500 feet on each side of the painted gores. Only one segment pair is displayed in Figure 13. However, similar patterns are observed in most weaving section in the Las Vegas freeway system where congestion is influenced by major arterial interchanges.

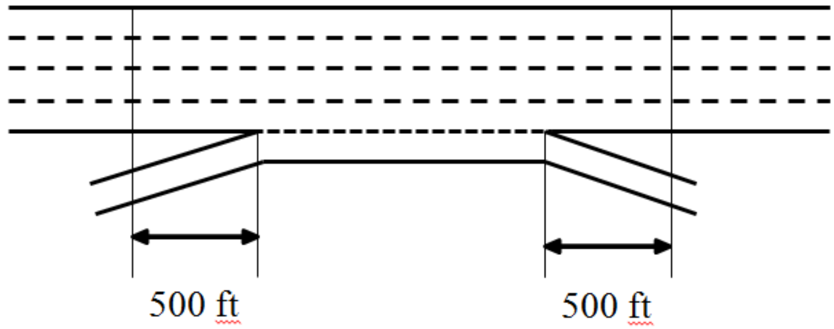

Figure 12 – Influence areas of weaving sections for density defined in HCM (TRB, 2010)

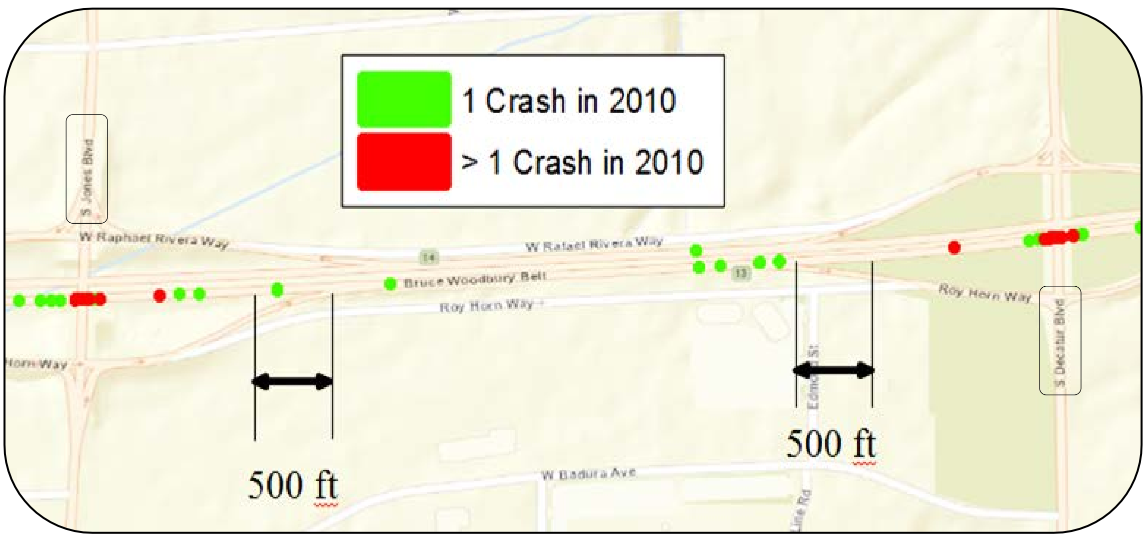

Figure 13 – Crash frequency of I-215 South between Jones Blvd. and Decatur Blvd.

Furthermore, the number of crashes for the west-bound EX-EN segment of I-215 at S. Jones Blvd. (interchange on the left-hand side of Figure 13) is 61 for 2010. The connecting EN-EX segment has 3 observed crashes. The connecting EX-EN segment at the S. Decatur Blvd. arterial interchange (right-hand side of Figure 13) has 67 crashes. Both EX-EN segments previously descried might be extreme cases of crashes observed when comparing to the connecting EN-EX segment. However, the pattern persists for all freeways in Las Vegas. The following Table 2 gives the total number of crashes for EX-EN segments paired with EN-EX segments.

Table 2 - Crash Totals for the Pooled and Paired Segment Data Sets

	EX-EN		EN-EX	
Entire Study Area				
Total Number of Segments	94		100	
Total Crashes for 2010	529	(53.8 %)*	454	(46.2 %)*
Paired Segments				
Total Number of Segments	70		70	
Total Crashes for 2010	467	(58.7 %)*	329	(41.3 %)*

* Percents given as portion of total crashes from respective data set

The larger number of crashes observed in EX-EN segments indicate the focus of freeway safety should not be on weaving sections. Most EN-EX segments in the Las Vegas freeway system can be considered weaving sections as they contain auxiliary lanes and have short segment lengths. These two variables are indicative to Type A weaving section defined in the HCM 2010 (about 0.5 miles in length and presence of auxiliary lanes). Weaving sections create situations for drivers where maneuvering for lane changes happens almost at the same time with other drivers within a short distance. This causes drivers to slow down during most of these interactions resulting in a shockwave of reduced speeds. The shockwave is then felt by the drivers upstream, some of which might not react in time resulting in crashes. The number of Type A weaving sections in the paired segment data set is 22 out of the 70 EN-EX segments and accounts for 107 crashes out of the 329 for 2010.

Note in Table 2, the pooled data set included EX-EN and EN-EX segments from the study area that totaled to 194.However, EX-EN and EN-EX segments were paired for SEM and resulted in 140 total segments. More on this reduction in explained in the next section.

5.2 – Descriptive Statistics

The three-stage least squares model used the exogenous and endogenous (crash rate) variables seen in Table 3. The reevaluation of the pooled data set resulted in the two data sets used for SEM, pair-segment (PS) and three-segment (TS) data sets. The pair-segment data set consists of EX-EN and connecting downstream EN-EX. For simplification of data set and model explanations, the pair-segment types are referred to as PS1 for the EX-EN segment and PS2 for the downstream EN-EX segment.

Table 3 – Descriptive Statistics for Pair-Segment Data Set

	EX-EN (PS1)				EN-EX (PS2)			
Variable	Mean	Min.	Max.	Std. Dev.	Mean	Min.	Max.	Std. Dev.
CRASHRATE	0.262	0	2.655	0.463	0.165	0	1.569	0.268
TLANES	3.23	2	5	0.73	3.30	2	6	0.84
SHOULDER	11.98	3.54	18.20	2.40	11.88	5.30	20.56	2.46
MEDIAN	15.77	3.07	45.78	10.23	15.27	4.02	45.96	10.23
AADT	122151	25500	291600	56368	136230	33000	298100	63313
GRADE	-0.09	-3.50	3.10	1.41	-0.03	-3.80	3.20	1.20
RADIUS	1448	1811	11088	2746	3244	1247	10768	3230
LENGTH	3175	797	5930	877	3877	897	14119	2163
AUX	-	-	-	-	0.63	0	1	0.49

The total of pooled segments reduced even more when modeling for three consecutive segments of EN-EX, downstream EX-EN and downstream EN-EX. The descriptive statistics for the three-segment data set are in Table 4. The three-segment data set and model equations are referred to as TS1 for EN-EX segment, TS2 for the downstream EX-EN segment and TS3 for the downstream EN-EX.

Table 4 – Description Statistics for Connecting Three-Segment Data Set

	EN-EX (TS1)				EX-EN (TS2)				EN-EX (TS3)			
Variable	Mean	Min.	Max.	Std. Dev.	Mean	Min.	Max.	Std. Dev.	Mean	Min.	Max.	Std. Dev.
CRASHRATE	0.145	0.000	1.057	0.214	0.282	0	2.655	0.496	0.166	0	1.569	0.279
TLANES	3.26	2.00	5.00	0.69	3.21	2	5	0.69	3.26	2	6	0.78
SHOULDER	12.03	6.35	20.56	2.44	12.16	3.54	18.20	2.38	12.10	6.30	19.90	2.11
MEDIAN	15.70	4.88	45.96	9.44	16.24	5.40	41.85	9.35	15.51	4.02	45.78	9.45
AADT	133540	33000	298100	58356	117926	25500	291600	53840	131709	33000	298100	62342
GRADE	-0.14	-3.80	3.60	1.26	-0.03	-3.10	3.10	1.36	0.00	-3.80	3.20	1.20
RADIUS	2708	2860	9958	3073	1601	1811	11088	2888	3032	2860	10768	3133
LENGTH	3791	1183	8648	1929	3306	1349	4734	658	4070	897	14119	2310
AUX	0.621	0.000	1.000	0.489	-	-	-	-	0.59	0	1	0.50

The reduction in the number of segments from the pool data set, to the paired segment data set and further reduced on the three-segment data set is due to not all EX-EN segments being followed by EN-EX. Other segment types, EN-EN and EX-EX, break up the sequence and are normally found at freeway-to-freeway interchanges and more complex arterial interchange layouts. When adding another EN-EX segment upstream from the paired segments, the number of trios decreases even further to 58 segment sets. A comparison of pair- and three-segments crash totals can be seen in Table 5.

Table 4 – Crash Total for the Pooled, Pair- and Three-Segment Data Sets

	EN-EX	EX-EN Downstream	EN-EX Downstream
Entire Study Area			
Total Number of Segments		94	100
Total Crashes for 2010		529	454
Paired Segments			
Total Number of Segments		70	70
Total Crashes for 2010		467	329
Paired Segments			
Total Number of Segments	58*	58	58
Total Crashes for 2010	251*	422	271

* Observations may already be used in EN-EX Downstream observation included in the same data set in order to analyze the most possible number of EN-EX to EX-EN to EN-EX segment observations

5.3 – Pair-Segment Model Results

SEM was estimated using "Three-Stage Least Squares" (3SLS) and can be seen in Table 6. Obtaining more efficient estimates from the observations is accomplished through the 3SLS method. The first stage involved regression of all exogenous variables. The fitted endogenous variables are used in the second stage. The regressed endogenous variables in conjunction with the instrumental variables are used to model the system of equations. The third stage involves iterating the covariance matrix until the estimated results converge in the generalized least squares process.

Table 5 – SEM Model for Pair-Segment Data Set

	EX-EN (PS1)				EN-EX (PS2)			
Variable	Coefficient	Standard Error	t-statistic	p-value	Coefficient	Standard Error	t-statistic	p-value
Constant	**1.215**	**0.267**	**4.550**	**0.000**	**0.480**	**0.177**	**2.718**	**0.009**
CRASHRATE (EN-EX)	**-0.359**	**0.146**	**-2.453**	**0.017**	-	-	-	-
SHOULDER	**-0.066**	**0.015**	**-4.325**	**0.000**	**-0.018**	**0.010**	**-1.790**	**0.079**
MEDIAN	**-0.020**	**0.009**	**-2.105**	**0.039**				
AADT	**0.000001**	**0.0000005**	**2.429**	**0.018**	**-0.0000006**	**0.0000004**	**-1.744**	**0.086**
RADIUS	**-0.00004**	**0.000018**	**-2.382**	**0.020**	**-0.00002**	**0.00001**	**-1.979**	**0.053**
RMSE	0.291				0.156			
McElroy's R2	0.605							

The modeling results in Table 6 were estimated using R Project software with "systemfit" installed package. The package was tested by Henningsen and Hamann (2007) for 3SLS estimate results comparable to estimates found in "Econometric Analysis" by Greene (2008).

The parameter estimates were found without truncating the data set by the different freeways as in the previous study done by Teng, et al. (2013). This way, the pair-segment model is more generalized for use with any freeway segmented in the same manner as in this study. Also note, the variables, both endogenous and exogenous, were not transformed to address heteroscedasticity among error terms. Transformations are not required due to the relaxing effect SEM has on heteroscedasticity.

In the pair-segment model, the following observed variables were found to be significant in the estimated EX-EN (PS1) model equation: shoulder width, median shoulder width, AADT and curve radius. Any increase for shoulder width and median shoulder width would decrease crash rate. Widening these areas of the freeway increases driver comfort for any lane changing maneuvering. Easing driver comfort decreases the chance for hesitation which can lead to mistakes, i.e., crashes. Also, increased curve radius results in a smoother curve which also decrease crash rate as horizontal alignment is handled better by the driver.

The increase of AADT has long been accredited to increases in crash rate. The increased volume experienced in the same freeway facility increases the chance for vehicle crashes to occur. The positive parameter estimated confirms this. The coefficient indicates that for every increase of AADT by 100,000 vehicles per day, while all other variables remain unchanged in the system, crash rate would increase by 0.1 million vehicle-miles traveled.

In the EN-EX (PS2) model equation, the shoulder width and curve radius follow the same ability to decrease crash rate if increased. However, the median shoulder width was not found significant at the 95% level. The AADT coefficient has the opposing effect of the EX-EN (PS1) equation. The negative parameter might be due to the EN-EX segment already experiencing high volumes compared to EX-EN. The entrance ramp exhibit volumes that add to the EN-EX segment and then a reduction in volume is experienced before the EX-EN segment begins due to exiting volumes. An increase of AADT in an already high volume situation would increase levels to near jam capacity reducing any crash rate increase by impeding all vehicle maneuvers. The shockwave theory is supported. The shockwave effects are felt upstream in the EX-EN segment increasing crash rate. The endogenous variable estimated is negative which also supports the theory. More studies are needed for peak hour analysis of the impact of AADT on crash rate for these connected segment types.

The Mc Elroy R^2 indicates goodness-of-fit of the model at 0.602. Root Mean Squared Error for both equations is relatively low considering the nature of the data. Measurements recorded using Google Earth and crash data are not as accurate as required for fundamental statistical modeling. However, the information gained when relating connecting freeway segment characteristics and crash rate is invaluable.

5.4 – Three-Segment Model Results

The SEM model for three sequential freeway segments, EN-EX, EX-EN downstream and EN-EX downstream, can be seen in Table 7. The number of segments decreased to 58 due to the limiting possibility of the three segments being uninterrupted by an EN-EN or EX-EX segment. Shoulder width and median shoulder width are the only variables significant for each estimated segment equation. Increases in these variables decrease crash rate for all equations for the same reason explained in the previous section.

Table 7 - SEM Model for Three-Segment Data Set

Variable	EN-EX (TS1) Coefficient	Std. Error	t-statistic	p-value	EX-EN (TS2) Coefficient	Std. Error	t-statistic	p-value	EN-EX (TS3) Coefficient	Std. Error	t-statistic	p-value
Constant	3.126	0.540	5.79	0.000	1.607	0.321	5.01	0.000	2.043	0.548	3.73	0.000
CRASHRATE (EN-EX)	-	-	-	-	-0.146	0.085	-1.71	0.087	-	-	-	-
TLANES												
SHOULDER	-0.128	0.033	0.03	0.000	-0.062	0.015	-4.19	0.000	-0.049	0.028	-1.73	0.084
MEDIAN	-0.076	0.016	-4.87	0.000	-0.040	0.011	-3.61	0.000	-0.060	0.019	-3.25	0.001
AADT					0.000002	0.0000007	3.03	0.002				
GRADE					0.127	0.0723	1.75	0.080				
RADIUS					-0.00006	0.000019	-3.20	0.001				
LENGTH	-0.00008	0.00003	-2.75	0.006								
RMSE	0.368				0.239				0.395			
McElroy's R2	0.545											

Only the EN-EX (TS1) model equation has segment length as a variable. The negative coefficient indicates that if segment length is increased by 100 feet, then annual crash rate would decrease by 0.008 million vehicle-miles traveled if all other variables in the system remain unchanged.

The EX-EN (TS2) estimated model equation exhibited an additional variable of average percent grade change when compared to the pair-segment model. The positive coefficient indicates that if average grade is increased by 1% then crash rate is increased by 0.127 when variables in the system remain unchanged.

The tradeoff from using pair-segment to three-segment model is goodness-of-fit. The Mc Elroy R^2 has decreased slightly from the paired segment model. Root Mean Squared Error has increased from the last model as well. Also, the endogenous parameter in the EX-EN (TS2) equation has decreased effect on crash rate then the previous model in terms of the coefficient. The interpretation of the endogenous variable is unclear on how it relates to the other equations. The increase in parameter estimates indicates the unobserved correlation between freeway segment types exist. Further, the use of SEM is useful in understanding the interrelatedness of crash rates across sequential freeway segments.

5.5 – Comparison of Model Estimates

The intent of 3SLS method is to produce asymptomatically more consistent parameter estimates. Table 8 shows the comparison of SEM paired model to multiple linear regression (MLR) of individual crash rate prediction models. One MLR equation for EX-EN and a separate model for the connecting downstream EN-EX were done without transformations for comparison purposes only. The estimates for the paired model have slightly more standard error when predicting their respective crash rate compared to MLR. However, more consistent estimators are seen with the addition of variable parameters when using SEM. Modeling crash rate simultaneously with the use of instrumental variables has provided more information for Las Vegas freeway segments. SEM usage for this study proves the interrelatedness of connecting EX-EN and EN-EX segments.

A comparison was not done for the three-segment model. A Chow-Fisher test was considered in order to combine regression equations for the upstream EN-EX (TS1) and downstream EN-EX (TS3) segments. If both EN-EX equations were combined, then a fair comparison would be warranted when considering the three-segment model with MLR and SEM paired segment model.

Table 8 – Comparison of Model Estimates

	EX-EN (PS1)				EN-EX (PS2)			
	MLR		3SLS		MLR		3SLS	
Variable	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
Constant	1.095	0.256	1.215	0.267	0.469	0.176	0.480	0.177
SHOULDER	-0.071	0.014	-0.066	0.015	0.023	0.011	-0.018	0.010
MEDIAN			-0.020	0.009				
AADT	1.27E-06	4.85E-07	1.16E-06	4.79E-07			-6.32E-07	3.63E-07
RADIUS	-4.47E-05	1.73E-05	-4.28E-05	1.80E-05			-2.12E-05	1.07E-05

Chapter 6 Conclusion

Development of SEM models consisting of connecting freeway segments was completed using pair-segment data set, EX-EN (PS1) and EN-EX (PS2), along with a three-segment data set, EN-EX (TS1), EX-EN (TS2) and EN-EX (TS3). The increase of significant model parameters from single equation MLR shows the correlation between crash rate prediction residuals exists between the connecting segment types. Also, the indication of unobserved variables when relating these freeway segments exists.

SEM modeling allows for freeway segments containing different characteristics, e.g., presents of auxiliary lane, to be examined together while resulting in a system of regression equations for crash prediction. Further study is needed to adhere to Highway Safety Manual (HSM) 2010 suggestions when considering the improvement of the models in this study.

The HSM defines a Safety Performance Function (SPF) as a regression based evaluation of the crash frequency of a roadway facility. The SPF is developed with the use of “expected average crash frequency” to account for the natural phenomenon of regression to the mean. For any facility, the extreme number of crashes observed in one year tends to decrease in the next year. The manual suggest examining more than three years in order to fully survey the up- and down- trends of crash frequency unrelated to the changes in physical characteristics. Once the average trend is observed, the corresponding crash frequency can be taken as the predictor variable.

Crash Modification Factors (CMF) are only developed after a treatment is installed and studied for safety effectiveness. The CMFs are then used along with the respective facility SPF for crash frequency evaluation. The results from SPF also include pedestrian and other transportation modes calculated safety assessment which is also factored in the SPF.

As for the models in this study, shoulder width plays a crucial role in preventing crashes. In the individual MLR, pair- and three-segment models, the shoulder width is significant. Each model shows by increasing the shoulder width the crash rate will reduce. The three-segment model parameters decrease crash rate for shoulder, median, radius and length distances for the segment equation in which they are significant for. As to be expected, AADT and average grade increases will increase crash rate in the EX-EN (TS2) segment.

The only counter-intuitive result is found in the pair-segment model. An increase in AADT increases crash rate in the EX-EN (PS1) segment equation. However, it decreases in the EN-EX (PS2) segment equation. As suggestion in the results chapter 5, peak hour volumes may play a role in adding to jam capacity in PS2 preventing any movement that restricts crashes.

Other suggestions for further study also involve AADT volumes. All previous studies reviewed rely on traffic volumes recorded in AADT from their respective transportation planning authorities. Although using AADT simplifies the modeling process, the generalized variable does not provide enough information. Archived ITS data is useful in analyzing various scenarios such as multiple crashes (in that one crash caused another) occurrence as Abdel-Aty et al. (2007), studied in Florida. The intent of the study was to reduce crash risk in real-time by imploring ITS system elements, i.e., ramp metering, Variable Messaging System, etc., after speed variation was detected. The archived data used for model calibration consisted of speed, volume and lane.

Works Cited

AASHTO, 2001. *A Policy on Geometric Design of Highways and Streets.* Fourth ed. Washington, D.C.: American Association of State Highway and Transportation Officials.

AASHTO, 2010. *Highway Safety Manual, 1st Edition, 2010.* First ed. Washington, D.C.: American Association of State Highway and Transportation Officials.

Abdel-Aty, M. et al., 2007. Crash Risk Assessment Using Intelligent Transportation Systems Data and Real-Time Intervention Strategies to Improve Safety an Freeways. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations,* 11(3), pp. 107-120.

Abdel-Aty, M., Pemmanaboina, R. & Hsia, L., 2006. Assessing Crash Occurrence on Urban Freeways by Applying a System of Interrelated Equations. *Transportation Research Record: Journal of the Transportation Research Board, No. 1953,* pp. 1-9.

Awuah-Baffour, R. et al., 1997. *Global positioning system with an attitude: Method for collecting roadway grade and superelevation data,* s.l.: Transportation Research Record.

Geedipally, S. R., Bonneson, J. A., Pratt, M. P. & Lord, D., 2012. *Safety Prediction Methodology and Analysis Tool for Frewways and Interchanges, NCHRP 17-45,* Washington, D.C.: Transportation Research Board.

Golob, T. F., 2003. Structural Equation Modeling for Travel Behavior Research. *Transportation Research Part B: Methodological,* 37(1), pp. 1-25.

Golob, T. F., Recker, W. W. & Alvarez, V. M., 2004. Safety aspects of freeway weaving sections. *Transportation Research Part A,* Volume 38, pp. 35-51.

Greene, W. H., 2008. *Econometric Analysis.* Sixth ed. Upper Saddle River, NJ: Pearson Prentice Hall.

Haleem, K., Gan, A. & Lu, J., 2013. Using multivariate adaptive regression splines (MARS) to develop crash modification factors for urban freeway interchange influence areas. *Accident Analysis and Prevention,* pp. Vol. 55, p. 12-21.

Hans, Z., Souleyrette, R. & Bogenreif, C., 2012. *Horizontal Curve Idenification and Evaluation,* Ames, IA: Midwest Transportation Consortium.

Henningsen, A. & Hamann, J. D., 2007. systemfit: A Package for Estimating Systems of Simultaneous Equations in R. *Journal of Statistical Software,* 23(4), pp. 1-40.

Medina, A. M. F. & Tarko, A. P., 2006. *Modeling the Endogenous Relationship Between Driver Behavoir and Highway Safety.* Washington, D.C., Transportation Research Board, p. 22.

Murray, M. P., 2005. *Econometerics A Modern Introduction.* First ed. Upper Saddle River, NJ: Pearson Prentice Hall.

Nevada Department of Transportation, 2012. *Nevada Traffic Crashes,* Carson City, NV: NDOT, Safety Engineering, Analysis Unit.

Nevada Department of Transportation, 2012. *The Annual Traffic Report 2011,* Carson City, NV: NDOT, Traffic Information Division.

O'Cinneide, D., 1998. *The Relationship Between Geometric Design Standards and Safety.* Washington, D.C., Transportation Research Board, pp. 44:1-7.

Park, B.-J., Fitzpatrick, K. & Lord, D., 2009. *Evaluating the Effects of Freeway Design Elements on Safety,* Washington, D.C.: Transportation Research Board.

Pilko, P., Bared, J. G., Edara, P. K. & Kim, T., 2007. *Safety Assessment of Interchange Spacing on Urban Freeways,* MCLean, VA: U.S. Department of Transportation, Federal Highway Administration.

Ray, B. L., Schoen, J., Jenior, P. & Knudsen, J., 2011. *Guidelines for Ramp and Interchange Spacing, NCHRP Report 687,* Washington, D.C.: Transportation Research Board.

Roess, R. P., Prassas, E. S. & McShane, W. R., 2011. *Traffic Engineering.* Fourth Edition ed. Upper Saddle River, NJ: Pearson Prentice Hall.

Teng, H., Mulokozi, E., Ramos, A. & Gibby, R., 2013. *Evaluation of Geometric Design Needs of Freeway Systems Based on Crash and Geometric Data,* s.l.: Submitted to TRB 93th Annual Meeting for Transportation Research Record.

Treat, J. R. et al., 1979. *Tri-Level Study of the Causes of Traffic Accidents: Final Report Volume II: Special Analyses,* Washington, D.C.: U. S. Department of Transportation, National Highway Traffic Safety Administration.

Wackerly, D. D., Mendenhall III, W. & Scheaffer, R. L., 2008. *Mathematical Statistics with Applications.* Seventh ed. Belmont, CA: Brooks/Cole.

Washington, S. P., Karlaftis, M. G. & Mannering, F. L., 2011. *Statistical and Econometric Methods for Transportation Data Analysis (Second ed.).* Boca Raton, FL: Chapman & Hall/CRC.

Xu, X., Kouhpanejade, A. & Saric, Z., 2013. Analysis of influencing factors identification of crash rates using tobit model with endogenous variable. *Promet - Traffic & Transportation*, 25(3), pp. 217-224.

Ye, X. et al., 2009. A simultaneous equations model of crash frequency by collision type for rural intersections. *Safety Science*, 47(3), pp. 443-452.

Zellner, A. & Theil, H., 1962. Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations. *Econometrica*, 30(1), pp. 54-78.

Zhang, Y., Li, Z., Liu, P. & Zha, L., 2011. Exploring contribution factors to crash injury severity at freeway diverge areas using ordered probit model. *Procedia Engineering*, pp. 21, 178-185.

CURRICULUM VITA

Anthony Ramos, E.I.

University of Nevada, Las Vegas

Department of Civil and Environmental Engineering

Howard R. Hughes College of Engineering

ramosa27@unlv.nevada.edu

Education:

Bachelor of Science in Engineering, University of Nevada, Las Vegas, 2012

Research Experience:

Graduate Research Assistant at the University of Nevada, Las Vegas Department of Civil and Environmental Engineering, 2012 - 2014

Special Honors and Awards:

UNLV Access Grant, Fall 2011, Spring 2012, and Spring 2013

Dean's List, Howard R. Hughes College of Engineering, UNLV, Spring 2011