Research Report
Research Project T1803, Task 31
Interaction between Roadway and Roadside Accidents

INTERACTION BETWEEN THE ROADWAY AND ROADSIDE - AN ECONOMETRIC ANALYSIS OF DESIGN AND ENVIRONMENTAL FACTORS AFFECTING SEGMENT ACCIDENT RATES

by

Songrit Chayanan
Dept of Civil and Environmental Engineering
University of Washington, Box 352700
Seattle, Washington 98195

MaryLou Nebergall
Washington State Department of Transportation
Box 47329
Olympia, Washington 98504-7329

Venky Shankar
Dept of Civil and Environmental Engineering
University of Washington, Box 352700
Seattle, Washington 98195

Naveen Juvva
Dept of Civil and Environmental Engineering
University of Washington, Box 352700
Seattle, Washington 98195

Yanfeng Ouyang
Department of Civil and Environmental Engineering
McLaughlin Hall
University of California, Berkeley
Berkeley, California

Washington State Transportation Center (TRAC)
University of Washington, Box 354802
University District Building
1107 NE 45th Street, Suite 535
Seattle, Washington 98105-4631

Washington State Department of Transportation Technical Monitor
John Milton, Assistant State Design Engineer
Design Office

Prepared for
Washington State Transportation Commission
Department of Transportation
and in cooperation with
U.S. Department of Transportation
Federal Highway Administration

March 2003

TECHNICAL REPORT STANDARD TITLE PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
WA-RD 562.1		

4. TITLE AND SUBTITLE	5. REPORT DATE
INTERACTION BETWEEN THE ROADWAY AND ROAD-SIDE —AN ECONOMETRIC ANALYSIS OF DESIGN AND ENVIRONMENTAL FACTORS AFFECTING SEGMENT ACCIDENT RATES	March 2003
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
S. Chayanan, M. Nebergall, V. Shankar, N. Juvva, Y. Ouyang	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. WORK UNIT NO.
Washington State Transportation Center (TRAC) University of Washington, Box 354802 University District Building; 1107 NE 45th Street, Suite 535 Seattle, Washington 98105-4631	11. CONTRACT OR GRANT NO. Agreement T1803, Task 31
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
Research Office Washington State Department of Transportation Transportation Building, MS 47370 Olympia, Washington 98504-7370 Jim Toohey, Project Manager, 360-407-0885	Final research report 14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

This study was conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. ABSTRACT

The purpose of this research was to explore the relationship between roadway and roadside accident rates for Washington State highways to improve the Washington State Department of Transportation's (WSDOT) process of modeling roadway and roadside accident rates and to arrive at possible improvements in the efficiency of WSDOT's safety project programming process.

The project tested the use of the seemingly unrelated regression estimation (SURE) model to model the roadway and roadside simultaneously. The theoretical advantage of the SURE approach is that it does not impose any *a priori* assumptions on the explicit linkage between roadway and roadside accident rates, and there is no theoretical support for explicit linkage, either. The data used to derive this model were from a random sample of 500 one-mile sections from the Washington State highway system. Traffic data included traffic volumes, truck compositions, AADT, traffic speeds, and other relevant information. Geometric data included lane, shoulder, median, curve, and intersection information. Historical weather data such as monthly precipitation and temperature were collected from the National Oceanic and Atmospheric Administration database.

In comparing significant explanatory variables between the roadway accident rate and roadside accident rate models, very few variables were common. This confirms that it is preferable to specify separate functional forms for roadway and roadside accident rates. Empirical results indicated that correlation between roadway and roadside accident rates was insignificant, indicating that efficiency gains from the SURE model would be minimal. The important finding from a programming standpoint is that jointly modeling the roadway and roadside simultaneously would not result in significant efficiency improvements in comparison to the current state-of-the-practice in Washington State.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Roadway, roadside, accident rates, safety programming	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22616

19. SECURITY CLASSIF. (of this report)	20. SECURITY CLASSIF. (of this page)	21. NO. OF PAGES	22. PRICE
None	None		

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Washington State Transportation Commission, Department of Transportation, or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

CONTENTS

TABLES

FIGURES

EXECUTIVE SUMMARY

The purpose of this research was to explore the relationship between roadway and roadside accident rates for Washington State highways to improve the Washington State Department of Transportation's (WSDOT) process of modeling roadway and roadside accident rates and to arrive at possible improvements in the efficiency of WSDOT's safety project programming process. While geometric, traffic, and environmental factors may have varied effects on roadway and roadside accident rates, it is reasonable to believe that these two accident rates for a given roadway section may be correlated because of unobserved effects common across the roadway and roadside. This relationship assumes significant relevance in the context of the efficiency of programming safety projects, as well as making reliable safety forecasts for policy making.

In the past, classical linear regression (CLR) by the method of ordinary least squares (OLS) has been used to model accident rates for roadway and roadside accidents. This approach has assumed a safety performance function that is related to average daily traffic, and accident propensity is modeled as an "exposure effect" in terms of the number of vehicles at risk. Conditioned on a set of exogenous regressors such as roadway geometrics, and environmental, traffic and human factors, CLR-OLS models are used to estimate roadway and roadside accident occurrence. In the process, the roadway context is modeled separately from the roadside context or is combined to be an overall accident rate. A theoretical argument against this approach is that while environmental and roadway factors are controlled for exogenously, deviations from the mean effects of

exogenous regressors may still occur, causing the error terms to be correlated in some fashion and resulting in potential loss of parameter efficiency.

Other factors support this argument. The state-of-the-practice around the nation and at WSDOT is to program safety projects for the roadway and the roadside on the basis of independent models. From a practical standpoint, the network-level impact of ignoring efficiency is the likelihood of inefficient safety improvement identification in the context of programming for roadway and roadside safety as a whole. Another argument is that combining the roadway and roadside accident rates into an overall accident rate to obtain an exposure rate might result in overestimation of the accident risk for the roadway and roadside accidents. This might also cover up the behavioral differences and correlation effects between roadway and roadside accidents, which could provide useful insights into risk programming. On the other side of the coin, modeling roadway and roadside accident rates separately might result in exclusion of important information that could decrease programming efficiency. Some of the exogenous factors that might be common to roadway and roadside accident rates could be missed by modeling them simultaneously. These common factors could make safety improvements for roadway and roadside more effective. The amount of efficiency gain and additional information from the joint modeling of the roadway and roadside could determine the number of programming turnovers: how many road locations are identified as being safe--but are actually hazardous from a roadway and roadside accident safety point of view.

Exploring the efficiency effect on parameter estimates requires a systems approach to roadway and roadside accident rate modeling. A logical extension of the CLR-OLS model is the seemingly unrelated regression estimation (SURE) model, after

Arnold Zellner. The theoretical advantage of the SURE approach is that it does not impose any *a priori* assumptions on the explicit linkage between roadway and roadside accident rates, and there is no theoretical support for explicit linkage, either (though necessary modeling procedures should/would be adopted to confirm the absence of an explicit linkage between roadway and roadside accident rates). However, it is reasonable to expect roadway and roadside processes to act separately through exogenous regressors but to be linked through their respective disturbances. One can view this indirect linkage as a random shock that occurs across the road section contemporaneously. If correlation between the respective disturbances is significant, using the SURE model should increase the efficiency of the parameter estimation. Also, all the above arguments about use of all the relevant information and overestimation of risk would come into effect. The important question being addressed in this study is whether empirical support exists to prove that sufficient inefficiencies occur in the way WSDOT currently programs safety projects for the roadway and the roadside.

The data used to derive this model were a random sample of 500 one-mile sections from the Washington State highway system. Geometric and traffic data were taken from WSDOT databases and aggregated by using a weighted average for the one-mile sections. The traffic data included traffic volumes, truck compositions, AADT, traffic speeds, and other relevant information. Geometric data included lane, shoulder, median, curve, and intersection information for the 500 sections selected for the model estimation. Historical weather data such as monthly precipitation and temperature were collected from the National Oceanic and Atmospheric Administration (NOAA) database, which provides a 240-station coverage of Washington State. The total number of roadway and roadside accidents within each of the 500 sections for year 1995 and part of 1996 was used in model estimation.

In comparing significant explanatory variables between the roadway accident rate and roadside accident rate models, very few variables were common. This confirms that it is preferable to specify separate functional forms for roadway and roadside accident rates. Empirical results indicated that correlation between roadway and roadside accident rates was insignificant, indicating that efficiency gains from the SURE model would be minimal. One possible explanation for this finding is that aggregation of section definitions at the one-mile level may have a canceling effect, thereby minimizing correlation. In addition, if environmental effects are accounted for, the correlation effect is mitigated. The important finding from a programming standpoint is that jointly modeling the roadway and roadside simultaneously does not result in significant efficiency improvements compared to the current state-of-the-practice in Washington State.

INTRODUCTION

Each year, around 45,000 accidents occur on Washington State highways. Some are simple fender-benders, while others result in fatalities. Depending on the frequency and severity of accidents on a given section of highway, the Washington State Department of Transportation (WSDOT) employs a weighted methodology to identify hazardous roadway and roadside locations on its network. The WSDOT methodology incorporates information from reported accident frequencies as well as severities, and it results in a net "hazard" index that is then used to rank locations for prospective safety improvements.

With limited resources and funds, however, the decision regarding what improvements to make and where to make them is further complicated by the current practice of identifying hazardous roadway and roadside locations separately. More efficient risk programming would enable WSDOT to identify factors that might be common to roadway and roadside accident rates and would thus make safety improvements more economical and effective. Also, the more reliable safety forecasts achieved from using all the available information would improve confidence in safety improvements implemented under limited funding.

The goal of this research is to determine whether efficiencies could be gained by employing a methodology that captured information from the roadside (the portion of the right-of-way outside the traveled way) and roadway (the portion of the highway between the "shoulder" stripes) simultaneously to determine predicted accident rates for both the roadway and roadside elements of a given WSDOT highway segment.

If efficiencies were to be gained from a "simultaneous" roadway-roadside model of accident prediction, they would indicate that crucial interactions between the roadway and the roadside are playing a significant role and not being accounted for in current practice. Such interactions may be directly observed and measured – for example, weather effects. On the other hand, unobserved factors, such as driver behavior on a particular segment, may still have a significant impact but not necessarily in a direct manner. Whether the interaction between the roadway and the roadside is direct or indirect, the pertinent research question is the identification of a statistical framework that accounts for this interaction and models events on the roadway and the roadside simultaneously. In examining this simultaneity, it was anticipated that this research study would extend beyond conventional accident-rate models that employ single-equation approaches and would provide a better modeling framework for roadway-roadside safety programming.

The WSDOT state-of-the-practice is to employ separate single-equation approaches to roadway and roadside elements. The **roadway** element is currently modeled through a series of "accident count" models, which account for geographic effects in addition to roadway geometrics (Milton and Mannering 1998). In that study, which has since evolved into a complete risk-programming model for roadways on the WSDOT network, geometrics and traffic factors were key variables. The effect of weather was not accounted for. Current **roadside** models in the WSDOT risk-programming framework employ a simple least-squares approach, through the use of Roadside (YEAR) software, to model encroachment rates on the roadside. As is the case with roadway models, the roadside framework is limited in its accounting of factors, most noticeably weather effects, which are omitted.

In sum then, the broader research question posed in this study is, while establishing a "simultaneous" framework to examine possible gains in programming efficiency, what are the effects of weather and other unobserved behavior factors that may be creating interactions between the roadway and the roadside? An investigation into these issues will potentially address several important issues in the larger risk programming context, including but not limited to, a) efficiency of safety improvements through current practice, and b) the impact of weather and the relative importance of weather advisories in the form of intelligent transportation systems or roadway signing. Future research studies can build on these insights in probing the risk programming efficiency issue further: what level and type of information are required to minimize "programming turnovers," that is, the proportion of identified safety improvement locations that continue to be problematic hazardous locations? The other aspect of this "turnover" issue is the identification of safety improvements that have the greatest potential to minimize turnover through permanent reduction or elimination of accidents. It is not the goal of this study to examine the "turnover" issue; however, it is believed that insights from this study will lay an important foundation for the construct of a methodology for examining that issue.

LITERATURE REVIEW

A review of accident modeling literature reflects the variety of methods that have been used to model accidents. The conventional method is to use linear regression to model accident rates, a continuous number (for example Mulinazzi and Michael 1969; Shah 1968). This is a straightforward method that models the number of accidents per million vehicle miles for a given roadway segment. However, accident frequency counts modeled using linear regression can result in inconsistent parameter estimates. There are benefits, however, to the "rate" method. It provides for well-behaved estimators and can in fact be extended to non-linear forms through appropriate transformations (log-linear for example). Estimation of accident counts by using this method can also be carried out through non-linear least squares, ensuring consistent parameter estimates. The alternative for modeling accident frequencies is to use count models such as Poisson and negative binomial (and their suitable variations) models.

The important aspect of the "rate" method relevant to this study is the established properties of techniques involving "simultaneity" issues when more than one equation is employed to model the roadway and the roadside. In the linear regression context, simultaneity is modeled through a system of equations. Considerable research on the properties of system-of-equations estimators exists in the econometric literature (see Theil 1971 for example). When examining risk-programming efficiencies, this established knowledge provides sufficient inferential leverage. It minimizes uncertainty about parameter behavior and allows us to focus on the relevant programming issues – whether it is beneficial to employ methods that necessarily have to account for interaction between the roadway and the roadside, and in so doing, whether they identify the factors significantly at play.

More recent methods for modeling accident frequencies have included models such as the Poisson and negative binomial (see for example Shankar, Mannering and Barfield 1995; Poch and Mannering 1996; Milton and Mannering 1998) and the zero inflated Poisson and zero inflated negative binomial (for example Shankar, Milton, and Mannering 1997). While the count models have been shown to be superior econometric modeling tools for accident frequencies (Jovanis and Chang 1986; Joshua and Garber 1990; Miaou and Lum 1993), their properties in simultaneous contexts are relatively not well established. In addition, accident rate analysis should not necessarily be abandoned. The rate method employs an "exposure" type approach by using average daily traffic (ADT) in the computation of the accident rate. In the absence of segment-specific "safety performance" information, the accident rate may tend to overemphasize low-volume segments when low accident counts are observed. While this may be interpreted as a mathematical artifact of using a lower number (ADT) in the denominator of the accident rate function, it may also provide cues about the safety potential for that segment. If the few accidents that occur are of higher severity, such a problem could be exacerbated with increasing traffic without timely intervention and attention to the "exposure" effect.

Bayesian estimation is another econometric technique that has been used in the field of highway safety. Bayesian analysis produces a density function for the desired information rather than a point estimate given by OLS or other "frequentist" methods. The main difference between the Bayesian and "frequentist" methods is how probability is used. The Bayesian theory views probability as a confirmation of beliefs. Therefore, numerical probability is the confidence the researcher has in various parameter values. However, the "frequentist" methods regard probability as the frequency with which an event would appear in repeated sampling.

The Bayesian approach is well summarized by Peter Kennedy (1998) with the following steps:

1. The researcher establishes a distribution known as the "prior" based on previous "expert" knowledge of the parameter desired by the analysis before reviewing the data.

2. Then, Bayes' theorem is used to combine the data with the prior, which produces a distribution known as the "posterior." This is the main product of the Bayesian analysis.

3. Optionally, the posterior can be combined with a loss or utility function, depending on whether the desired result is minimizing loss or maximizing expected utility, respectively.

Kennedy also lists several advantages of the Bayesian approach over the classical methods:

1. The researcher's belief about the desired parameters and how this is affected by the data is one of the main issues in this approach, and several alternative hypotheses or models can be evaluated.

2. The prior can incorporate in a standardized method the extraneous information that would traditionally be left unused.

3. Through the selection of the loss function, the approach can be tailored to fit the scope and intention of the analysis, which lends itself to working well with decision analysis.

4. The prior and the sample data are the sole justification for the Bayesian approach. There is no need to justify the performance of the estimator in hypothetical repeated samples.

Despite these advantages, the reason that the Bayesian approach is not used frequently is the practicality of the approach. It is very complex and relies on human judgment to establish the prior. Also, complex software packages are needed to combine the prior and the data to produce the posterior.

Gary Davis, for example (2000), uses Bayesian estimation to calculate accident rates while taking into consideration the potential error in the estimated total traffic value. Traditional methods for calculating accident rates assume that the total traffic at a location or through a section is a known value. However, average annual daily traffic (AADT) numbers for a location are typically calculated on the basis of a sampling of traffic counts at that location. Using the Gibbs sampling approach, Davis' paper presents a Bayesian method to be used in estimating the accident rate while taking into consideration the potential error introduced by the traffic counts. On the basis of a two-day traffic and accident count at 17 sites in Minnesota, Davis found that using traditional methods of estimating accident rates underestimated the error of the rates. Use of the Bayesian method, which incorporated the traffic count errors, increased the standard deviation of the accident rates by anywhere from 12 to 40 percent, depending on the site. In other words, the traditional method makes the estimation seem more accurate than it really is.

Persaud, Lyon, and Nguyen (1999) explored a method for employing the empirical Bayes method to identify sites in need of safety improvement. They pointed out that conventional methods often wrongly identify "unsafe" sites on the basis of randomly high accident counts. This incorrect identification is due to the regression-to-the-mean assumption used in conventional estimation methods. However, they recognized that the data required to implement the method they present are not available to most highway agencies. As data collection resources improve, the Bayesian approach may become more practical. Nevertheless, as in case of the Poisson and negative binomial models, simultaneous Bayesian approaches (to model the roadway and the roadside simultaneously) are not well understood and hence not the most appropriate tools for examining our research objective.

In thinking about logical data inclusions in a simultaneous-versus-independent models approach, it is possible that different factors affect roadway and roadside accidents, respectively. A count model approach to examining this dimension was employed by Lee and Mannering (2000). However, that study did not examine possible interaction or correlation between the roadside and the roadway.

METHODOLOGICAL APPROACH – ACCIDENT RATE ANALYSIS

Econometric modeling has been used extensively for statistical modeling of accident rates in the past few years. However, little has been done to research the correlation between roadway and roadside accident rates. Frequently, the number of accidents used in determining accident rates is an aggregated total number of accidents, regardless of type of accident, or is limited to a specific type or severity (Milton and Mannering 1998; Shankar, Milton and Mannering 1997). While this may be the simplest approach, it is clearly probable that roadway and roadside accident rates may be correlated although influenced by different factors. To lump them into one category of overall accident rate may be too limiting, yet independently analyzing them may result in loss of useful information. Use of a single accident rate (sum of roadway and roadside accident rates) also leads to overestimation of the level of risk for some roadway sections. On the other hand, analysis of roadway and roadside accidents separately could decrease efficiency. The impact of this loss is loss in statistical efficiency of parameters in the model, at the very least.

A few issues are noteworthy at this point. There is something called endogeneity and contemporaneous correlation between the disturbance terms of the roadway and roadside accident rates, which have not yet been discussed. **Endogeneity** results from a bi-directional relationship in a linear regression: independent and dependent variables affecting each other. **Contemporaneous correlation** exists when the disturbance terms in the system of equations are correlated. Existence of endogeneity and contemporaneous correlation are not entirely ruled out in the case of roadway and roadside equations. Endogeneity alone can be tested using the Hausman test (Greene 2000). Presence of endogeneity alone will have to be accounted for by the use of the two-stage least squares estimator (2SLS), which is the asymptotically best instrumental variables estimator. Use of the instrumental variable (IV) and indirect least squares (ILS) methods will result in over-identification problems in estimating all the parameters of the system of equations. The Two-stage least squares method consists of using, as the instruments for the suspected endogenous variable, the predicted values in a regression of the endogenous variable on all the available exogenous variables. In the additional presence of contemporaneous correlation between the disturbance terms of the system of equations, 2SLS is not beneficial, and one must resort to the use of 3SLS for consistent estimates. The 3SLS estimation entails the computation of the GLS estimator and also an estimate of the asymptotic variance-covariance matrix for the estimator. In the case of not using these remedies for endogeneity and contemporaneous correlation, the estimates obtained with the CLR-OLS will suffer from simultaneous equation bias (Greene 2000).

The other scenario is the presence of possible contemporaneous correlation between the disturbance terms of the roadway and roadside accident rates. The comprehensive approach to this analysis is the use simultaneous equations for continuous data. Some empirical work, as described below, has been conducted in the field of traffic safety using Bayesian methods, but this work has noted that data limitations and limitations in the joint modeling of the roadway and roadside exist. The common, conventional approach is to assume an explicit bi-directional relationship between accident rates for the roadway and the roadside. This method can be used to incorporate the correlation between the two types of accident rates in a simple, more easily obtainable manner as opposed to Bayesian methods. For example,

$$\mathbf{Y}_{i,RW} = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Y}_{i,RS} \boldsymbol{\theta} + \boldsymbol{\varepsilon}_i \tag{1.1}$$

$$\mathbf{Y}_{i,RS} = \mathbf{Z}_i \boldsymbol{\gamma} + \mathbf{Y}_{i,RW} \boldsymbol{\lambda} + \boldsymbol{\eta}_i \tag{1.2}$$

where $\mathbf{Y}_{i,RS}$ is the roadside accident rate and $\mathbf{Y}_{i,RW}$ is the roadway accident rate, $\mathbf{X}_i$ and $\mathbf{Z}_i$ are vectors of traffic, geometric, and environmental factors, $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\lambda}$ are vectors of estimable parameters, and $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\eta}_i$ are error terms assumed to be normally distributed.

In the above equations, although the equations are not simultaneous and therefore do not interact, they may be connected because their error terms are related. Zellner (1962) first developed this estimation method, the seemingly unrelated regression estimation (SURE) model, using investment functions of General Electric and Westinghouse. This modeling framework was employed in this research project.

The SURE model is defined as follows:

$$y_i = X_i\beta_i + \varepsilon_i,\ \ i = 1,\ldots,M \tag{2.1}$$

where

$$\varepsilon = [\varepsilon'_1, \varepsilon'_2, \ldots, \varepsilon'_M,] \tag{2.2}$$

and

$$E[\varepsilon] = 0 \tag{2.3}$$

$$E[\varepsilon\varepsilon'] = V \tag{2.4}$$

A total of N observations are used to estimate the parameters of the M equations. The disturbances are assumed to be uncorrelated across observations, so

$$E[\varepsilon_{it}\varepsilon_{js}] = \sigma_{ij}, \text{ if } t = s, \text{ 0 otherwise.} \tag{3.1}$$

The disturbance formulation is

$$E[\varepsilon_i \varepsilon_j'] = \sigma_{ij} I_T$$

or

$$E[\boldsymbol{\varepsilon\varepsilon}'] = \mathbf{V} = \begin{bmatrix} \sigma_{11} I & \sigma_{12} I & \Lambda & \sigma_{1M} I \\ \sigma_{21} I & \sigma_{22} I & \Lambda & \sigma_{2M} I \\ & \mathrm{M} & & \\ \sigma_{M1} I & \sigma_{M2} I & \Lambda & \sigma_{MM} I \end{bmatrix} \tag{3.2}$$

Individually, each equation is a classical regression. Thus, the parameters could be estimated consistently, but not necessarily efficiently, by ordinary least squares (OLS). The generalized regression model applies to the stacked model,

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathrm{M} \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{X_1} & \mathbf{0} & \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{X_2} & \Lambda & \mathbf{0} \\ & & \mathrm{M} & \\ \mathbf{0} & \mathbf{0} & \Lambda & \mathbf{X_M} \end{bmatrix} \begin{bmatrix} \square_1 \\ \square_2 \\ \mathrm{M} \\ \square_{\square} \end{bmatrix} + \begin{bmatrix} \square_1 \\ \square_2 \\ \mathrm{M} \\ \square_{\square} \end{bmatrix} \tag{4.1}$$

$$= \mathbf{X\beta} + \boldsymbol{\varepsilon}.$$

Therefore, the efficient estimator is generalized least squares (Greene 2000). For the t^{th} observation, the M × M covariance matrix of the disturbances is

$$\Sigma = \begin{bmatrix} \sigma_1 & \sigma_{12} & \Lambda & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \Lambda & \sigma_{2M} \\ & \mathrm{M} & & \\ \sigma_{M1} & \sigma_{M2} & \Lambda & \sigma_{MM} \end{bmatrix} \tag{4.2}$$

so, in (3.2),

$$V = \Sigma \otimes I \tag{4.3}$$

and

$$V^{-1} = \Sigma^{-1} \otimes I \tag{4.4}$$

Thus, the generalized least squares (GLS) estimator is found to be

$$
\begin{aligned}
\beta &= [\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\
&= [\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{X}]^{-1}\mathbf{X}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{y}
\end{aligned}
$$

$$
= \begin{bmatrix} \sigma_{11}\mathbf{X}_1'\mathbf{X}_1 & \sigma_{12}\mathbf{X}_1'\mathbf{X}_2 & \cdots & \sigma_{1M}\mathbf{X}_1'\mathbf{X}_\mathrm{M} \\ \sigma_{21}\mathbf{X}_2'\mathbf{X}_1 & \sigma_{22}\mathbf{X}_2'\mathbf{X}_2 & \cdots & \sigma_{2M}\mathbf{X}_2'\mathbf{X}_\mathrm{M} \\ & & \vdots & \\ \sigma_{M1}\mathbf{X}_\mathrm{M}'\mathbf{X}_1 & \sigma_{M2}\mathbf{X}_\mathrm{M}'\mathbf{X}_2 & \cdots & \sigma_{MM}\mathbf{X}_\mathrm{M}'\mathbf{X}_\mathrm{M} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^{M}\sigma_{1j}\mathbf{X}_1'\mathbf{y}_j \\ \sum_{j=1}^{M}\sigma_{2j}\mathbf{X}_2'\mathbf{y}_j \\ \vdots \\ \sum_{j=1}^{M}\sigma_{Mj}\mathbf{X}_M'\mathbf{y}_j \end{bmatrix} \tag{4.5}
$$

when the ij^{th} element of $\boldsymbol{\Sigma}^{-1}$ is denoted as σ_{ij}. The first matrix in (4.5) is the asymptotic covariance matrix for the GLS estimator. However, this assumes that $\boldsymbol{\Sigma}$ in (4.2) is known, which is usually not the case; hence, the common recourse is to employ the feasible generalized least squares (FGLS) estimators. The FGLS estimator is denoted by

$$
\hat{\hat{\boldsymbol{\beta}}} = [\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\hat{\Sigma}^{-1}\mathbf{y}. \tag{4.6}
$$

where the variance matrix is estimated. This estimator is different from the OLS estimator because the equations are linked. However, because they are only linked through the disturbances, the model turns out to be a *seemingly unrelated regression estimation (SURE) model*. Gains in efficiency by using the GLS instead of the OLS model are data dependent. In the following, Greene (2000) has presented some specific cases in which there is no efficiency gained by using GLS (p.616):

1. If the equations are not related through the error terms so that $\sigma_{ij} = 0$ for $i \neq j$, then no efficiency is gained and GLS is equal to OLS.

2. If the explanatory variables for all the equations are the same, $\mathbf{X}_i = \mathbf{X}_j$, then GLS is equal to OLS.

3. If the regressors in one equation or a block of equations are a subset of those in another, no efficiency is gained for the smaller subset.

Greene has also related some guidelines for maximum efficiency gains. Greater efficiency gains from GLS are found with the following:

1. a greater correlation of the disturbances
2. less correlation between the **X** matrices.

In our study, the stacked equation for the roadway-roadside system consists of two equations, with a set of explanatory factors including roadway and roadside geometrics, traffic factors, and weather factors. The employment of the SURE approach stems from our base hypothesis that roadway and roadside accident occurrences are not related directly but indirectly through unobserved effects that are common. For example, a one-mile segment of highway may have driver behavior factors that commonly affect the roadway and roadside within that one-mile length cross-section. There is no theoretical support for an explicit bi-directional relationship between accident propensities on the roadway and the roadside, as equations 1.1 and 1.2 suggest.

EMPIRICAL SETTING

The Washington State highway system contains over 7,000 centerline miles of state highways. A random sample of one-mile sections was used to estimate and calibrate the accident rate models. The sections for which the data were collected included all classes of roads, from arterials to interstates. Accident data for the years 1995 and 1996 was used to estimate the models. Segment averages over the two-year period were used to compute the dependent variable, accident rate per million vehicle miles. Sections with construction related accidents during 1995 or 1996 were not included, as accurate geometric and traffic data during construction were not available. It is possible that selectivity bias may have occurred by omitting these sections, since they were being improved. However, we believe the bias to be minimal because the sample characteristics without the construction sections were still consistent with the population characteristics. Highway sections that are routinely closed during portions of the year were also omitted, as were sections with incomplete geometric or traffic data.

A graphical view of the accident rate distribution for the roadway and roadside is presented in figures 1 and 2. Figure 1 shows the number of sections that have a combination of zero and non-zero accident rates for the roadway and roadside. It can be seen that an equal number of sections have either zero or non-zero accident rates for both roadway and roadside scenarios. It is less common to have non-zero roadside accident rates for roadway sections that have zero accident rates than to have zero roadside accident rates for sections that have non-zero roadway accident rates. One cannot even briefly conclude that safe roadway sections generally imply corresponding safe roadside sections, or that unsafe.

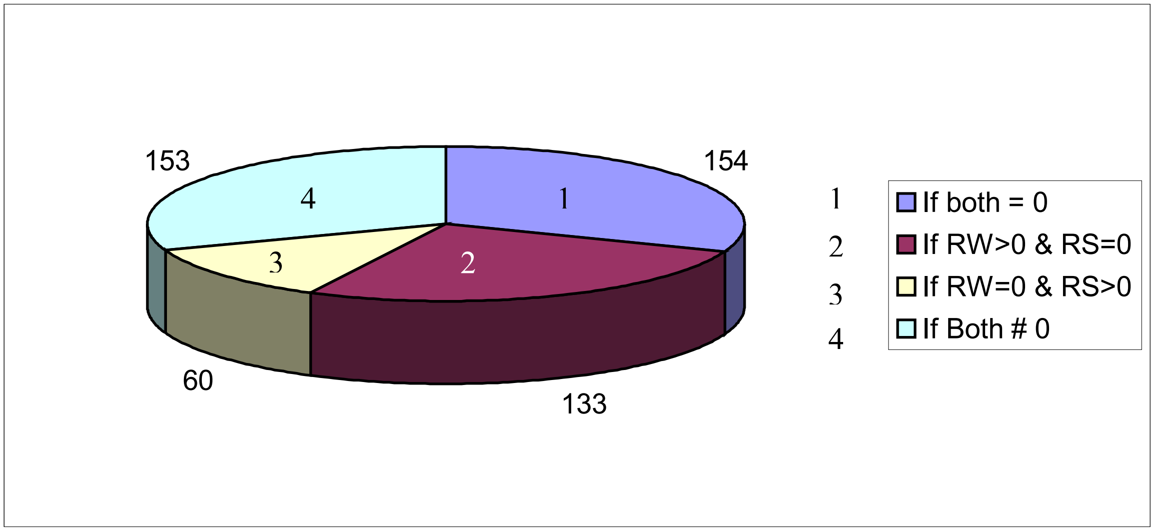

Figure 1. Pie chart showing the share of sections with combinations of zero and non-zero roadway and roadside accident rates.

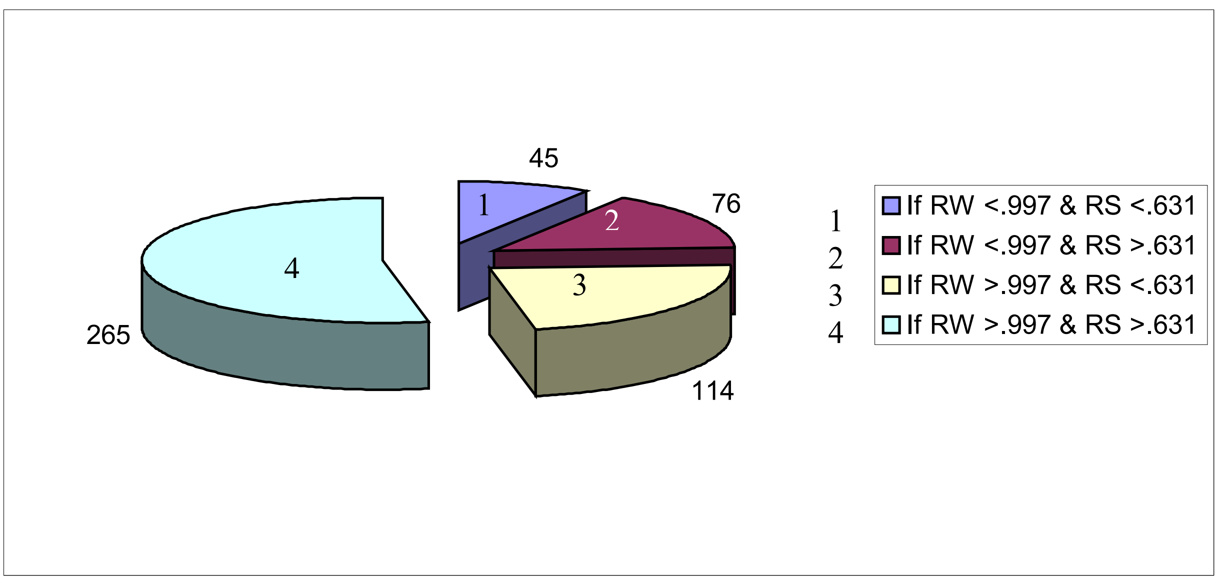

Figure 2: Pie chart showing the sections with combinations of high and low roadway and roadside accident rates

roadway sections do not necessarily imply unsafe roadside sections from the chart. This would only be clear after modeling the roadside and roadway accident rates as seemingly unrelated regression equations. The other pie chart (Figure 2) shows similar characteristics for the distribution of roadway and roadside sections that have combinations of high and low roadway and roadside accident rates. It is evident from the figure that a very high share of the sections have a combination of high roadway and roadside accident rates, and a very small share of sections have a combination of low roadway and roadside accident rates

The geometric and traffic data for each highway section were taken from the WSDOT highway geometric/traffic database. This database contains information for all the state highways divided into varying highway sections lengths, with each section having homogenous characteristics. The data available from this database include geometric data such as roadway widths, lane widths, number of lanes, shoulder widths, horizontal curve information, median widths, and barrier types; traffic data such as average annual daily traffic (AADT), truck volume as a percentage of AADT, and peak hour volumes as a percentage of AADT; and other data such as roadway classification, surfacing type, terrain, access restrictions and legal speed limit. Because modeling varying section lengths could introduce unnecessary heteroskedasticity, this project used one-mile sections for theoretical and practical reasons. From a practical standpoint, using one-mile sections is easier for program applications. Theoretically, using one-mile sections cancels out aggregation effects, which might introduce unnecessary heteroskedasticity arising from unequal sections. Heteroskedasticity would result in inefficient estimates of the parameters. Approximately 690 one-mile segments were sampled from the 7,000-mile system. A 10 percent sample was the target, ensuring adequate representation of all functional classes, including interstates andprincipal, minor and collector arterials. Out of the 690 one-mile segments, 500 were used for model estimation, and the remainder were used for out-of-sample prediction testing. The geometric and traffic data for the one-mile sections were aggregated by using a weighted average from the section lengths listed in the database. Minimum, maximum, and weighted averages were recorded for data such as roadway widths, shoulder widths, traffic volumes, number of lanes, grades, curvature, and speed limits.

Historical weather data were collected for each one-mile roadway section. The data were taken from 240 National Oceanic and Atmospheric Administration (NOAA) weather stations located around Washington State. The level of detail for the weather data that were to be used for the model estimation had to be agreed upon. Microscopic data would give more information and capture all the variations in temperature, precipitation, and snowfall, whereas yearly or seasonal averages would sum up the weather characteristics of the one-mile section but would not show the variations in the weather conditions at the microscopic level. We decided upon the monthly/seasonal averages, as microscopic weather data were too detailed for the model and would render the model too complex to interpret. Also, collection and compilation of microscopic weather data would be too painstaking for their use in the model. The monthly/seasonal averages were considered to be at the right level of detail for the model and advantageous for interpretation and prediction.

A comprehensive geographic information system (GIS) procedure was employed to obtain the weather data for the sections. In summary, the location of each one-mile section was determined and matched to the nearest weather station. The weather data available for the weather stations were then compiled for the corresponding one-mile section. A detailed account of the procedure employed for matching the one-mile sections to their nearest weather stations is presented below:

The GIS software ArcView 3.2 was used to map sections to weather stations. The Washington State GIS files were obtained, and the extensions for coordinate plotting and finding the nearest neighbor were loaded into the extension folder. The state GIS map was loaded and mapped in the Base Map. A table with State Route (SR) information was created and exported as delimited text to form the SR attribute table. This table was used along with the roadway section table to map state route identification (Ids) to state route numbers in the SR table (this operation could be performed manually or, better, with the help of Microsoft Excel). The resulting information table held information for SR number, SR ID, beginning SR- milepost, and ending SR milepost. The table was named and referred to as "Roadsections.txt". The road sections table was then loaded into the ArcView, and it was assured that the sections were arranged in the order of SR and beginning SR milepost. Then, the weather stations table was added to ArcView. This table contained information about weather station identification and longitude and latitude in degrees decimal. The road sections table and weather table were then plotted on the SR highway map in ArcView. To perform the operations in ArcView, the plots for the roadway sections and weather stations had to be in the form of shapefiles, the form compatible with ArcView (or even other GIS software). Therefore, these table text themes were converted to shapefiles. Using the Coordinate utility extension, the coordinates for the shapefiles were converted from the latitude-longitude to planar X,Y coordinates. This was for the convenience of analysis. Once the roadway sections and weather shapefiles had been created and the coordinate system had been defined by the planar X,Y coordinates, the Nearest Neighbor extension was used to map the one-mile sections to their nearest weather station. The weather information from the stations was used for the corresponding nearest sections. Unfortunately, the weather information was not available for some of the stations. Many alternative options were considered, and using the weather information from the second nearest station was found to be the best alternative in these cases. So stations with missing information (in the first attempt) were removed from the list of weather stations, and the mapping was done with the remaining stations. This resulted in mapping the sections to the second nearest station if the nearest station lacked weather data. Three iterations were required to arrive at the final list of nearest stations and the corresponding weather attributes. The resulting table had the nearest station, its distance from the midpoint of the section, and the coordinates of the station.

This procedure was found to be highly effective in mapping out the sections with their nearest weather stations for the required weather information. This procedure also accommodated alternative tactics for mapping the second or third nearest neighbor stations when data for the nearest station were missing.

For some weather stations, data had been collected since the 1930s. In general, 30-year histories were available for all stations. The data gathered were the average maximum monthly temperature, average minimum monthly temperature, average monthly total precipitation, average monthly snowfall, and average monthly snow depth for all 12 months, and the annual averages for temperatures, precipitation, snowfall, and snow depth. Note that 1995 and 1996 weather data were not significantly different from the longer 30-year historical averages. This offered an advantage from the standpoint of prediction. Given that the year-specific weather observations were statistically similar to the 30-year averages, it was unlikely that time effects specific to those years would be significant in the estimated models. From a prediction standpoint, this is advantageous because historical averages can be used for future years.

Monthly weather data were too disaggregate to be of added value (in the statistical sense) to the models. In particular, consecutive months had similar values for temperature and precipitation, which showed up as significant in the models, but with parameters that were statistically indifferent from each other. For this reason, the data were aggregated to clarify the variability at the seasonal level. Maximum average temperature, minimum average temperature, total precipitation, total snowfall, and total snow depth for winter (November – March), spring (April – June), summer (July and August) and autumn (September and October) months were computed for possible specification in the accident rate models. These combinations were chosen on the basis of their similar values for each category, i.e., when variation within that period was minimal.

To determine the relationship between roadway and roadside accidents rates and the roadway geometrics, traffic conditions, and weather factors, accident counts for the one-mile roadway section were taken from the Washington State accident database. The total number of reported roadway and roadside accidents within each section during 1995 and part of 1996 were determined from the database. Only 7 months of data were available for 1996. More recent accident and geometric data were not available.

Accidents occurring within the traveled way were classified as roadway accidents. Roadside accidents were defined as accidents reported to have occurred off the roadway. By definition, this included accidents occurring off the traveled way as a result of encroachments, where objects such as trees, fences, ditches, guardrail, signposts, utility poles, and bodies of water were struck, as well as vehicle rollovers off the roadway.

The accident rates for both roadway and roadside accidents were calculated for each year with the following equation:

$$Accident\ rate = \frac{\#\ accidents \times 10^6}{\#\ days \times AADT \times 1.0\ mile} \tag{5.1}$$

where "# accidents" for 1995 was the total number of accidents for the year and for 1996 was the total for the first seven months. The "# days" was 365 for 1995 and 212 for 1996. The AADT used in the accident rate calculation was a weighted average of the AADT for the section. Thus, the accident rate calculated according to the above formula was the number of accidents in the section per million veh-km.

A sample of 500 one-mile sections was used to estimate the coefficients for the OLS and SURE models. Statistics for the sample can be found in Table 1. A cursory glance at the table presents a few statistics about the accident rates and the roadway types. The mean roadway accident rate was found to be 0.99, and the mean roadside accident rate was found to be 0.63. Around 34 percent of the observations come from the data for principal arterials, 32 percent of the roads were collectors, 24 percent of them were minor arterials, and 10% of them were interstates.

Table 1 Sample Statistics (not all variables are shown)

Variables	Mean	Std. Dev.	Min	Max
General variables				
Principal Arterial indicator (1 if the section is part of a principal arterial, 0 otherwise)	0.344	0.4755	0	1
Collector indicator (1 if the section is part of a collector, 0 otherwise)	0.322	0.4677	0	1
Minor Arterial indicator (1 if the section is part of a minor arterial, 0 otherwise)	0.236	0.425	0	1
Interstate indicator (1 if the section is part of an interstate, 0 otherwise)	0.098	0.2976	0	1
Urban indicator (1 if the section is in urban area, 0 otherwise)	0.126	0.332182	0	1
Traffic Variables				
Percent change in AADT (max-min)/min	0.1354	0.418709	0	4.2
AADT per the number of lanes	2653.08	3405.04	72	29906.8
Access type indicator 1 (1 if access type is most restrictive or fully controlled, 0 otherwise)	0.264	0.441241	0	1
Access type indicator 2 (1 if access type is partially controlled, 0 otherwise)	0.672	0.469955	0	1
Access type indicator 3 (1 if access type is not controlled, 0 otherwise)	0.06	0.237725	0	1
Average speed limit in miles per hour	54.516	8.60056	25	70
Natural logarithm of AADT per the number of lane	7.31241	1.11025	4.27667	10.3058
Minimum AADT in section	7151.23	13034.9	144	117272
Single Truck percentage	6.2562	2.76825	0	16.8
Truck total percentage	14.5638	9.51823	0	71.5
Truck-Train percentage	2.3834	4.13737	0	43.4
Total truck percentage indicator (1 if total truck%>25, 0 otherwise)	0.108	0.3107	0	1
Total truck share indicator (1 if total truck % < 5), 0 otherwise	0.058	0.234	0	1
Parking indicator (1 if parking is permitted at sometime during the day, 0 otherwise)	0.074	0.262033	0	1
Peak hour indicator (1 if peak hour >12%, 0 otherwise)	0.116	0.320546	0	1
Weather Variables				
Average temperature (of max and min's)	49.7708	2.85552	37.6	55.3
Low temperature indicator (1 if average temperature < 46 ºF, 0 otherwise)	0.108	0.310691	0	1
High temperature Indicator (1 if average temperature > 52 ºF, 0 otherwise)	0.174	0.379489	0	1
Average difference between minimum and maximum monthly temperatures for spring months (Apr-Jun)	24.3819	4.21683	13.333	31.967
Average difference between minimum and maximum monthly temperatures for summer months (Jul-Aug)	28.8825	5.62932	12.05	39.45
Average difference between minimum and maximum monthly temperatures for winter months (Nov-Mar)	15.1612	2.10547	9.42	19.92
Total precipitation for autumn months (Sep-Oct)	4.3074	3.51382	0.8	17
Total precipitation for Winter and Spring months	26.5682	21.3029	5.5	97.9
Total precipitation for summer months (Jul-Aug)	1.7288	1.04495	0.3	5.8

Table 1. Sample Statistics (continued)

Variables	Mean	Std.Dev.	Min	Max
Average annual maximum temperature	60.4544	2.92946	45.1	66
Minimum average min monthly temperature for winter months (Nov-Mar)	25.984	6.50315	10.5	36.8
Number of months with total precipitation > 9 inches	0.584	1.45436	0	6
Total annual precipitation indicator (1 if total annual precipitation > 50 inches, 0 otherwise)	0.236	0.425047	0	1
Number of months when minimum temperature < 33 ºF	3.288	2.2608	0	8
Indicator no minimum temperature months below 33 ºF	0.166	0.3724	0	1
Number of months when snow depth > 0	1.848	1.8847	0	7
Roadway Geometric Variables				
Horizontal curve central angle in degrees	29.5502	27.3606	0	168.5
Horizontal curve indicator (1 if there is horizontal curve in section, 0 otherwise)	0.792	0.406283	0	1
Horizontal curve central angle indicator (1 if horizontal curve central angle >= 50, 0 otherwise)	0.206	0.404835	0	1
Maximum shoulder width feet	6.67	3.65401	0	35
Median indicator 1 (1 if median begins or ends in this section, 0 otherwise)	0.016	0.125601	0	1
Right turn lane indicator (1 if the section has right turn lane, 0 otherwise)	0.1	0.3003	0	1
Average right shoulder width in feet	5.5586	2.91793	0	12
Radius interval Indicator 3 (1 if radius interval is greater than 3, 0 otherwise)	0.148	0.355456	0	1
Intersection indicator (1 if the section has intersections, 0 otherwise)	0.71	0.454217	0	1
Number of intersections in the section	2.202	2.6907	0	16
Indicator if the number of intersections in a section > 5	0.102	0.3029	0	1
Lane difference indicator (1 if min and max number of lanes in section are not equal, 0 otherwise)	0.042	0.20079	0	1
Special lane indicator (1 if a special lane exists in the section, 0 otherwise)	0.07	0.255403	0	1
Two-way left turn indicator (1if a two-way left-turn lane in section, 0 otherwise)	0.04	0.196155	0	1
Left turn lane indicator (1 if the section has left turn lane, 0 otherwise)	0.174	0.379489	0	1
Terrain indicator 1 (1 if terrain is all level or level & rolling, 0 otherwise)	0.23	0.421254	0	1
Indicator if terrain is all rolling, rolling & level or rolling & mountainous	0.746	0.4357	0	1
Indicator if terrain is all mountainous or mountainous & rolling	0.066	0.2485	0	1
Indicator if terrain changes in the section	0.042	0.2008	0	1
Median barrier indicator (1 if the median barrier type changes within the section, 0 otherwise)	0.024	0.1532	0	1
Median barrier type indicator (1 if the median barrier type is depressed, 0 otherwise)	0.132	0.3338	0	1

Table 1. Sample Statistics (continued)

Variables	Mean	Std.Dev	Min	Max
Interaction Variables				
Interaction variable between horizontal curves and posted speed greater than 50 mph (1 if horizontal curve and speed > 50 mph, 0 otherwise)	0.566	0.49612	0	1
Interaction between horizontal curves and intersections indicator (1 if there are horizontal curves and intersections in section, 0 otherwise)	0.546	0.49837	0	1
Interaction variable between total precipitation for Summer months and level terrain	0.3932	0.87530	0	4.1
Interaction variable between total precipitation for Winter months and level terrain	4.983	12.5073	0	81
Interaction variable between snow depth and horizontal curve (1 if snow depth > 0 for any month and horizontal curve exists in the section, 0 otherwise)	0.482	0.5001	0	1
Interaction variable between interaction and posted speed higher than 50 mph (if intersection exists in the section and speed > 50 mph, 0 otherwise)	0.44	0.4968	0	1
Number of indicators in sections with posted speed limit > 45 mph	1.348	1.8115	0	12
Number of intersections in sections with posted speed limit <= 45 mph	0.854	2.503	0	16
Indicator if peak hour > 12% and total truck % >25	0.014	0.1176	0	1
Indicator if total truck % > 25 and the section has rolling or mountainous terrain	0.094	0.2921	0	1
Indicator if snow depth > 0 and for any month and a horizontal curve is present in the section	0.482	0.5	0	1
Indicator if radius interval is nonzero and right shoulder width is less than 3 feet	0.164	0.3706	0	1

A few important aspects about the descriptive statistics of the variables collected for the 500 one-mile sections are discussed in this section. As shown in Table 1, the ADT per lane varied from 72 to a high of 29,907, with a mean of 2653, indicating a high share of roads with high ADTs. The percentage of change in ADT for the component sections within the one-mile section varied from 0 to 4.2, with a mean change of 0.135 percent, implying that most of the sections did not have a high proportion of ADT change. However, with the given scale of ADT, it still worked out to be a high number for change in ADT. The natural logarithm of ADT per lane variable was formed to avoid possible heteroskedasticity due to the high scale of the 'ADT per lane' values. It showed a mean of 7.3 with a minimum and maximum of 1.1 and 4.3, respectively. The minimum ADT ranged from 144 to 117,272, with a mean ADT of 7151. Variables were created for the truck percentages. Single, truck-train, and total truck percentages were found to be an average of 6.26, 2.38, and 14.56, respectively. It is interesting to note that the total truck percentage was as high as 71.5 percent in at least one section, while the truck-train composition reached as high as 43.4 percent in some sections. Also, the share of total trucks in the traffic was higher than 25 percent in 54 sections, and it dipped below 5 percent in 29 sections. Regarding the peak hour,11.6 percent of the sections had a peak hour that accounted for more than 12 percent. The data indicated that only 7.4 percent of the sections allowed parking at some time of the day. This indicates that even among some of the local arterials, the sections that did not allow parking were included, which means that there was not an over- or under-representation of the roads with respect to parking. The access type indicator revealed that 67.2 percent of the sections were partly controlled, 26.4 percent of them were fully controlled, and a very low 6 percent of them were not controlled. The sample contained all classes of roads, ranging from arterials to interstates, and the presence of minor and major collector arterials that were not usually controlled resulted in a high share of partially accessible roadway sections. The average speed limit for the sections was found to vary from 25 miles per hour to 70 miles per hour, with a mean speed limit of 54.5 miles per hour. This indicates the presence of a great number of high-speed sections in the database. However, these sections were not necessarily interstates, as shown by the significant presence of other classes of roadway sections from the other indicators.

About roadway geometrics, the horizontal curve central angle indicator revealed that about 80 percent of the sections had horizontal curves with a central angle of higher than 50. In addition, the central angle for the horizontal curves varied from 0 to 168.5, with an average angle of 29.5, indicating a high presence of low angle curves. The high mean of 0.792 for the horizontal curve indicator meant that 79.2 percent of the sections had one or more horizontal curves within the section. There were a few sections with no shoulders, as well as sections with an average shoulder width of up to 12 feet. The maximum shoulder width for the component sections within the one-mile section ranged from 0 to 35 feet. As few as eight sections out of the 500 sections had a median ending or beginning within the section. The median barrier type changed within 2.4 percent of the sections, and about 13.2 percent of the sections had a median barrier type that was depressed. Regarding turning lane indicators, 20 sections had one or more two-way left turn lanes within the section, 35 of them had a special lane, 50 had a right turn lane, and 87 had left turn lanes. The presence or absence of turning lanes will be later shown to have a significant impact on the roadway accident rate. On an average, 2.2 intersections were found to be present in a single section. Interestingly, a very high 71 percent of the sections had at least one intersection within them. The intersection indicator, along with the presence of horizontal curves and posted speed limits, significantly explains roadway accident rates, as will be shown later. The sections that had posted speed limits of greater than 45 mph had around 1.3 intersections per section on an average, while those with posted speed limits of less than 45 mph had only 0.85 intersections per section. The table shows that around 4 percent of the sections had a terrain change within the section. Also, about 75 percent of the sections were all rolling, or rolling and mountainous, or rolling and level; 23 percent of the sections were all level or level and rolling, and only about 7 percent of the sections were all mountainous or mountainous and rolling.

The average temperature, which was the mean of maximums and minimums of temperature within a single section, was found to vary from 37.6 to 55.3, with an average of around 50 degrees. About 11 percent (55) of the sections had an average temperature of less than 46ºF, indicating a significant number (though not very high) of sections in low temperature regions in the State of Washington. The high temperature indicator shows that 87 sections had an average temperature of greater than 52ºF. This means that the rest of the sections fell into the category of average temperature between 46ºF and 52ºF. The number of months when minimum temperature was below 33ºF was found to be 3.3 on an average. And for about 17 percent of the sections, the minimum temperature did not go below 33ºF. We also looked at variables depicting the difference between the maximum and minimum monthly temperatures for different months in a single season. The mean values for these differences were found to be 24.4 for spring months, 28.9 for summer months, and 15.2 for winter months. The number of sections that had a total annual precipitation of more than 50 inches was 118, and the number of months for which the total precipitation was greater than

9 inches ranged from 0 to 6, with a mean of 0.584. The average annual maximum temperature ranged from 45 to 66ºF, with a mean average maximum temperature of approximately 60ºF. There was not significant variation in the value of this variable among the sections. The precipitation variables revealed that the total precipitation for the autumn months (Sep-Oct) was 4.3 inches, while that for winter and spring months together was 26.6 inches, and for summer months (July-Aug) it was 1.73 inches. The number of months when the snow depth was greater than 0 inches was 1.85 on average, while just less than half the number of sections had some amount of snow for any month and also have a horizontal curve.

We also developed a few interaction variables that might possibly explain the correlation between the disturbances across the roadway and roadside accident rates. These variables were developed from interacting the different categories of variables, such as traffic variables, geometrics, and weather variables. The interaction variable between the horizontal curve and posted speed limits showed that more than half the sections had a horizontal curve in them and a posted speed limit of greater than 50 mph. About 55 percent of the sections had both a horizontal curve and an intersection within the section. This variable is potentially significant in explaining the roadway accident rate as it influences the difficulty of driving at curves. About 240 sections had a horizontal curve and snow present, while 220 sections had an intersection present and the posted speed limit exceeding 50 mph. The interaction indicator between peak hour and truck percentages showed that only 1.4 percent of the sections had both peak hour greater than 12 percent and total truck composition greater than 25 percent. Another indicator showed that around 10 percent of the observations had sections with a total truck percentage of greater than 25 and a rolling or mountainous terrain. The interaction variable between horizontal curve and right shoulder showed that 16.4 percent of the sections had a non-zero radius interval and a right shoulder width of less than 3 feet. This variable indicates the level of difficulty in maneuvering the curve and the amount of leeway for loss of control.

MODEL ESTIMATION

Most of the explanatory data found in Table 1, as well as some data interactions not listed, were used in the software program named Esther (Ulfarsson 2001) to choose the significant variables for the OLS model. Esther automates the process of removing insignificant variables one at a time until only significant variables are left in the model. Various scenarios were tested to estimate the possible form of relationship between the roadway and roadside accident rates. As mentioned in the Methodology section, endogeneity and contemporaneous correlation were possible. 2SLS and 3SLS estimators were obtained, but they did not prove to be beneficial, as the models showed lack of endogeneity. The roadside accident rate variable did not find significance in the roadway equation and vice-versa, indicating that the extent of endogeneity was not significant. Also, the OLS estimates were still consistent, indicating that the existence of endogeneity and its coexistence with contemporaneous correlation between the disturbance terms were not significant in the scenario of roadway and roadside accident modeling.

Once the best specifications had been found, SURE models were developed and compared with their OLS counterparts for efficiency differences. As mentioned previously, if significant correlation existed in the unobserved effects common to the roadway and the roadside, the SURE model would be bound to provide more efficient parameters. The upshot of this is that the right set of explanatory variables would be identified. In fact, it might help identify variables that were thought to be insignificant in single-equation OLS models. Another prospect of gaining efficiency would be if the set of variables affecting roadway accidents was significantly different from those affecting roadside accident occurrences.

Table 2 lists the coefficients, standard errors, and t-statistics for both the OLS and SURE models for roadway accident rates, and Table 3 lists the same for roadside accident rates. The efficiency gained from using the SURE model over the OLS model was about 0.62 percent at the greatest. A couple of factors may explain the minimal efficiency gains. First, as mentioned previously, if the **X** matrices were identical or included mostly the same variables, no efficiency would be gained. This was certainly not the case since very little overlap occurred between the two models. Second, there might not be any significant correlation between the roadway accident rate and roadside accident rate. Most likely, this is the reason for the lack of efficiency gains. The physical factors affecting roadway and roadside accidents are different. Including historical weather data may have increased the explanatory power of the models, making the SURE model unnecessary. Accounting for weather factors explicitly in the model specification minimized unobserved effects common to the roadway and the roadside, and possibly minimized the advantage of the SURE estimator. Absent the weather effects, the gain in efficiency in other parameters was 1.54 percent at the largest. We loosely termed this an efficiency gain, because in fact the parameters for the remaining variables were biased because of omitted variable effects. In comparing significant explanatory variables between the roadway accident rate and roadside accident rate models, only one variable was common. Clearly, different factors affect the accident rates, including weather. While this was not consistent with our common expectation that weather factors would be similar between roadway and roadside accident phenomena, it underscores the importance of interactions between weather and geometrics.

Table 2 Comparison of ordinary least squares and seemingly unrelated regression models for roadway accident rates with weather variables

Variables	OLS			SURE		
	Estimated coefficient	Standard error	t-statistic	Estimated coefficient	Standard error	t-statistic
Equation 1: Roadway equation (dependent variable)						
Constant	5.14644	1.08079	4.762	5.14555	1.06424	4.835
Percent change in AADT (max-min)/min	0.61939	0.19031	3.255	0.62029	0.18743	3.309
Natural Log of (average AADT/ numbers of lanes)	-0.34699	0.10836	-3.202	-0.34710	0.10670	-3.253
Horizontal curve indicator (1 if there is horizontal curve in section, 0 otherwise)	-0.77440	0.28659	-2.702	-0.77047	0.28219	-2.730
Interaction variable between horizontal curves and posted speed greater than 50 mph (1 if horizontal curve and speed > 50 mph, 0 otherwise)	1.22923	0.31994	3.842	1.21759	0.31502	3.865
Intersection indicator (1 if the section has intersections, 0 otherwise)	1.02192	0.31196	3.276	1.01337	0.30716	3.299
Interaction variable between intersection and posted speed higher than 50 mph (if intersections and speed > 50 mph, 0 otherwise)	-0.99533	0.31073	-3.203	-0.98579	0.30595	-3.222
Right turn lane indicator (1 if the section has right turn lane, 0 otherwise)	0.86944	0.28844	3.014	0.87049	0.28399	3.065
Two-way left turn lane indicator (1if a two-way left-turn lane in section, 0 otherwise)	1.01096	0.42529	2.377	1.01946	0.41874	2.435
Access type indicator 2 (1 if access type is partially controlled, 0 otherwise)	-0.49608	0.18045	-2.749	-0.49039	0.17767	-2.760
Percentage of truck-train	0.07445	0.03602	2.067	0.07575	0.03547	2.136
Percentage of total trucks	-0.03748	0.01697	-2.209	-0.03793	0.01671	-2.270
Average right shoulder width in feet	-0.12119	0.04917	-2.465	-0.11935	0.04841	-2.465
Maximum shoulder width feet	0.06019	0.03423	1.758	0.05945	0.03371	1.764
Average difference between minimum and maximum monthly temperatures for summer months (Jul-Aug)	-0.04221	0.01609	-2.623	-0.04217	0.01585	-2.662
Number of observation	500			500		
R^2	0.153			0.153		
Adjusted R^2	0.129			0.129		

Table 3 Comparison of ordinary least squares and seemingly unrelated regression models for roadside accident rates with weather variables

Variables	**OLS**			**SURE**		
	Estimated coefficient	Standard error	t-statistic	Estimated coefficien t	Standard error	t-statistic
Equation 2 : Roadside equation (dependent variable)						
Constant	0.97991	2.22330	0.441	1.02046	2.18910	0.466
Average speed limit in miles per hour	-0.02217	0.00858	-2.584	-0.02213	0.00845	-2.620
Percent change in AADT (max-min)/min	0.34290	0.15184	2.258	0.34314	0.14954	2.295
AADT/ numbers of lanes	-0.00017	0.00006	-3.080	-0.00017	0.00006	-3.116
Minimum AADT in section	0.00003	0.00001	1.966	0.00003	0.00001	1.982
Horizontal curve central angle in degrees	0.00620	0.00225	2.758	0.00616	0.00221	2.783
Special lane indicator (1 if a special lane is in the section, 0 otherwise)	0.57241	0.24083	2.377	0.57009	0.23713	2.404
Parking indicator (1 if parking is permitted at sometime during the day, 0 otherwise)	-0.58830	0.25537	-2.304	-0.59155	0.25144	-2.353
Average annual maximum temperature	-0.09682	0.04980	-1.944	-0.09678	0.04904	-1.973
Average temperature (of max and min's)	0.14065	0.05913	2.379	0.13970	0.05822	2.399
Low temperature indicator (1 if average temperature < 46 ºF, 0 otherwise)	0.63228	0.28377	2.228	0.62542	0.27940	2.238
High temperature Indicator (1 if average temperature > 52 ºF, 0 otherwise)	-0.43284	0.23893	-1.812	-0.42565	0.23526	-1.809
Total precipitation for autumn months (Sep-Oct)	0.43403	0.13653	3.179	0.42948	0.13443	3.195
Total precipitation for Winter and Spring months	-0.05290	0.01807	-2.928	-0.05221	0.01779	-2.935
Total precipitation for summer months (Jul-Aug)	-0.41247	0.17408	-2.369	-0.40991	0.17140	-2.392
Number of observation R^2 Adjusted R^2	500 0.116 0.091			500 0.116 0.091		

Employing the same weather factors would not only reduce the explanatory power of the model but also impose undue restrictions on interactions with geometrics, thereby skewing our understanding of the multidimensional impact of weather. In addition, it is also clear that weather effects added to the specification, albeit at the level of seasonal weather station observations, did improve the specifications significantly enough to justify OLS-based structures as opposed to simultaneous structures such as the SURE model.

ROADWAY ACCIDENT MODEL

The tabular results for the model estimates of roadway accident rate models are presented in Table 2.

The variables related to average annual daily traffic (AADT) significant in the roadway model were percentage of change in AADT and the logarithm of per-lane AADT. As might be expected, percentage of change in AADT within a segment was positively correlated with roadway accident rate. The average change in AADT was about 0.135 percent, with the maximum change being over 4.2 percent. The logarithm of AADT per lane was negatively correlated with roadway accident rate. (The logarithm of average AADT per lane was a surrogate for density of traffic.) The negative correlation of the traffic density indicator may suggest that low traffic density increases roadway accident rates when all other factors are taken into account. The negative coefficient for the variable indicating the presence of a horizontal curve in section suggests that the roadway accident rate decreases in the presence of a horizontal curve, but increases if the posted speed at the section is greater than 50 mph.

As might be expected, the intersection indicator (1 if a section contains one or more intersections) had a positive coefficient. Intersections increase the number of conflict points and increase the frequency of accidents. This may also be the reason that the right turn lane indicator and two-way left-turn lane indicator had positive coefficients. Turn lanes are usually thought to decrease the frequency and severity of accidents; however, they may increase the number of rear-end accidents and may also serve as an indicator for a congested area. This is especially true when turn lanes overflow into through lanes.

The combined effect of the posted speed being greater than 50 mph and intersection(s) in a section may reduce roadway accident rates. The negative coefficient of this variable may have resulted from higher standards being applied to high-speed intersections. The severity of accidents at high-speed intersections may still increase while accident rates may not.

Note that a different focus on speed limits (greater or less than 50 mph) can result in different impacts to roadway accident rates. Roadway accident rates increase if horizontal curves exist in the section and the posted speed exceeds 50 mph.

Partially controlled access in sections presented an interesting outcome. A negative coefficient for this indicator implies that the roadway accident rate decreases when partially controlled access is employed in a section. One would expect fully controlled access to have the same effect. Lack of sufficient data on fully controlled access may have partly contributed to its insignificance, while partially controlled access is significant relative to no-control. Partially controlled access type accounted for 67 percent of the total from the sample, while full-control and no control access were 27 percent and 6 percent, respectively.

The presence of trucks on roadways may have both positive and negative impacts on roadway accident rates. The high truck-train percentage on a roadway section increased the accident rate, while the high total truck percentage on a roadway decreased, the accident rate. This is conceivable because passenger cars are involved in accidents more frequently than are trucks. Therefore, for a roadway that has a high total truck percentage, the accident rate should be low unless the major part of the total truck percentage on that roadway consists of truck-trains.

Two shoulder width variables were significant and must be analyzed together. If the shoulder widths are constant for the entire section, the cumulative effect is negative. This is rather intuitive. As the shoulder width increases, the roadway accident rate decreases. However, if the maximum shoulder width is much larger than the average shoulder width, a positive effect on the accident rate may be seen, which is probably related to the shoulder width changes.

The only weather variable affecting the roadway accident rate was the average difference between the maximum and minimum monthly temperatures during the summer months (July and August). As the difference increased, the accident rate decreased. Since the weather data used were averaged over many years, this should not be applied to yearly temperature trends but to climates in general. Greater temperature differentials are most likely seen in the more mountainous and desert regions in eastern Washington, where much of the traffic is recreational. This may account for the negative impact on accident rate.

A graphical comparison of OLS and SURE roadway accident models is presented in figures 3a and 3b. There was not much difference in the coefficients and the standard errors (and as a result, the t-statistics) of the SURE model in comparison to those of the OLS

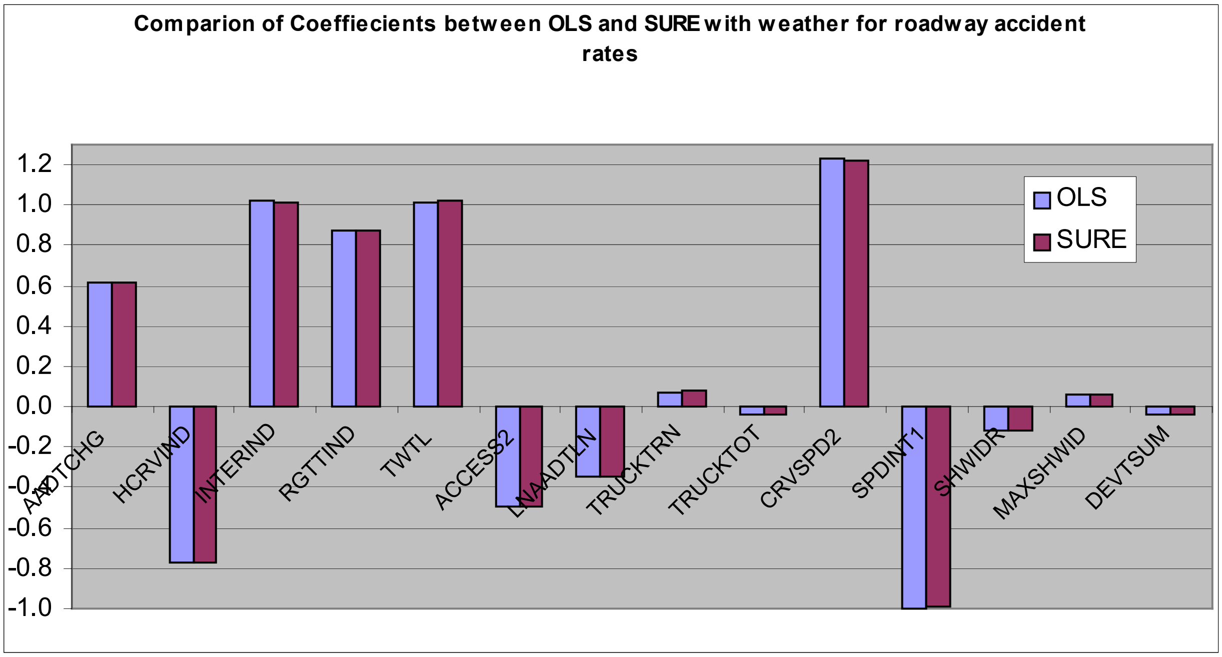

Figure 3a. Comparison of coefficients between OLS and SURE with weather for roadway accident rates

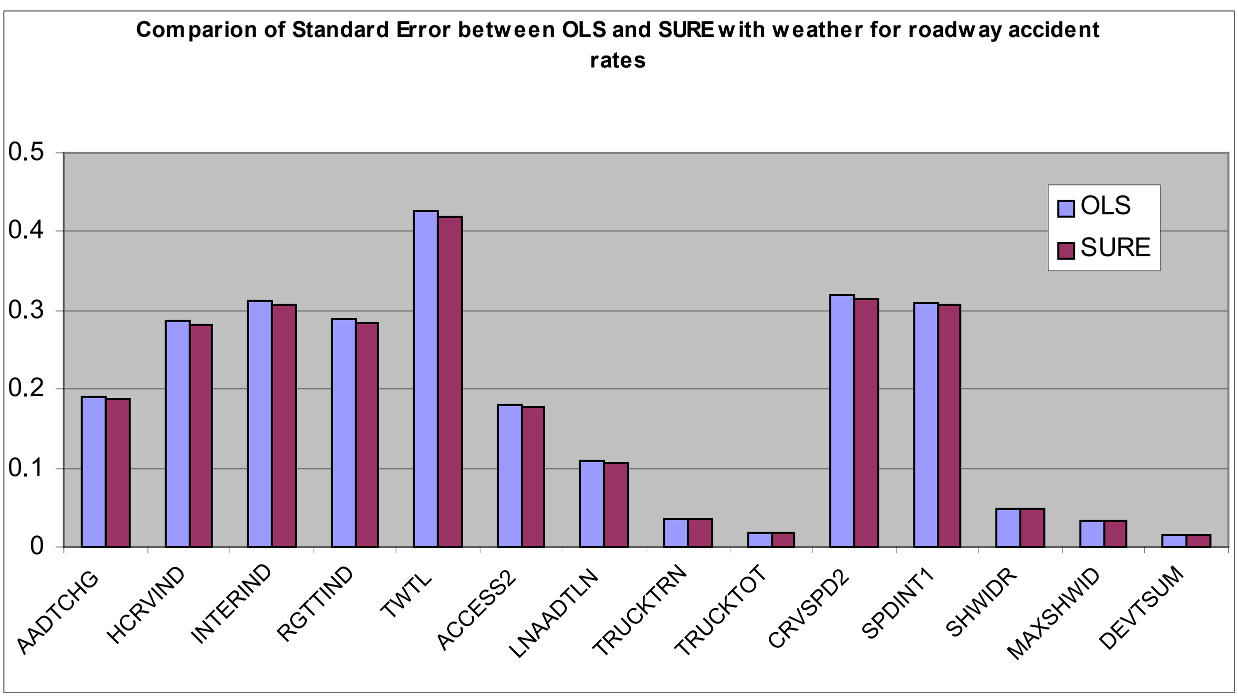

Figure 3b. Comparison of standard errors between OLS and SURE with weather for roadway accident rates

model. The standard errors of the coefficients in the SURE model were definitely lower than those in the OLS model, but the difference was not worth pointing out. Almost all the variables that made up the OLS model constituted the SURE model as well. This proves that there is not enough evidence to conclude that an increase in safety project programming efficiency would result by using the SURE approach to simultaneously model roadway and roadside accident rates.

ROADSIDE ACCIDENT RATE MODEL

The tabular results for the model estimates of roadside accident rate models are presented in Table 3.

As with roadway accident rate modeling, no significant improvement in roadside accident rate modeling efficiency was gained by using the SURE approach for simultaneous roadway and roadside accident rate modeling. The results for graphical comparison of coefficients and standard errors for the various variables across OLS and SURE are presented in figures 4a and 4b.

The coefficient of the average speed limit was negative, which means the higher the speed limit is, the lower the roadside accident rate should be. It can be interpreted that roadside accidents may not frequently occur on Interstates, where the posted speed limit can be very high, because their geometrics are designed to prevent the occurrence of roadside accidents. On low speed limit roadways such as collectors or arterials, the roadside accident rate can be higher because of a lower design level for geometric or speed limit violations.

Three AADT variables were significant in the model. The first was percentage of change in AADT over the section. For 71.8 percent of the sample, this variable had a value

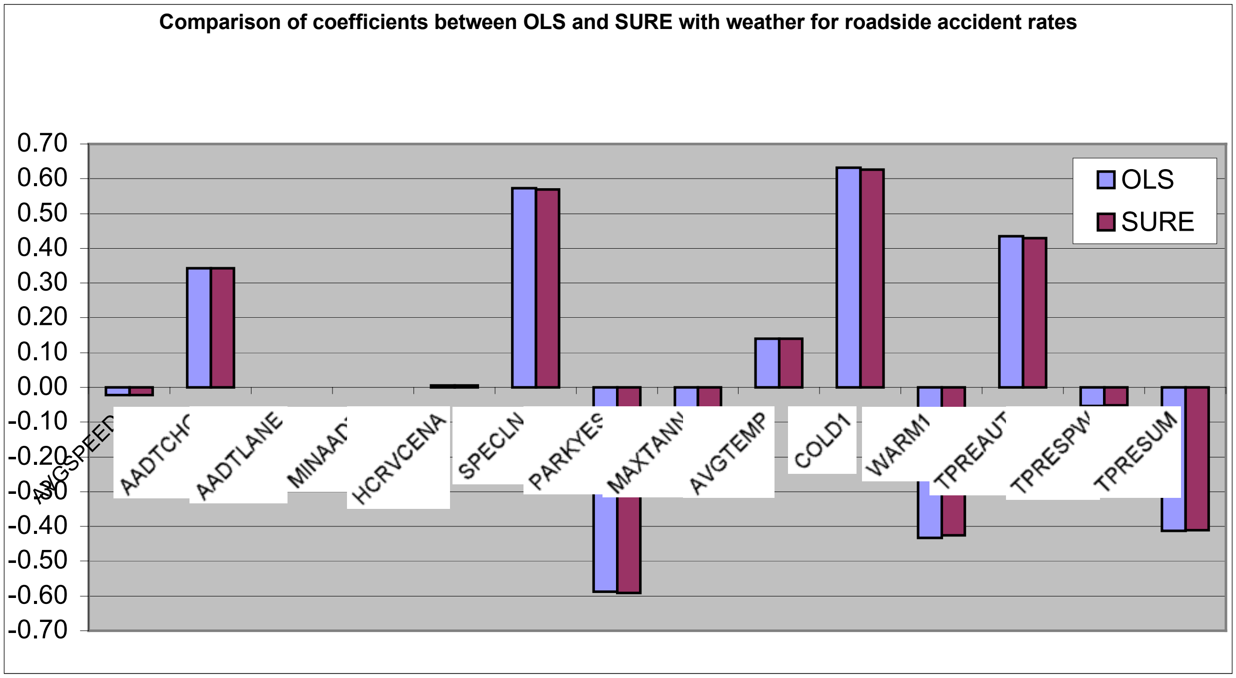

Figure 4a. Comparison of coefficients between OLS and SURE with weather for roadside accident rates

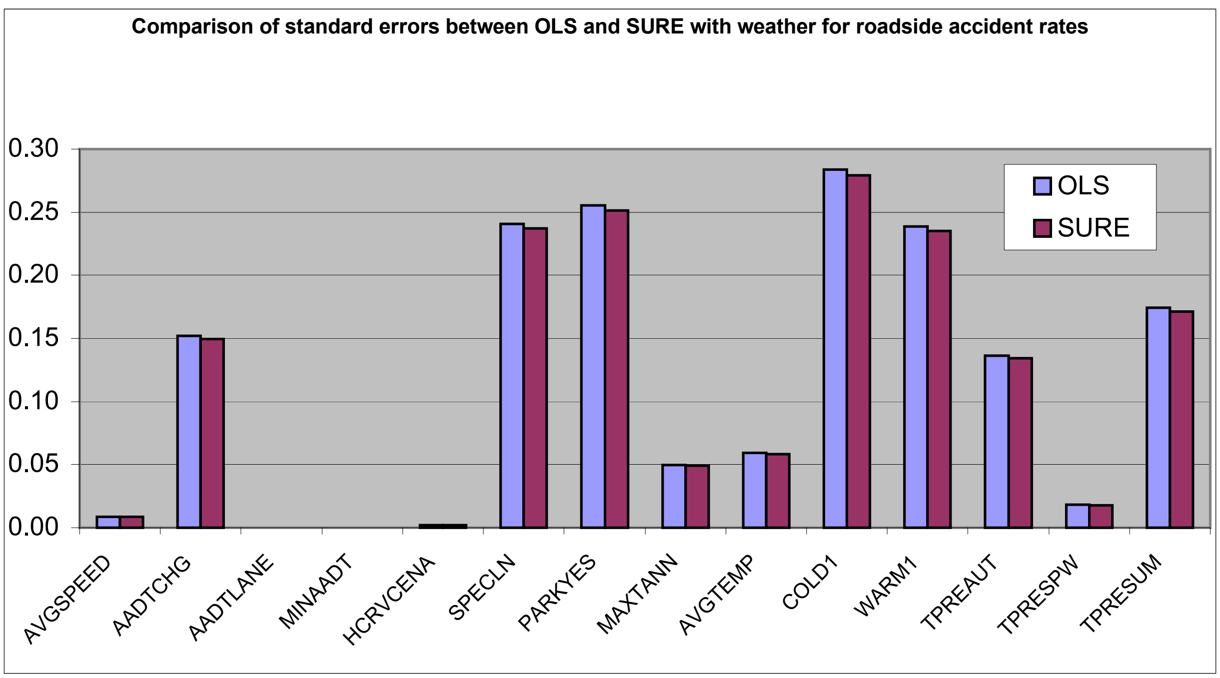

Figure 4b. Comparison of standard errors between OLS and SURE with weather for roadside accident rates

of zero. As might be expected, as the change increased, so did the accident rate as a result of volume differentials. In addition, as the minimum AADT in the section increased, the roadside accident rate increased. Finally, AADT per lane (total AADT divided by total number of lanes) had a negative effect on roadside accident rates.

Horizontal curve central angle was a continuous variable containing the largest central angle in the one-mile section. A tangent section had a value of zero. As might be expected, this variable's coefficient was positive. The explanation is that the further around a curve a driver must go, the more exposure s/he has to leaving the roadway in a roadside accident

The indicator for presence of special lanes had a positive coefficient. While comprising only 7.0 percent of the sections, the effect of special lanes on the roadside accident rate was significant, mainly because of their functionality. Special lanes can be bicycle lanes, climbing lanes, chain-up lanes, slow vehicle turnouts, and shoulder uses.

On sections where parking was permitted during the day, roadside accident rates decreased. This is likely because people drive more cautiously when parking is permitted. In addition, accidents that do happen off the traveled way involving parked vehicles would not be classified as roadside accidents.

Many weather related variables were significant in the roadside model. As average maximum annual temperature (average of monthly maximums) increased, roadside accident rates decreased. Contrarily, as average temperature (averages monthly minimum and maximum) increased, the accident rate decreased. In addition, if average daily temperature fell below 46ºF, roadside accident rates increased, while average daily temperatures exceeding 52ºF were associated with a decrease in accidents.

Seasonal precipitation indicators were also significant in their impact on roadside accident rates. Autumn months increased roadside accident rates, while winter, spring and summer months decreased roadside accident rates. The effect due to winter precipitation is a surprising finding.

Further research would help probe deeper into issues regarding risk programming efficiency. This concerns programming turnover, which determines the number of sections that are still hazardous but have not been improved. These raise equity issues that influence the funding for safety programs. Other methodologies could be researched for their potential to improve the efficiency of risk programming for roadways and roadsides.

THE EFFECTS OF WEATHER ON MODELING THE ROADWAY AND ROADSIDE

We suspected that the weather characteristics decreased the potential efficiency to be gained from using the SURE approach to modeling the simultaneity between the roadway and roadside. We investigated the impact that weather might have on roadway and roadside accident rates.

The results for the OLS and SURE models for roadway accident rates, with the exclusion of the weather variables, are shown in figures 5a and 5b, while those for the roadside accident rates are presented in figures 6a and 6b.

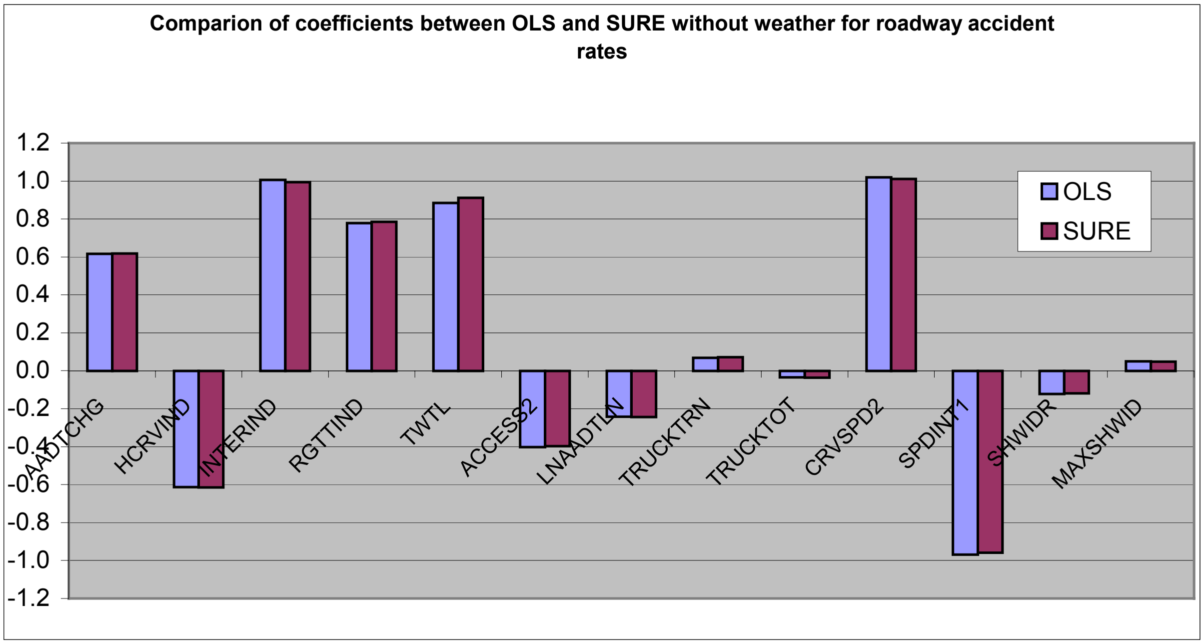

Figure 5a. Comparison of coefficients between OLS and SURE without weather for roadway accident rates

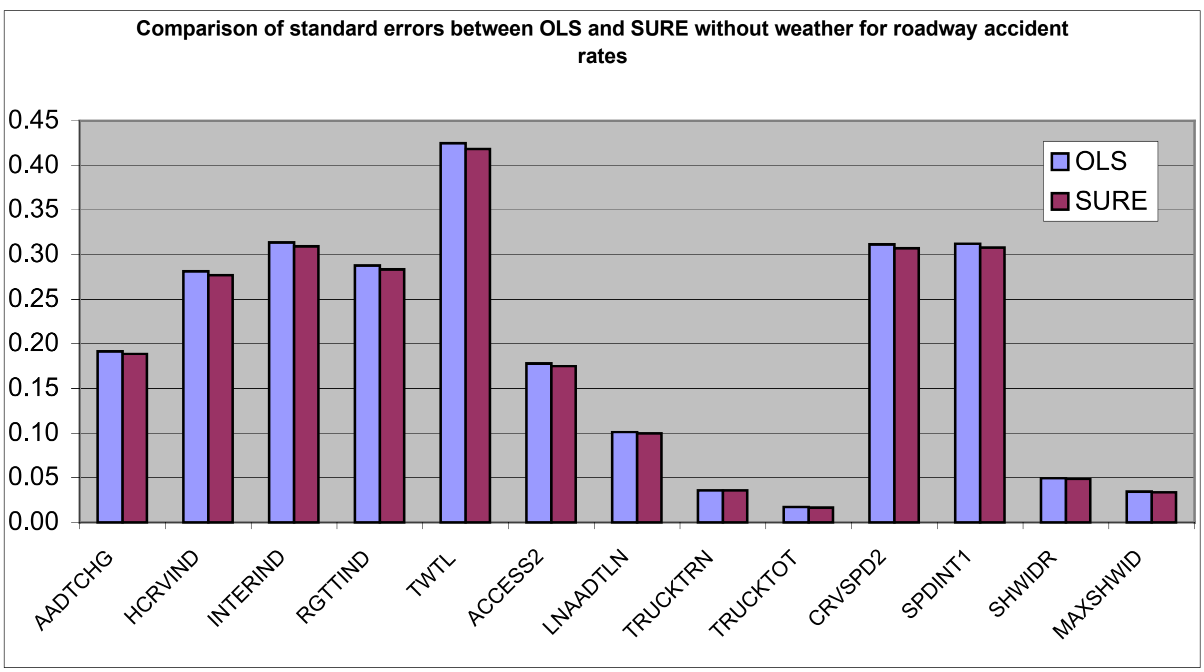

Figure 5b. Comparison of standard errors between OLS and SURE without weather for roadway accident rates

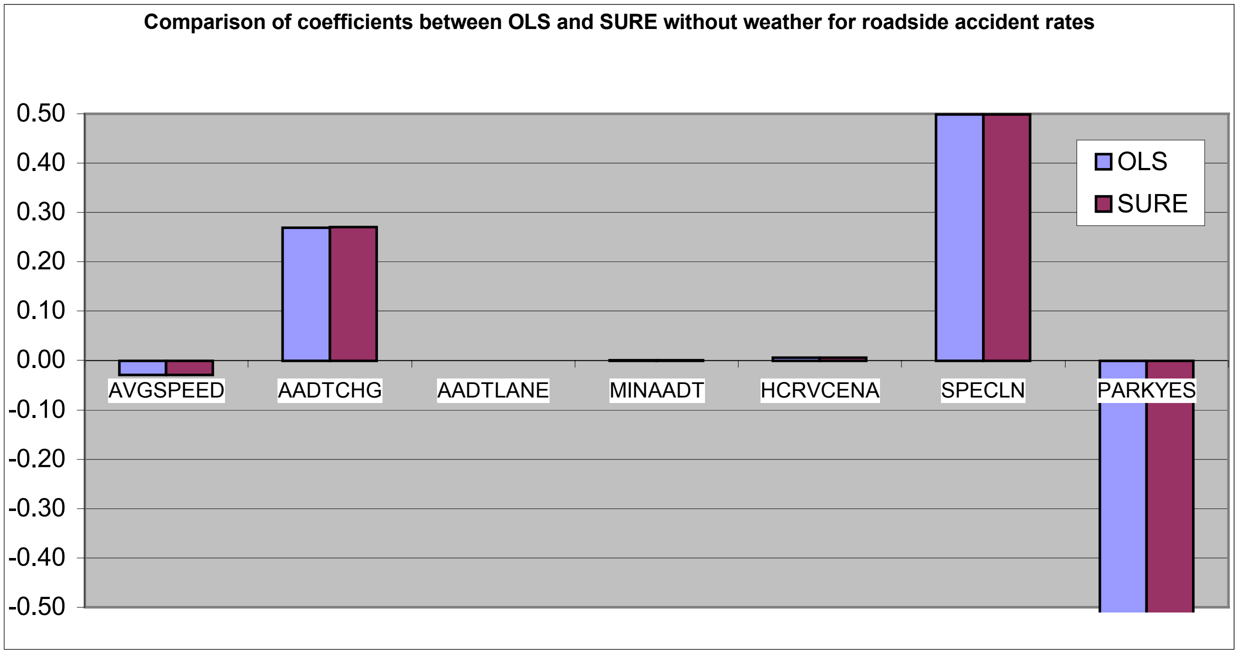

Figure 6a. Comparison of coefficients between OLS and SURE without weather for roadside accident rates

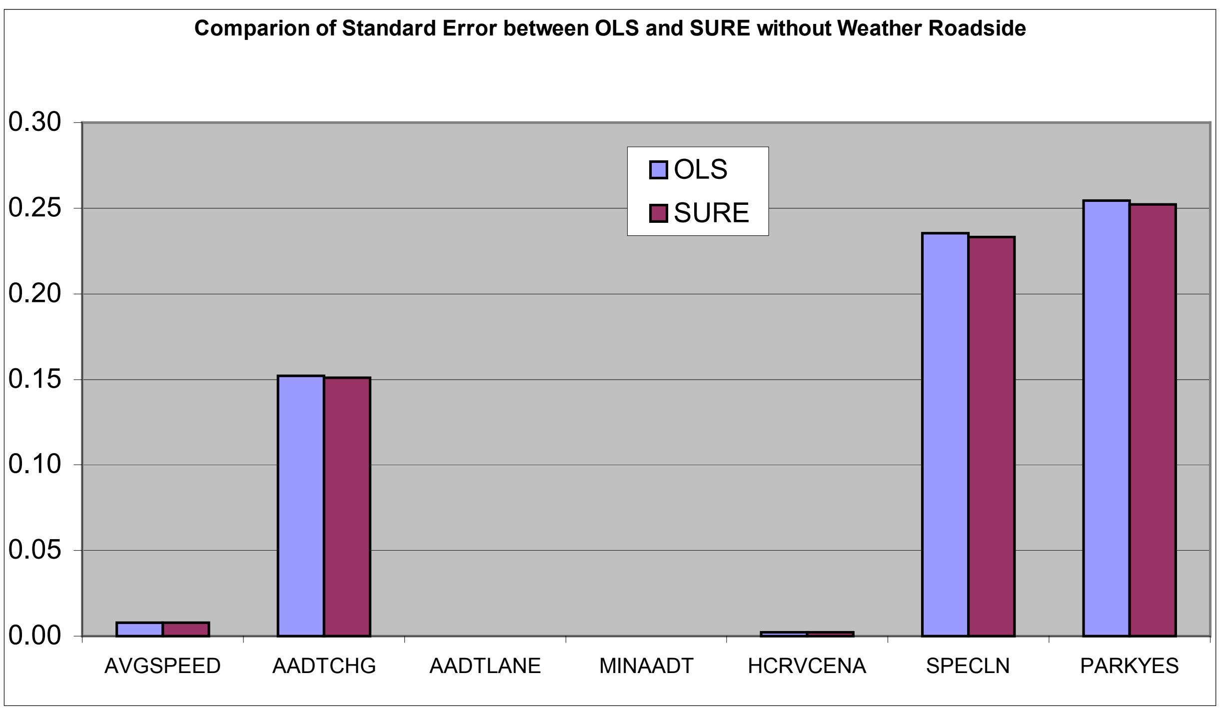

Figure 6b. Comparison of standard errors between OLS and SURE without weather for roadside accident rates

Two issues were considered. First was whether excluding weather variables produced any improvement in efficiency for roadway and roadside accident rate programming. A comparison of the coefficients and variables for the OLS and SURE models without weather variables revealed no significant improvement in efficiencies. This discards the possibility that weather effects decreased potential improvements in efficiency resulting from use of the SURE models.

The second issue was the extent of influence that weather had on roadway and roadside accident rates. Table 4 shows the OLS and SURE estimation results without weather variables for roadway accident rates. Table 5 shows the same for roadside accident rates. A comparison of the OLS models (for roadway and roadside) with and without weather showed no significant bias in the coefficients of the OLS model without the weather variables. This is possibly due to the fact that the variables were correlated with the omitted weather variables. However, the constant term was significantly biased. Also, the missing weather information resulted in a lower R-squared value. The same results were found for the SURE models without the weather information.

Table 4 Comparison of ordinary least squares and seemingly unrelated regression models for roadway accident rates without the weather characteristics

Variables	OLS			SURE		
	Estimated coefficient	Standard error	t-statistic	Estimated coefficient	Standard error	t-statistic
Equation 1: Roadway equation (dependent variable)						
Constant	3.1480021	0.7712475	4.082	3.1665318	0.7597386	4.168
Percent change in AADT (max-min)/min	0.6163098	0.191456	3.219	0.6187632	0.1887464	3.278
Natural Log of (average AADT/ numbers of lanes)	-0.612472	0.2815488	-2.175	-0.615089	0.2773146	-2.218
Horizontal curve indicator (1 if there is horizontal curve in section, 0 otherwise)	1.0053625	0.3137728	3.204	0.9937514	0.3090241	3.216
Interaction variable between horizontal curves and posted speed greater than 50 mph (1 if horizontal curve and speed > 50 mph, 0 otherwise)	0.7790023	0.2880941	2.704	0.7848745	0.283716	2.766
Intersection indicator (1 if the section has intersections, 0 otherwise)	0.8840204	0.4250743	2.08	0.911729	0.4186051	2.178
Interaction variable between intersection and posted speed higher than 50 mph (if intersections and speed > 50 mph, 0 otherwise)	-0.403041	0.1779921	-2.264	-0.39672	0.1752924	-2.263
Right turn lane indicator (1 if the section has right turn lane, 0 otherwise)	-0.241307	0.1011979	-2.385	-0.243613	9.97E-02	-2.444
Two-way left turn lane indicator (1if a two-way left-turn lane in section, 0 otherwise)	6.78E-02	3.61E-02	1.875	7.10E-02	3.56E-02	1.993
Access type indicator 2 (1 if access type is partially controlled, 0 otherwise)	-3.46E-02	1.70E-02	-2.029	-3.56E-02	1.68E-02	-2.119
Percentage of truck-train	1.0198525	0.3116946	3.272	1.0109746	0.3069555	3.294
Percentage of total trucks	-0.968925	0.3124398	-3.101	-0.958708	0.3077111	-3.116
Average right shoulder width in feet	-0.122499	4.95E-02	-2.477	-0.119411	4.87E-02	-2.451
Maximum shoulder width feet	4.97E-02	3.42E-02	1.453	4.86E-02	3.37E-02	1.442
Number of observation R^2 Adjusted R^2	500 0.141 0.118			500 0.141 0.118		

Table 5 Comparison of ordinary least squares and seemingly unrelated regression models for roadside accident rates without the weather characteristics

Variables	OLS			SURE		
	Estimated coefficient	Standard error	t-statistic	Estimated coefficient	Standard error	t-statistic
Equation 2 : Roadside equation (dependent variable)						
Constant	2.2262424	0.474107	4.696	1.02046	2.18910	0.466
Average speed limit in miles per hour	-2.94E-02	8.02E-03	-3.667	-0.02213	0.00845	-2.620
Percent change in AADT (max-min)/min	0.2689286	0.152068	1.768	0.34314	0.14954	2.295
AADT/ numbers of lanes	-1.30E-04	5.50E-05	-2.363	-0.00017	0.00006	-3.116
Minimum AADT in section	2.22E-05	1.44E-05	1.549	0.00003	0.00001	1.982
Horizontal curve central angle in degrees	5.62E-03	2.21E-03	2.549	0.00616	0.00221	2.783
Special lane indicator (1 if a special lane is in the section, 0 otherwise)	0.498807	0.235348	2.119	0.57009	0.23713	2.404
Parking indicator (1 if parking is permitted at sometime during the day, 0 otherwise)	-0.598752	0.254557	-2.352	-0.59155	0.25144	-2.353
Number of observation	500			500		
R^2	0.074			0.074		
Adjusted R^2	0.061			0.061		

CONCLUSIONS AND RECOMMENDATIONS

This research developed models for roadway and roadside accident rates for Washington State highways. Analysis determined that different factors affect the accident rates of the roadway and roadside .

Correlation between the roadway and roadside was also examined. In particular, the need for an efficient "systems" approach incorporating roadway and roadside effects simultaneously was examined. The systems approach, namely the SURE model, was found to produce minimal gains in efficiency over the single-equation OLS models. Thus, the existing programming techniques adopted by the WSDOT can be concluded to be reasonably efficient for the roadway and roadside accident programming.

The correlation of weather data to roadway and roadside accident rates was also examined. Temperature and precipitation were found to be significant at the seasonal level. Weather variables were significant in both the roadway and roadside models. Only one temperature variable was significant in the roadway model, while many more variables significantly explained the accident rate in the roadside model. Additional research with more in-depth climate information could add more explanatory power to highway safety models.

Additional work in the area of roadside inventory is needed as well to improve roadside models. Minimal data are available regarding roadside conditions. To improve upon the explanatory power of the roadway and roadside models, data about side slopes and lengths of guardrail through-sections is needed. As information inventory systems improve in the area of transportation, this information is more likely to be available. In the larger context of safety and risk programming at the statewide level, the results from this study indicate that the current system of programming the roadway and the roadside separately is not inefficient. However, other pragmatic concerns, such as the efficiency and evaluation of various safety program components such as collision prevention and collision reduction, need to be addressed.

REFERENCES

Davis, G. Estimating Traffic Accident Rates While Accounting for Traffic-Volume Estimation Error, *Transportation Research Record 1717*, 94-101, 2000.

Greene, W. *Econometric Analysis*, Fourth Edition, Prentice Hall, 2000.

Joshua, S. and Garber N. Estimating truck accident rate and involvements using linear and Poisson regression models, *Transportation Planning and Technology*, 15, 41-58, 1990.

Jovanis, P. and Chang, H.-L. Modeling the relationship of accidents to miles traveled, *Transportation Research Record* 1068, 42-51, 1986.

Kennedy, P. *A Guide to Econometrics,* Fourth Edition. Cambridge: MIT Press, 1998.

Lee, J. and Mannering, F. Econometric analysis of the effect of roadway geometric and roadside features on run-off-roadway accident frequencies and severities, doctoral dissertation, Department of Civil and Environmental Engineering, University of Washington, Seattle, WA., 2000

Miaou, S.-P. and Lum H. Modeling vehicle accidents and highway geometric design relationships, *Accident Analysis and Prevention*, 25(6), 689-709, 1993.

Milton, J. and Mannering, F. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies, *Transportation* 25, 395-413, 1998.

Mulinazzi, T.E. and Michael, H.L. Correlation of Design Characteristics and Operational Controls with Accident Rates on Urban Arterials, *Purdue University Road School Proc*., 1969.

Persaud, B., Lyon, C, and Nguyen, T. Empirical Bayes Procedure for Ranking Sites for Safety Investigation by Potential for Safety Improvement, *Transportation Research Record 1665*, 1999.

Poch, M. and Mannering, F. Negative binomial analysis of intersection-accident frequencies, *Journal of Transportation Engineering*, 122(3), 105-113, 1996.

Shah, K. A Methodology to Relate Traffic Accidents to Highway Design Characteristics, Ohio University, 31 August 1968.

Shankar, V., Mannering, F. and Barfield, W. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies, *Accident Analysis and Prevention*, 27(3), 371-389, 1995.

Shankar, V., Milton, J. and Mannering, F. Modeling accident frequencies as zero-altered probability processes: An empirical inquiry, *Accident Analysis and Prevention*, 29(6), 829-837, 1997.

Theil, H. Principles of Econometrics. New York: John Wiley and Sons, 1971.

Ulfarsson, G. *Esther*. Version 1.1. Unpublished software, 2001.

Zellner, A. An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias, *Journal of the American Statistical Association*, 57 (298), 1962.

BIBLIOGRAPHY

Gujarati, D., *Basic Econometrics*, Fourth Edition, McGraw Hill, 2003

Hadi, M., Aruldhas J., Chow L.-F. and Wattleworth J., Estimating safety effects of cross-section design for various highway types using negative binomial regression, *Transportation Research Record* 1500, 169-177, 1995.

Mak, K. Safety effects of roadway design decisions-roadside, *Transportation Research Record* 1512, 16-21, 1995.

Miaou, S.-P. Estimating vehicle roadside encroachment frequencies by using accident prediction models, *Transportation Research Record* 1599, 64-71, 1997.

NCHRP Research Results Digest 220, Strategies for improving roadside safety, Transportation Research Board, National Research Council, Washington, DC, November 1997.

Ogden, K. The effects of paved shoulders on accidents on rural highways, *Accident Analysis and Prevention*, 29(3), 353-362, 1997.

Okamoto, H. and Koshi, M. A Method to Cope with the Random Errors of Observed Accident Rates in Regression Analysis, *Accident Analysis and Prevention* 25 (4), 317-332, 1989.

Persaud, B., Retting, R., and Lyon, C. Guidelines for Identification of Hazardous Highway Curves, *Transportation Research Record 1717*, 14-18, 2000.

Viner, J. Rollovers on sideslopes and ditches, *Accident Analysis and Prevention*, 27 (4), 483-491, 1995.

Zegeer, C. and Council, F. Safety relationships associated with cross-sectional roadway elements, *Transportation Research Record* 1512, 29-36, 1995.