2008

Modeling crash severity and speed profile at roadway work zones

Zhenyu Wang
University of South Florida

Modeling Crash Severity and Speed Profile at Roadway Work Zones

by

Zhenyu Wang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Civil and Environment
College of Engineering
University of South Florida

Major Professor: Jian John Lu, Ph.D.
Edward Miezejewski, Ph.D.
Sastry Putcha, Ph.D.
Huaguo Zhou, Ph.D.
George Yanev, Ph.D.

Date of Approval:
March 25, 2008

Keywords: logit model, ordered logit regression, parallel assumption, simulation, support vector regression

Acknowledgments

I am indebted to my academic advisor, Dr. Jian John Lu, for his guidance, patience, encouragement and support throughout my five-year Ph.D. study. I thank Dr. Edward Miezejewski, Dr. Sastry Putcha, Dr. Huaguo Zhou, and Dr. George Yanev for serving on my committee and providing their valuable suggestions. Dr. Xin Ye used to be one of my colleagues and close friends at USF. At the early stage of this study, I greatly benefits from his help and discussion with him. I also want to express my gratitude to my colleague, Pan Tao. Tao gave me an important assistant for data collection and reduction.

Table of Contents

List of Tables

List of Figures

Modeling Crash Severity and Speed Profile at Roadway Work Zones

Zhenyu Wang

ABSTRACT

Work zone tends to cause hazardous conditions for drivers and construction workers since work zones generate conflicts between construction activities and the traffic, therefore aggravate the existing traffic conditions and result in severe traffic safety and operational problems. To address the influence of various factors on the crash severity is beneficial to understand the characteristics of work zone crashes. The understanding can be used to select proper countermeasures for reducing the crash severity at work zones and improving work zone safety. In this dissertation, crash severity models were developed to explore the factor impacts on crash severity for two work zone crash datasets (overall crashes and rear-end crashes). Partial proportional odds logistic regression, which has less restriction to the parallel regression assumption and provides more reasonable interpretations of the coefficients, was used to estimate the models. The factor impacts were summarized to indicate which factors are more likely to increase work zone crash severity or which factors tends to reduce the severity.

Because the speed variety is an important factor causing accidents at work zone area, the work zone speed profile was analyzed and modeled to predict the distribution of speed along the distance to the starting point of lane closures. A new learning machine algorithm, support vector regression (SVR), was utilized to develop the speed profile model for freeway work zone sections under various scenarios since its excellent generalization ability. A simulation-based experiment was designed for producing the speed data (output data) and scenario data (input data). Based on these data, the speed profile model was trained and validated. The speed profile model can be used as a reference for designing appropriate traffic control countermeasures to improve the work zone safety.

Chapter One

Introduction

1.1 Background

Work zone is defined in the 1994 Highway Capacity Manual as "an area of highway in which maintenance and construction operations are taking place that impinge on the number of lanes available to moving traffic or affect the operational characteristics of traffic flowing through the area." Two work zone types on multi-lane highways are shown in Figure 1.1 and their definitions are given as follows:

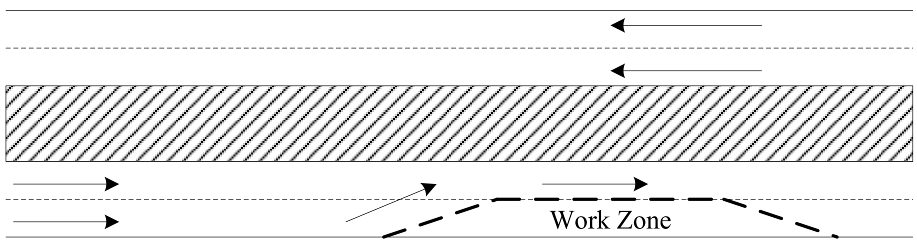
(1) Lane Closure

When one or more lanes in one direction are closed, there is little or no disruption to traffic in the opposite direction.

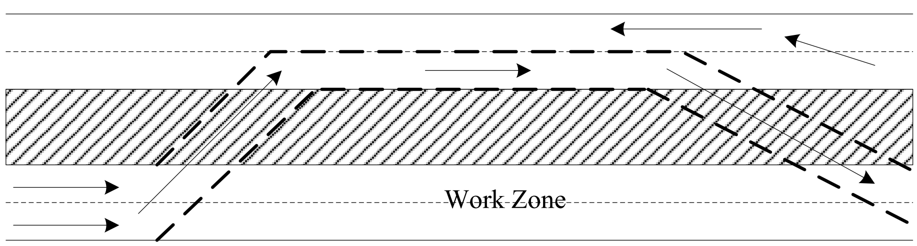
(2) Crossover

When one roadway approach is closed and the traffic which normally uses that roadway is crossed over the median and two-way traffic is maintained on another roadway approach.

As the description in the Manual on Uniform Traffic Control Devices (MUTCD) 2003, a work zone is typically marked by signs, channelizing devices, barriers, pavement markings, and/or work vehicles. It extends from the first warning sign or high-intensity rotating, flashing, oscillating, or strobe lights on a vehicle to the “END ROAD WORK” sign or the last Temporary Traffic Control device. Most work zones are divided into four areas: the advance warning area, the transition area, the activity area, and the termination area (see Figure 1.2).



Lane Closure



Crossover

Figure 1.1 Work Zone Types

The advance warning area is the section of highway where road users are informed about the upcoming work zone or incident area. The transition area is that section of highway where road users are redirected out of their normal path. Transition areas usually involve strategic use of tapers, which because of their importance are discussed separately in detail. The activity area is the section of the highway where the work activity takes place. It is comprised of the work space, the traffic space, and the buffer space. The work space is that portion of the highway closed to road users and set aside for workers, equipment, and material, and a shadow vehicle if one is used upstream. Work spaces are usually delineated for road users by channelizing devices or, to exclude vehicles and pedestrians, by temporary barriers. The termination area shall be used to return road users to their normal path. The termination area shall extend from the downstream end of the work area to the last TTC device such as “END ROAD WORK” signs, if posted.

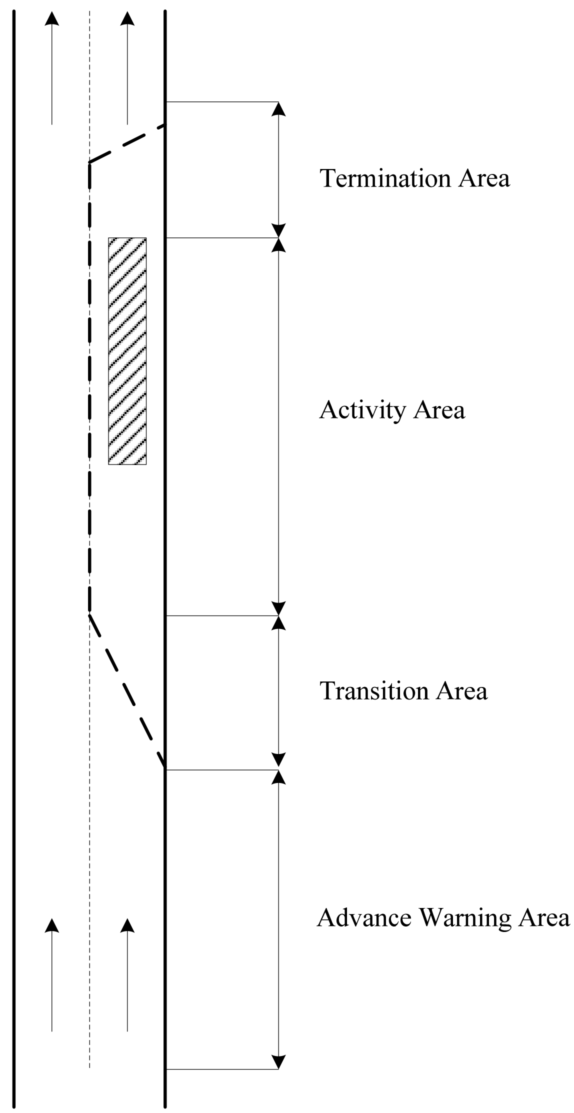


Figure 1. 2 Work Zone Layouts (Lane Closure)

Work zone tends to cause hazardous conditions for vehicle drivers and construction workers since work zones generate conflicts between construction activities and the traffic, and therefore aggravate the existing traffic conditions and result in severe traffic safety problems. Improving safety at work zones has become one of the overwhelming challenges that traffic engineers and researchers have to confront. Nationally, great efforts have been devoted to improve the safety and mobility of work zone traffic (Bai and Li, 2004). Congress addressed the work zone safety issue in the Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991 and National Highway System designation Act of 1995 (FHWA, 1998). In addition, Federal Highway Administration (FHWA) and American Association of State Highway and Transportation Officials (AASHTO) have been developing comprehensive highway work zone safety guidelines and programs. Many state Departments of Transportation (DOTs) have funded various projects to improve work zone safety in their states. Other concerned organizations or research communities have also participated in this campaign and devoted their contributions by conducting meaningful researches on various work zone safety issues.

Regardless of these efforts, there is little indication of significant improvements in work zone safety in Florida. Work zone crash rates by work zone travel mileage are not precisely known, but statistics of work zone fatalities have shown a serious traffic safety problem. Annual work zone fatal crashes rose from 79 in 2002 to 128 in 2005, with the increase of fatalities from 99 in 2002 to 148 in 2005. It was estimated that the direct cost of highway work zone crashes was as high as $6.2 billion per year between 1995 and 1997 with an average cost of $3,687 per crash (Mohan and Gautam 2002). The alarming numbers indicate an urgent need for improving every work zone-safety related field including traffic control and information, project management, public education, and regulation/policy making.

Actual crash data is an important source for identifying safety problems and developing effective countermeasures. Investigating the characteristics of work zone crashes is the first step towards improving work zone safety. The Investigation enables researchers to identify unique work zone safety problems. Accordingly, appropriate countermeasures could be developed to reduce the harm to both construction workers and drivers.

Crash severity is an important criterion to evaluate the social and economic impacts of work zone crashes. Fatalities result in great losses in economy even in life, while no injury crashes just lead to the property damage only. Different traffic factors (driver, vehicle, environmental, and roadway features) have different influences on crash severity levels. To address the diversity of influences of traffic factors is beneficial to understand the characteristics of work zone crashes more deeply. And the analysis results can be used to develop the proper countermeasures for eliminating the factors that deteriorate work zone safety.

Vehicle speed is a critical topic in highway design and operations because of relationships to crash probability and severity. This issue is further complicated in construction work zones due to speed reductions from conditions existing prior to the work zone and transitions into the work zone. Additional design features at work zones such as temporary traffic barriers, reduced lane widths, and crossover sections may influence vehicle speeds. Vehicle operating speed is a factor affecting a variety of work zone design and management decisions, including those related to geometric and roadside features and possible regulatory (posted) speed reduction (Taylor et al, 2007).

The relationships between travel speed and accident rates indicate that accident rates increase as speed variance increases (Migletz and Graham, 1991). A large speed variance coupled with hazardous conditions at work zones (e.g., workers' presence, lane closure, and narrow lane) may lead to higher accident rates at work zones (Maze et al., 2000). Speed variance was generally higher at work zones than that at common roadway sections. Therefore, it could be conjectured that by reducing the speed variance, that is, by having vehicles travel at about the average speed, accident rates would decrease at work zones.

Speed profiles, which are a representation of speeds as a function of position, can be used to assess the characteristics of speed variances at work zones. Accordingly, the understanding of speed profiles is beneficial to developing proper countermeasures to reduce the speed variance. It is well known that vehicle operating speed is affected by various traffic factors, such as driver, vehicle, environment, and roadway features. However, most of the existing operating speed profile models use ordinary linear regression methods (McFadden et al. 2001). The assumptions and limitations inherent to linear regression may at the very least complicate model formulation and, if not corrected for, the deviations from these assumptions can adversely affect the efficacies of such models (Taylor et al, 2007).

1.2 Objectives

This dissertation intends to reach two major objectives:

(1) To develop crash severity models for addressing the influences of traffic factors (driver, vehicle, environmental, and roadway features) on the work zone crash severity.

(2) To develop work zone speed profile models by utilizing a new regression methodology based on learning machine theory. The new methodology has the capability to overcome the issues associated with existing models.

1.3 Scope

The crash severity models will be developed for understanding the relationship between work zone crash severity and various factors. Two kinds of logit regressions will be utilized for the model estimation: ordered logit regression and partial proportional odds logit regression. The crash severity models using ordered logit regression will be estimated with the stepwise variable selection process. And then the violation of the parallel odds assumption will be tested. If the assumption is violated, the partial proportional odds logit regression will be implemented to re-estimate the crash severity models. The results of two kinds of models will be compared and explained.

A simulation-based experiment will be designed for providing speed profile data for developing speed profile models. The speed profile model describes the relationship between speed pattern and various traffic factors. The model based on a new learning machine algorithm will be trained and tested. The result of validation will be used to assess the effectiveness of the new algorithm on predicting work zone speed profile. In this dissertation, only the work zone section on two-way (one direction) freeways will be considered for the development of the speed profile model.

1.4 Outline

The remainder of the dissertation is organized as follows. Chapter 2 initially reviews the past researches on work zone crashes, crash severity modeling, and work zone speed. Chapter 3 briefly introduces ordered logit regression and partial proportional odds logit regression, including their forms, assessing criteria, and interpretation methods. The parallel assumption and its testing method are also introduced. Chapter 4 presents the estimation results of crash severity models, and interpretation of the factors that have significant influence on work zone crash severity is also provided. Chapter 5 offers the introduction of the simulation-based experiment design and the new learning machine algorithm, Support Vector Regression. Chapter 6 gives the training and testing result of the speed profile model. Conclusions and contributions are summarized and some recommendations for future research are provided in Chapter 7.

Chapter Two

Literature Review

2.1 Previous Researches on Work Zone Crashes

Many studies have been conducted to analysis on highway work zone crashes over past years in several states. These researches focused on examining the characteristics of work zone crashes, and evaluated the effectiveness of traffic control countermeasures on traffic safety at work zones.

Bai and Li (2004) conducted a study to the investigate the characteristics of work zone fatal crashes in Kansas and dominant contributing factors to these crashes in the work zones so that effective safety countermeasures could be developed and implemented in the near future. A total 157 crashes during 1992 and 2004 were examined using descriptive analysis and regression analysis. They found that

(1) Male drivers cause about 75% of the fatal work zone crashes in Kansas; Drivers between 35 and 44 years old, and older than 65, are the high-risk driver groups in work zones;

(2) The daytime non-peak hours (10:00 a.m. – 4:00 p.m.) are the most hazardous time period in work zones;

(3) Work zones on rural roads with speed limit from 51 mph to 70mph or located on complex geometric alignments are high risk locations;

(4) Most fatal crashes are multi-vehicle crashes, and head-on, angle-side impact, and rear-end are the three most frequent collision types for the multi-vehicle crashes;

(5) Inefficient traffic controls and human errors contributed to most fatal work zone crashes, and Inattentive driving and misjudgment/disregarding traffic control are the top contributing factors for work zone fatal crashes.

In Taxes, Hill et al. (2003) analyzed the characteristics of work zone fatalities and then evaluated the effectiveness of existing work zone traffic safety countermeasures based on 376 work zone fatal crashes in Texas from January 1, 1997 to December 31, 1999. In this study, three comparisons were conducted between daytime versus nighttime, male drivers versus female drivers, and commercial-truck-involved versus non-commercial-truck-involved. Then logistic regression was implemented to examine the effectiveness of traffic counter measures such as using an officer/flagman and using a stop/go signal. Results of this study indicated that there was a significant difference in crash type and driver error between daytime crashes and nighttime crashes. This difference also existed between driver genders. In addition, commercial truck related crashes were more likely to involve multiple vehicles. According to the logistic regression results, the use of an officer/flagman or a stop/go signal would reduce the chance of having a crash by 68% or 64% respectively.

Ullman et al. (2006) conducted a study on the safety effects of night work activity upon crashes at two types of construction projects in Texas. The first project type involved both day and night work (hybrid project), whereas the other project type performed only at night. Researchers determined the change in crash likelihood during periods of active night work, active day work (if applicable), and during times of work inactivity day and night. Some conclusions were derived from this study:

(1) Crashes increased significantly during periods of work activity than during periods of work inactivity;

(2) A large crash increase at night was expected because the night work more likely involved lane closure than the day work;

(3) For the hybrid project, crashes increased at night more than at day.

Garber and Zhao (2002) studied the distribution of work zone crashes in Virginia in terms of severity, crash type, and road type over four different locations within the work zone referred to as the advance warning area, transition area (taper), longitudinal buffer area, activity area, and termination area. In total, 1484 work zone related crashes during 1993 and 1999 were analyzed. The results indicate that the activity area is the predominant location for work zone crashes for all crash types, and the rear-end crashes are the predominant type of crashes except for the terminate area, where the proportion of angle crashes is significantly higher than other types.

A study on the typical characteristics of multistate work zone crashes was conducted by Chambless et al. (2002) to perform a set of comprehensive comparisons of computerized work zone and non-work zone crash data in Alabama, Michigan, and Tennessee. The Information Mining for Producing Accident Countermeasure Technology (IMPACT) module of Critical Analysis Reporting Environment (CARE) software developed by University of Alabama was used in this study to process the statistical analysis to obtain the conclusions:

(1) 63% of work zone crashes take place on interstate, US, and state roads, as compared to 37% of non-work zone crashes.

(2) 48% of work zone crashes occur on 45- and 55-mph speed zones, as opposed to 34% of non-work zone crashes.

(3) "Misjudging stopping distance/following too close" accounted for 27% of the "prime contributing crash circumstances" for work zone crashes as opposed to 15 percent for non-work zone crashes.

In the study conducted by Mohan and Gautam (2002), the various injury types and their cost estimates were analyzed. As the results, researchers found that

(1) The average direct cost of a motorist's injury is estimated at $3,687;

(2) An overturned vehicle has the largest average cost of $12,627, followed by a rear-end collision averaging $5,541; and

(3) Rear-end collisions are the most common (31%) vehicle crashes, followed by "hit-small-object" collisions at 11% of the total motor vehicle crashes.

Ha and Nemeth (1995) conducted a study in an effort to identify the major cause- and effect relationships between work zone crashes and traffic controls in order to make the first step towards development of effective work zone traffic control strategies. They analyze the crash data during 1982 and 1986 at nine sites in Ohio, and focused on the impacts of factors such as inadequate or confusing traffic control, edge drop or soft shoulder, traffic slowdowns, lane changing or merging, guardrails, and alcohol impairment on work zone crashes. Results of the study indicates that

(1) The predominant type of crash was rear-end;

(2) Improper traffic control was one of the safety problems in construction zones;

(3) Involvement of trucks in crashes at crossovers was significant;

(4) Work zone crashes were slightly less severe than other types of crashes; and

(5) Although work zone crashes increased at nights, they actually decreased in proportion to all crashes.

Pigman and Agent (1990) studied the traffic data and traffic control devices of 20 highway work zones for 3 years (1983-1986) in Kentucky, and found that

(1) Most work zone crashes occur on interstate roads;

(2) Work zone crashes are more server than other crashes, especially in night or truck involved;

(3) The dominant crash type is rear-end and same-direction-sideswipe; and

(4) The dominant contributing factor is following to close.

Hall and Lorenz (1989) investigated the crashes at work zones in New Mexico from 1983 to 1985 by comparing the difference of crashes before- and during-construction at same road sections. They concluded that the proportion of crashes caused by following too close was much higher in during-work zone periods than in before-work zone periods. Another conclusion was that improper traffic control was the prevalent problem causing high crash rates in work zones.

2.2 Previous Researches on Crash Severity Modeling

There has been considerable number of studies on the development of injury or crash severity models, even though none of them address work zone crashes. The general models developed in the past to identify the most important parameters which are crucial in reducing or increasing the level of injury severity of the passengers, drivers or crashes are discussed in this section.

Holdridge et al. (2004) performed a study to analyze the in-service performance of roadside hardware on the entire urban State Route system in Washington State by developing multivariate statistical models of injury severity in fixed-object crashes using discrete outcome theory. The objective is to provide deeper insight into significant factors that affect crash severities involving fixed roadside objects, through improved statistical efficiency along with disaggregate and multivariate analysis. The developed models are multivariate nested logit models of injury severity and they are estimated with statistical efficiency using the method of full information maximum likelihood. The results show that leading ends of guardrails and bridge rails, along with large wooden poles increase the probability of fatal injury. The face of guardrails is associated with a reduction in the probability of evident injury, and concrete barriers are shown to be associated with a higher probability of lower severities. The presented models show the contribution of guardrail leading ends toward fatal injuries. It is therefore important to use well-protecting vehicles from crashes with rigid poles and tree stumps, as these are linked with greater severities and fatalities.

A study was conducted in 1995 to develop a statistical model explaining the relationships between certain driver characteristics and behaviors, crash severity, and injury severity (Kim et al, 1995). Applying the techniques of categorical data analysis to comprehensive data on crashes in Hawaii during 1990, authors built a structural model relating driver characteristics and behaviors to type of crash and injury severity. The structural model helped to clarify the role of driver characteristics and behaviors in the causal sequence leading to more severe injuries. Odds multipliers that are how much dies each factor increase or decrease the odds of more severe crash types or injuries were estimated in this study. It was found that the driver behaviors of alcohol and drug use and lack of seal belt use greatly increased the odds of more severe crashes and injuries. Driver errors were found to have small effect, while personal characteristics of age and sex generally insignificant, as found in this study.

Another study by Mercier et al used logistic regression to assess whether age or gender or both influenced severity of injuries suffered in head-on automobile collusions on rural highways (Mercier et al. 1997). The initial hypothesis that, because of physiological changes, and possibly other changes related to aging including loss if bone density, older drivers and passengers would suffer more severe injuries when involved in head-on crashes was utilized first. It was later found through logistical regression that 14 individual and interactive variables were strongly related to injury severity. Individual variables included age of driver or passenger, position of the vehicle, and the form of the protection used, along with a set of interactive variables. The importance of age related effects in injury severity was verified in this study by hierarchical and principal components logistic regression models, amplifying findings of exploratory stepwise logistic analysis. Variations in findings resulted when the population was divided by gender. Age remained as a very important factor in predicting injury severity for both men and women, but the use of lap and shoulder restraints was found to be more beneficial for men than women, while deployed air bags seemed more beneficial for women than men.

2.3 Previous Researches on Work Zone Speed

A study conducted by Taylor et al (2007) developed a speed profile model for construction work zone on high speed highways using neural networks and made available for use by practitioners through a MS EXCEL interface. The model inputs include horizontal and vertical alignment variables, cross section dimensions and traffic control features. A linear reference system is used for model input and output. Three categories-cars, trucks, and all vehicles were used in this study. Models for the 15^{th} speed, mean speed, and 86^{th} speed were developed. The results of Measured Squared Error in $(km/h)^2$ indicate the neural network is able to predict the vehicle operating speeds to a good accuracy.

Jiang (1999) conducted a study to analyze the traffic flow characteristics of freeway work zones based on the traffic data collected from Indiana four-lane freeways. The study found that vehicle speeds at work zones under uncongested conditions remained stable and close to the work zone speed limit of 55 miles per hour, while they dropped 31.6% to 56.1% from the normal work zone speeds during congestion.

Sisiopiku, Et al. (1999) investigated the effectiveness of various standard speed limits in work zones and the effects of related factors on work zones. The mainly factors influencing work zone speed included number of lanes in work zone, worker presence, and type of lance closure. From the study, five findings were: (1) under free flowing traffic conditions, speeds in work zones were higher than the posted speed limits. (2) Motorists were responsive to the request for reduced speeds when traversing work zones. However, the observed speed reduction was only a fraction (55 to 75%) of that requested. (3) Speed reduction appeared to be highly correlated with the number of open lanes. (4)

The effects of the presence of workers were difficult to isolate. The analysis supported the conclusion that worker presence did not always make a difference in motorist speeds. (5) Lower speeds were associated with less formidable types of lane closure/separation. Overall, the results showed that motorists' speeds within work zones exceeded posted speed limits and were associated with the number of open lanes. Regardless of the posted work zone speed limit, motorists selected higher speeds when more lanes were open to traffic. This indicated that attempts to increase the road capacity during construction were likely to result in higher speeds through the work zones when free flow conditions existed.

Rouphail and Tiwari (1985) studied flow characteristics in freeway lane closures in Illinois. The work activity descriptors were numerated. The sum of the numerical codes was termed as the activity index (AI) of the work zone. The work activity data were collected manually in five-minute intervals corresponding to the speed flow observations. The results from the project showed that on an average the observed mean speed at lane closure was 3 mph lower than the predicted mean speed. The difference in mean speeds was found to increase with increasing AI but the difference was less than 1 mph. The difference in speed increased significantly as the proximity of work zone moved to within 6 ft of the traffic lane.

Sun and Benekohal (2004) investigated platooning and gap characteristics of short-term and long-term freeway work zones. The finding showed that the average gap was the shortest when a car was following another car. The next shortest gap was when a car was following a truck. The gap was longer when a truck was following a car or a truck. When a truck was following a car or a truck, the gap sizes were not as different as when a car was following a car or a truck. This indicated that car drivers were more sensitive to what type of vehicles they were following than the truck drivers. Additionally, the researcher also found that the gaps at short-term work zone were longer than the gaps at long term work zone for the same combination of leader and follower.

Chapter Three

Modeling Methodology for Crash Severity Models

3.1 Crash Severity Models

A set of crash severity models were developed in this study to identify the variables that were significantly influential on the injury severity degree of work zone crashes. These models utilized the crash severity of work zone crashes as a dependent variable and described the relationship between the injury severity and a set of explanatory variables. Crash severity, defined as the most severe injury sustained by a person involved in the crash, is scaled into five major levels shown in Table 3.1. Obviously, crash severity is an ordinal (ordered) categorical variable ranked from the least severe level (no injury) to the most severe level (fatal).

Table 3.1 Definition of Crash Severity

Level	Definition
1	No Injury
2	Possible Injury
3	Non-Incapacitating Injury
4	Incapacitating Injury
5	Fatal (within 30 days)

The models were developed based on the history crash data; so that they explain the effects of various factors on crash severity given that the person is involved in a crash. In other words, the models estimate the probability of a certain crash severity when a crash has been happened. Several statistical methods are used to estimate the model with ordinal categorical outcomes. In the rest of the chapter, the statistical methods were discussed in detail.

3.2 Ordinal (Ordered) Logit Regression

3.2.1 Introduction to the Regression

When the scale of a multiple category outcome is ordinal scale rather than nominal scale, Ordinal Logit Regression (OLR), also called as Ordered Logit Regression, is used to describe the relationship between the outcomes and a set of explanatory variables. In contrast to the multinomial logit regression, the ordinal logit regression can reflect the ordinal features of the model outcomes.

Assume that the ordinal outcome variable, Y, can take on K values coded $1,2,...,K$. The probability of Y adopting a specific value can be defined by

$$p_j = \Pr((Y = j \mid \mathbf{x}), \qquad j = 1,2,...,K \tag{3.2.1}$$

where $\mathbf{x}$ is the vector of explanatory variables. The ordinal logit regression model can be written as

$$\begin{aligned} Logit(p_1 + ... + p_j) &= \ln\left[\frac{\Pr(Y > j \mid \mathbf{x})}{\Pr(Y \le j \mid \mathbf{x})}\right] = \ln\left[\frac{\Pr(Y > j \mid \mathbf{x})}{1 - \Pr(Y > j \mid \mathbf{x})}\right] \\ &= \alpha_j + \mathbf{x}\boldsymbol{\beta}, \qquad j = 1,2,...,K-1 \end{aligned} \tag{3.2.2}$$

where α_j is the jth constant coefficient (interpret); $\boldsymbol{\beta}$ is the vector of slope coefficients associated with the explanatory variables.

With this model, the probability of the probability of a larger response, $Y > j$, is compared to an equal or smaller response, $Y \le j$. The cumulative probability can be calculated as

$$\Pr(Y \le j \mid \mathbf{x}) = \sum_{1}^{j} \Pr(y = j \mid \mathbf{x}) = \frac{1}{1 + \exp(\alpha_j + \mathbf{x}\boldsymbol{\beta})}, \quad j = 1,2,\ldots,K-1 \qquad (3.2.3)$$

or

$$\Pr(Y > j \mid \mathbf{x}) = 1 - \Pr(Y \le j \mid \mathbf{x}) = \frac{\exp(\alpha_j + \mathbf{x}\boldsymbol{\beta})}{1 + \exp(\alpha_j + \mathbf{x}\boldsymbol{\beta})}, \quad j = 1,2,\ldots,K-1 \qquad (3.2.4)$$

where exp() is the exponential function.

3.2.2 Parallel Regression Assumption

From Equation 3.2.2, it can be concluded that the ordinal logit regression assumes common slope parameters associated with the predictor variables. This model is also knows as proportional-odds model because the ration of the odds of the event $Y \le j$ is independent of the category j. In other words the odds ratio is constant for all categories.

For example, Figure 3.1 plots the cumulative probability curves when there are four ordered categories, resulting in three curves. To see why these curves are parallel, a value point of the outcome probability is picked. At this point, the following equation can be derived as

$$\frac{\partial \Pr(Y \le 1 \mid \mathbf{x})}{\partial x} = \frac{\partial \Pr(Y \le 2 \mid \mathbf{x})}{\partial x} = \frac{\partial \Pr(Y \le 3 \mid \mathbf{x})}{\partial x}$$

It is in sense that the regression curves are parallel.

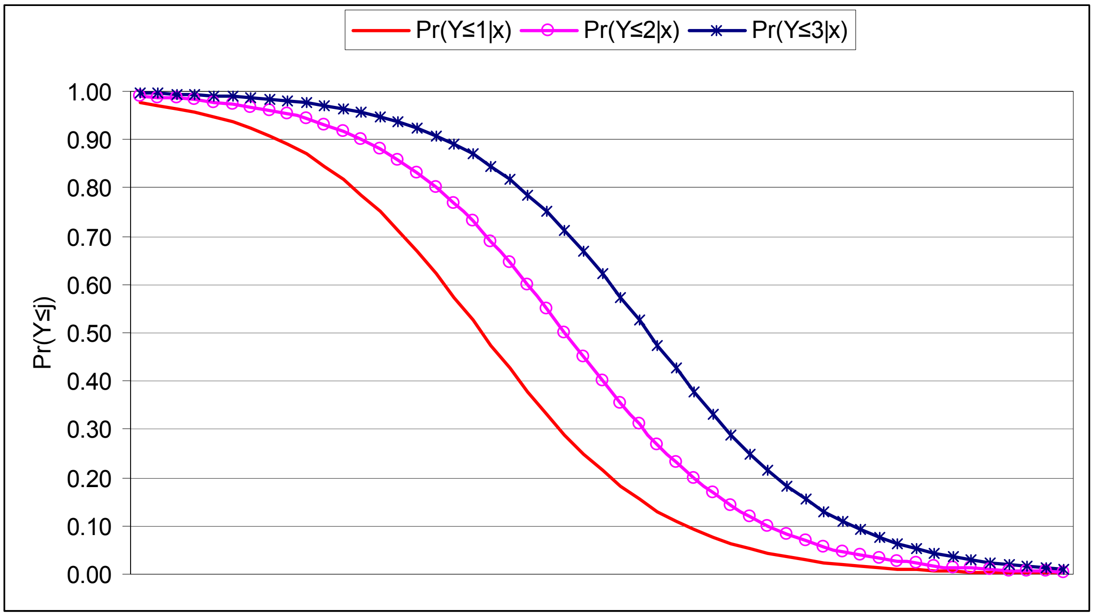


Figure 3. 1 Illustration of the Parallel Regression Assumption

A Wald test is proposed by Brant to assess the parallel assumption of the ordinal regression. This test allows both an overall test that the coefficients for all variables are equal and tests of the equality of the coefficients for individual variables.

For overall test, $K-1$ binary regressions are constructed as following:

$$z_j = \begin{cases} 1 & Y > \mathrm{j} \\ 0 & Y \le j \end{cases}, \qquad j = 1,2,\ldots,K-1 \tag{3.2.5}$$

so we have

$$Logit[\Pr(z_j \mid \mathbf{x})] = \alpha_j + \mathbf{x}\boldsymbol{\beta}_{\mathbf{j}}, \qquad j = 1,2,\ldots,K-1 \tag{3.2.6}$$

the hypothesis of the overall test is

$$H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \ldots = \boldsymbol{\beta}_{K-1} = \boldsymbol{\beta} \tag{3.2.7}$$

A Wald test statistic is derived as *chi-square* with $(K-2)M$ degrees of freedom, where M is the number of explanatory variables. For the *mth* individual variable, the null hypothesis is

$$H_0^m : \beta_{m,1} = \beta_{m,2} = ... = \beta_{m,K-1} = \beta_m \tag{3.2.8}$$

The resulting test statistic follows *chi-square* distribution with *K-2* degrees of freedom. If the probability of these tests (*p-value*) is less than a value (usually is 0.05), the hypothesis is rejected; in other words, there are strong evidences for the violation of the assumption for overall variables or individual ones.

3.3 Partial Proportional Odds Regression

3.3.1 Generalized Ordered Logit Regress

A key problem with the ordinal logit regression is that the parallel assumption is often violated. It is common for one or more coefficients of the explanatory variables (β) to differ across values of the outcome (j). In this situation, the ordinal logit regression is overly restrictive. For passing over the restriction, a generalized ordered logit regression was proposed by Clogg and Shihadeh (1994). This regression can be written as

$$\Pr(Y > j \mid \mathbf{x}) = g(\mathbf{x}\boldsymbol{\beta}_j) = \frac{\exp(\alpha_j + \mathbf{x}\boldsymbol{\beta}_j)}{1 + \exp(\alpha_j + \mathbf{x}\boldsymbol{\beta}_{j)}}, \qquad j = 1,2,..., K-1 \tag{3.3.1}$$

The probabilities that Y will take on each of the values is equal to

$$\begin{aligned} &\Pr(Y = 1 \mid \mathbf{x}) = 1 - g(\mathbf{x}\boldsymbol{\beta}_1) \\ &\Pr(Y = j \mid \mathbf{x}) = g(\mathbf{x}\boldsymbol{\beta}_{j-1}) - g(\mathbf{x}\boldsymbol{\beta}_j), \qquad j = 2,3,...K-1 \\ &\Pr(Y = K \mid \mathbf{x}) = g(\mathbf{x}\boldsymbol{\beta}_{K-1}) \end{aligned} \tag{3.3.2}$$

When $K = 2$, the generalized logit regression is equivalent to the binary logit regression. When $K > 2$, the regression becomes equivalent to a series of binary logistic regressions where categories of the response variable are combined.

The generalized ordered logit regression gives freedom to each coefficient of variables across the outcome values. When the parallel assumption is violated for only some of variables but not for all variables, the regression estimates far more parameters than is really necessary. For overcoming the less restriction, a partial proportional odds logit regression was proposed by Peterson & Harrel (1990).

3.3.2 Partial Proportional Odds Logit Regression (POLR)

In the partial proportional odds regression, some of the regression coefficients can be same for all outcome values where the parallel assumption for the variables associated with the coefficients is not violated. Other coefficients can differ if their associated assumptions are violated.

We assume that N independent random observations are sampled and that the responses of these observations on an ordinal variable Y are classified in K categories with $Y = 1,2,..,K$. The cumulative probabilities is

$$\Pr(Y_i > j \mid \mathbf{x_i}) = \frac{\exp(\alpha_j + \mathbf{x_i^a}\boldsymbol{\beta}^a + \mathbf{x_i^n}\boldsymbol{\beta}_j^n)}{1 + \exp(\alpha_j + \mathbf{x_i^a}\boldsymbol{\beta}^a + \mathbf{x_i^n}\boldsymbol{\beta}_j^n)}, \qquad j = 1,2,\ldots,K-1 \tag{3.3.3}$$

where $\mathbf{x_i^a}$ is a vector containing the values of observation i on that subset of explanatory variables for which the parallel assumption is not violated; $\boldsymbol{\beta}^a$ is the vector of coefficients associated with the non-violated variables, and is same across values of Y . $\mathbf{x_i^n}$ is a vector containing the values of observation i on that subset of explanatory variables for which the parallel assumption is violated; $\boldsymbol{\beta}_j^n$ is the vector of coefficients associated with the violated variables, and differs across the response values.

For estimating the constants (α) and coefficients (β) through Maximum Likelihood Estimation (MLE), the log-likelihood function for the model is defined as

$$\ln L = \sum_{i=1}^{N} \sum_{j=1}^{K-1} I_{i,j} \ln\{\Pr(Y = j \mid \mathbf{x_i})\} \tag{3.3.4}$$

where $I_{i,j}$ is an indicator variable for observation i such that $I_{i,j} = 1$ if $Y_i = j$, and $I_{i,j} = 0$ if $Y_i \neq j$.

3.3.3 Criteria for Assessing the Model

3.3.3.1 $z-$Test

$z-$Test is used to test the statistical significance of individual estimated coefficients of the ordered logit regression or the partial proportional odds logit regression. For MLE, estimators are distributed asymptotically normally. This means that as sample size increases, the sampling distribution of an ML estimator becomes approximately normal. So the hypothesis is $H_0 : \beta_m = 0$, and the *z*-Statistic follows the standard normal distribution $N(0,1)$ given as

$$z = \frac{\hat{\beta}_m}{\hat{\sigma}_{\hat{\beta}_m} / \sqrt{n}} \tag{3.3.5}$$

where β_m is the *mth* coefficient of the model, and $\hat{\beta}_m$ is the estimator of β_m;

$\hat{\sigma}_{\hat{\beta}_m}$ is the estimator of standard deviation of the coefficient β_m;

n is number of observations.

If H_0 is true, the coefficient β_m of the model is not statistically significant. If H_0 is rejected at a confidence level (usually is 0.05), the coefficient β_m is significant to the response.

3.3.3.2 Likelihood Ratio (LR) Test

It is often useful to take an overall significance test for all coefficients of the model, that means, to test if all coefficients are simultaneously equal to zero or not. The hypothesis can be written as $H_0 : \boldsymbol{\beta} = 0$. Such hypothesis can be tested with likelihood ratio (LR) test, which can be thought of as a comparison between the estimates obtained after the constraints implied by the hypothesis ($\boldsymbol{\beta} = 0$) have been imposed to the estimates obtained without the constraints.

To define the test, let model M_β be the unconstrained model that includes constants (α_m) and slope coefficients (β_m). Let model M_α be the constrained model that excludes all slop coefficients. To test the hypothesis, the test statistic is used:

$$G^2(M_\beta) = 2\ln L(M_\beta) - 2\ln L(M_a) \qquad (3.3.6)$$

where $\ln L()$ is the log-likelihood function defined in Equation 3.3.4. If the null hypothesis is true, the test statistic is distributed as *chi-square* with degrees of freedom equal to the number of slope coefficients. If the test statistic falls to the rejection region, p value is less than a confidence level (usually is 0.05), then the null hypothesis is rejected. It can be concluded that not all slope coefficients are equal to 0, in other words, at least one explanatory variable has significant influence on the model response.

3.3.3.3 Pseudo-R^2

To assess the goodness-of-fit of the model, which is a statistical model that describes how well the model fits a set of observations, the Pseudo-R^2 is provided as

$$R^2 = 1 - \frac{\ln L(M_\beta)}{\ln L(M_a)} \qquad (3.3.7)$$

where M_β is the unconstrained model with all slope coefficients; M_a is the constrained model with only constants; and $\ln L()$ is the log-likelihood function. If the unconstrained model does much better than the constrained model, this value will be close to 1. If the unconstrained model does not explain much at all, the value will be close to zero. In this study, the purpose focused on exploring the influence of explanatory variables on the response. So this value was just taken as a reference.

3.3.4 Interpretation of Model Coefficients

The generalized ordered logit model is often interpreted in terms of odds ratios for cumulative probabilities. The odds that a response is j or less versus greater than j given $\mathbf{x}$ can be derived from Equation 3.3.1 as

$$\Omega_j(\mathbf{x}) = \frac{\Pr(Y > j \mid \mathbf{x})}{1 - \Pr(Y > j \mid \mathbf{x})} = \frac{\Pr(Y > j \mid \mathbf{x})}{\Pr(Y \le j \mid \mathbf{x})} = \exp(\alpha_j + \mathbf{x}\boldsymbol{\beta}_j) \tag{3.3.8}$$

To determine the effect of a change in $\mathbf{x}$ from $\mathbf{x_s}$ to $\mathbf{x_e}$, the odds ratio at $\mathbf{x_s}$ versus $\mathbf{x_e}$ is

$$\frac{\Omega_j(\mathbf{x_s})}{\Omega_j(\mathbf{x}_e)} = \frac{\exp(\alpha_j + \mathbf{x}_s\boldsymbol{\beta}_j)}{\exp(\alpha_j + \mathbf{x}_e\boldsymbol{\beta}_j)} = \exp([\mathbf{x}_s - \mathbf{x}_e]\boldsymbol{\beta}_j)$$

When only a single variable (x_m) changes by δ, then

$$\frac{\Omega_j(\mathbf{x}, x_m + \delta)}{\Omega_j(\mathbf{x}, x_m)} = \exp(\delta \times \beta_{m,j}) \tag{3.3.9}$$

With an increase of δ in x_m, the odds of a response that is less than or equal to j are changed by the factor $\exp(\delta \times \beta_{m,j})$, holding all other variables constant. If x_m changes by 1, the odds ratio equals

$$\frac{\Omega_j(\mathbf{x}, x_m + 1)}{\Omega_j(\mathbf{x}, x_m)} = \exp(\beta_{m,j}) \qquad (3.3.10)$$

Because the exponential function is a monotonic increasing function, if $\beta_{m,j}$ is greater than zero, the odds ratio is greater than 1; and if $\beta_{m,j}$ is negative, the odds ratio falls into (0, 1). Positive coefficients mean that higher values on the explanatory variables make higher values on the dependent variable more likely.

For the ordinal logit regression, the odds ratio is same across all values of the response. So the interpretation of coefficients is same for all responses values. In contrast, for the partial proportional odds logit regression, some coefficients have same interpretation if the parallel assumption is not violated for these coefficients, and violated-variables have various interpretations across the response values.

Chapter Four

Estimation Results of Crash Severity Models

4.1 Data Preparation

The dataset used for analysis and model estimation was extracted from the Florida Crash Analysis Reporting (CAR) system. This system is maintained by the Florida Department of Transportation (FDOT) and contains comprehensive information of Florida motor vehicle collisions, and that of the involved vehicles and persons. The dataset contains all work zone crash data from 2001 to 2005 which were identified by the variable FIRST_ROAD_ CONDITION_CRASH_COD equal to 04 (Road under Repair/Construction). This original dataset was downloaded from the FDOT mainframe, and was transformed to SPSS data file for data arrangement and data reduction.

From the original dataset, some variables were selected for data analysis. These variables could be measured at ordinal scale, nominal scale, or continuous scale. For handling the data in an easy way, except for three continuous variables, all categorical variables were transformed to binary data. Some records have empty values for continuous variables. These records were deleted from the dataset. The description of the original variables is given in Appendix A.

4.2 Data Description

Table 4.1 gives the description of the selected variables for the crash severity model development. The response variable is crash severity level, which can be ranked at 5 levels in ascending order from no injury to fatal. The explanatory variables can be classified into 4 categories: driver-related factors, environmental-related factors, crash-related factors, and roadway-related factors.

Table 4.1 Description of Selected Variables for Model Development

Variable	Description	Type	Value	Definition
ACCISEV	Crash Severity Level	Ordinal	1	No Injury
			2	Possible Injury
			3	No-Capacitating Injury
			4	Incapacitating Injury
			5	Fatal
		Driver-Related		
ALCHDRUG	If driver was under influence of alcohol or drugs	Binary	0	No
			1	Yes
AGE	Driver's Age	Categorical	1	Young (<24)
			2	Adult (≥25 and <65)
			3	Old (≥65)
		Environmental-related		
DAYLIGHT	If the crash occurred during day light condition	Binary	0	No
			1	Yes
BDWTHER	If weather was clear	Binary	0	No
			1	Yes

Table 4.1 (continued)

Variable	Description	Type	Value	Definition
	Roadway-related			
SPEC_SECT	If road section was specific type (Intersection, Interchange …)	Binary	0 1	No Yes
SURF_DRY	If road surface was dry	Binary	0 1	No Yes
GR_CUR	If there was a curve or grade at the crash location	Binary	0 1	No Yes
TRAF_CONT	If there was a traffic control strategy	Binary	0 1	No Yes
VIS_OBS	If there was a vision obstruction at the crash location	Binary	0 1	No Yes
URBAN	If the crash occurred in a urban area	Binary	0 1	No Yes
FREEWAY	If the crash occurred in a freeway	Binary	0 1	Surface Road Freeway
SURWIDTH	Road Surface Width	Continuous		
MAXSPEED	Speed Limit	Continuous		
SECTADT	Annual Average Daily Traffic	Continuous		
	Crash-Related			
HVINV	If heavy vehicle was involved	Binary	0 1	No Yes

Tables 4.2 and 4.3 illustrate the statistic description of the variables. In Table 4.2, the minimum value, maximum value, range, mean, and standard deviation of the three continuous variables are provided. Surface width is the width of roadway except for shoulders with the mean of 28.75 feet and the range of 80 feet. The minimum speed limit at work zones is 15miles/h and the maximum speed limit is 70 miles/h. The mean speed limit is 52.75 miles/h. The AADT has a large range from 250 vehicles per day to over 300 thousand vehicles per day.

The distribution of crash severity is given in Table 4.3. The percentage of the severity level descends with the increase of crash severity. The slight injury crashes (ACCISEV=1, 2, 3) hold 90.7% of the total work zone crashes, and Incapacitating Injury crashes only holds the percentage of 7.9%, followed by the fatal crashes of 1.5%. There is 8.0% of work zone crashes involved alcohol or drugs. And about 66% of work zone crashes occurred under good weather or good light conditions. 46% of the locations where work zone crashes occurred are under the influence of specific section, like bridge, intersection, interchange, or railway cross. Most of work zone crashes occurred where road surface is dry (83.1%), road section has not grade or curve (79.1%), and vision condition is good (89.9%). The percentage of heavy vehicle involvement, urban area, and freeway is 15%, 41%, and 85% respectively.

The distribution of driver's age groups is 9.4% for old drivers, 24.2% for young drivers, and 66.4% for adult drivers. The most frequent crash type is rear-end with 37.7% percent, followed by angle crash (11.9%) and side swipe crash (11.0%).

Table 4.2 Descriptive Statistic of Continues Variables

Variable	N	Minimum	Maximum	Range	Mean	Std. Deviation
SURWIDTH	16868	8	88	80	28.75	8.845
MAXSPEED	16868	15	70	55	52.75	11.457
SECTADT	16868	250	302,000	301,750	63,149.89	50,543.123

Table 4.3 Frequencies of Discrete Variables

Variable	Value	Frequency	Percent
Sample Size: 16,868			
ACCISEV	1	7,654	45.4%
	2	4,268	25.3%
	3	3,368	20.0%
	4	1,325	7.9%
	5	253	1.5%
ALCHDRUG	0	15,521	92.0%
	1	1,347	8.0%
AGE	1	4,083	24.2%
	2	11,202	66.4%
	3	1,583	9.4%
DAYLIGHT	0	5,642	33.4%
	1	11,226	66.6%
BDWTHER	0	11,173	66.2%
	1	5,695	33.8%
SPEC_SECT	0	9,112	54.0%
	1	7,756	46.0%
SURF_DRY	0	2,854	16.9%
	1	14,014	83.1%
GR_CUR	0	13,345	79.1%
	1	3,523	20.9%
TRAF_CONT	0	5,479	32.5%
	1	11,389	67.5%
VIS_OBS	0	15,163	89.9%
	1	1,705	10.1%
URBAN	0	2,663	15.8%
	1	14,205	84.2%
FREEWAY	0	9,930	58.9%
	1	6,938	41.1%
CRASHTYPE	0	6,646	39.4%
	1	6,355	37.7%
	2	2,009	11.9%
	3	1,858	11.0%
HVINV	0	14,341	85.0%
	1	2,527	15.0%

4.3 Overall Work Zone Crash Severity Model

4.3.1 Estimation Procedure

This section presents the estimation results of the work zone crash severity model for all work zone crashes. At first, cross tabulation analysis was performed to check the distribution of explanatory variables across injury severity levels and ensure enough observations in each cell. And AGE variable was transformed to three dummy variables: YOUNG_AGE (AGE=0), MIDDLE_AGE (AGE=1), and OLD_AGE (AGE=2). After then, the ordinal logit regression model was developed using the OLOGIT procedure available in the STATA software package. Stepwise model selection was carried out where the significant levels for entry into the model was 0.05 and it for removal from the model is 0.15. Variables were entered into and removed from the model in such a way that each forward selection step was followed by one or more backward elimination steps. The stepwise selection procedure terminated when further variable can be added into the model, or if the variable just entered into the model is the only variable removed in the subsequent backward elimination. Thirdly, the Brant test was performed to test if the parallel regression assumption was violated. The Brant procedure in the STATA was used to execute this test with the confidence level 0.05. Finally, if the assumption was violated, the partial odds regression model was developed with the selected explanatory variables in the first step. The GOLOGIT2 procedure in the STATA was carried out for the estimation. This procedure is developed by Richard Williams to estimate generalized ordered logit models for ordinal dependent variables, including three special cases of the generalized model: the proportional odds/parallel lines model, the partial proportional odds model, and the logistic regression model. Hence, the GOLOGIT2 can estimate models that are less restrictive than the proportional odds /parallel lines models (whose assumptions are often violated) but more parsimonious and interpretable than those estimated by a non-ordinal method, such as multinomial logistic regression.

4.3.2 Cross Tabulations between Explanatory Variables and Crash Severity

In order to obtain a better understanding about the selected explanatory variables, cross tabulations of binary or categorical variables with crash severity were developed and given in Tables 4.4 and 4.5.

Table 4.4 Cross Tabulation between Explanatory Variables and Crash Severity

		Crash Severity					
Frequency Row %	Value	1	2	3	4	5	Total
DAYLIGHT	0	2502	1318	1176	500	146	5642
		44.3%	23.4%	20.8%	8.9%	2.6%	100.0%
	1	5152	2950	2192	825	107	11226
		45.9%	26.3%	19.5%	7.3%	1.0%	100.0%
	Total	7654	4268	3368	1325	253	16868
		45.4%	25.3%	20.0%	7.9%	1.5%	100.0%
BDWTHER	0	5119	2775	2242	871	166	11173
		45.8%	24.8%	20.1%	7.8%	1.5%	100.0%
	1	2535	1493	1126	454	87	5695
		44.5%	26.2%	19.8%	8.0%	1.5%	100.0%
	Total	7654	4268	3368	1325	253	16868
		45.4%	25.3%	20.0%	7.9%	1.5%	100.0%
SPEC_SEC	0	4105	2248	1823	766	170	9112
		45.1%	24.7%	20.0%	8.4%	1.9%	100.0%
	1	3549	2020	1545	559	83	7756
		45.8%	26.0%	19.9%	7.2%	1.1%	100.0%
	Total	7654	4268	3368	1325	253	16868
		45.4%	25.3%	20.0%	7.9%	1.5%	100.0%
SURF_DRY	0	1256	751	593	216	38	2854
		44.0%	26.3%	20.8%	7.6%	1.3%	100.0%
	1	6398	3517	2775	1109	215	14014
		45.7%	25.1%	19.8%	7.9%	1.5%	100.0%
	Total	7654	4268	3368	1325	253	16868
		45.4%	25.3%	20.0%	7.9%	1.5%	100.0%

Table 4.4 (Continued)

Frequency Row %	Value	Crash Severity 1	2	3	4	5	Total
TRAF_CONT	0	2508	1395	1068	442	66	5479
		45.8%	25.5%	19.5%	8.1%	1.2%	100.0%
	1	5146	2873	2300	883	187	11389
		45.2%	25.2%	20.2%	7.8%	1.6%	100.0%
	Total	7654	4268	3368	1325	253	16868
		45.4%	25.3%	20.0%	7.9%	1.5%	100.0%
VIS_OBS	0	6942	3812	2985	1190	234	15163
		45.8%	25.1%	19.7%	7.8%	1.5%	100.0%
	1	712	456	383	135	19	1705
		41.8%	26.7%	22.5%	7.9%	1.1%	100.0%
	Total	7654	4268	3368	1325	253	16868
		45.4%	25.3%	20.0%	7.9%	1.5%	100.0%
URBAN	0	1053	577	626	323	84	2663
		39.5%	21.7%	23.5%	12.1%	3.2%	100.0%
	1	6601	3691	2742	1002	169	14205
		46.5%	26.0%	19.3%	7.1%	1.2%	100.0%
	Total	7654	4268	3368	1325	253	16868
		45.4%	25.3%	20.0%	7.9%	1.5%	100.0%
FREEWAY	0	4362	2563	2045	813	147	9930
		43.9%	25.8%	20.6%	8.2%	1.5%	100.0%
	1	3292	1705	1323	512	106	6938
		47.4%	24.6%	19.1%	7.4%	1.5%	100.0%
	Total	7654	4268	3368	1325	253	16868
		45.4%	25.3%	20.0%	7.9%	1.5%	100.0%
HVINV	0	6069	3843	3041	1192	196	14341
		42.3%	26.8%	21.2%	8.3%	1.4%	100.0%
	1	1585	425	327	133	57	2527
		62.7%	16.8%	12.9%	5.3%	2.3%	100.0%
	Total	7654	4268	3368	1325	253	16868
		45.4%	25.3%	20.0%	7.9%	1.5%	100.0%
ALCHDRUG	0	7070	4029	3112	1160	150	15521
		45.6%	26.0%	20.1%	7.5%	1.0%	100.0%
	1	584	239	256	165	103	1347
		43.4%	17.7%	19.0%	12.2%	7.6%	100.0%
	Total	7654	4268	3368	1325	253	16868
		45.4%	25.3%	20.0%	7.9%	1.5%	100.0%

Table 4.4 (Continued)

		Crash Severity					
Frequency Row %	Value	1	2	3	4	5	Total
GR_CUR	0	6058	3428	2628	1064	167	13345
		45.4%	25.7%	19.7%	8.0%	1.3%	100.0%
	1	1596	840	740	261	86	3523
		45.3%	23.8%	21.0%	7.4%	2.4%	100.0%
	Total	7654	4268	3368	1325	253	16868
		45.4%	25.3%	20.0%	7.9%	1.5%	100.0%
AGE	0	1674	1107	911	333	58	4083
		41.0%	27.1%	22.3%	8.2%	1.4%	100.0%
	1	5307	2779	2101	847	168	11202
		47.4%	24.8%	18.8%	7.6%	1.5%	100.0%
	2	673	382	356	145	27	1583
		42.5%	24.1%	22.5%	9.2%	1.7%	100.0%
	Total	7654	4268	3368	1325	253	16868
		45.4%	25.3%	20.0%	7.9%	1.5%	100.0%

4.3.3 Estimation Results

The estimation of results of the ordinal logit regression is given in Table 4.6. The sample size is 16,868 observations, and the Likelihood Ratio (LR) test statistic falls into the rejection area ($p - value = 0 < 0.05$) . That means the overall explanatory variables of the model have significant influence on the responses (crash severity levels) at a statistical significance level 0.05. Except for VIS_OBS, all slope coefficients are significant at a confidence level 0.05. Although the *p-value* of VIS_OBS is little greater than 0.05, the variable was still included in the model since more variables increase the explanation ability of the model.

Because the response variable has 5 levels, four models were fitted with same slope coefficients and different constants (interprets). Model I indicates the probability ratio of the high injury severity levels (greater than no injury) to the lowest injury severity level (no injury); Model II presents the probability ratio of the higher injury severity (greater than possible injury) to the low severity levels (possible injury and no injury); Model III and Model IV denote the probability ratios at a resemble way. In the STATA, the cut point on the latent variable is estimated as a substitute of model constant coefficient (α_j). Actually, the cut point is equal to the reversed value of the constant.

Table 4.5 Estimation of Ordinal Logit Regression for Work Zone Crash Severity Model

Ordered logistic regression | Number of observation = 16868
LR chi2(11) = 669.52
Prob > chi2 = 0.0000
Log likelihood = -21438.289 | Pseudo R2 = 0.0154

ACCISEV	Coef.	Std. Err.	z	$p > \lvert z \rvert$	[95% Conf. Interval]	
DAYLIGHT	-0.0883	0.0317	-2.7900	0.0050	-0.1504	-0.0263
GR_CUR	0.0850	0.0360	2.3600	0.0180	0.0145	0.1556
VIS_OBS	0.0895	0.0470	1.9100	0.0570	-0.0026	0.1815
URBAN	-0.2839	0.0450	-6.3100	0.0000	-0.3721	-0.1958
FREEWAY	-0.4392	0.0462	-9.5000	0.0000	-0.5299	-0.3486
MAXSPEED	0.0227	0.0021	10.8300	0.0000	0.0186	0.0268
ALCHDRUG	0.2927	0.0575	5.0900	0.0000	0.1799	0.4054
HVINV	-0.7720	0.0448	-17.2400	0.0000	-0.8598	-0.6842
MIDDLE_AGE	-0.1434	0.0304	-4.7200	0.0000	-0.2030	-0.0838
/cut1 (Model I)	0.2862	0.1306			0.0303	0.5420
/cut2 (Model II)	1.3845	0.1310			1.1277	1.6413
/cut3 (Model III)	2.8030	0.1327			2.5430	3.0631
/cut4 (Model IV)	4.7298	0.1446			4.4463	5.0132

Table 4.7 illustrates the results of the Brant Test of the parallel regression assumption. Four binary logistic models were constructed to perform the Wald test on the identification of slope coefficients across the four binary models. From Table 4.7, the p-values of DAYLIGHT, GR_CUR, URBAN, ALCHDRUG, and HVINV are less than 0.05. That means the hypothesis that these slope coefficients are identical across the models is rejected at a confidence level 0.05. So it can be concluded that a significant test statistic provides evidence that the parallel regression assumption has been violated for

these variables. However, the assumption is not violated for FREEWAY, MAXSPEED, MIDDLE_AGE, and VIS_OBS.

Table 4.6 Results of Brant Test of Parallel Regression Assumption

Binary Logistic Model				
	$Y>1$	$Y>2$	$Y>3$	$Y>4$
DAYLIGHT	-0.0481	-0.1544	-0.1918	-0.5071
GR_CUR	0.0487	0.1230	0.0755	0.6317
VIS_OBS	0.1191	0.0793	-0.0456	-0.1883
URBAN	-0.1519	-0.3567	-0.5290	-0.6107
FREEWAY	-0.4609	-0.4375	-0.3933	-0.6015
MAXSPEED	0.0234	0.0224	0.0219	0.0401
ALCHDRUG	-0.0075	0.3636	0.8972	2.0421
HVINV	-0.8667	-0.5855	-0.2988	0.6685
MIDDLE_AGE	-0.1461	-0.1638	-0.1038	-0.1752
Cons	-0.4887	-1.3731	-2.7409	-5.9643
Brant Test of Parallel Regression Assumption				
Variable	*Chi-Square*	*p-value*	DF	
ALL	499.88	0.0000	27	
DAYLIGHT	17.3300	0.0010	3	
GR_CUR	24.1900	0.0000	3	
VIS_OBS	3.6800	0.2980	3	
URBAN	31.3600	0.0000	3	
FREEWAY	1.9100	0.5910	3	
MAXSPEED	4.2600	0.2340	3	
ALCHDRUG	217.2100	0.0000	3	
HVINV	110.2200	0.0000	3	
MIDDLE_AGE	1.8700	0.5990	3	

Tables 4.8 and 4.9 present the estimate results of the partial proportional regression for work zone crashes. The model estimation is presented in Table 4.8, and the statistic criteria for assessing the partial proportional odds model are given in Table 4.9. As same as the ordinal logit regression, four models were estimated. The variables for which the parallel assumption is not violated (FREEWAY, MAXSPEED, MIDDLE_AGE, and VIS_OBS) have same coefficients across the models and those for which the parallel assumption is violated (DAYLIGHT, GR_CUR, URBAN, ALCHDRUG, and HVINV) have different coefficients.

Table 4.7 Estimation Results of Coefficients of Partial Proportional Odds Regression

ACCISEV	Coef.	Std. Err.	z	$p > \|z\|$	[95% Conf.	Interval]
Model I: $j = 1$						
DAYLIGHT	-0.0433	0.0347	-1.2500	0.2120	-0.1113	0.0247
GR_CUR	0.0445	0.0392	1.1300	0.2570	-0.0325	0.1214
VIS_OBS	0.0900	0.0471	1.9100	0.0560	-0.0024	0.1823
URBAN	-0.1434	0.0491	-2.9200	0.0030	-0.2395	-0.0472
FREEWAY	-0.4522	0.0462	-9.7800	0.0000	-0.5428	-0.3616
MAXSPEED	0.0233	0.0021	11.0900	0.0000	0.0192	0.0274
ALCHDRUG	-0.0170	0.0606	-0.2800	0.7790	-0.1358	0.1017
HVINV	-0.8677	0.0460	-18.8600	0.0000	-0.9578	-0.7775
MIDDLE_AGE	-0.1431	0.0305	-4.6900	0.0000	-0.2028	-0.0833
Cons	-0.4956	0.1263	-3.9300	0.0000	-0.7431	-0.2482
Model II: $j = 2$						
DAYLIGHT	-0.1440	0.0374	-3.8500	0.0000	-0.2174	-0.0707
GR_CUR	0.1347	0.0422	3.1900	0.0010	0.0520	0.2174
VIS_OBS	0.0900	0.0471	1.9100	0.0560	-0.0024	0.1823
URBAN	-0.3474	0.0494	-7.0300	0.0000	-0.4443	-0.2505
FREEWAY	-0.4522	0.0462	-9.7800	0.0000	-0.5428	-0.3616
MAXSPEED	0.0233	0.0021	11.0900	0.0000	0.0192	0.0274
ALCHDRUG	0.3713	0.0620	5.9900	0.0000	0.2498	0.4928
HVINV	-0.5917	0.0539	-10.9800	0.0000	-0.6973	-0.4861
MIDDLE_AGE	-0.1431	0.0305	-4.6900	0.0000	-0.2028	-0.0833
Cons	-1.4436	0.1268	-11.3800	0.0000	-1.6922	-1.1949
Model III: $j = 3$						
DAYLIGHT	-0.1835	0.0577	-3.1800	0.0010	-0.2966	-0.0705
GR_CUR	0.0734	0.0646	1.1400	0.2560	-0.0532	0.2000
VIS_OBS	0.0900	0.0471	1.9100	0.0560	-0.0024	0.1823
URBAN	-0.5022	0.0658	-7.6300	0.0000	-0.6312	-0.3733
FREEWAY	-0.4522	0.0462	-9.7800	0.0000	-0.5428	-0.3616
MAXSPEED	0.0233	0.0021	11.0900	0.0000	0.0192	0.0274
ALCHDRUG	0.8945	0.0792	11.3000	0.0000	0.7393	1.0497
HVINV	-0.3199	0.0815	-3.9300	0.0000	-0.4796	-0.1603
MIDDLE_AGE	-0.1431	0.0305	-4.6900	0.0000	-0.2028	-0.0833
Cons	-2.8041	0.1369	-20.4800	0.0000	-3.0724	-2.5357
Model IV: $j = 4$						
DAYLIGHT	-0.5151	0.1344	-3.8300	0.0000	-0.7785	-0.2518
GR_CUR	0.5186	0.1260	4.1200	0.0000	0.2717	0.7655
VIS_OBS	0.0900	0.0471	1.9100	0.0560	-0.0024	0.1823
URBAN	-0.6882	0.1284	-5.3600	0.0000	-0.9399	-0.4364
FREEWAY	-0.4522	0.0462	-9.7800	0.0000	-0.5428	-0.3616
MAXSPEED	0.0233	0.0021	11.0900	0.0000	0.0192	0.0274
ALCHDRUG	1.8792	0.1393	13.4900	0.0000	1.6062	2.1522
HVINV	0.4900	0.1361	3.6000	0.0000	0.2233	0.7567
MIDDLE_AGE	-0.1431	0.0305	-4.6900	0.0000	-0.2028	-0.0833
Cons	-4.9755	0.1961	-25.3800	0.0000	-5.3598	-4.5913

Table 4.8 Statistic Criteria for Assessing
Partial Proportional Odds Regression

Partial Proportional Odds Regression	Number of obs = 16868
	LR chi2(24) = 1087.66
	Prob > chi2 = 0.0000
Log likelihood = -21229.217	Pseudo R2 = 0.0250

4.3.4 Interpretation

The crash severity model estimated by the ordinal logit regression has same slope coefficients across all *K-1* severity levels. For example, the coefficient for DAYLIGHT is -0.0883, which means that the presence of day light (DAYLIGHT=1) tends to reduce the injury severity of work zone crashes, and the odds ratio ($\exp(-0.0883) = 0.9155$) is same for all pairs of the comparisons: 2, 3, 4, 5 versus 1; 3, 4, 5 versus 1, 2; 4, 5 versus 1, 2, 3; and 5 versus 1, 2, 3, 4. The Table 4.10 gives the odds ratio for each explanatory variable in the ordinal logit models.

Table 4.9 Odds Ratio of Explanatory Variables in the Ordinal Logit Models

	Model I	Model II	Model III	Model IV
Variable	$\frac{\Pr(Y>1)}{\Pr(Y=1)}$	$\frac{\Pr(Y>2)}{\Pr(Y\le 2)}$	$\frac{\Pr(Y>3)}{\Pr(Y\le 3)}$	$\frac{\Pr(Y=5)}{\Pr(Y\le 4)}$
DAYLIGHT	0.9155	0.9155	0.9155	0.9155
GR_CUR	1.0887	1.0887	1.0887	1.0887
VIS_OBS	1.0936	1.0936	1.0936	1.0936
URBAN	0.7528	0.7528	0.7528	0.7528
FREEWAY	0.6446	0.6446	0.6446	0.6446
MAXSPEED	1.0230	1.0230	1.0230	1.0230
ALCHDRUG	1.3400	1.3400	1.3400	1.3400
HVINV	0.4621	0.4621	0.4621	0.4621
MIDDLE_AGE	0.8664	0.8664	0.8664	0.8664

Based on this table, some interpretations can be concluded as follows:

(1) the presence of day light tends to reduce the crash severity of work zone crashes;

(2) if the crash location is in urban area or in freeway, the injury severity of work zone crashes is also more likely to decrease;

(3) if there is a curve or grade at the crash location, an increase in the injury severity of work zone crashes is expected more likely;

(4) the presence of vision obstruction leads to a high probability of the occurrence of more severe work zone crashes;

(5) a high speed limit tends to increase the crash severity of work zone crashes;

(6) if alcohol or drug is involved, the work zone crash severity is more likely to increase;

(7) the involvement of heavy vehicles tends to reduce the work zone crash severity; and

(8) young and old drivers (MIDDLE_AGE=0) tends to conduct more severe work zone crashes.

In contrast to the ordinal models, some of the slope coefficients in the partial regression model are different across severity levels. So the interpretation for coefficients in the partial proportional odds logit regression model is different to those in the ordinal logit model. The odds ratios of explanatory variables for the partial proportional odds logit regression model are given in Table 4.11.

Table 4.10 Odds Ratio of Explanatory Variables in the Partial Regression Models

Variable	Model I $\frac{\Pr(Y>1)}{\Pr(Y=1)}$	Model II $\frac{\Pr(Y>2)}{\Pr(Y\le 2)}$	Model III $\frac{\Pr(Y>3)}{\Pr(Y\le 3)}$	Model IV $\frac{\Pr(Y=5)}{\Pr(Y\le 4)}$
DAYLIGHT	-	0.8659	0.8324	0.5974
GR_CUR	-	1.1442	-	1.6797
VIS_OBS	1.0942	1.0942	1.0942	1.0942
URBAN	0.8664	0.7065	0.6052	0.5025
FREEWAY	0.6362	0.6362	0.6362	0.6362
MAXSPEED	1.0236	1.0236	1.0236	1.0236
ALCHDRUG	-	1.4496	2.4461	6.5483
HVINV	0.4199	0.5534	0.7262	1.6323
MIDDLE_AGE	0.8667	0.8667	0.8667	0.8667

Because VIS_OBS, FREEWAY, MAXSPEED, and MIDDLE_AGE do not violate the parallel regression assumption, the odds ratios for them are identical across the four models (Model I to Model IV). So the interpretations for these variables are similar to the corresponding ones in the ordinal logit models. The interpretations are given as:

(1) the presence of vision obstructions at crash location tends to increase the injury severity;

(2) if crashes occurred in freeways rather than surface roads, injury severity is likely to be reduced;

(3) a higher speed limit tends to a higher injury severity; and

(4) young and old drivers (MIDDLE_AGE=0) tends to increase the work zone crash severity.

But for DAYLIGHT, GR_CUR, URBAN, ALCHDRUG, and HVINV, the interpretations are different across the four models. The interpretation for these variables is given as follows:

(1) The presence of day light (DAYLIGHT=1)

The factor in Model I has no significant influence on the crash severity (p-value=0.2120>0.05). For other three models, this factor tends to reduce the crash severity (odds ratios are less than 1). But its effects on injury severity change in ascending order (associated odds ratios in descending order) as an increase of severity levels from possible injury to fatality. The factor most tends to reduce the probability of fatality (Y=5) to no fatality (Y=1, 2, 3, and 4), followed by the probability ratio of more severe injury (Y= 4 and 5) to less severe injury (Y=1, 2, and 3). The factor least tends to reduce the probability ratio of injury (*Y*=3, 4, and 5) to possible or no injury (*Y*=2 and 1). Figure 4.1 indicates the variety of odds ratios.

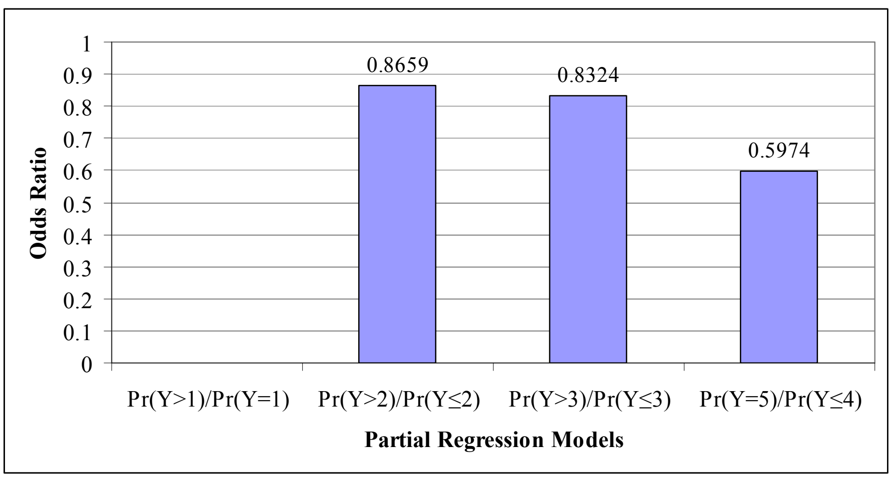


Figure 4.1 Effects of Presence of Day Light

(2) The crash location in urban area (URBAN=1)

This factor is significant in all models. This factor has ascending tendency to reduce the injury severity of work zone crashes across the four models from I to IV as shown in Figure 4.2.

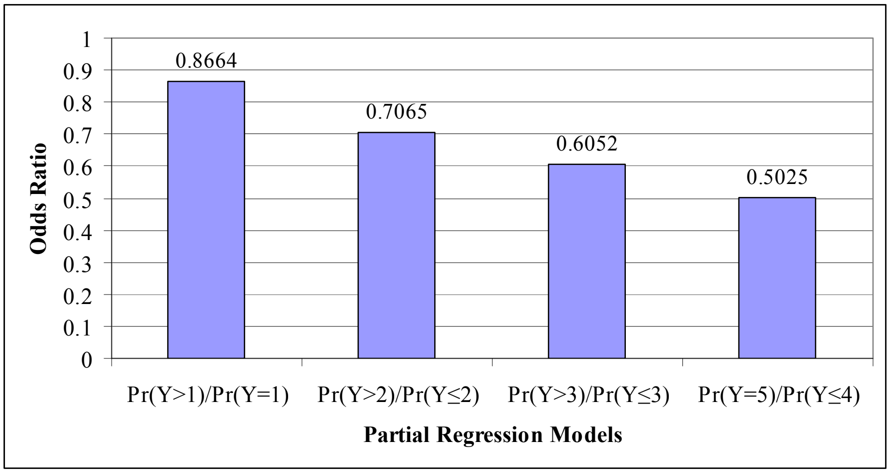


Figure 4.2 Effects of Urban Area

(3) A curve or grade at the crash location (GR_CUR=1)

This factor has significant influence in Model II and Model IV. That the odds for these two models are greater than 1 indicates that the factor tends to increase the injury severity at ascending order from Model II to Model IV. From Figure 4.3, we know that the factor most tends to increase the probability ratio of fatality to no fatality, followed by the one of injury to possible or no injury.

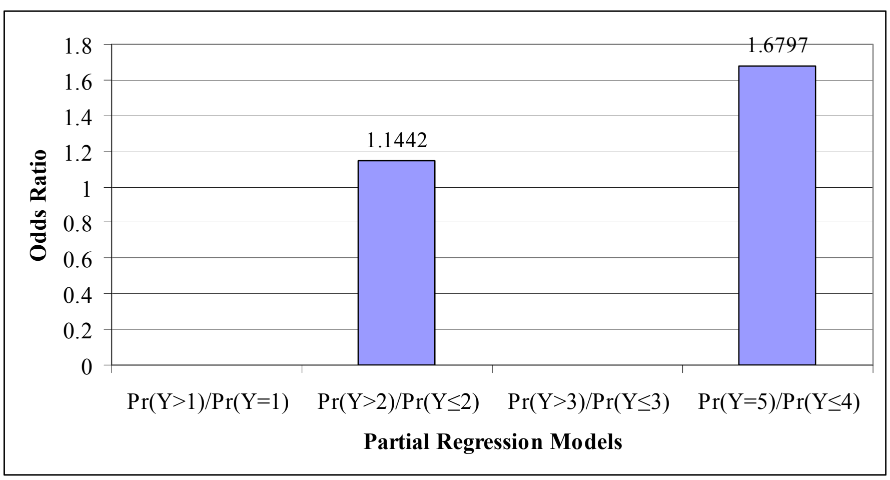


Figure 4.3 Effects of Curve or Grade

(4) The involvement of alcohol or drugs (ALCHDRUG=1)

This factor has no significant influence on the probability ratio of injury to no injury (*p-value*>0.05). But it tends to increase the injury severity across other severity levels. Its odds ratio for the fatality to no fatality model is highest, almost double greater than the second one. It can be concluded that the involvement of alcohol or drugs is a very important factor to conduct a fatal work zone crash.

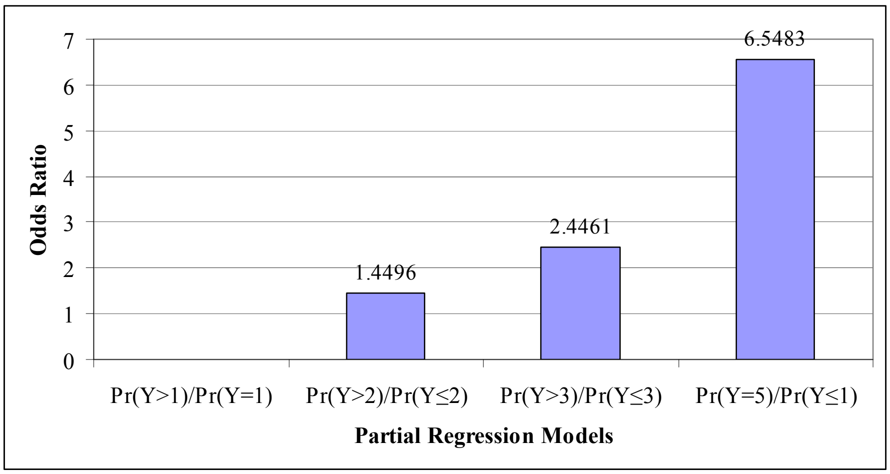


Figure 4.4 Effects of Involvement of Alcohol or Drugs

(5) The involvement of Heavy Vehicle (HVINV=1)

The factor tends to increase the probability ratio of fatality to no fatality. That means it is an important factor to introduce a work zone fatal crash. But for other severity levels, the factor tends to reduce the crash severity.

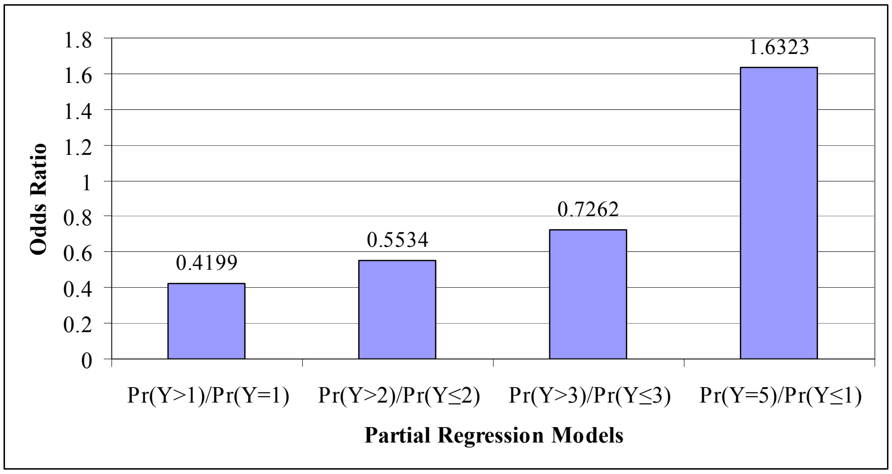


Figure 4.5 Effects of Involvement of Heavy Vehicle

4.4 Crash Severity Model for Rear-end Work Zone Crashes

Figure 4.6 presents the distribution of work zone crashes over crash types. The most dominant crash type is rear-end with a percentage of 38%, followed by angle (12%) and side swipe (11%). In this section, the crash severity model for rear-end work zone crashes was developed to investigate the influence of explanatory variables on the injury severity of rear-end work zone crashes.

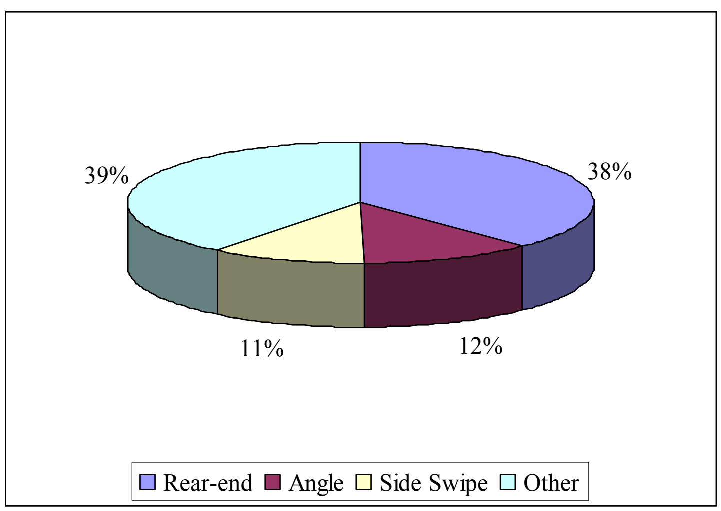


Figure 4.6 Distribution of Crash Type

4.4.1 Description of Rear-end Dataset

There are 6355 observations extracted from the original dataset for rear-end crash analysis. To explore the rear-end dataset, we found the number of fatal crashes was too small (0.6% of total crashes). For ensuring enough observations for each severity value, the severity levels 4 and 5 were combined. The updated description of injury severity of work zone rear-end crashes is given in Table 4.12. The cross tabulations of categorical variables are shown in Table 4.13, while Table 4.14 presents the description of continuous variables.

Table 4.11 Description of Response Variable for Rear-end Dataset

Value	Description	Frequency	Distribution
1	No Injury	2523	39.7%
2	Possible Injury	2150	33.8%
3	No-Capacitating Injury	1251	19.7%
4	Incapacitating Injury or Fatal	431	6.8%

Table 4.12 Cross Tabulations between Explanatory Variables and Crash Severity for Rear-end Dataset

		Crash Severity				
Frequency Row %	Value	1	2	3	4	Total
DAYLIGHT	0	691	549	329	148	1717
		40.2%	32.0%	19.2%	8.6%	100.0%
	1	1832	1601	922	283	4638
		39.5%	34.5%	19.9%	6.1%	100.0%
	Total	2523	2150	1251	431	6355
		39.7%	33.8%	19.7%	6.8%	100.0%
BDWTHER	0	1766	1417	864	294	4341
		40.7%	32.6%	19.9%	6.8%	100.0%
	1	757	733	387	137	2014
		37.6%	36.4%	19.2%	6.8%	100.0%
	Total	2523	2150	1251	431	6355
		39.7%	33.8%	19.7%	6.8%	100.0%
SPEC_SEC	0	1494	1287	809	313	3903
		38.3%	33.0%	20.7%	8.0%	100.0%
	1	1029	863	442	118	2452
		42.0%	35.2%	18.0%	4.8%	100.0%
	Total	2523	2150	1251	431	6355
		39.7%	33.8%	19.7%	6.8%	100.0%
SURF_DRY	0	344	318	181	58	901
		38.2%	35.3%	20.1%	6.4%	100.0%
	1	2179	1832	1070	373	5454
		40.0%	33.6%	19.6%	6.8%	100.0%
	Total	2523	2150	1251	431	6355
		39.7%	33.8%	19.7%	6.8%	100.0%
GR_CUR	0	2025	1708	960	347	5040
		40.2%	33.9%	19.0%	6.9%	100.0%
	1	498	442	291	84	1315
		37.9%	33.6%	22.1%	6.4%	100.0%
	Total	2523	2150	1251	431	6355

Table 4.12 (Continued)

Frequency Row %	Value	Crash Severity 1	2	3	4	Total
TRAF_CONT	0	773	740	422	150	2085
		37.1%	35.5%	20.2%	7.2%	100.0%
	1	1750	1410	829	281	4270
		41.0%	33.0%	19.4%	6.6%	100.0%
	Total	2523	2150	1251	431	6355
		39.7%	33.8%	19.7%	6.8%	100.0%
VIS_OBS	0	2381	2020	1152	402	5955
		40.0%	33.9%	19.3%	6.8%	100.0%
	1	142	130	99	29	400
		35.5%	32.5%	24.8%	7.3%	100.0%
	Total	2523	2150	1251	431	6355
		39.7%	33.8%	19.7%	6.8%	100.0%
URBAN	0	303	226	199	85	813
		37.3%	27.8%	24.5%	10.5%	100.0%
	1	2220	1924	1052	346	5542
		40.1%	34.7%	19.0%	6.2%	100.0%
	Total	2523	2150	1251	431	6355
		39.7%	33.8%	19.7%	6.8%	100.0%
FREEWAY	0	1265	1175	683	232	3355
		37.7%	35.0%	20.4%	6.9%	100.0%
	1	1258	975	568	199	3000
		41.9%	32.5%	18.9%	6.6%	100.0%
	Total	2523	2150	1251	431	6355
		39.7%	33.8%	19.7%	6.8%	100.0%
HVINV	0	2224	1987	1150	354	5715
		38.9%	34.8%	20.1%	6.2%	100.0%
	1	299	163	101	77	640
		46.7%	25.5%	15.8%	12.0%	100.0%
	Total	2523	2150	1251	431	6355
ALCHDRUG	0	2323	2034	1173	373	5903
		39.4%	34.5%	19.9%	6.3%	100.0%
	1	200	116	78	58	452
		44.2%	25.7%	17.3%	12.8%	100.0%
	Total	2523	2150	1251	431	6355
		39.7%	33.8%	19.7%	6.8%	100.0%

Table 4.12 (Continued)

Frequency Row %	Value	Crash Severity 1	2	3	4	Total
AGE	0	639	565	353	100	1657
		38.6%	34.1%	21.3%	6.0%	100.0%
	1	1718	1424	780	292	4214
		40.8%	33.8%	18.5%	6.9%	100.0%
	2	166	161	118	39	484
		34.3%	33.3%	24.4%	8.1%	100.0%
	Total	2523	2150	1251	431	6355
		39.7%	33.8%	19.7%	6.8%	100.0%

Table 4.13 Descriptive Statistic of Continues Variables

Variable	N	Minimum	Maximum	Range	Mean	Std. Deviation
SURWIDTH	6355	10	88	78	29.58	9.298
MAXSPEED	6355	20	70	50	53.60	10.586
SECTADT	6355	1800	302000	300200	73032.64	55231.438

4.4.2 Estimation Results for Rear-end Dataset

As same as for the overall work zone crash dataset, an ordinal logit regression model was developed for rear-end dataset using the STATA software. Three dummy variables (YOUNG_AGE, MIDDLE_AGE, and OLD_AGE) were also derived from AGE. Stepwise procedure was implemented to select explanatory variables which are significant to the response. The parallel regression assumption was examined by the Brant test, and the partial proportional odds logit regression model was estimated. Because the injury severity for the rear-end dataset has 4 levels rather than 5, 3 models (Model I, II, III) were estimated for both of the two regressions. The estimation results of ordinal logit models are given in Tables 4.15. From the table, we know that overall or individual coefficients are significantly not equal to zero. Table 4.16 offers the result of Brant test. It indicates that the parallel regression assumption is violated for HVINV, FREEWAY, and MAXSPEED.

Table 4.14 Estimation of Ordinal Logit Regression for Crash Severity Model (Rear-end Dataset)

Ordered logistic regression	Number of obs. = 6355
	LR chi2(11) = 134.97
	Prob > chi2 = 0.0000
Log likelihood = -7786.3796	Pseudo R2 = 0.0086

ACCISEV	Coef.	Std. Err.	z	$p > \|z\|$	[95% Conf. Interval]	
SPEC_SECT	-0.2824	0.0547	-5.1600	0.0000	-0.3897	-0.1751
HVINV	-0.1679	0.0814	-2.0600	0.0390	-0.3275	-0.0082
GR_CUR	0.1905	0.0582	3.2700	0.0010	0.0764	0.3046
TRAF_CONT	-0.1401	0.0496	-2.8200	0.0050	-0.2373	-0.0429
FREEWAY	-0.6623	0.0710	-9.3300	0.0000	-0.8015	-0.5232
MAXSPEED	0.0248	0.0033	7.4700	0.0000	0.0183	0.0313
OLD_AGE	0.2064	0.0866	2.3800	0.0170	0.0367	0.3761
/cut1 (Model I)	0.4231	0.1651			0.0994	0.7467
/cut2 (Model II)	1.8874	0.1669			1.5604	2.2144
/cut3 (Model III)	3.5043	0.1730			3.1652	3.8434

Table 4.15 Results of Brant Test of Parallel Regression Assumption (Rear-end Dataset)

Binary Logistic Models			
	$Y > 1$	$Y > 2$	$Y > 3$
SPEC_SECT	-0.2460	-0.3262	-0.5546
HVINV	-0.3527	-0.0185	0.6342
GR_CUR	0.1916	0.2265	-0.0129
TRAF_CONT	-0.1781	-0.0982	-0.0995
FREEWAY	-0.5727	-0.7765	-0.8052
MAXSPEED	0.0179	0.0342	0.0329
OLD_AGE	0.1971	0.2445	0.1516
Cons	-0.0682	-2.3832	-3.8827

Brant Test of Parallel Regression Assumption			
Variable	*Chi-Square*	*p-value*	DF
All	97.3500	0.0000	14
SPEC_SECT	5.7800	0.0550	2
HVINV	51.6800	0.0000	2
GR_CUR	4.2000	0.1220	2
TRAF_CONT	1.7600	0.4150	2
FREEWAY	6.0600	0.0480	2
MAXSPEED	16.5400	0.0000	2
OLD_AGE	0.4900	0.7820	2

Tables 4.17 and 4.18 show the estimation results of the partial proportional logit regression.

Table 4.16 Estimation Results of Coefficients of Partial Proportional Odds Regression (Rear-end Dataset)

ACCISEV	Coef.	Std. Err.	z	$p > \|z\|$	[95% Conf. Interval]	
Model I: $j = 1$						
SPEC_SECT	-0.2878	0.0550	-5.2300	0.0000	-0.3956	-0.1799
HVINV	-0.3470	0.0854	-4.0600	0.0000	-0.5145	-0.1796
GR_CUR	0.1880	0.0581	3.2300	0.0010	0.0741	0.3020
TRAF_CONT	-0.1444	0.0497	-2.9100	0.0040	-0.2418	-0.0470
FREEWAY	-0.6607	0.0708	-9.3400	0.0000	-0.7994	-0.5220
MAXSPEED	0.0201	0.0035	5.7800	0.0000	0.0133	0.0269
OLD_AGE	0.2086	0.0868	2.4000	0.0160	0.0384	0.3787
Cons	-0.1503	0.1753	-0.8600	0.3910	-0.4939	0.1933
Model II: $j = 2$						
SPEC_SECT	-0.2878	0.0550	-5.2300	0.0000	-0.3956	-0.1799
HVINV	-0.0022	0.0951	-0.0200	0.9810	-0.1885	0.1841
GR_CUR	0.1880	0.0581	3.2300	0.0010	0.0741	0.3020
TRAF_CONT	-0.1444	0.0497	-2.9100	0.0040	-0.2418	-0.0470
FREEWAY	-0.6607	0.0708	-9.3400	0.0000	-0.7994	-0.5220
MAXSPEED	0.0318	0.0037	8.4900	0.0000	0.0244	0.0391
OLD_AGE	0.2086	0.0868	2.4000	0.0160	0.0384	0.3787
Cons	-2.2778	0.1918	-11.8800	0.0000	-2.6536	-1.9020
Model III: $j = 3$						
SPEC_SECT	-0.2878	0.0550	-5.2300	0.0000	-0.3956	-0.1799
HVINV	0.6380	0.1363	4.6800	0.0000	0.3708	0.9052
GR_CUR	0.1880	0.0581	3.2300	0.0010	0.0741	0.3020
TRAF_CONT	-0.1444	0.0497	-2.9100	0.0040	-0.2418	-0.0470
FREEWAY	-0.6607	0.0708	-9.3400	0.0000	-0.7994	-0.5220
MAXSPEED	0.0340	0.0057	6.0200	0.0000	0.0229	0.0451
OLD_AGE	0.2086	0.0868	2.4000	0.0160	0.0384	0.3787
Cons	-4.1088	0.3054	-13.4500	0.0000	-4.7073	-3.5102

Table 4.17 Statistic Criteria for Assessing Partial Proportional Odds Regression (Rear-end Dataset)

Partial Proportional Odds Regression	Number of obs = 16868
	LR chi2(24) = 206.70
	Prob > chi2 = 0.0000
Log likelihood = -7750.5142	Pseudo R2 = 0.0132

4.4.3 Interpretation

The odds ratios for the ordinal logit models are given in Table 4.19, and those for the partial proportional odds regression model are given in Table 4.20.

Table 4.18 Odds Ratios for the Ordinal Logit Models (Rear-end Dataset)

Variable	Model I $\frac{\Pr(Y>1)}{\Pr(Y=1)}$	Model II $\frac{\Pr(Y>2)}{\Pr(Y\le 2)}$	Model III $\frac{\Pr(Y>3)}{\Pr(Y\le 3)}$
SPEC_SECT	0.7540	0.7540	0.7540
HVINV	0.8454	0.8454	0.8454
GR_CUR	1.2099	1.2099	1.2099
TRAF_CONT	0.8693	0.8693	0.8693
FREEWAY	0.5157	0.5157	0.5157
MAXSPEED	1.0251	1.0251	1.0251
OLD_AGE	1.2292	1.2292	1.2292

Table 4.19 Odds Ratios for the Partial Regression Models (Rear-end Dataset)

Variable	Model I $\frac{\Pr(Y>1)}{\Pr(Y=1)}$	Model II $\frac{\Pr(Y>2)}{\Pr(Y\le 2)}$	Model III $\frac{\Pr(Y>3)}{\Pr(Y\le 3)}$
SPEC_SECT	0.7499	0.7499	0.7499
HVINV	0.7068	-	1.8927
GR_CUR	1.2068	1.2068	1.2068
TRAF_CONT	0.8655	0.8655	0.8655
FREEWAY	0.5165	0.5165	0.5165
MAXSPEED	1.0203	1.0323	1.0346
OLD_AGE	1.2320	1.2320	1.2320

Since only HVINV and MAXSPEED violate the parallel regression assumption, their coefficients are different across the injury severity levels. Others without the violation have same coefficients. The interpretations for the slope coefficients are given as follows:

(1) The crash location under the influence of intersection, interchanges, or other special sections (SPEC_SECT=1)

This factor is more likely to reduce the injury severity of work zone rear-end crashes.

(2) A curve or grade at the crash location (GR_CUR=1)

This factor tends to increase the injury severity of work zone rear-end crashes.

(3) The presence of traffic control measure at crash location (TRAF_CONT=1)

The injury severity of work zone rear-end crashes is more likely to be reduce due to this factor.

(4) The crash location at freeway sections (FREEWAY=1)

This factor is likely to reduce the injury severity of work zone rear-end crashes.

(5) Old drivers (OLD_AGE=1)

Old drivers tends to increase the injury severity of work zone rear-end crashes.

(6) Heavy vehicle involvement (HVINV=1)

This factor has no significant influence on the Model II. But its odds for Model III is greater than 1, in other words, the factor is likely to conduce fatal or incapacitating injury when a rear-end crash occurs at work zone area.

(7) Speed Limit (MAXSPEED)

A higher speed limit is likely to result in a more severe rear-end crash at work zone area.

4.5 Summary

In this chapter, two kinds of logit regression models for overall work zone crash dataset and rear-end crash dataset were developed respectively. The ordinal logit regression model has same slope coefficients across different severity levels. But the parallel regression assumption is violated for the two samples. The partial proportional odds logit regression model has less restrictive on the assumption. In this model, some variables for which the assumption is not violated have same slope coefficients across crash severity levels, while some variables that do not meet the assumption have different coefficients. And more specific interpretations are given to the variables that do not meet the assumption.

By summarizing the estimation results of the partial proportional odds logit model, Table 4.12 indicates the significant influences of the variables on the work zone crash severity. "⟵" indicates a variable is more likely to reduce the work zone crash severity if the variable adopts "1" (for binary variable) or increase by a positive value (for continuous variable). By contraries, "⟶" denotes a variable tends to increase the work zone crash severity.

Table 4.20 Summary of the Influence of the Explanatory Variables

Variable	Overall Work Zone Crash Severity Model				Rear-end Work Zone Crash Severity Model		
	$\frac{\Pr(Y > 1)}{\Pr(Y = 1)}$	$\frac{\Pr(Y > 2)}{\Pr(Y \le 2)}$	$\frac{\Pr(Y > 3)}{\Pr(Y \le 3)}$	$\frac{\Pr(Y = 5)}{\Pr(Y \le 4)}$	$\frac{\Pr(Y > 1)}{\Pr(Y = 1)}$	$\frac{\Pr(Y > 2)}{\Pr(Y \le 2)}$	$\frac{\Pr(Y > 3)}{\Pr(Y \le 3)}$
DAYLIGHT		←	←	←			
GR_CUR		→		→	→	→	→
VIS_OBS	→	→	→	→			
URBAN	←	←	←	←			
FREEWAY	←	←	←	←	←	←	←
MAXSPEED	→	→	→	→	→	→	→
ALCHDRUG		→	→	→			
HVINV	←	←	←	→	←		→
MIDDLE_AGE	←	←	←	←			
OLD_AGE					→	→	→
SPEC_SECT					←	←	←
TRAF_CONT					←	←	←

Chapter Five

Modeling Methodology for Work Zone Speed Profile Models

5.1 Introduction

In this study, work zone speed profile is defined as the speed distribution over the distance to the starting point of lane closure. Figure 5.1 shows a typical speed profile in open lanes at freeway work zones. When vehicles are running at open lanes and start to close to a work zone, their speed may be reduced due to the disturbance of lane changes from closed lanes or the backward queue formed by the capacity reduction. The speed continues to decrease until reaching a steady low value. After then, vehicles start to accelerate up to the normal speed. Apparently, the speed profile model is a nonlinear function of the position to the start point of lane closure. It is difficult to develop a uniform equation to describe the characteristics of speed profile using traditional statistical methodologies, like linear regression. In this study, a new learning machine algorithm, Support Vector Regression, was implemented to develop a uniform equation for describing the relationship between speed profile and various factors.

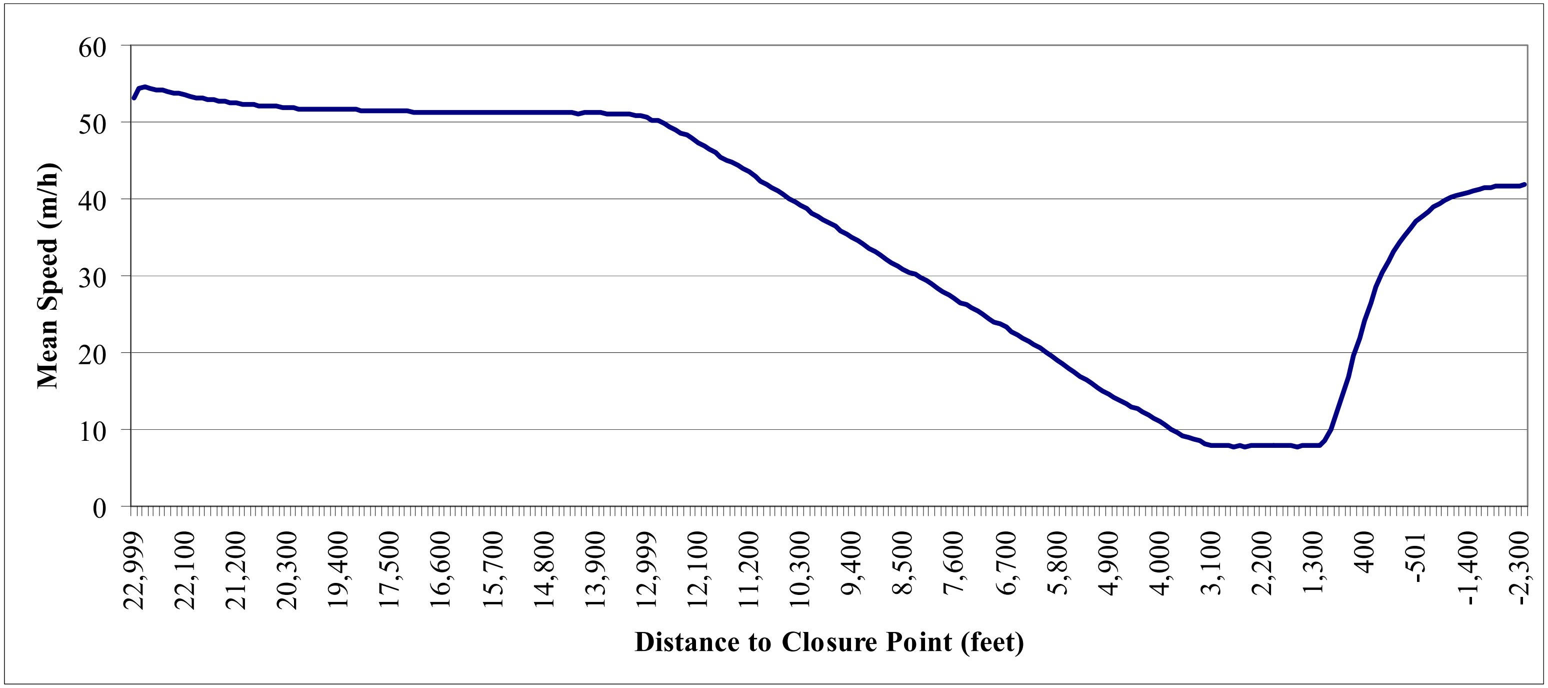


Figure 5.1 Speed Profile at Work Zones (Open Lanes)

5.2 Simulation-based Experiment Design

A simulation-based experiment was designed for the data collection. Rather than field experiment, simulation-based experiment has several advantages: (1) computer simulation could reduce the cost of data collection; (2) traffic factors are easy to be changed according to researcher's needs; (3) measures of effectiveness could be handled automatically by programming. Especially for work zone study, lane closure scenarios are difficult found in field, the simulation-based experiment provides a feasible method to generate various traffic scenarios so that the speed profile models can be constructed based on a more comprehensive dataset. But simulation-based experiment also has some limitations: (1) some factors cannot be realized in current simulation software; (2) even based on a calibrated model, the error between the simulation environment and the real world.

The micro-simulation software package CORSIM 5.1 was selected to create the experiment in this study. This package, originally developed by FHWA, has been used and validated for traffic operations research in past 20 years. CORSIM has the ability to simulate the freeway section with the integrated FRESIM module. The lane closure can be realized in FRESIM module by simulating an incident on a lane. A series of incidents are created along the lanes during a long period (greater than the simulation time), and the traffic on the same lanes is blocked from the range of the incidents. This method does not take into consideration of the taper section prior to the lane closure.

5.2.1 Simulation Model

In Figure 5.2, a 30,000 feet freeway section was setup in CORSIM with FRESIM module for simulating a real work zone section. Because this study focused on the four divided lanes freeway, two lanes with one closed lane on one direction was configured in the simulation model. Along the open lane, 31 traffic detectors were installed at an 800 feet interval to collect the measure of speed. An incident was simulated at the close lane to realize the lane closure. The closure point could be changed from A to G with different length of closure zone.

5.2.2 Model Calibration

The goodness of the results of computer-based traffic micro simulation is based on a well calibrated model which makes the model could reflect the real world more accurately. In general, the calibration configures internal factors related to traffic flow characteristics according to small size of observed values in field. Because this study did not aim at representing any real freeway segment or project, a calibrated simulation model developed in a previous research for freeway work zone was adopted.

Park and Won (2006) developed a systematic procedure for microscopic simulation model calibration and validation, which was successfully applied to freeway work zone case studies. In the procedure, a genetic algorithm optimization program is implemented to find an optimal calibration parameter set from the feasible parameter ranges. The optimal calibration parameter set for freeway work zones, shown in Table 5.1, which was adopted in this study.

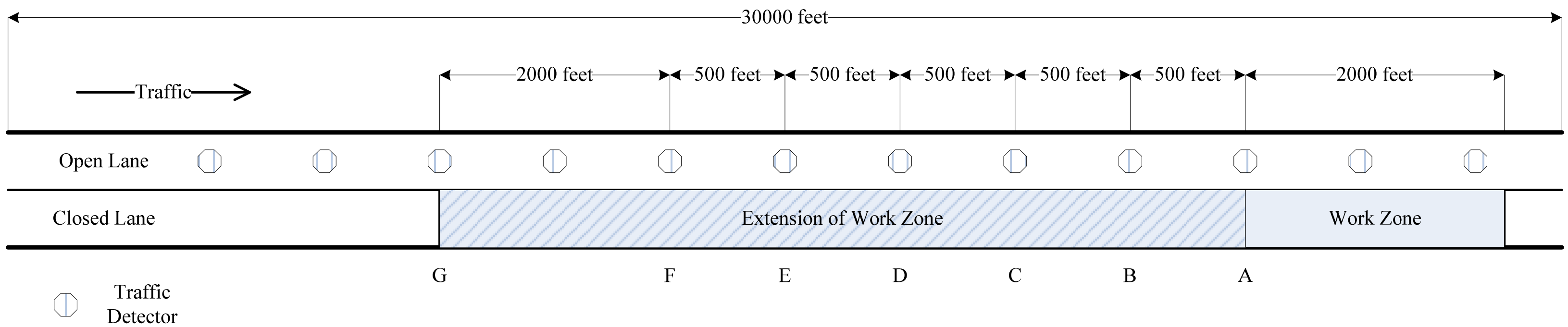


Figure 5.2 Simulation Model

Table 5.1 Calibration Parameters

Parameter	Default Value	Altered Value
Entry Vehicles Headway Distribution	Uniform	Enlarge
Car following sensitivity Index	1	1
Pitt car following constant (ft)	10	3
Lag acceleration (sec)	0.3	1.2
Lag deceleration (sec)	0.3	0.5
Time to complete a lane-change maneuver (sec)	2.0	1.0
Gap acceptance parameter	3	4
Percent of drivers desiring to yield to merging vehicles (%)	20	20
Multiplier for desire to make a discretionary lane change	0.5	0.4
Advantage threshold for discretionary lane change	0.4	0.8
Minimum separation for generation of vehicles (sec)	1.6	1.3
Distribution of free flow speed by driver type Index	1	2

5.2.3 Input Variables and Simulation Scenarios

Through the review of past research, the variables listed in Table 5.2 were selected to form various simulation scenarios. In the study, a typical work zone configuration was selected: two-lane freeway (one direction) with one lane closed. The default value of the volume distribution over lanes, 50:50, was adopted in this study, and there is no difference between left lane closed and right lane closed. In this study, only right lane closure was considered.

Free flow speed (FFS) is defined in HCM 2000 as the mean speed of passenger cars that could be accommodated under low to moderate flow rates on a uniform freeway segment under prevailing roadway and traffic conditions. A measure of FFS is the speed where the average headway is greater 4 seconds between two successive vehicles. The "highest" (ideal) type of basic freeway section is one in which the free-flow speed is 70 mph or higher. But the maximum FFS value is 70mph in CORSIM. In this study, the levels of FFS in freeway were categorized as three types: 70mph, 65mph, and 55mph. Since the simulation time constraints, the level of FFS 65mph was not presented.

A survey (Kamyab and Maze, Et al. 2001) conducted in 1999 to the state transportation agencies and toll authorities throughout the country showed that most participating agencies reported reducing speed limits by 10 mph below the normal posted speed during construction activities. In this study, based on the FFS in freeway sections, the reduction of FFS in work zone is fixed as 10mph.

Work zone grade is another important factor which affects the speed because of the presence of grades would exacerbate any flow constriction that would otherwise exist, particularly in the presence of heavy vehicles. In this study, 3 levels of work zone grade are selected: -5, 0, +5.

Heavy vehicle occupy more space on the roadway than passenger cars. Moreover, heavy vehicles accelerate slowly and their presence makes other drivers more apprehensive, and they need more operation time to shift lane in freeway. These factors reduce the overall capacity of the work zone. In this study, percentage of heavy vehicle is categorized into four levels: 0%, 5%, and 15%.

The entry volume for different scenarios should cover a wide range to evaluate the variable early merge comprehensively. Krammes and Lopez (1994) recommended that the short-term work zone lane closure capacity is 1600 pcphpl. For estimating the speed profile under congested and uncongested conditions, the range of entry volume is adopted from 800 pcph to 4000 pcph (one approach, two lanes). For reducing the simulation time, the entry volume is selected at 6 levels.

The length of lane closure is also selected as an input variable. A longer work zone would reduce the capacity of the freeway, and conducted a backward queue to the upstream which affects the upstream speed profile. In this study, 7 levels of the length of work zone were selected.

Table 5.2 Input Variables

Factor	Level
Work Zone Configuration	2 lanes with one closed
FFS	70mph, 65mph, 55mph
Work Zone FFS Reduction	10mph
Work Zone Grade	-5,0,+5
Percentage of Heavy Vehicle	0,5%,15%
Entry Volume (one direction)	800pcph, 1600pcph, 2400pcph, 2800pcph,3200pcph,4000pcph
Length of Lane Closure	2000feet, 2500feet, 3000feet, 3500feet, 4000feet, 4500feet, 7500feet

Simulation scenarios were performed by the combinations of each level of the variables. In total, 3 (FFS) ×3(Grade) ×3(HV %) ×6(Volume) ×7(Length of lane closure) = 1134 simulation scenarios were performed in this study. A Visual Basic.NET program was developed to generate these scenarios through revise the CORSIM (.trf) input file.

5.2.4 Data Collection

Because the CORSIM simulation is stochastic, the results from different simulations with a same input files will not be identical. To reduce the stochastic errors and get a stable result, it is necessary to run simulation for many times instead of only once. But too many runs will result in increase in the simulation time and the amount of output data. So the default value of 10 run times for each traffic scenario was adopted in this study, because it satisfied the precision of results and did not increase the simulation time greatly. The analysis time period for each run was 15 minutes.

The speed data was collected from 31 traffic detectors on the open lane at a 1 minute time interval with each run. For each scenario, the sample size is 15 (sample size per run) × 10 (run times) × 31 (detectors) = 4650. A program was developed to read these data from the CORSIM (.out) output files, and calculate the mean speed value of each detector for each scenario. Finally, each scenario had 31 observations.

5.3 Support Vector Regression

5.3.1 Introduction to Learning Machine

The classical regression statistical techniques like linear regression were based on the very strict assumption that probability distribution models or probability-density functions are known. Unfortunately, in many practical situations, there is not enough information about the underlying distributions laws, and distribution-free regression is needed that does not require knowledge of probability distributions.

In practical world, there are some systems are very complicated. People can only observe input and the corresponding output, but do not understand the relationship inside. The relationship between input and output deduced by learning from experimental data (samples or observations) is expected not only to be good fit to the samples, but also to have good generalization ability. This learning mechanism is called statistical learning machine, and its concept is shown in Figure 5.3. System is the research object which generates output y given input x. $\hat{y}$, the output of the learning machine, is the predicted value of y. The objective of the learning machine is to estimate the relationship between y and x to minimize the difference between y and $\hat{y}$.

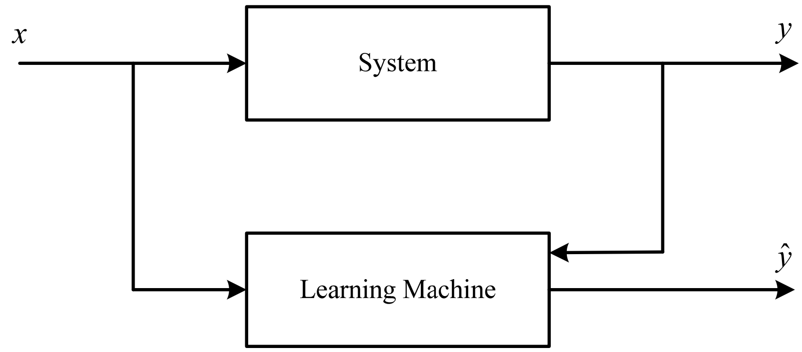


Figure 5.3 Concept of Learning Machine

Suppose we are given training data $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2)..., (\mathbf{x}_n, y_n)\}$. An approximating function $f(\mathbf{x}, \mathbf{w})$, which approximates of the underlying dependency between the input and output, minimizes the expected risk. $\mathbf{w}$ is the vector of parameters of the approximating function. The risk function is calculated as

$$R(f) = \int L(y, \hat{y}) dP(\mathbf{x}, y) = \int L(y, f(\mathbf{x}, \mathbf{w})) dP(\mathbf{x}, y) \tag{5.3.1}$$

where $P(\mathbf{x}, y)$ is a joint probability distribution equal to $P(\mathbf{x})P(y \mid \mathbf{x})$. $L(y, f(\mathbf{x}, \mathbf{w}))$ is the loss function, which represents the measure of the error introduced by the $f(\mathbf{x}, \mathbf{w})$. In regression, y is continues variable, two functions in use are the square error (L_2 norm),

$$L(y, f(\mathbf{x}, \mathbf{w})) = (y - f(\mathbf{x}, \mathbf{w}))^2 \tag{5.3.2}$$

and the absolution error (L_1 norm)

$$L(y, f(\mathbf{x}, \mathbf{w})) = |y - f(\mathbf{x}, \mathbf{w})| \tag{5.3.3}$$

5.3.2 Empirical Risk Minimization and Structural Risk Minimization

Learning can be considered a problem of finding the best estimator f using available data. However, the joint probability distribution is unknown, and the distribution-free learning must be performed based only on the training data pairs. With the only source of information a data set, the classical learning algorithm adopts the principle of empirical risk minimization (ERM):

$$\text{Min } R_{emp}[f] = \frac{1}{n}\sum_{i=1}^{n} L(y_i, f(\mathbf{x}_i, \mathbf{w})) \tag{5.3.4}$$

According to the classical law of large numbers ensures that the empirical risk R_{emp} converges to the expected risk R as the number of data points tends to infinity:

$$\lim_{n\to\infty}(|R(f) - R_{emp}(f)|) = 0 \tag{5.3.5}$$

Because ERM does not suggest how to find a constructive procedure for model design, the learning algorithms based on the principle of ERM (like ANN) may conduct an over-fitting problem and thus bad generalization properties.

When the training data is finite, the expected can be written as ():

$$R(f) \le R_{emp}(f) + \Phi \tag{5.3.6}$$

where Φ is confidence interval which is a monotonic decreasing function of the sample size over the complexity of the structure of the approximating function ($\Phi(\frac{n}{h})$). When n tends toward infinity, the confidence interval is tends to zero, so the estimator f_{emp} by minimizing $R_{emp}(f)$ is converged to the true estimator f by minimizing the $R(f)$. And the more complex the approximating function is, the larger confidence interval is. According to Equation 5.3.6, let function set $S = \{f(\mathbf{x}, \mathbf{w})\}$ be divided into a sequence of nested subsets ranked by corresponding confidence interval Φ.

$$S_1 \subset S_2 \subset \cdots \subset S_k \subset \cdots \subset S \tag{5.3.7}$$

Within each function subset, the confidence interval is same. And a superset has a lager Φ than its subset.

Structural risk minimization (SRM) is a novel inductive principle for learning from finite training data sets, and is shown in Figure 5.4. The basic idea of SRM is:

(1) To choose, from the sequence of the nested subsets of models (approximating functions), a subset (S_2) of the right complexity to describe the training data;

(2) To decide the best model by minimizing the empirical risk within the selected subset (S_2).

.

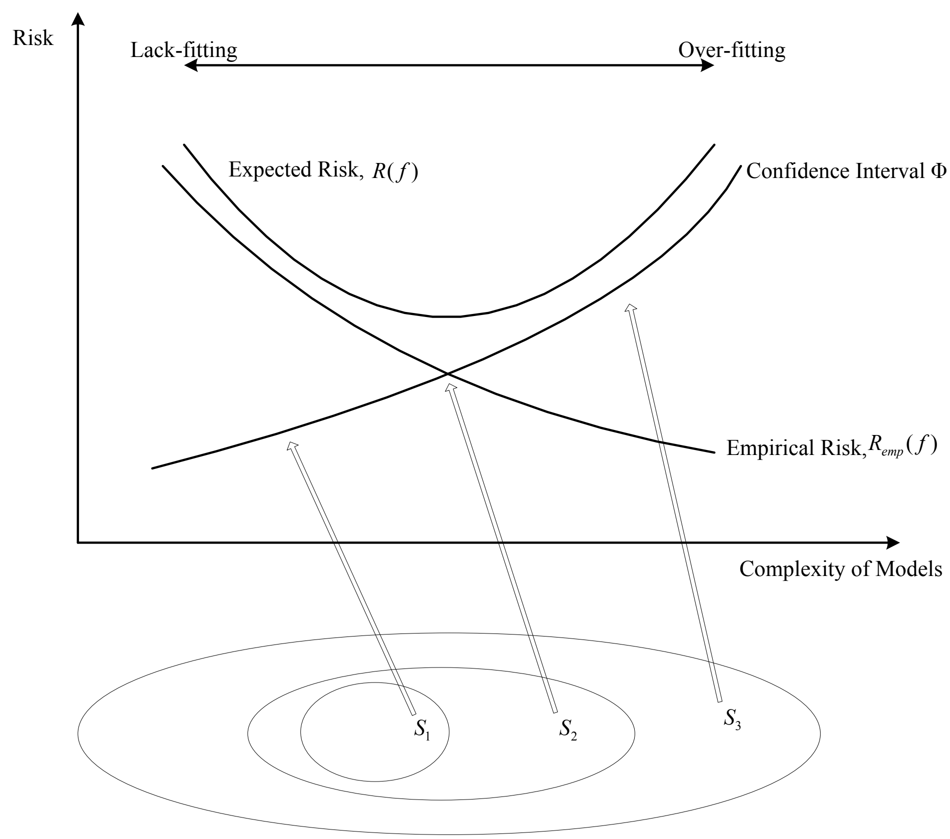


Figure 5.4 Concept of Structural Risk Minimization

5.3.3 Support Vector Regression

The support vector regression (SVR) is a nonlinear learning machine based on the principle of SRM for functional approximation. The learning machine is given n training data, from which it attempts to learn the input-output relationship (dependency or function) $f(\mathbf{x})$. A training data set $D = \{[\mathbf{x_i}, y_i] \in R^d \times R, \ \ i = 1,2,...,n\}$ consists of n pairs $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)$, where inputs $\mathbf{x} \in R^n$ are d-dimensional vectors, and the system responses $y \in R$ are continuous values. The SVR considers the approximating functions of the general form:

$$f(\mathbf{x}, \mathbf{w}) = \sum_{j=1}^{K} w_j \varphi_j(\mathbf{x}) \tag{5.3.8}$$

To introduce all relevant concepts of SVR in a gradual way, the linear form is considered first.

$$f(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T \mathbf{x} + \mathbf{b} \tag{5.3.9}$$

where $\mathbf{b}$ is the bias vector $\mathbf{b} \in R$.

Rather than the square error (L_2 norm) or the absolution error (L_1 norm) (Equations 5.3.2 and 5.3.3), SVR adopts $\varepsilon - insensitivity$ function introduced as the loss function. This SVR form is also called as ε -SVR. The $\varepsilon - insensitivity$ function is given as

$$L^{\varepsilon}(y, f(\mathbf{x}, \mathbf{w})) = |y - f(\mathbf{x}, \mathbf{w})|_{\varepsilon} = \begin{cases} 0 & |y - f(\mathbf{w}, \mathbf{x})| \leq \varepsilon \\ |y - f(\mathbf{x}, \mathbf{w})| - \varepsilon & \text{otherwise} \end{cases} \tag{5.3.10}$$

The loss function is equal to zero if the difference between the predicted $f(\mathbf{x}, \mathbf{w})$ and the observation is less than ε . In order to perform SVR, a new empirical risk is introduced:

$$R_{emp}^{\varepsilon}(f) = \frac{1}{n}\sum_{i=1}^{n}\left|y - f(\mathbf{x},\mathbf{w})\right|_{\varepsilon} \qquad (5.3.11)$$

In formulating an SV algorithm for regression, the objective is to minimize the expected risk

$$R(f) = \frac{1}{2}\|\mathbf{w}\|^2 + C\left(\sum_{i=1}^{n}\left|y_i - f(\mathbf{x}_i,\mathbf{w})\right|_{\varepsilon}\right) \qquad (5.3.12)$$

where $\|\mathbf{w}\|$ is the parameter vector norm which reflects the complexity of the model. From Equation 5.3.10 and Figure 5.5, it follows that for all training data outside an ε tube,

$\left|y - f(\mathbf{x},\mathbf{w})\right| - \varepsilon = \xi$ for data "above" an ε tube,

$\left|y - f(\mathbf{x},\mathbf{w})\right| - \varepsilon = \xi^*$ for data "below" an ε tube.

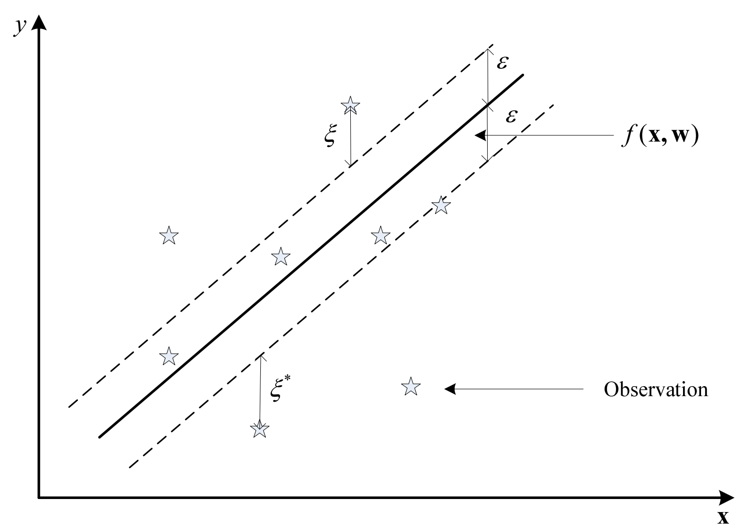


Figure 5.5 ε-tube Parameters used in SVR (linear kernel)

Thus, minimizing the risk in 5.3.12 is equivalent to minimizing the risk (Vapnik 1995, 1998)

$$R(\mathbf{w},\xi,\xi^*) = \frac{1}{2}\|\mathbf{w}\|^2 + C\left(\sum_{i=1}^{n}\xi_i + \sum_{i=1}^{n}\xi_i^*\right) \qquad (5.3.13)$$

subject to

$$\begin{cases} y_i - \mathbf{w}^T\mathbf{x}_i - \mathbf{b} \le \varepsilon + \xi_i \\ \mathbf{w}^T\mathbf{x}_i + \mathbf{b} - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, \quad i = 1,2,..n \end{cases} \tag{5.3.14}$$

where ξ_i and ξ_i^* are slack variables.

For minimizing the risk $R(\mathbf{w}, \xi, \xi^*)$, a Lagrange function is constructed from the objective function in Equation 5.3.13 and the constraints in Equations 5.3.14, by introducing a dual set of variables. The primal variables Lagrange function is given:

$$\begin{aligned} L = \frac{1}{2}\|\mathbf{w}\|^2 &+ C\sum_{i=1}^{n}(\xi + \xi^*) - \sum_{i=1}^{n}(\eta_i\xi_i + \eta_i^*\xi_i^*) \\ &- \sum_{i=1}^{n}\alpha_i(\varepsilon + \xi_i - y_i + \mathbf{w}^T\mathbf{x}_i + b) \\ &- \sum_{i=1}^{n}\alpha_i^*(\varepsilon + \xi_i^* + y_i - \mathbf{w}^T\mathbf{x}_i - b) \end{aligned} \tag{5.3.15}$$

where $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$ are Lagrange multipliers and equal to or greater than zero. The partial derivatives of L with respect to the primal variables have to vanish for optimality.

$$\begin{cases} \dfrac{\partial L}{\partial b} = \sum_{i=1}^{n}(\alpha_i^* - \alpha_i) = 0 \\ \dfrac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)x_i = 0 \\ \dfrac{\partial L}{\partial \xi_i^{(*)}} = C - \alpha_i^{(*)} - \eta_i^{(*)} = 0 \end{cases} \tag{5.3.16}$$

where $\xi_i^{(*)}, \alpha^{(*)}, \eta^{(*)}$ refer to ξ_i and ξ_i^*, α_i and α_i^*, η_i and η_i^* respectively. Substituting Equation 5.3.16 into Equation 5.3.15 yields the dual optimization problem,

$$\text{maximize} \quad \begin{cases} -\dfrac{1}{2}\sum_{i,j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle \mathbf{x}_i, \mathbf{x}_j \rangle \\ -\varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n}y_i(\alpha_i - \alpha_i^*) \end{cases} \tag{5.3.17}$$

subject to $\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0$ and $\alpha_i, \alpha_i^* \in [0, C]$

From equation 5.3.16, we have

$$\mathbf{w} = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\mathbf{x}_i$$

Thus

$$f(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\langle \mathbf{x}_i, \mathbf{x} \rangle + b \tag{5.3.18}$$

w can be completely described as a linear combination of the training patterns x_i. The x_i corresponding to $(\alpha_i - \alpha_i^*) \neq 0$ are called as support vectors (SVs). In a sense, the complexity of a function's representation by SVs depends only on the number of SVs. For the general form of the approximating function, kernel function is introduced to substitute the dot product $\langle \mathbf{x}_i, \mathbf{x} \rangle$. Thus,

$$f(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(\mathbf{x}_i, \mathbf{x}) + b \tag{5.3.19}$$

Except for the linear kernel function, Gaussian radial basis function (RBF) is also a very important kernel function

$$K(\mathbf{x}_i, \mathbf{x}) = \exp(-\gamma \left\| x - x_j \right\|^2) \tag{5.3.20}$$

5.3.4 Procedure to Apply SVR

LIBSVM is an integrated software developed by Chang and Lin (the National Taiwan University) for support vector classification, regression (ε-SVR) and distribution estimation. In this study, the software is used to estimate the approximate function and evaluate the effectiveness of the SVR. The proposed procedure to apply SVR using LIBSVM is given as follows (Hsu, Chang and Lin, 2008)

(1) Transform data to the format of the LIBSVM software

(2) Conduct simple scaling on the data

(3) Consider the RBF kernel

(4) Determine the parameter ε, C and γ

(5) Perform the data training

(6) Test the model with the test data

LIBSVM requires that each data instance is represented as a vector of real numbers. The whole dataset is split into two parts: training dataset and testing dataset. The former is used for model training while the latter is used for model validation.

Scaling data before applying SVR is very important (Sarle 1997, Part 2 of Neural Networks FAQ). The main advantage of scaling is to avoid attributes in greater numeric ranges dominate those in smaller numeric ranges. Another advantage is to avoid numerical difficulties during the calculation. Because kernel values usually depend on the inner products of feature vectors, large attribute values might cause numerical problems. Two most common linearly scaling methods are given as:

$$
\begin{aligned}
x_{scaled}^{[-1,1]} &= 2 \times \frac{x - x_{\min}}{\left| x_{\max} - x_{\min} \right|} - 1 \\
x_{scaled}^{[0,1]} &= \frac{x - x_{\min}}{\left| x_{\max} - x_{\min} \right|}
\end{aligned}
\tag{5.3.21}
$$

Apparently, we have to use the same rule as training data to scale testing data before model validation.

The RBF function is a reasonable first choice for the kernel function. The RBF function nonlinearly maps samples into a higher dimensional space can handle the case when the relation between class labels and attributes is nonlinear. Furthermore, the RBF kernel has less numerical difficulties.

There are three parameters while using RBF kernels: ε, C and γ. The ε-insensitivity decides the range of the admissible error for model training. A small ε value will lead to a small empirical risk, but may result in over-fitting and increase the training time. In contrast, a big ε value could reduce the training cost, but may bring a low accuracy in data training and predicting. In general, the ε value is selected from 0.01 to 0.1.

The penalty factor C affects the training accuracy and the predicting ability. As the increase of C, the approximating error decreases and the training cost increases. When the value of C reaches a certain big value, the approximating error may stop decreasing, even start increasing due to over-fitting. The kernel parameter γ is also a factor which has influence on the approximating error. A larger γ will result in a complex model, thus may lead to over-fitting problem, while a small γ will reduce the flatness of approximating function curve, and the approximating error. The C and γ are correlated.

The best pair of C and γ are needed to decide before model training. In this study, the pairs of exponentially growing sequences of C and γ are tried to indentify a good pair.

The model training is processed using LIBSVM with training dataset and the selected model parameters. A validation procedure is used with the testing data to evaluate the accuracy of the trained model. In LIBSVM, the Mean Square Error (*MSE*) and Squared Correlation Coefficient (R^2) is adopted as the accuracy criteria.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - f(\mathbf{x}_i)\right)^2 \tag{5.3.22}$$

$$\begin{cases} R^2 = \dfrac{\text{SST - SSE}}{\text{SST}} = \dfrac{\text{SSR}}{\text{SST}} = 1 - \dfrac{\text{SSE}}{\text{SST}} \\ \text{SSE} = \sum_{\text{i=1}}^{\text{n}}\left(y_i - f(\mathbf{x}_i)\right)^2 \\ \text{SSR} = \sum_{\text{i=1}}^{\text{n}}\left(f(\mathbf{x}_i) - \bar{y}\right)^2 \\ \text{SST} = \sum_{\text{i=1}}^{\text{n}}\left(y_i - \bar{y}\right)^2 \end{cases} \tag{5.3.23}$$

In statistics, the mean squared error or *MSE* of an estimator is one of many ways to quantify the amount by which an estimator differs from the true value of the quantity being estimated. An MSE of zero means that the predictions are perfect to approximate the observations.

R^2 is a statistic that will give some information about the goodness of fit of a model. In regression, the R^2 coefficient of determination is a statistical measure of how well the regression line approximates the real data points. An R^2 of 1.0 indicates that the regression line perfectly fits the data.

Chapter Six

Experiment Results of Speed Profile Models

6.1 Data Preparation

The data for analysis were collected from the output files of the simulation experiment and translated into SPSS software for data reduction. In total, 31,031 observations for 1001 scenarios were selected as the dataset for model development. The variables included in the dataset are given in Table 6.1. The values of the scenario factors are shown in Table 5.2. Another input variable is the location of detectors. Because the start point of lane closure is changed as the variety of the closure length, the original values were translated into relative values which are measured as the distance of detectors to the start point.

Table 6.1 Definition of Variables for Speed Profile Model

Level	Variable	Definition	Type
Response	SPEED	Mean speed at detector points	Continuous
Scenario	GRADE	Work zone grade	Continuous
	CLOSELENGTH	Length of closure zone	Continuous
	VOLUME	Upstream volume	Continuous
	HV	Heavy vehicle percentage	Continuous
	FFS	Free flow speed	Continuous
Space	LOCATION	The distance to the start point of closure lane	Continuous

The comparison table of LOCATION to CLOSELENGTH is shown in Table 6.2. A positive value means the corresponding detector is located before the start point of lane closure, while a negative value means the detector is located after the point.

Table 6.2 Comparison Table of LOCATION to CLOSELENGTH

	LOCATION		
LOSELENGTH	Min	Max	Number
2000feet	499feet	25499feet	31
2500feet	-1feet	24999feet	31
3000feet	-501feet	24499feet	31
3500feet	-1001feet	23999feet	31
4000feet	-2001feet	22999feet	31
7500feet	-5001feet	22999feet	31

The dataset was split randomly into two parts: the training dataset was used for model training, and the testing dataset was used for model validation. The statistical descriptions of the two datasets are given in Tables 6.3 and 6.4 respectively. These input variable were both rescaled into [-1, 1] and the output variable was rescaled to [0, 1]

Table 6.3 Descriptive Statistics for Training Dataset

	Number	Minimum	Maximum	Mean	Std. Deviation
Grade	22010	-2.00	2.00	0.01	1.62
CloseLength	22010	0.00	5500.00	1859.86	1667.16
Volume	22010	800.00	4000.00	2453.52	1047.65
HV	22010	0.00	15.00	5.70	5.90
FFS	22010	55.00	70.00	62.46	6.12
Location	22010	-5001.00	25499.00	10801.33	7580.99
Speed	22010	1.13	70.05	55.26	13.32

Table 6.4 Descriptive Statistics for Testing Dataset

	Number	Minimum	Maximum	Mean	Std. Deviation
Grade	9021	-2.00	2.00	-0.08	1.65
CloseLength	9021	0.00	5500.00	1850.52	1725.43
Volume	9021	800.00	4000.00	2461.86	1018.39
HV	9021	0.00	15.00	5.45	5.71
FFS	9021	55.00	70.00	62.42	6.10
Location	9021	-5001.00	25499.00	10810.68	7594.26
Speed	9021	5.21	70.05	55.50	13.21

6.2 Analysis on Speed Profiles

The speed profile (pattern) at work zones is more complex than that at common freeway section. Figure 6.1 shows the speed profile with different entry volumes. It can be known that when the entry volume was less than the capacity (1600 suggested in HCM 2000), the speed profile is similar and approximated the free flow speed. When the entry volume is obviously greater than the capacity, the speed profile becomes different much to that at common freeway section.

Inspecting the figure, the speed profiles are described as follows:

(1) When the traffic flow is under uncongested traffic conditions, the speed along the freeway is controlled by the FFS (speed limit). The difference between the FFS and measured speed is little. But when vehicles are entering in work zone, due to the FFS reduction at the closure zone, the speed profile has a descent within lane closure area.

(2) Being far from the start point of lane closure, the traffic flow on the open lane is not disturbed by the lane changes from the closed lane. The speed profile in this section is almost the same as that in normal traffic flow in freeway.

(3) When the traffic flow is closing to the start point, it is disturbed by the vehicle lane shifting. The speed of the vehicles is descending up to a small value. The start point of speed descent is going far from the work zone as the increase of the volume.

(4) When the speed reaches a small and steady value, the backward queue is formed. The larger the entry volume is, the longer the queue length becomes.

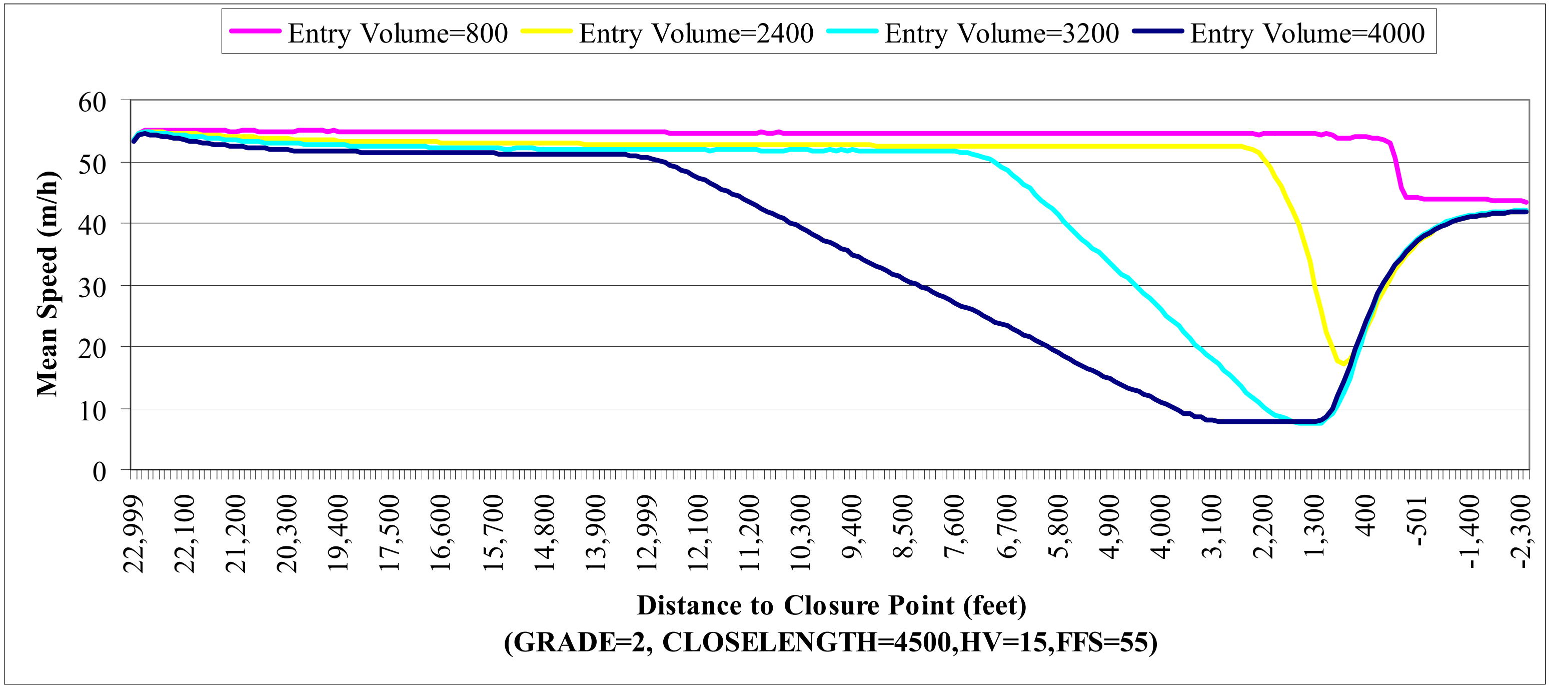


Figure 6.1 Speed Profile under Different Congestion Condition

(5) After the start point of lane closure, vehicles start to accelerate up to the work zone FFS.

6.3 Results of Modeling Training

The model training was a time-consuming process, and the prediction accuracy depended on the parameters selection. For reducing the training time, the parameters ε-insensitivity factor, penalty factor C and kernel function parameter γ were selected from a limited set which was the combination of ε=0.1 or 0.01, C=1, 1000, 2000, or 3000, and γ=0.16667, 1, 2, or 3. All these combinations were used for model training and testing, and the best parameter combination was determined by the minimum *MSE* value. The final values of the parameters are $\varepsilon = 0.01$, C=3000, and γ=2. As a comparison, the model trained with the default parameters was also provided. The training results are given in Table 6.5.

Table 6.5 Results of Model Training and Validation

		Final Model	Default Model	Definition
Parameter	Kernel Function	RBF	RBF	
	ε	0.01	0.1	ε-insensitivity factor
	C	3000	1	penalty factor
	γ	2	0.166667	kernel function parameter
Training Results	total_SV	9481	3368	number of support vectors
	rho	-0.717976	-0.649306	bias term
Validation Results	*MSE(*rescaled)	0.00396214	0.0104952	mean squared error (rescaled)
	MSE (mile/h)^2	18.82	49.85	mean squared error
	R^2	0.892287	0.731187	squared correlation coefficient

The final model has more support vectors than the default model since its ε-tube is more narrow than that of the default model. The *MSE* of the final model is a small than that of default value (MSE is rescaled value). And its R^2 is close to 1. That indicates the predicted values are good at fitting the observations; in other words, the final model has good prediction ability.

The selected comparisons of the predicted speed profile to the observed speed profile for various scenarios are given in Figure Tables 6.2 through 6.10. These scenarios were selected from the testing dataset. From these figures, it can be concluded that the final speed profile model can fit the observations perfectly for all sections under different scenarios. But the model with the default parameters has acceptable approximating characteristics only for the sections where the traffic flow is normal.

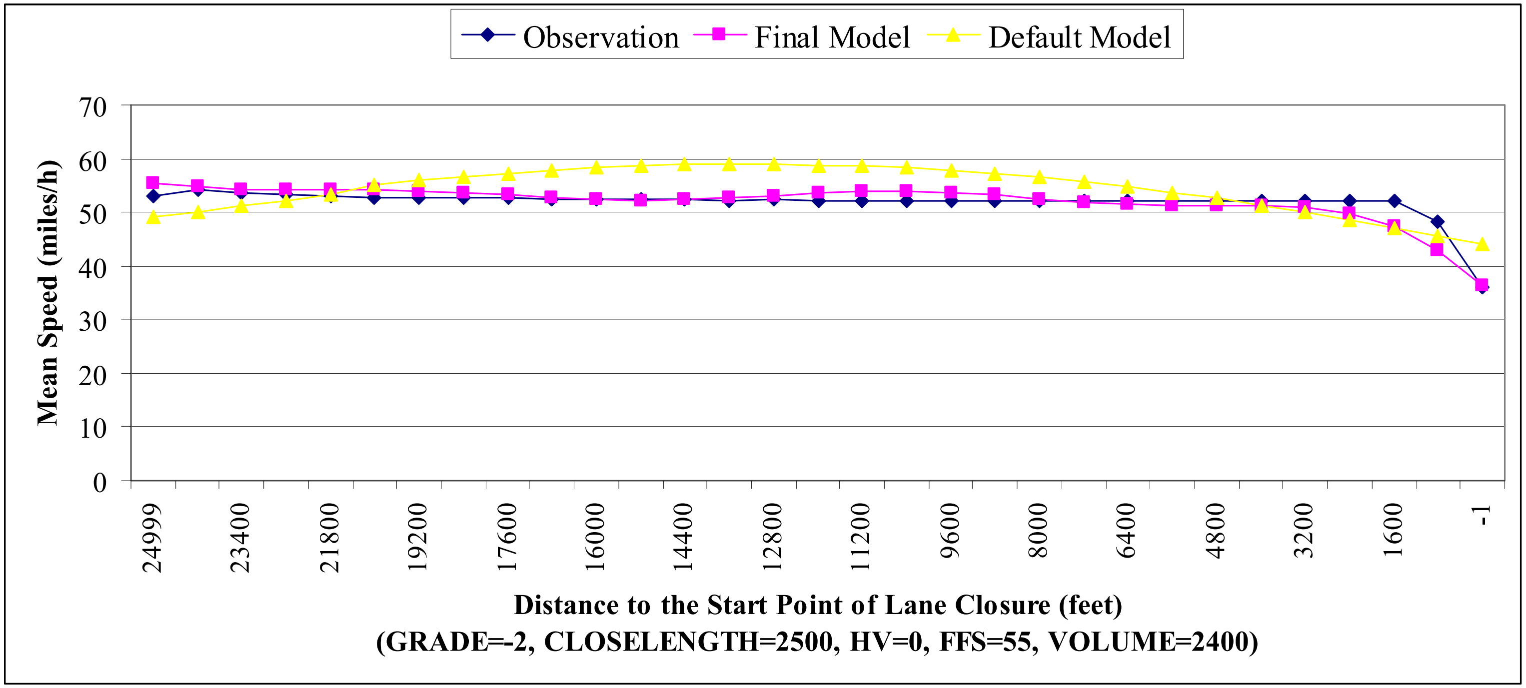


Figure 6.2 Comparison of Speed Profile Models to Observations (low FFS and low VOLUME)

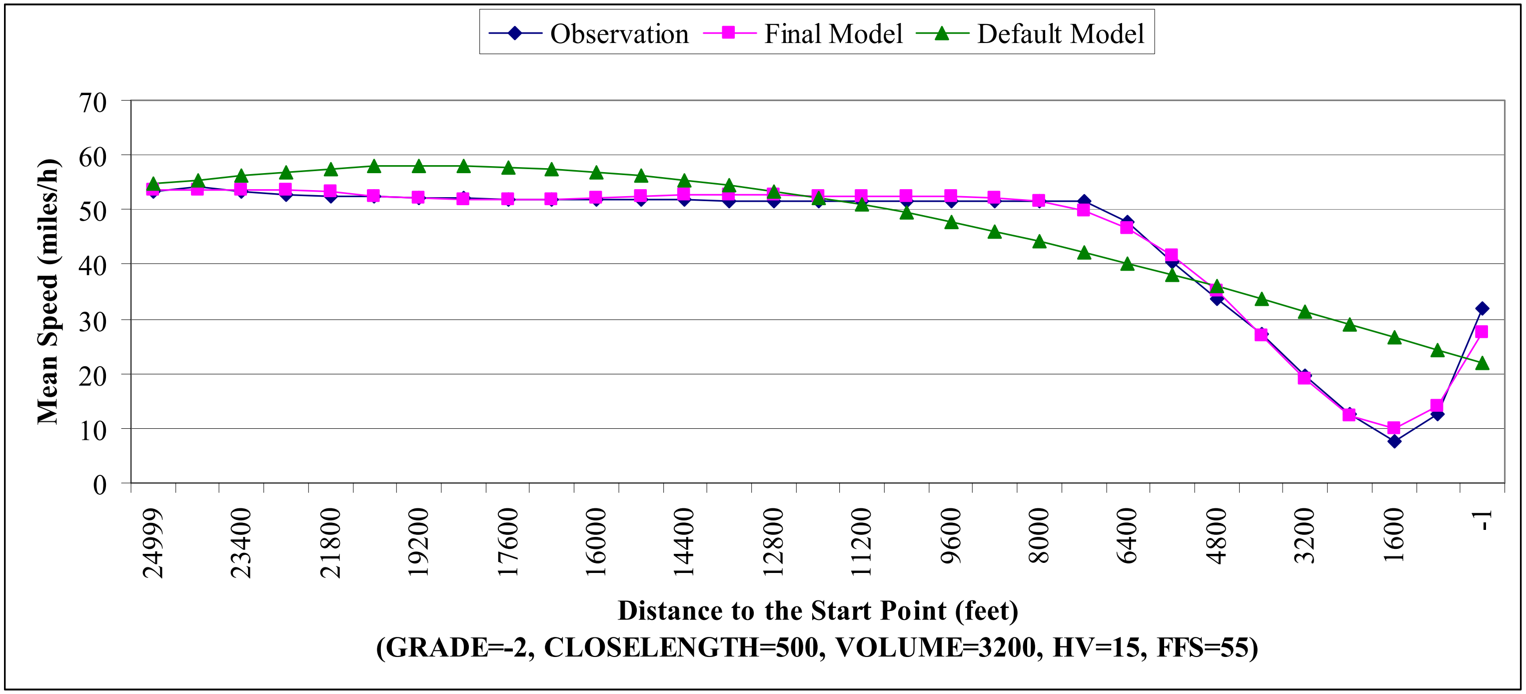


Figure 6.3 Comparison of Speed Profile Models to Observations (low FFS and medium VOLUME)

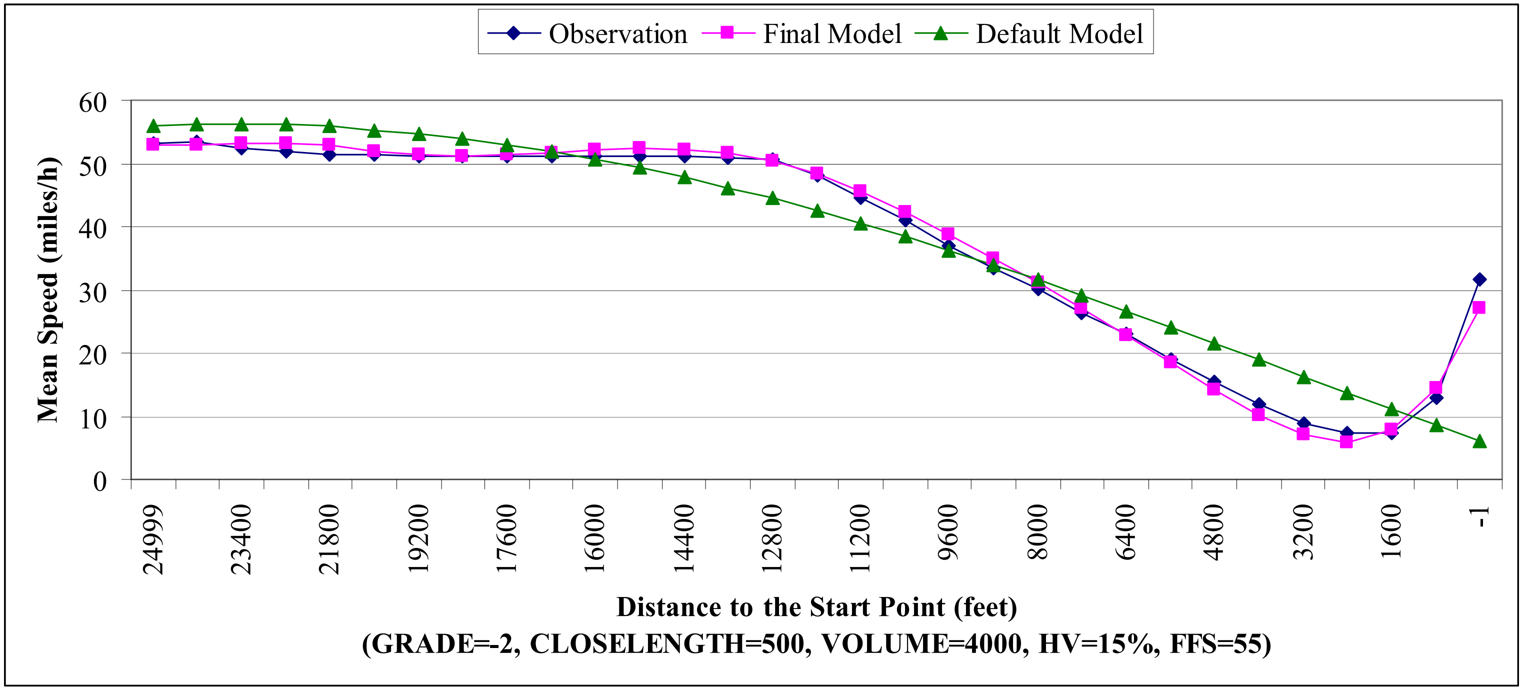


Figure 6.4 Comparison of Speed Profile Models to Observations (low FFS and high VOLUME)

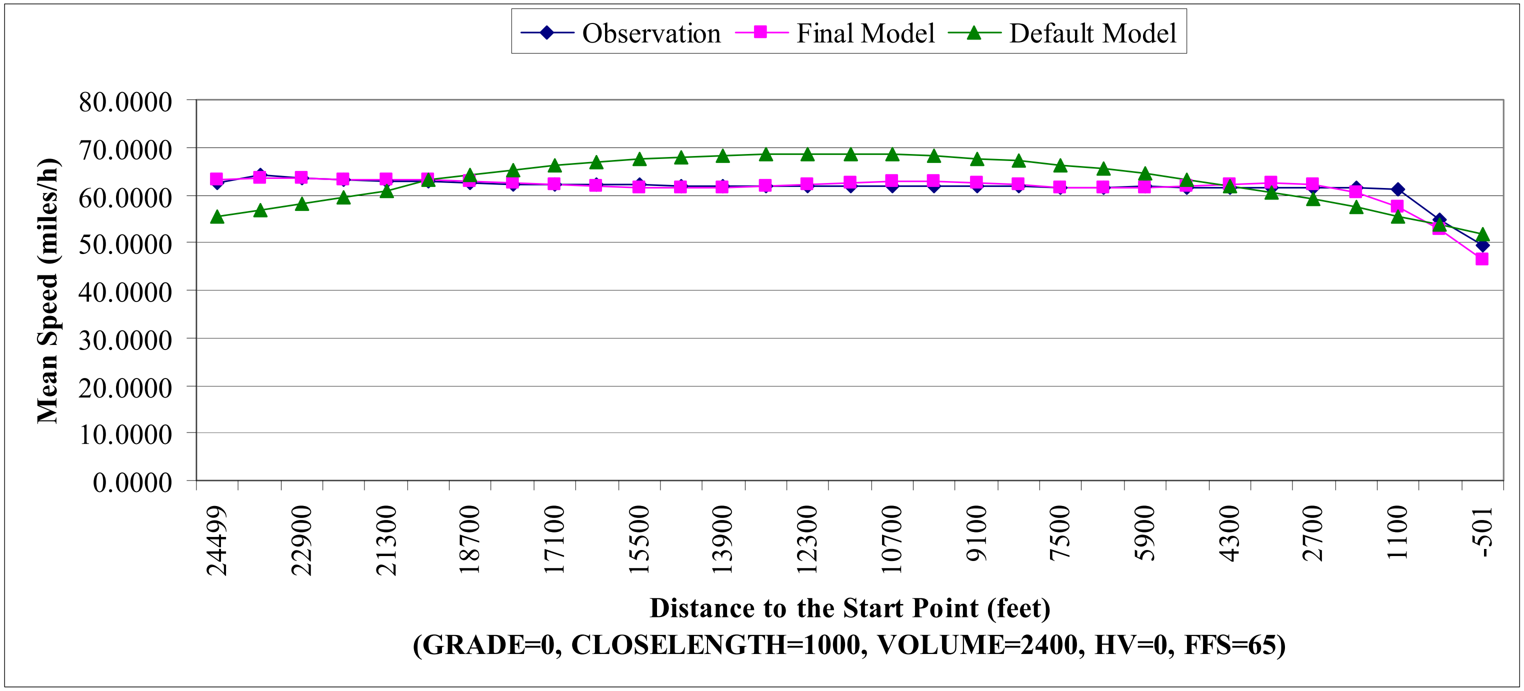


Figure 6.5 Comparison of Speed Profile Models to Observations (medium FFS and low VOLUME)

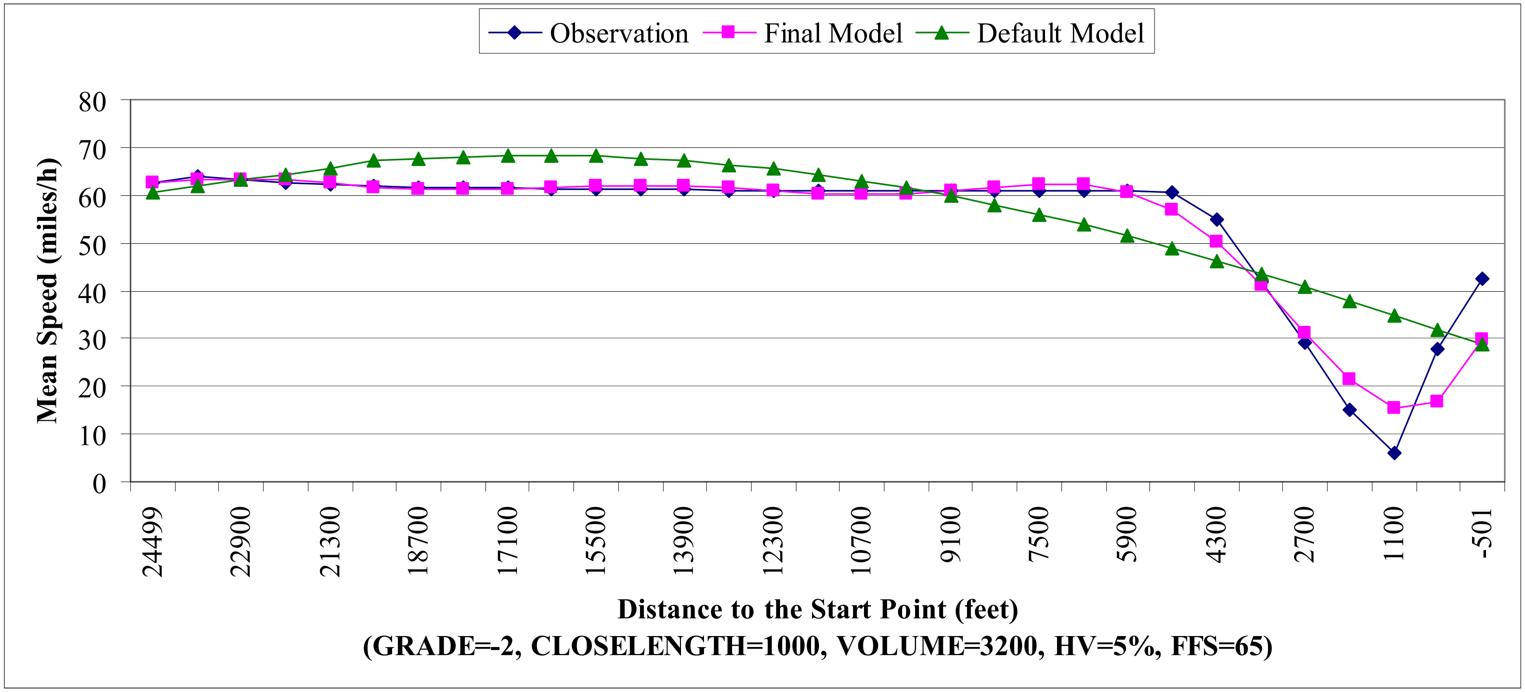


Figure 6.6 Comparison of Speed Profile Models to Observations (medium FFS and medium VOLUME)

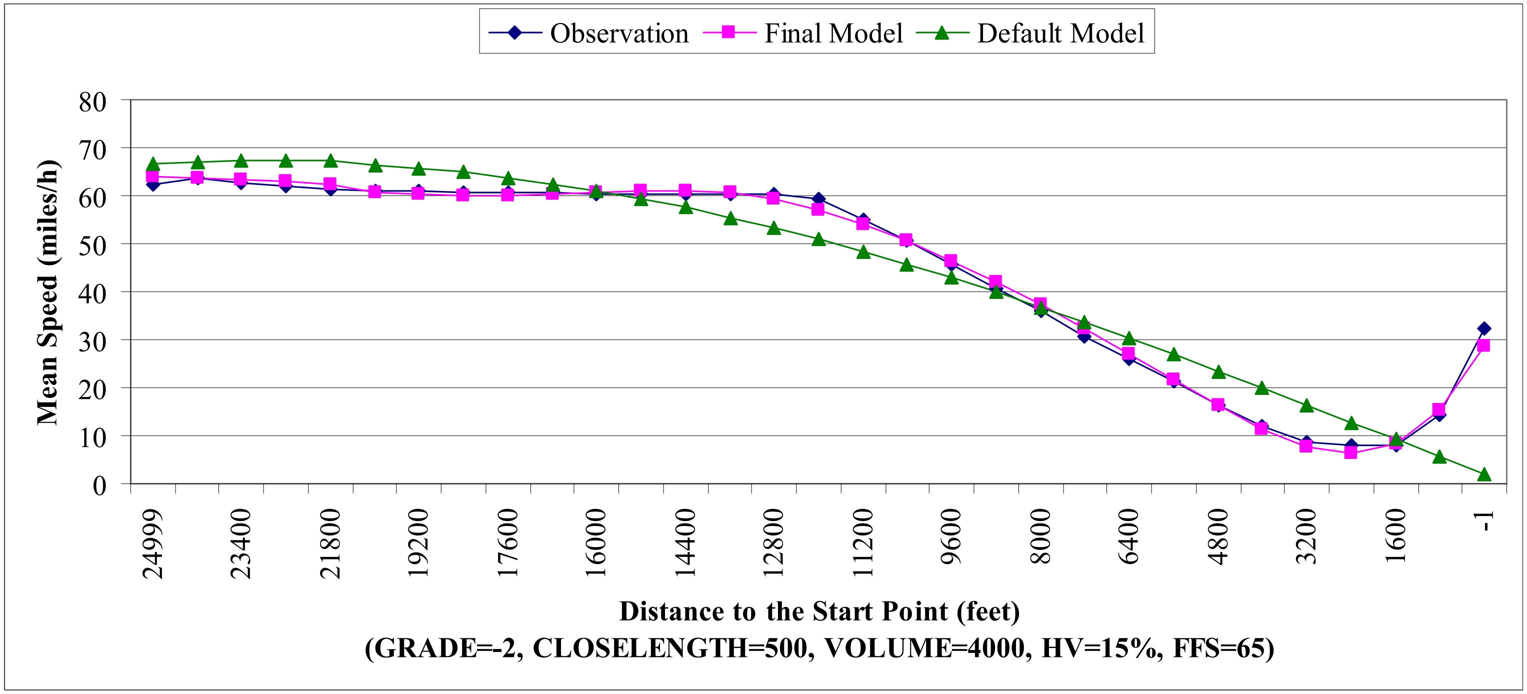


Figure 6.7 Comparison of Speed Profile Models to Observations (medium FFS and high VOLUME)

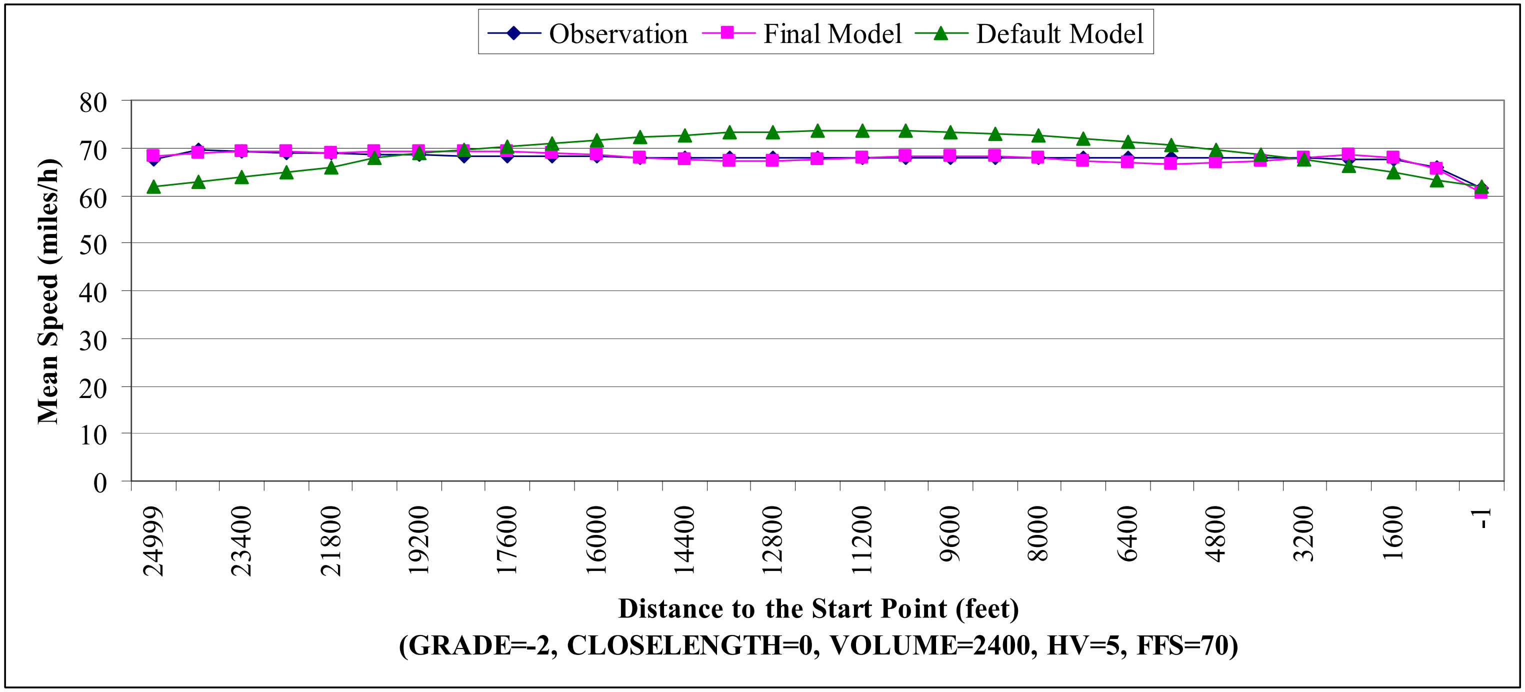


Figure 6.8 Comparison of Speed Profile Models to Observations (high FFS and low VOLUME)

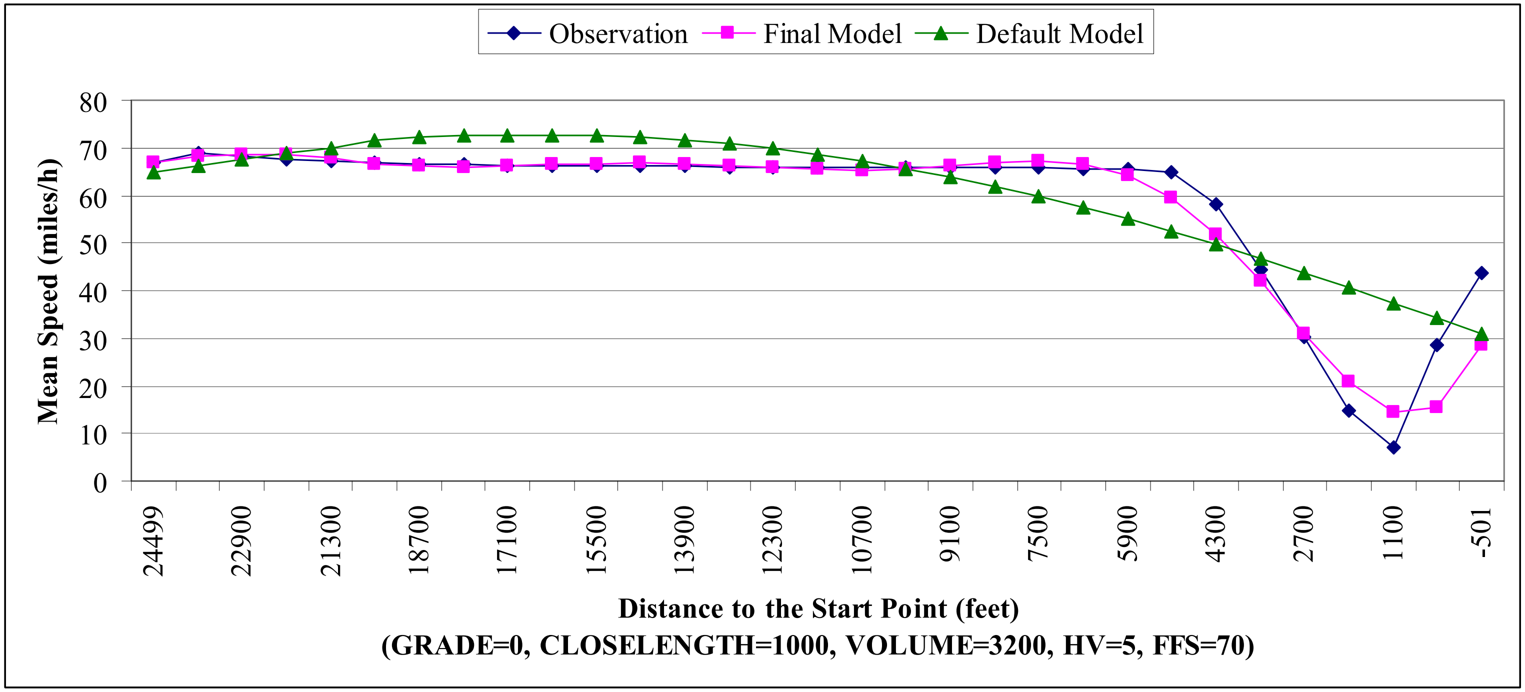


Figure 6.9 Comparison of Speed Profile Models to Observations (high FFS and medium VOLUME)

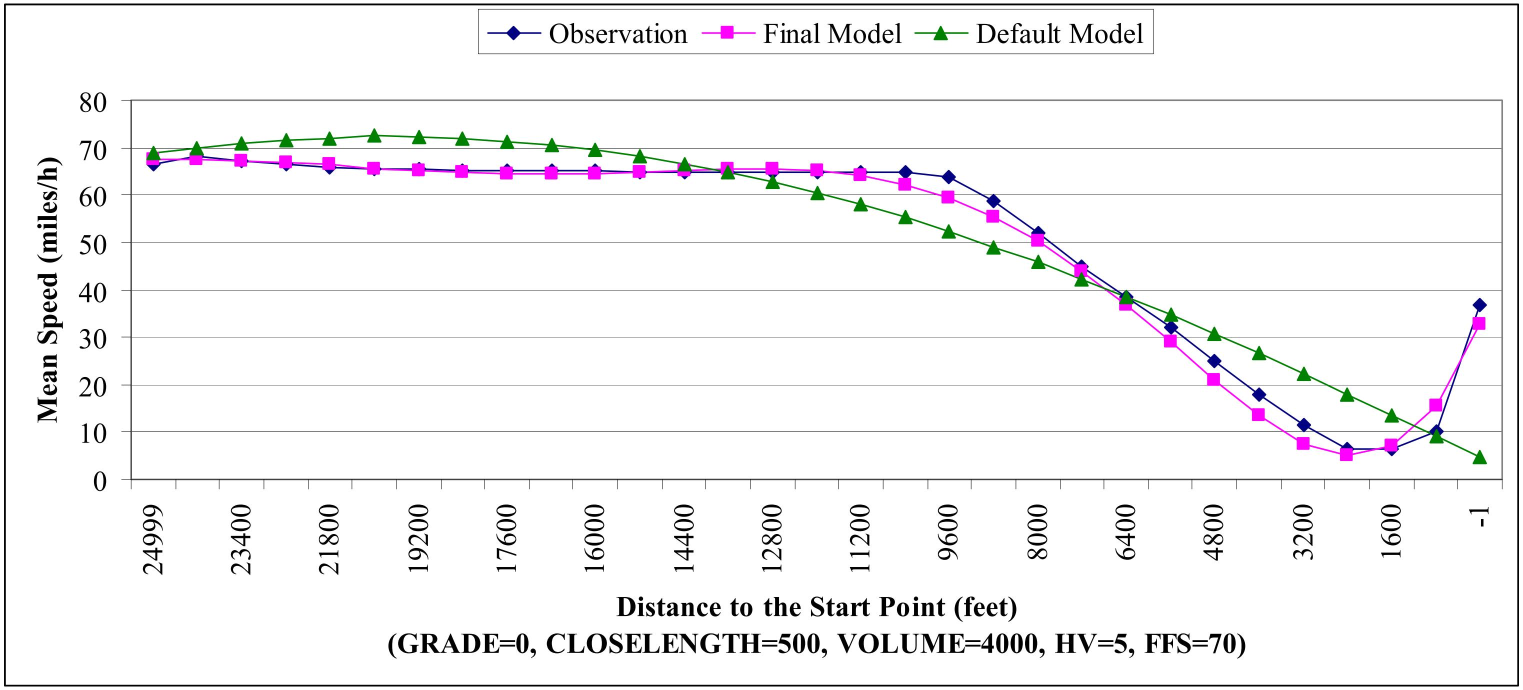


Figure 6.10 Comparison of Speed Profile Models to Observations (high FFS and high VOLUME)

Chapter Seven

Conclusions and Discussions

7.1 Conclusions

This dissertation focused on modeling crash severity and speed profile at work zones. Crash severity is an important criterion reflecting the cost of work zone crashes in social and economy, and affected by various factors including driver's characteristics, vehicle characteristics, environmental factors, and roadway features. To understand the influence of these factors on the crash severity can be used to select proper countermeasure to reduce the crash severity at work zones and decrease the loss of construction/maintenance on roadway. A new modeling regression for ordinal output, partial proportional odds logit regression, was used to estimate the crash severity models for two crash datasets: overall work zone crashes and rear-end work zone crashes. Based on the results of crash severity modeling and analysis, some conclusions can be obtained:

(1) The parallel regression assumption is always violated when the ordinal logit regression is utilized to estimate the work zone crash severity model. The partial proportional odds logit regression which has less restrict to the assumption can given more accurate and more detailed explanation on the impacts of factors on the work zone crash severity.

(2) For over all work zone crashes, the presence daylight, the location at urban area or at freeway is more likely to reduce the severity of work zone crashes, while grade or curve of roadway section, vision obstruction, high speed limit, alcohol involvement, and young or old drivers tends to increase the severity of work zone crashes. The involvement of heavy vehicle is likely increase the probability of fatal crashes at work zones, but tends to reduce the severity of injury only work zone crashes.

(3) For rear-end work zone crashes, the factors that tend to reduce the work zone crash severity include traffic controls, the influence of special roadway section, and the freeway section. The factors that have reversed impacts include old drivers, grade or curve at roadway section, and high speed limit. Heavy vehicle involvement is more likely increase the probability of fatality or incapacitating injury, while reduce the probability of injury crashes rather than that of no injury crashes.

Work zone speed profile (pattern) is the mean value of the distribution of vehicle speed over the distance to the start point of lane closure at work zones. Predicting work zone operating speeds under various scenarios is a useful precursor to appropriate regulatory and design decisions for work zones. The speed profile model for the open lane on two-lane with one-lane closed (one direction) freeway was developed with a new learning machine algorithm, Support Vector Regression. Based on the results of analysis and model development, the conclusion can be summarized as:

(1) The speed profile is a typical non-linear complicate system which is difficult to be described by a linear regression. The SVR has great capability to

provide a uniform model for expressing the complicate relationship between speed profile and various traffic factors.

(2) Based on the validation results, predictions of the speed profile model with selected parameters approximate observations perfectly under various scenarios. That means the SVR model has good generalization ability for work zone speed profile; in other words, the SVR model can predict accurately the work zone speed rather than the training data.

7.2 Contributions to the Field

7.2.1 Methodological Contribution

On the crash severity analysis aspect, this dissertation is dedicated into utilizing the partial proportional odds regression, a new logit regression method for ordinal outputs, to address the relationship between crash severity and various factors. This regression can avoid the parallel regression assumption, and provide more detailed interpretations of the factor coefficients. The ability for explaining the factor impacts on crash severity is beneficial to understand the characteristics of traffic crashes. The partial proportional odds regression can be used to analyze other crash data rather than work zone crashes.

On the speed profile modeling aspect, this dissertation is dedicated into utilizing the support vector regression algorithm to estimate the speed profile model. Except for the capability to describe the complication non-linear system, SVR has excellent prediction ability. Due to the features of SVR, SVR is applicable to modeling traffic systems which are always non-linear complicate systems. The conclusions of this dissertation can be used as the guidance for the application of SVR in transportation research field.

7.2.2 Practical Contribution

The interpretations of work zone crash severity model can be used to understand the impacts of factors on work zone safety. This understanding is benefit of addressing safety problem at work zones and selecting proper countermeasures to reduce crash severity and improve work zone safety.

Because the simulation experiment is based on a calibrated CORSIM model, the speed profile model can be utilized as a reference for highway design and operational analysis at work zones. It also can be used to help traffic engineers to understand the speed profile at work zone area, and to implement proper traffic control systems to improve the work zone safety and operational performance.

7.3 Future Research Direction

It is far from the end to come up with a full understanding of the factor impacts on work zone crash severity. Due to the limitation of crash data collection, some useful variables were missed, such as gender of drivers, work zone types, traffic control countermeasures at work zones, and so on. In feature, crash work zone severity models which integrate the missed variables can provide more accurate and powerful interpretations of the factor impacts on the work zone severity. In this dissertation, only rear-end work zone crashes were analyzed. For provide more specific understanding of the characteristics of work zone crashes, other types of work zone crashes should be modeled and explained. In addition, the interpretation of some factors is different from common sense. For example, the involvement of heavy vehicle tends to reduce the severity for work zone injury only crashes. A deeper research should be conducted for giving a more accurate explanation on this phenomenon.

The effectiveness of SVR training is based on the parameter selection. In this dissertation, the selection performed by a simple method within a small range. Although the final result is good, the model developed in this dissertation is not guaranteed to be the best estimation. In feature study, an optimal parameter selection process for SVR should be developed. This process should adopt a search algorithm for global optimization (like genetic algorithm) to find the best combination of parameters for SVR training.

List of References

Al-Kaisy, A., Zhou, M., and Hall, F. New insights into freeway capacity at work zones: empirical case study. Transportation Research Record 1710, Transportation Research Board, Washington, D.C., pp. 154–160., 2001.

Bai, Y. and Li, Y. Determining Major Causes of Highway Work Zone Accidents in Kansas. Publication K-TRAN: KU-05-1. Kansas Department of Transportation, 2004.

Benekohal, R. F., Elzohairy, Y. M. and Saak, J. E., A Comparison of Delay from HCS, Synchro, PASSER II, PASSER IV and CORSIM for an Urban Arterial, Transportation Research Board, National Research Council, Washington, D.C., 2002.

Benekohal, R. F., Kaja-Mohideen, A., and Chitturi, M., A Methodology for Estimating Operating Speed and Capacity in Work Zones, Transportation Research Board, National Research Council, Washington, D.C., 2004.

Benetohan, R., R. Orloski, and A. Hashni. Drivers Opinion on Work Zone Traffic
Blincoe, L. J. The Economic Costs of Motor Vehicle Crashes, 1994. National Highway Traffic Safety Administration, Washington, D.C., 1996.

Bloomberg, L.., Swenson, M., and Haldors, B., Comparison of Simulation Models and the HCM, Transportation Research Board, National Research Council, Washington, D.C., 2003.

Chambless, J., Chadiali, A. M., Lindly, J. K, and McFadden, J. Multistate Work zone Crash Characteristics. ITE Journal, Vol. 72, pp46 – 50., 2002.

Chien, Steven I-Jy and Chowdhury, M. S., Simulation-Based Estimates of Delays at Freeway Work Zones, Transportation Research Board, National Research Council, Washington, D.C., 2002.

Clogg, C. C. and Shihadeh, E. S. Statistical Models for Ordinal Variables. Thousand Oaks, CA., 1994.

Cohen, S. L. An Approach to Calibration and Validation of Traffic Simulation Models, Transportation Research Board, National Research Council, Washington, D.C., 2004.

Crowther, B. C., A Comparison of CORSIM and INTEGRATION for the Modeling of Stationary Bottlenecks. Master Thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 2001.

Dixon, K. K., Hummer, J. E., and Rouphail, N. M., Comparison of Rural Freeway Work Zone Queue Length Estimation Techniques: A Case Study, Transportation Research Board, Washington, D. C, 1998.

Fan R. E., Chen, P.-H. and Lin, C.-J. Working set selection using the second order information for training SVM. Journal of Machine Learning Research. No.6, 1889-1918, 2005.

FHWA. Manual on Uniform Traffic Control Devices for Streets and Highways. 2003 Edition, Federal Highway Administration. U.S. Department of Transportation, 2003.

Garber, N. J. and Zhao, M. Crash Characteristics at Work Zones. Publication VTRC 02-R12, Virginia Transportation Research Council, Charlottesville, Virginia, 2002.

Ha, T. and Nemeth, Z. A. Detailed Study of Accident Experience in Construction and Maintenance Zones. Transportation Research Record, No. 1590, TRB, National Research Council, Washington, D.C., pp38 – 45, 1995.

Hall, J. W. and Lorenz, V. M. Characteristics of Construction Zone Crashes. Transportation Research Record No. 1230, TRB, National Research Council, Washington, D.C., pp20 – 27., 1989.

Hill, R. W. Statistical Analysis of Fatal Traffic Accident Data. Master's Thesis, Texas Tech University. 2003.

Holdridge, J. M., Shankar, V. N. and Ulfarsson, G. F. The Crash Severity Impacts of Fixed Roadside Objects. Journal of Safety Research. No.2, 139-147., 2005

Hosmer, D. W. and Lemeshow, S. Applied Logistic Regression, Second Edition, John Wiley and Sons, New York, 1998.

Hsu, C., Chang, C. C., and Lin, C. J. A Practical Guide to Support Vector Classification. http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf., 2008

Jiang, X. J., and Adeli, H. Freeway Work Zone Traffic Delay and Cost Optimization Model. Journal of Transportation Engineering. Vol. 129, No. 3. 230-241., 2003.

Jiang, Y. Traffic Capacity, Speed and Queue-Discharge Rate of Indiana's Four-Lane Freeway Work Zones, Transportation Research Record No. 1657, Transportation Research Board, National Research Council, Washington D.C., 1999.

Kamyab, A., Maze, T.H., Gent, S. and Schrock, S. Work Zone Speed Control and Management by State Transportation Agencies and Toll Authorities, Transportation Research Board, National Research Council, Washington, D.C., 2001.

Kim, K. N., Richardson, J., and Li, L. Personal and Behavioral Predictors of Automobile Crash and Injury Severity. Accident Analysis and Prevention, Vol. 27, No. 4, 469-481., 1995

Krammes, R.A., and G. O. Lopez. Updated Capacity Values For Short-Term Freeway Work Zone Lane Closures. Transportation Research Record No.1442, Transportation Research Board, National Research Council, Washington, D.C., 1994.

Martineli, D. R., and Xu, D. Delay estimation and optimal length for four-lane divided freeway work zones. J. Transp. Engrg., ASCE, 122(2), 114–122., 1996.

Maze, T. and Kamyab, A. Work Zone Simulation Model, Companion Report for Traffic Management Strategies for Merge Areas in Rural Interstate Work Zones. Center for Transportation Research and Education, Iowa State University, September, 1999.

Maze, T. Kamyab, A. and Schrock, S. Evaluation of Work Zone Speed Reduction Measures. Final Contract Report. Iowa Department of Transportation, IA, 2000.

McFadden, J., Yang, W. T., and Durrans, S. R. Application of Artificial Neural Networks to Predict Speeds on Two-Lane Rural Highways. Transportation Research Record. No. 1751, Transportation Research Board, National Research Council, Washington D.C., 2005.

McCoy, P. T., and Pesti, G. Dynamic Late Merge Control Concept for Work Zones on Rural Freeways. Transportation Research Board, National Research Council, Washington, D.C., 2001.

McCoy, P. T., Pang, M., and Post, E. R. Optimum length of two-lane, two-way, no-passing traffic operation in construction and maintenance zones on rural four-lane divided highways. Transp. Res.Rec. 773, Transportation Research Board, Washington, D.C., 20–24., 1980.

Memmott, J. L., and Dudek, C. L. Queue and cost evaluation of work zones (QUEWZ). Transportation Research Record. No.979, Transportation Research Board, Washington, D.C., 12–19., 1984.

Mercier, C. R., Shelley, M. C., Rimkus, J. B., and Mercier, J. M., Age and Gender as Predictors of Injury Severity in Head-on Highway Vehicular Crashes. Transportation Research Record No. 1581, Transportation Research Board, National Research Council, Washington D.C., 1997.

Migletz, J. and Graham, J. L. Work Zone Speed Limit Procedure. 78th Annual Meeting of the Transportation Research Board CD-ROM, Transportation Research Board, National Research Council, Washington, D.C., 1999.

Migletz, J., and J.L. Graham. Work Zone Speed Control Procedures: Proceedings of the Symposium on Work Zone Traffic Control. Report FHWA-TS-91-003. Turner-Fairbank Highway Research Center, Mclean, VA, 1991.

Mohan, S. B. and Gautam, P. Cost of Highway Work Zone Injuries. Practical Periodical on Structural Design and Construction. Vol. 7, No.2, 2002, pp68 – 73., 2002.

Mousa, R.M., Rouphail, N.M. and Azadivar, F. Integrating Microscopic Simulation and Optimization: Application to Freeway Work Zone Traffic Control. Transportation Research Record 1254, 1990.

Park, B. B. and Qi, H. Development and Evaluation of a Calibration and Validation Procedure for Microscopic Simulation Models. Final Contract Report, Virginia Transportation Research Council, 2004.

Peterson, B. and Harrell, F. E. Partial Proportional Odds Models for Ordinal Response Variables. Application Statistics. No.2. 205-217., 1990.

Pigman, J. G. and Agent, K. R. Highway Crashes in Construction and Maintenance Work Zones. Transportation Research Record, No. 1270, TRB, National Research Council, Washington, D.C., pp12 – 21., 1990.

Richard, S. H., and Dudek, C. L. Implementation of work-zone speed control measures. Transportation Research Record, No. 1086, Transportation Research Board, Washington, D.C., 36–42., 1986.

Rouphail, N.M., and Tiwari, G. Flow Characteristics at Freeway Lane Closures. Transportation Research Record No.1035, Transportation Research Board, National Research Council, Washington, D.C., 1985.

Sarle, W.S. Part 2 of Neural Networks FAQ, ftp://ftp.sas.com/pub/neural/FAQ.html 1997

Sisiopiku, V. P., Lyles, R. W., Krunz, M., Yang, Q., Akin, D., and Abbasi, M. Study of Speed Patterns in Work Zones, Transportation Research Board, National Research Council, Washington, D.C., 1999.

Spainhour, L. K. and Wootton, I. A. Modeling Fault in Fatal Pedestrian Crashes Using Various Data Sources. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.

Suresh, R., Analysis of Freeway Weaving Areas Using Corridor Simulator and Highway Capacity Manual, Master's Thesis, Master Thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 1997.

Talyor, D. R., Muthiah, S., Kulakowski, B. T., Mahoney, K. M. and Porter, R. J. Artificial Neural Network Speed Profile Model for Construction Work Zone on High-Speed Highways. Journal of Transportation Engineering. Vol.133, No.3. 198-204., 2007.

Tarko, A., Kainpakapatman, S., and Wasson, J. Modeling and Optimization of the Indiana Lane Merge Control System on Approaches to Freeway Work Zones. Final Report, FHWA/IN/JTRP-97/12. Purdue University, West Lafayette, Iniana., 1998.

Thomas, G.B. Optimal Detector Location on Arterial Streets for Advanced Traveler Information Systems. Doctoral Dissertation, Arizona State University, Tempe, Arizona, Dec. 1998.

Tian, Z., Urbanik, T., Engelbrecht, R.J., and Balke, K.N. Variations in Capacity and Delay Estimates from Microscopic Traffic Simulations Models. Transportation Research Board, National Research Council, Washington, D.C., 2002.

Transportation Research Board, Highway Capacity Manual, 4th edition, National Research Council, Washington, D.C., 2000.

Transportation Research Board. National Cooperative Highway Research Program (NCHRP) Report No. 350, Transportation Research Board, National Research Council, 1993.

Ullman, G. L., Finley, M. D., and Ullman B. R. Analysis of Crash at Active Night Work Zone in Texas. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C., 2006.

Wang, J., Hughes, W. E., Council, F. M., and Paniati, J. F. Investigation of Highway Work Zone Crashes: What We Know and What We Don't Know. Transportation Research Record, No.1529, Transportation Research Board, National Research Council, Washington, D.C., 1996.

Appendices

Appedix A: Variables and Codes of Work Zone Crash

Table A-1 Variable of Work Zone Fatal Crash

Variable	Description	Type
YEAR	The year of work zone fatal crash	Nominal
TIME	The time of work zone fatal crash	Nominal
AGE	The age of driver at fault	Ordinal
VEHMOVEMENT	The movement of vehicle at fault before accident	Nominal
CRASHTYPE	The type of crash	Nominal
VEHICLETYPE	Heavy vehicle involved?	Nominal
FUNCLASS	The function of roads	Nominal
TRWAYCHR	Road Characteristics (level / curve?)	Nominal
MAXSPEED	The speed limit	Continue
SECTADT	The AADT of the section of work zones	Continue
TYPESUR	The type of road surface	Nominal
SITELOCA	Site Location	Nominal
LIGHTCONDITION	Light condition	Nominal
WEATHERCONDITION	Weather condition	Nominal
ROADSURFACE	Road surface condition	Nominal
VISION	Vision Obstructed	Nominal
RDACCESS	Access control type	Nominal
SURWIDTH	The width of roads	Continue
CONTRIBUTINGFACTORS	The contributing factors	Nominal
TRAFCONT	Traffic Control	Nominal

Appendix A (Continued)

Table A-2 Codes for TIME

Codes	Description
1	6:00-10:00
2	10:00-16:00
3	16:00-20:00
4	20:00-6:00

Table A-3 Codes for AGE

Codes	Description
1	<19
2	20-24
3	25-34
4	35-44
5	45-54
6	55-64
7	>65

Table A-4 Codes for VEHMOVEMENT

Codes	Description
01	STRAIGHT AHEAD
02	SLOWING/STOPPED/STALLED
03	MAKING LEFT TURN
04	BACKING
05	MAKING RIGHT TURN
06	CHANGING LANES
07	ENTERING/LEAVING PARKING SPACE
08	PROPERLY PARKED
09	IMPROPERLY PARKED
10	MAKING U-TURN
11	PASSING
12	DRIVERLESS OR RUNAWAY VEH.
77	ALL OTHERS
88	UNKNOWN

Appendix A (Continued)

Table A-5 Codes for CRASHTYPE

Codes	Description
01	COLL. W/MV IN TRANS. REAR-END
02	COLL. W/MV IN TRANS. HEAD-ON
03	COLL. W/MV IN TRANS. ANGLE
04	COLL. W/MV IN TRANS. LFT-TURN
05	COLL. W/MV IN TRANS. RGT-TURN
06	COLL. W/MV IN TRANS. SIDESWIP
07	COLL. W/MV IN TRANS. BAKD INTO
08	COLL. W/PARKED CAR
09	COLLISION WITH MV ON ROADWAY
10	COLL. W/ PEDESTRIAN
11	COLL. W/ BICYCLE
12	COLL. W/ BICYCLE (BIKE LANE)
13	COLL. W/ MOPED
14	COLL. W/ TRAIN
15	COLL. W/ ANIMAL
16	MV HIT SIGN/SIGN POST
17	MV HIT UTILITY POLE/LIGHT POLE
18	MV HIT GUARDRAIL
19	MV HIT FENCE
20	MV HIT CONCRETE BARRIER WALL
21	MV HIT BRDGE/PIER/ABUTMNT/RAIL
22	MV HIT TREE/SHRUBBERY
23	COLL. W/CONSTRCTN BARRICDE/SGN
24	COLL. W/TRAFFIC GATE
25	COLL. W/CRASH ATTENUATORS
26	COLL. W/FIXED OBJCT ABOVE ROAD
27	MV HIT OTHER FIXED OBJECT
28	COLL. W/MOVEABLE OBJCT ON ROAD
29	MV RAN INTO DITCH/CULVERT
30	RAN OFF ROAD INTO WATER
31	OVERTURNED
32	OCCUPANT FELL FROM VEHICLE
33	TRACTOR/TRAILER JACKNIFED
34	FIRE
35	EXPLOSION
36	DOWNHILL RUNAWAY
37	CARGO LOSS OR SHIFT
38	SEPARATION OF UNITS
39	MEDIAN CROSSOVER
77	ALL OTHER (EXPLAIN)

Appendix A (Continued)

Table A-6 Codes for VEHICLETYPE

Codes	Description
00	UNKNOWN/NOT CODED
01	AUTOMOBILE
02	PASSENGER VAN
03	PICKUP/LIGHT TRUCK (2 REAR TIR)
04	MEDIUM TRUCK (4 REAR TIRES)
05	HEAVY TRUCK (2 OR MORE REAR AX)
06	TRUCK TRACTOR (CAB)
07	MOTOR HOME (RV)
08	BUS (DRIVER + 9 - 15 PASS)
09	BUS (DRIVER + > 15 PASS)
10	BICYCLE
11	MOTORCYCLE
12	MOPED
13	ALL TERRAIN VEHICLE
14	TRAIN
15	LOW SPEED VEHICLE
77	OTHER
88	PEDESTRIAN NO VEHICLE

Table A-7 Codes for TRWAYCHR

Codes	Description
1	STRAIGHT-LEVEL
2	STRAIGHT-UPGRADE/DOWNGRADE
3	CURVE-LEVEL
4	CURVE-UPGRADE/DOWNGRADE

Table A-8 Codes for TYPESUR

Codes	Description
01	SLAG/GRAVEL/STONE
02	BLACKTOP
03	BRICK/BLOCK
04	CONCRETE
05	DIRT
77	ALL OTHER

Appendix A (Continued)

Table A-9 Codes for SITELOCA

Codes	Description
01	NOT AT INTERSECTION/RRX/BRIDGE
02	AT INTERSECTION
03	INFLUENCED BY INTERSECTION
04	DRIVEWAY ACCESS
05	RAILROAD CROSSING
06	BRIDGE
07	ENTRANCE RAMP
08	EXIT RAMP
09	PARKING LOT/TRAFFIC WAY
10	PARKING LOT AISLE OR STALL
11	PRIVATE PROPERTY
12	TOLL BOOTH
13	PUBLIC BUS STOP ZONE
77	ALL OTHER

Table A-10 Codes for LIGHTCONDITION

Codes	Description
01	DAYLIGHT
02	DUSK
03	DAWN
04	DARK (STREET LIGHT)
05	DARK (NO STREET LIGHT)
88	UNKNOWN

Table A-11 Codes for WEATHERCONDITION

Codes	Description
01	CLEAR
02	CLOUDY
03	RAIN
04	FOG
77	ALL OTHER
88	UNKNOWN

Appendix A (Continued)

Table A-12 Codes for ROADSURFACE

Codes	Description
01	DRY
02	WET
03	SLIPPERY
04	ICY
77	ALL OTHER
88	UNKNOWN

Table A-13 Codes for VISION

Codes	Description
01	VISION NOT OBSCURED
02	INCLEMENT WEATHER
03	PARKED/STOPPED VEHICLE
04	TREES/CROPS/BUSHES
05	LOAD ON VEHICLE
06	BUILDING/FIXED OBJECT
07	SIGNS/BILLBOARDS
08	FOG
09	SMOKE
10	GLARE
77	ALL OTHER (EXPLAIN)

Table A-14 Codes for RDACCESS

Codes	Description
1	FULL
2	PARTIAL
3	NONE

Appendix A (Continued)

Table A-15 Codes for CONTRIBUTINGFACTORS

Codes	Description
01	NO IMPROPER DRIVING/ACTION
02	CARELESS DRIVING
03	FAILED TO YEILD RIGHT OF WAY
04	IMPROPER BACKING
05	IMPROPER LANE CHANGE
06	IMPROPER TURN
07	ALCOHOL-UNDER INFLUENCE
08	DRUGS-UNDER INFLUENCE
09	ALCOHOL DRUGS-UNDER INFLUENCE
10	FOLLOWED TOO CLOSELY
11	DISREGARDED TRAFFIC SIGNAL
12	EXCEEDED SAFE SPEED LIMIT
13	DISREGARDED STOP SIGN
14	FAILED TO MAINTAIN EQUIP/VEHIC
15	IMPROPER PASSING
16	DROVE LEFT OF CENTER
17	EXCEEDED STATED SPEED LIMIT
18	OBSTRUCTING TRAFFIC
19	IMPROPER LOAD
20	DISREGARDED OTHER TRAFFIC CONT
21	DRIVING WRONG SIDE/WAY
22	FLEEING POLICE
23	VEHICLE MODIFIED
24	DRIVER DISTRACTION
77	ALL OTHER (EXPLAIN)

Table A-16 Codes for TRAFCONT

Codes	Description
01	NO CONTROL
02	SPECIAL SPEED ZONE
03	SPEED CONTROL SIGN
04	SCHOOL ZONE
05	TRAFFIC SIGNAL
06	STOP SIGN
07	YIELD SIGN
08	FLASHING LIGHT
09	RAILROAD SIGNAL
10	OFFICER/GUARD/FLAGMAN
11	POSTED NO U-TURN
12	NO PASSING ZONE
77	ALL OTHER

Appedix B: Sample of CORSIM Input File

```
Created by TSIS Wed Sep 26 14:26:18 2007 from TNO Version 61
Work Zone Simulation                                                        0
12345678 1 2345678 2 2345678 3 2345678 4 2345678 5 2345678 6 2345678 7 234567
Zhenyu Wang                          6  202007USF                        0   1
     1   1   1  10     9927 0000  22            81419        8219   24007   2
 900                                                                  3
              60                                                     4
  0      0   0   0   0   0   0   0   0   0   0                           5
  94  86before                                                          10
  94  86  85 30000 2                    1                                19
  86  85  84  5000 2                   1                                19
  85  84  83  5000 2                   1                                19
  84  83  82  5000 2                   1                                19
  83  82  81  5000 2                   1                                19
  82  81  80  5000 2                   1                                19
  81  80  50 20000 2                    1                                19
  93  94  86 50000 2                    1                                19
  92  93  94 50000 2                    1                                19
  91  92  93 50000 2                    1                                19
  90  91  92 40000 2                    1                                19
8001  90  91     0 2                   1                                19
  80  508002  2000 2                    1                                19
  94  86 0 0   0 11055                              100      20
  86  85 0 0   0 11055                              100      20
  85  84 0 0   0 11055                              100      20
  84  83 0 0   0 11055                              100      20
  83  82 0 0   0 11055                              100      20
  82  81 0 0   0 11055                              100      20
  81  80 0 0   0 11045                              100      20
  93  94 0 0   0 11055                              100      20
  92  93 0 0   0 11055                              100      20
  91  92 0 0   0 11055                              100      20
  90  91 0 0   0 11055                              100      20
8001  90 0 0   0 11055                                       20
  80  50 0 0   0 11055                              100      20
  94  86  85 100                                       25
  86  85  84 100                                       25
  85  84  83 100                                       25
  84  83  82 100                                       25
  83  82  81 100                                       25
  82  81  80 100                                       25
  81  80  50 100                                       25
```

```
93  94  86 100                                    25
  92  93  94 100                                   25
  91  92  93 100                                   25
  90  91  92 100                                   25
8001  90  91 100                                    25
  80  508002 100                                    25
  94  86   2   1          0  94                    28
  94  86   1   1          0  94                    28
  94  86   2 100           0  94                    28
  94  86   1 100           0  94                    28
  94  86   2 200           0  94                    28
  94  86   1 200           0  94                    28
  94  86   2 300           0  94                    28
  94  86   1 300           0  94                    28
  94  86   2 400           0  94                    28
  94  86   1 400           0  94                    28
  94  86   2 500           0  94                    28
  94  86   1 500           0  94                    28
  94  86   2 600           0  94                    28
  94  86   1 600           0  94                    28
  94  86   2 700           0  94                    28
  94  86   1 700           0  94                    28
  94  86   2 800           0  94                    28
  94  86   1 800           0  94                    28
  94  86   2 900           0  94                    28
  94  86   1 900           0  94                    28
  94  86   21000           0  94                    28
  94  86   11000           0  94                    28
  94  86   21100           0  94                    28
  94  86   11100           0  94                    28
  94  86   21200           0  94                    28
  94  86   11200           0  94                    28
  94  86   21300           0  94                    28
  94  86   11300           0  94                    28
  94  86   21400           0  94                    28
  94  86   11400           0  94                    28
  94  86   21500           0  94                    28
  94  86   11500           0  94                    28
  94  86   21600           0  94                    28
  94  86   11600           0  94                    28
  94  86   21700           0  94                    28
  94  86   11700           0  94                    28
  94  86   21800           0  94                    28
  94  86   11800           0  94                    28
```

94	86	21900		0	94	28
94	86	11900		0	94	28
94	86	22000		0	94	28
94	86	12000		0	94	28
94	86	22100		0	94	28
94	86	12100		0	94	28
94	86	22200		0	94	28
94	86	12200		0	94	28
94	86	22300		0	94	28
94	86	12300		0	94	28
94	86	22400		0	94	28
94	86	12400		0	94	28
94	86	22500		0	94	28
94	86	12500		0	94	28
94	86	22600		0	94	28
94	86	12600		0	94	28
94	86	22700		0	94	28
94	86	12700		0	94	28
94	86	22800		0	94	28
94	86	12800		0	94	28
94	86	22900		0	94	28
94	86	12900		0	94	28
86	85	2	1	0	86	28
86	85	1	1	0	86	28
86	85	2	100	0	86	28
86	85	1	100	0	86	28
86	85	2	200	0	86	28
86	85	1	200	0	86	28
86	85	2	300	0	86	28
86	85	1	300	0	86	28
86	85	2	400	0	86	28
86	85	1	400	0	86	28
85	84	2	1	0	85	28
85	84	1	1	0	85	28
85	84	2	100	0	85	28
85	84	1	100	0	85	28
85	84	2	200	0	85	28
85	84	1	200	0	85	28
85	84	2	300	0	85	28
85	84	1	300	0	85	28
85	84	2	400	0	85	28
85	84	1	400	0	85	28
84	83	2	1	0	84	28
84	83	1	1	0	84	28

84	83	2	100	0	84	28
84	83	1	100	0	84	28
84	83	2	200	0	84	28
84	83	1	200	0	84	28
84	83	2	300	0	84	28
84	83	1	300	0	84	28
84	83	2	400	0	84	28
84	83	1	400	0	84	28
83	82	2	1	0	83	28
83	82	1	1	0	83	28
83	82	2	100	0	83	28
83	82	1	100	0	83	28
83	82	2	200	0	83	28
83	82	1	200	0	83	28
83	82	2	300	0	83	28
83	82	1	300	0	83	28
83	82	2	400	0	83	28
83	82	1	400	0	83	28
82	81	2	1	0	82	28
82	81	1	1	0	82	28
82	81	2	100	0	82	28
82	81	1	100	0	82	28
82	81	2	200	0	82	28
82	81	1	200	0	82	28
82	81	2	300	0	82	28
82	81	1	300	0	82	28
82	81	2	400	0	82	28
82	81	1	400	0	82	28
81	80	2	1	0		28
81	80	1	1	0		28
81	80	2	100	0		28
81	80	1	100	0		28
81	80	2	200	0		28
81	80	1	200	0		28
81	80	2	300	0		28
81	80	1	300	0		28
81	80	2	400	0		28
81	80	1	400	0		28
81	80	2	500	0		28
81	80	1	500	0		28
81	80	2	600	0		28
81	80	1	600	0		28
81	80	2	700	0		28
81	80	1	700	0		28

81	80	2 800	0		28
81	80	1 800	0		28
81	80	2 900	0		28
81	80	1 900	0		28
81	80	21000	0		28
81	80	11000	0		28
81	80	21100	0		28
81	80	11100	0		28
81	80	21200	0		28
81	80	11200	0		28
81	80	21300	0		28
81	80	11300	0		28
81	80	21400	0		28
81	80	11400	0		28
81	80	21500	0		28
81	80	11500	0		28
81	80	21600	0		28
81	80	11600	0		28
81	80	21700	0		28
81	80	11700	0		28
81	80	21800	0		28
81	80	11800	0		28
81	80	21900	0		28
81	80	11900	0		28
93	94	2 1	0	93	28
93	94	1 1	0	93	28
93	94	2 100	0	93	28
93	94	1 100	0	93	28
93	94	2 200	0	93	28
93	94	1 200	0	93	28
93	94	2 300	0	93	28
93	94	1 300	0	93	28
93	94	2 400	0	93	28
93	94	1 400	0	93	28
93	94	2 500	0	93	28
93	94	1 500	0	93	28
93	94	2 600	0	93	28
93	94	1 600	0	93	28
93	94	2 700	0	93	28
93	94	1 700	0	93	28
93	94	2 800	0	93	28
93	94	1 800	0	93	28
93	94	2 900	0	93	28
93	94	1 900	0	93	28

93	94	21000	0	93	28
93	94	11000	0	93	28
93	94	21100	0	93	28
93	94	11100	0	93	28
93	94	21200	0	93	28
93	94	11200	0	93	28
93	94	21300	0	93	28
93	94	11300	0	93	28
93	94	21400	0	93	28
93	94	11400	0	93	28
93	94	21500	0	93	28
93	94	11500	0	93	28
93	94	21600	0	93	28
93	94	11600	0	93	28
93	94	21700	0	93	28
93	94	11700	0	93	28
93	94	21800	0	93	28
93	94	11800	0	93	28
93	94	21900	0	93	28
93	94	11900	0	93	28
93	94	22000	0	93	28
93	94	12000	0	93	28
93	94	22100	0	93	28
93	94	12100	0	93	28
93	94	22200	0	93	28
93	94	12200	0	93	28
93	94	22300	0	93	28
93	94	12300	0	93	28
93	94	22400	0	93	28
93	94	12400	0	93	28
93	94	22500	0	93	28
93	94	12500	0	93	28
93	94	22600	0	93	28
93	94	12600	0	93	28
93	94	22700	0	93	28
93	94	12700	0	93	28
93	94	22800	0	93	28
93	94	12800	0	93	28
93	94	22900	0	93	28
93	94	12900	0	93	28
93	94	23000	0	93	28
93	94	13000	0	93	28
93	94	23100	0	93	28
93	94	13100	0	93	28

Appendix B (Continued)

93	94	23200	0	93	28
93	94	13200	0	93	28
93	94	23300	0	93	28
93	94	13300	0	93	28
93	94	23400	0	93	28
93	94	13400	0	93	28
93	94	23500	0	93	28
93	94	13500	0	93	28
93	94	23600	0	93	28
93	94	13600	0	93	28
93	94	23700	0	93	28
93	94	13700	0	93	28
93	94	23800	0	93	28
93	94	13800	0	93	28
93	94	23900	0	93	28
93	94	13900	0	93	28
93	94	24000	0	93	28
93	94	14000	0	93	28
93	94	24100	0	93	28
93	94	14100	0	93	28
93	94	24200	0	93	28
93	94	14200	0	93	28
93	94	24300	0	93	28
93	94	14300	0	93	28
93	94	24400	0	93	28
93	94	14400	0	93	28
93	94	24500	0	93	28
93	94	14500	0	93	28
93	94	24600	0	93	28
93	94	14600	0	93	28
93	94	24700	0	93	28
93	94	14700	0	93	28
93	94	24800	0	93	28
93	94	14800	0	93	28
93	94	24900	0	93	28
93	94	14900	0	93	28
92	93	2 1	0	92	28
92	93	1 1	0	92	28
92	93	2 100	0	92	28
92	93	1 100	0	92	28
92	93	2 200	0	92	28
92	93	1 200	0	92	28
92	93	2 300	0	92	28
92	93	1 300	0	92	28

92	93	2 400	0	92	28
92	93	1 400	0	92	28
92	93	2 500	0	92	28
92	93	1 500	0	92	28
92	93	2 600	0	92	28
92	93	1 600	0	92	28
92	93	2 700	0	92	28
92	93	1 700	0	92	28
92	93	2 800	0	92	28
92	93	1 800	0	92	28
92	93	2 900	0	92	28
92	93	1 900	0	92	28
92	93	21000	0	92	28
92	93	11000	0	92	28
92	93	21100	0	92	28
92	93	11100	0	92	28
92	93	21200	0	92	28
92	93	11200	0	92	28
92	93	21300	0	92	28
92	93	11300	0	92	28
92	93	21400	0	92	28
92	93	11400	0	92	28
92	93	21500	0	92	28
92	93	11500	0	92	28
92	93	21600	0	92	28
92	93	11600	0	92	28
92	93	21700	0	92	28
92	93	11700	0	92	28
92	93	21800	0	92	28
92	93	11800	0	92	28
92	93	21900	0	92	28
92	93	11900	0	92	28
92	93	22000	0	92	28
92	93	12000	0	92	28
92	93	22100	0	92	28
92	93	12100	0	92	28
92	93	22200	0	92	28
92	93	12200	0	92	28
92	93	22300	0	92	28
92	93	12300	0	92	28
92	93	22400	0	92	28
92	93	12400	0	92	28
92	93	22500	0	92	28
92	93	12500	0	92	28

92	93	22600	0	92	28
92	93	12600	0	92	28
92	93	22700	0	92	28
92	93	12700	0	92	28
92	93	22800	0	92	28
92	93	12800	0	92	28
92	93	22900	0	92	28
92	93	12900	0	92	28
92	93	23000	0	92	28
92	93	13000	0	92	28
92	93	23100	0	92	28
92	93	13100	0	92	28
92	93	23200	0	92	28
92	93	13200	0	92	28
92	93	23300	0	92	28
92	93	13300	0	92	28
92	93	23400	0	92	28
92	93	13400	0	92	28
92	93	23500	0	92	28
92	93	13500	0	92	28
92	93	23600	0	92	28
92	93	13600	0	92	28
92	93	23700	0	92	28
92	93	13700	0	92	28
92	93	23800	0	92	28
92	93	13800	0	92	28
92	93	23900	0	92	28
92	93	13900	0	92	28
92	93	24000	0	92	28
92	93	14000	0	92	28
92	93	24100	0	92	28
92	93	14100	0	92	28
92	93	24200	0	92	28
92	93	14200	0	92	28
92	93	24300	0	92	28
92	93	14300	0	92	28
92	93	24400	0	92	28
92	93	14400	0	92	28
92	93	24500	0	92	28
92	93	14500	0	92	28
92	93	24600	0	92	28
92	93	14600	0	92	28
92	93	24700	0	92	28
92	93	14700	0	92	28

92	93	24800	0	92	28
92	93	14800	0	92	28
92	93	24900	0	92	28
92	93	14900	0	92	28
91	92	2 1	0	91	28
91	92	1 1	0	91	28
91	92	2 100	0	91	28
91	92	1 100	0	91	28
91	92	2 200	0	91	28
91	92	1 200	0	91	28
91	92	2 300	0	91	28
91	92	1 300	0	91	28
91	92	2 400	0	91	28
91	92	1 400	0	91	28
91	92	2 500	0	91	28
91	92	1 500	0	91	28
91	92	2 600	0	91	28
91	92	1 600	0	91	28
91	92	2 700	0	91	28
91	92	1 700	0	91	28
91	92	2 800	0	91	28
91	92	1 800	0	91	28
91	92	2 900	0	91	28
91	92	1 900	0	91	28
91	92	21000	0	91	28
91	92	11000	0	91	28
91	92	21100	0	91	28
91	92	11100	0	91	28
91	92	21200	0	91	28
91	92	11200	0	91	28
91	92	21300	0	91	28
91	92	11300	0	91	28
91	92	21400	0	91	28
91	92	11400	0	91	28
91	92	21500	0	91	28
91	92	11500	0	91	28
91	92	21600	0	91	28
91	92	11600	0	91	28
91	92	21700	0	91	28
91	92	11700	0	91	28
91	92	21800	0	91	28
91	92	11800	0	91	28
91	92	21900	0	91	28
91	92	11900	0	91	28

91	92	22000	0	91	28
91	92	12000	0	91	28
91	92	22100	0	91	28
91	92	12100	0	91	28
91	92	22200	0	91	28
91	92	12200	0	91	28
91	92	22300	0	91	28
91	92	12300	0	91	28
91	92	22400	0	91	28
91	92	12400	0	91	28
91	92	22500	0	91	28
91	92	12500	0	91	28
91	92	22600	0	91	28
91	92	12600	0	91	28
91	92	22700	0	91	28
91	92	12700	0	91	28
91	92	22800	0	91	28
91	92	12800	0	91	28
91	92	22900	0	91	28
91	92	12900	0	91	28
91	92	23000	0	91	28
91	92	13000	0	91	28
91	92	23100	0	91	28
91	92	13100	0	91	28
91	92	23200	0	91	28
91	92	13200	0	91	28
91	92	23300	0	91	28
91	92	13300	0	91	28
91	92	23400	0	91	28
91	92	13400	0	91	28
91	92	23500	0	91	28
91	92	13500	0	91	28
91	92	23600	0	91	28
91	92	13600	0	91	28
91	92	23700	0	91	28
91	92	13700	0	91	28
91	92	23800	0	91	28
91	92	13800	0	91	28
91	92	23900	0	91	28
91	92	13900	0	91	28
91	92	24000	0	91	28
91	92	14000	0	91	28
91	92	24100	0	91	28
91	92	14100	0	91	28

91	92	24200	0	91	28
91	92	14200	0	91	28
91	92	24300	0	91	28
91	92	14300	0	91	28
91	92	24400	0	91	28
91	92	14400	0	91	28
91	92	24500	0	91	28
91	92	14500	0	91	28
91	92	24600	0	91	28
91	92	14600	0	91	28
91	92	24700	0	91	28
91	92	14700	0	91	28
91	92	24800	0	91	28
91	92	14800	0	91	28
91	92	24900	0	91	28
91	92	14900	0	91	28
90	91	2 1	0	90	28
90	91	1 1	0	90	28
90	91	2 100	0	90	28
90	91	1 100	0	90	28
90	91	2 200	0	90	28
90	91	1 200	0	90	28
90	91	2 300	0	90	28
90	91	1 300	0	90	28
90	91	2 400	0	90	28
90	91	1 400	0	90	28
90	91	2 500	0	90	28
90	91	1 500	0	90	28
90	91	2 600	0	90	28
90	91	1 600	0	90	28
90	91	2 700	0	90	28
90	91	1 700	0	90	28
90	91	2 800	0	90	28
90	91	1 800	0	90	28
90	91	2 900	0	90	28
90	91	1 900	0	90	28
90	91	21000	0	90	28
90	91	11000	0	90	28
90	91	21100	0	90	28
90	91	11100	0	90	28
90	91	21200	0	90	28
90	91	11200	0	90	28
90	91	21300	0	90	28
90	91	11300	0	90	28

90	91	21400	0	90	28
90	91	11400	0	90	28
90	91	21500	0	90	28
90	91	11500	0	90	28
90	91	21600	0	90	28
90	91	11600	0	90	28
90	91	21700	0	90	28
90	91	11700	0	90	28
90	91	21800	0	90	28
90	91	11800	0	90	28
90	91	21900	0	90	28
90	91	11900	0	90	28
90	91	22000	0	90	28
90	91	12000	0	90	28
90	91	22100	0	90	28
90	91	12100	0	90	28
90	91	22200	0	90	28
90	91	12200	0	90	28
90	91	22300	0	90	28
90	91	12300	0	90	28
90	91	22400	0	90	28
90	91	12400	0	90	28
90	91	22500	0	90	28
90	91	12500	0	90	28
90	91	22600	0	90	28
90	91	12600	0	90	28
90	91	22700	0	90	28
90	91	12700	0	90	28
90	91	22800	0	90	28
90	91	12800	0	90	28
90	91	22900	0	90	28
90	91	12900	0	90	28
90	91	23000	0	90	28
90	91	13000	0	90	28
90	91	23100	0	90	28
90	91	13100	0	90	28
90	91	23200	0	90	28
90	91	13200	0	90	28
90	91	23300	0	90	28
90	91	13300	0	90	28
90	91	23400	0	90	28
90	91	13400	0	90	28
90	91	23500	0	90	28
90	91	13500	0	90	28

```
90  91   23600           0  90                                        28
  90  91   13600           0  90                                        28
  90  91   23700           0  90                                        28
  90  91   13700           0  90                                        28
  90  91   23800           0  90                                        28
  90  91   13800           0  90                                        28
  90  91   23900           0  90                                        28
  90  91   13900           0  90                                        28
  81  80 2                      0   2000      0 2000      001200        29
8001  900800  00   0  100                            50 50           50
     1  30                            20   1               64
82838485869091929394                                               67
 125 115 105  95  85  75  65  55  45  35   3                           68
  16  16  16  16  12   5                                    69
  10  13   4  20   4   8  80  80  80  80  80  80  80  80  80 100  15         70
  90  93  95  98  99 101 102 105 107 110                                147
  0                                                    170
8001      0      0                                          195
8002    5846      0                                           195
  50    5700      0                                         195
  80    5500      0                                         195
  81    5100      0                                         195
  82    5000      0                                         195
  83    4900      0                                         195
  84    4800      0                                         195
  85    4700      0                                         195
  86    4600      0                                         195
  90     200      0                                        195
  91    1000      0                                         195
  92    2000      0                                         195
  93    3000      0                                         195
  94    4000      0                                         195
  1   0   0                                          210
```

About the Author

Zhenyu Wang received a Bachelor's Degree in Electronic Engineering from Taiyuai University of Technology, Taiyuan, China in 1996 and a M.S. degree in Civil Engineering from Chang'An University in 1999. He continued to study for a Ph.D. degree in Transportation Systems of Civil Engineering at the University of South Florida in 2003. While in the Ph.D. program at the University of South Florida, Mr. Ye was very active in transportation research. He has completed several research projects and authored one journal publication.