NOTE TO USERS

This reproduction is the best copy available.

UMI®

OBSERVATIONAL BEFORE AND AFTER SAFETY STUDY OF INSTALLING SIGNALS AT RURAL INTERSECTONS: USING THE EMPIRICAL BAYES (EB) AND CONVENTIONAL METODS

By
Ravindranauth Bhim
BEng, Ryerson University, Toronto, 2003

A thesis
Presented to Ryerson University
in partial fulfillment of the
requirements for the degree of
Master of Applied Science
in the Program of Civil Engineering

Toronto, Ontario, Canada, 2005
© Ravindranauth Bhim

PROPERTY OF
RYERSON UNIVERSITY LIBRARY

UMI Number: EC53006

All rights reserved

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI®

UMI Microform EC53006
Copyright 2008 by ProQuest LLC
All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Borrower's Page

Ryerson University requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

<u>Name</u>	<u>Address</u>	<u>Signature</u>	<u>Date</u>

Observational Before and After Safety Study of Installing Signals at Rural Intersections: Using the Empirical Bayes (EB) and Conventional Methods

Master of Applied Science, 2005

By Ravindranauth Bhim

Department of Civil Engineering

Ryerson University

Abstract

Signalizing an intersection usually results in a reduction in right-angle and left-turn crashes, and an increase in rear-end crashes. This study used the conventional and Empirical Bayes (EB) before and after methods on 45 treated sites (converted from stop to signal control) in California and Minnesota to estimate the safety effect of having signals installed. The results confirm the belief that right-angle and left-turn crashes are reduced and rear-end crashes increase. However, these effects cannot be used to quantitatively assess the benefit gained from the reduction of right-angle and left-turn crashes against the increase in rear-end crashes, simply because crash types have different severities. By performing an economic examination of the safety effects, this study was able to show that by installing signals on 45 treated sites, there was a positive aggregate economic benefit of $155,883,978 which represents a 69 percent reduction in cost. This translated into $616,142 per site-year.

Acknowledgement

I would to like express my appreciation and gratitude to my supervisor, Dr. Bhagwant Persaud for his guidance, encouragement and patience during the course of this thesis. Also, I would like to thank Craig Lyon for his help. Both of these individuals were instrumental in the completion of this thesis.

I would also like to thank the National Cooperative Highway Research Program (NCHRP) for their funding under the NCHRP 17-25 project titled "Crash Reduction Factors for Traffic Engineering and ITS Improvements".

I would like to thank my oral examination committee members Dr. Mike Chapman, Dr. Khaled Sennah, Dr. Bhagwant Persaud and Dr. Songnian Li (Chair) for their thoroughness in reviewing this thesis.

A special thank you to my parents, Mr. and Mrs. Lallman Bhim, for their financial and moral support and my sister, Pryia, for her encouragement.

Finally, my sincere thanks to Jennifer Singh for her support, love, patience, and encouragement.

Dedication

To My Grandparents – Mr. and Mrs. Bheem Ragnauth

Table of Contents

Notation

The following notations are used in this report:

AADT	Annual average daily traffic
CURE	Cumulative Residual
EB	Empirical Bayes
GOF	Goodness-of-fit
HSIS	Highway Safety Information System
Maj	The AADT on the major approach of the intersection
Min	The AADT on the minor approach of the intersection
MUTCD	Manual on Uniform Traffic Control Devices
NB	Negative Binomial
pdf	Probability density function
RTM	Regression-to-mean
SPF	Safety Performance Functions
π	The expected accident frequency at an entity had a specific treatment NOT been implemented
λ	The expected accident frequency in the after a specific treatment has been implemented
K	Accident count at some entity
κ	The expected accident frequency at an some entity
α	The weighted average of the empirical Bayes approach
r_d	Ratio of duration, i.e. the ratio of the after period to the before period.
$C_{i,y}$	The ratio of model estimate for entity 'i' in year y to model estimate for entity 'i' in year 1.
δ	The expected reduction in accident frequency.
θ	The index of effectiveness, usually illustrated in percentages.

List of Tables

List of Figures

CHAPTER 1: Introduction

At-grade highway intersections are inevitable due to conflicting traffic streams. This makes them a very critical component in a transportation system. Hence the need for safety is paramount, especially at high-speed intersections, such as those in a rural jurisdiction. In Ontario alone, 224,642 collisions were reported in the Ministry of Transposition 2002 annual road safety report (MTO 2002) inclusive of 48,116 and 60,618 intersection and intersection-related accidents respectively; a staggering 48 percent of reported collisions. In 2003, the U.S Department of Transportation reported that 9,213 fatalities occurred due to intersection related collisions. Clearly there is a need for safer operating conditions at intersections. To this end, the National Cooperative Highway Research Program (NCHRP) is currently working on a research project titled "Crash reduction Factors for Traffic Engineering and ITS Countermeasures (NCHRP 17-25)". The research in this thesis was performed as part of that project.

1.1 Statement of Problem

Installing traffic signals at intersections has been one of the many methods used in traffic engineers' constant quest for more safety and better efficiency. The Manual on Uniform Traffic Control Devices (MUTCD 2003) is one of the main standards providing guidance to traffic engineers on the use of traffic control devices. In Part 4 of the MUTCD, guidance on when one should consider installing a traffic signal is explained under "warrants". Warrants alone are not the only criteria in determining if a signal is necessary or not. Whether or not the warrants are met, the MUTCD (2003) suggests that an engineering study be performed to investigate whether the overall safety and efficient operation of the intersection will improve if a signal were to be installed.

Eight warrants are proposed in Part 4 of the MUTCD; they are based on eight-hour vehicular volume, four-hour vehicular volume, peak hour volume, pedestrian volume, school crossing needs, coordination of the signal system, crash experience, and the characteristics of the roadway network. The warrant that pertains to road safety, and particularly to this study, is warrant 7. The focus of this warrant is directly related to the crash experience sustained at an intersection.

The MUTCD does not actually propose a procedure for executing the suggested engineering study. However, work done by McGee et al. (2003) proposed a procedure for estimating the change in expected crash frequency and costs if a traffic signal is installed at an intersection. This procedure was based on urban/suburban intersections. Rural intersections are different from urban and sub-urban intersections in several aspects with the main difference being the design speeds, which are higher than those in urban and sub-urban areas. Other factors are driver demography, traffic volumes, business activity, etc. To facilitate the development of an engineering procedure for determining the need for a traffic signal at a rural stop controlled intersection, one must know what would be the safety benefit of installing a traffic signal in a rural environment. The methods adapted in this research for answering this question are discussed in the following section.

1.2 Objective of Study

The main objective of this study is to evaluate the safety effect of installing traffic signals at rural stop controlled intersections. The safety effect of a treatment at an entity is evaluated by finding the difference between the accident frequency that would have occurred had the entity not been treated and the accident frequency that occurred on the entity after treatment. In this study, four methods were compared for evaluating the effects of a treatment (installing a traffic signal). These are:

- The simple (Naive) before and after study,
- The Naive before and after study with correction for traffic flows,
- The Naive before and after with safety performance functions (SPFs), and
- The Empirical Bayes (EB) before and after method.

For all methods crash effects are estimated. For the EB method, an analysis of the economic effects is performed to accommodate the differential impacts on different crash types and severities.

1.3 Structure of Report

This thesis is divided into 8 chapters. This Chapter (1) outlines the motives and objectives for this research; these are supported by an in-depth literature review covered in Chapter 2. The methodology used in estimating safety is reviewed in Chapter 3 with a discussion on the theoretical approaches to the EB and conventional procedures used. Chapter 3 also has a discussion on fitting a functional form in multivariate regression models that are fundamental to the EB methodology. Chapter 4 addresses the data used in this research, while a description of the accident prediction models fitted for use in the application of the EB procedure is covered in Chapter 5. The economic analysis of the change in crashes of various types is in Chapter 7, while Chapter 8 has the conclusions derived from this study.

CHAPTER 2: Literature Review

Evaluating the safety effect of the installation of traffic signals at rural stop controlled intersections is the main focus of this research. Thus, a thorough literature review on the estimation of the safety effect of installing signals was conducted, along with a review of methods for estimating safety at both signal and stop controlled intersections. The latter is pertinent to the EB methodology used in this study.

2.1 Accident Pattern after Signalization

King and Goldblatt (1975) conducted a cross-sectional study to illustrate the relationship between accident types and intersection controls. They analyzed data obtained from 250 signal and stop controlled intersections across several states in the United States of America. For each state, data were disaggregated at four main levels; geographic area (north, south, etc), type of area (central business district, rural etc), major and minor approach volumes, and control type. Three methods of analysis were applied to the disaggregated data; Analysis of Variance, Multiple Linear Regression Analysis, and Hypothesis Testing. The following are findings pertinent to this research as concluded by the authors:

1. If a traffic signal were to be installed, a reduction in right-angle accidents and an increase in rear-end accidents will result,
2. No definitive answer could be given as to whether the installation of a traffic signal will have an overall reduction of accidents at intersections,
3. Right-angle accidents were not directly reduced after a signal was installed, and

4. Accident rates, in number of accidents per million entering vehicles, are higher at signalized intersections but this is offset by the lower severity per accident.

However, cross-sectional studies, such as this is, are not a reliable method for evaluating changes in safety because of its underlying assumptions. Cross-sectional studies facilitate safety comparison between two distinct sets of entities. For example, one set can comprise of signalized intersections and the other set contains stop controlled intersections. The difference in safety between these two distinct sets of entities is not entirely due to the difference in control type since it can also be due to differences in factors such as geometric configuration. Therefore, the findings of King and Goldblatt are questionable.

Rural intersections are characterized by their high design speeds which clearly has an effect on the severity of crashes. This distinct characteristic of rural intersections implies that one should expect differences in accident pattern by types and severity from urban intersections. A study performed by Hanna et al. (1976) attempted to quantify this difference using 232 rural intersections located in Virginia, United States of America. The authors used accident rates (number of accidents per million entering vehicles) at each intersection to compare the relationship of traffic control and intersection geometrics to accident types. Based on the data, 36 percent of accidents were classified as rear-end and 43 percent of accidents as angle. More rear-end accidents occurred at signalized intersections while more angle accidents occurred at stop-controlled intersections. No substantive findings came to light regarding differences in safety between urban and rural intersections. The findings of this study by Hanna et al. (1976) are also questionable because they are based on a cross-sectional evaluation.

Datta and Dutta (1990) focused on quantifying accident experience after traffic signals were installed. A simple before and after method was adopted using 102 intersections with signals installed during 1978 and 1983 in Michigan, United States of America. A paired t-test at 0.05 significance level was also performed on the before and after mean crash rates to check if they were statistically significant. The findings were:

1. After signals were installed, there was a 19.2 percent reduction in total accident,
2. Rear-end accident rates increased by 53 percent,
3. Right angle accident rates were 57 percent lower after signals were installed, and
4. Head-on and left-turn accident rates increased by 50 percent.

Datta (1991) carried out another study, using the same data set, on the safety effect of left-turn lanes at intersections after the installation of signals. A similar approach was used, that is, a paired t-tests at 0.05 significance level for the before and after mean accident rates. The analysis was disaggregated into three groups; locations with left-turn lanes, locations without left-turn lanes, and locations where left-turn lanes were installed with signal installation. The researcher found that total accidents were reduced at intersections where a left turn lane was added at the same time a signal was installed. A similar finding was obtained for right-angle accidents. The number of rear-end crashes increased after signals were installed at these intersections.

The findings from both Datta (1990) and Datta (1991) somewhat concur with the findings of King (1975) and Hanna et al. (1976) in that there was a reduction in the total and right angle crashes, and an increase in rear-end crashes. However, the methodology adapted by Datta (1990 and 1991) is more favoured when the objective is to evaluate changes in safety following the signalizing of intersections. This is because Datta (1990 and 1991) used sites that were actually converted to signal in their before and after study while the other two studies used two distinct sets of sites used in a cross-sectional evaluation. Even so, the simple before and after methodology used by Datta and Dutta is also questionable because of the assumption that the before number of crashes at an entity is a good estimate of the number of crashes in the after period had the treatment not been implemented. The problem with this assumption is that had there been no treatment, accident frequency could have easily increased or decreased due to regression to the mean (RTM) (Persaud 2001; Hauer 1997; Hauer and Persaud 1983; etc)., AADT trends or changes in factors such as weather, driver demography etc. Indeed, if an entity was selected to be treated because of its high crash counts alone, there is a good chance regression to the mean is at play (Hauer 1997). It is paramount, therefore, that when evaluating the safety effects of a particular treatment, one must be able to account for these changes in safety unrelated to the treatment. The Empirical Bayes (EB) procedure, developed by Hauer (1997), does just that.

2.2 The Empirical Bayes (EB) Procedure

Over the last decade, significant advancements were made towards estimating safety effect of treatments. The groundbreaking Empirical Bayes before-and after procedure by Hauer (1997) help safety analysts account for more than just RTM in their safety analyses. It accounts for time trend in accident counts, and trends in AADT and other non-treatment factors that may cause changes in accidents. The EB procedure does this by the joining of two clues (Hauer 1997); the pre-treatment accident counts on the treated entity and the expected accident frequency at sites that share the same traits as the treated entity. The underlying theory of the EB procedure is documented in Chapter 3. There are several studies done prior to this research that adapted the EB procedure (See e.g., Hauer et al. 2002; Persaud et al. 1997; Council et al. 2005; etc.).

One such study was by McGee et al. (2003) where the EB methodology was used to develop a procedure for the MUTCD, for estimating the likely change in safety when the installation of a traffic signal is contemplated. The study used data from several states in the United States of America and from Toronto, Canada. The treatment (installation of signals) sites comprised of 22 three-legged and 100 four-legged intersections. Corresponding reference populations were compiled and regression models were developed for applying the EB method. The study found the same pattern for right-angle and rear-end accidents as did King (1975), Hanna et al. (1976), and Datta (1990 and 1991). However, the safety effects were lower simply because the EB method accounted for RTM, AADT trend and causal factors.

In the EB methodology the expected accident frequency at similar sites to the treated entity is estimated from a regression model (also known as a Safety Performance Function), that is calibrated from entities with similar traffic volumes and characteristics as the treated entity. Therefore, the need for reliable regression models is critical. A more detail discussion on safety performance functions can be found in Chapter 3.

CHAPTER 3: Research Methodology

The safety of roadway segments and intersections is always a major concern. Therefore, the question of what is safety, and how it should be measured is important. Hauer (1997) defines safety as "the number of accidents by kind and severity, expected to occur on an entity during a specified period". Relating this definition of safety to entities of interest, road segments or intersections, the following expression is derived:

$$E\{Accident\ Frequency\ at\ Entity\} = E\{Frequency\ of\ Accidents\ at\ Entity\ /\ Unit\ of\ Time\}$$

Another measure of safety usually found in literature is the expected accident rate (Hauer 1997) and is defined as:

$$Accident\ Rate = \frac{Accident\ Frequency}{Exposure\ /\ Unit\ of\ Time} \tag{3.1}$$

where exposure is the traffic flow through an entity during a specified duration (usually per year since the exposure is usually measure in AADT).

However, this measure of safety is not theoretically sound because of the non-linear relationship between the number of accidents and traffic flow. Accident rate assumes that the number of accidents that occurred at an entity is proportional to its AADT. The problem with this assumption is that safety performance functions (SPFs) developed in numerous sources (Bauer 1997; Bauer 1998; Hauer 1988; Lord 2000 and 2005; etc) show that the relationship between accidents and AADT in not linear. A SPF developed by Persaud et al. (2001) for estimating total crashes per year for signalized intersections in Maryland will be used to illustrate this non-linear relationship. The SPF was developed with an AADT ranging from 365 - 3,133 for the minor approach and 8,625 - 52,144 for the major approach and is of the form:

$$Total\ Crashes\ /\ year = \alpha\left(total\ entering\ AADT\right)^{\beta} \tag{3.2}$$

where α = 8.5706E-04 and β = 0.804 are regression coefficients

The SPF given in Equation 3.2 is depicted in Figure 3.1 using a range of AADT (total entering) from 1,000 – 35,000.

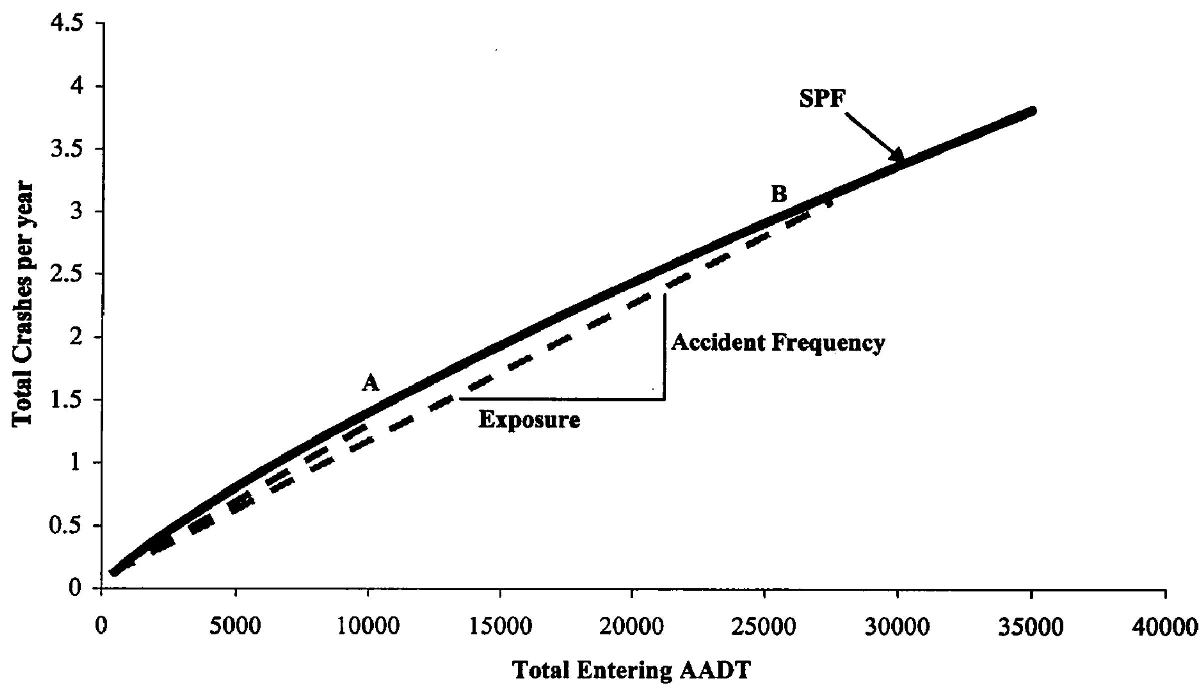


Figure 3.1 – Safety Performance Function for signalized intersections in Maryland, US.

The nonlinear shape shown in Figure 3.1 supports the point that accident frequency indeed exhibits a non-linear relationship with traffic flow. In Figure 3.1, accident rate is the slope at any point on the SPF since the slope is simply the ratio of the change in accident to the change in AADT. Therefore, if one is to use accident rate as a measure of safety, they are simply saying that at low AADT (point A) it is more hazardous than at higher AADT (point B) due to the difference in slope of the two dashed lines (red and green). This is a false estimate of safety since the slope at any point could be equal with different levels of AADT. Therefore, accident rate should not be used in estimating safety at an entity and one should use accident frequency instead.

Knowing what safety is quantitatively, safety analysts are able to measure the effectiveness of treatments on various entities. The effectiveness of a particular treatment is usually estimated by taking the difference between the *safety in the after period had the treatment not been applied*, and what the *safety of the treated entity in the after period was*. So we are faced with two fundamental questions that need answers:

1. What would have been the safety of the entity in the after period had the treatment not been implemented, π?
2. What is the safety after the treatment was implemented, λ?

Two methods are usually used to quantify the effectiveness of a treatment once λ and π:

1. Reduction in Expected Number of Crashes, δ
2. Index of Effectiveness, θ

where

$$\delta = \pi - \lambda \tag{3.3}$$

and

$$\theta = \frac{\lambda}{\pi} \qquad (3.4)$$

δ is simply the reduction in crashes after the treatment was implemented whereby, if δ is positive it simply implies that the treatment was effective and if negative that the treatment is detrimental to safety. The estimate of θ also informs safety analysts about how safe a treatment is, that is if θ is less than 1, the treatment is considered to be effective since the number of crashes after the treatment is lower than the estimated number of crashes in the after period had the treatment not been implemented. Conversely, if θ is larger than one, the treatment is considered to be harmful. θ is also used to estimate the percent reduction or increase in crashes after a treatment is implemented as $(1-\theta)*100$.

Hauer (1997) has shown that simply taking the ratio of λ to π will usually result in a biased estimate of θ and proposed an unbiased estimate of θ instead (Equation 3.5).

$$\theta = \frac{\frac{\lambda}{\pi}}{\left[1 + \frac{VAR\{\pi\}}{\pi^2}\right]} \qquad (3.5)$$

The variances for θ and δ are computed by Equations 3.6 and 3.7 respectively:

$$VAR\{\delta\} = VAR\{\pi\} - VAR\{\lambda\} \qquad (3.6)$$

and

$$VAR\{\theta\} = \frac{\theta^2\left[\left(\frac{VAR\{\lambda\}}{\lambda^2}\right)+\left(\frac{VAR\{\pi\}}{\pi^2}\right)\right]}{\left[1+\frac{VAR\{\pi\}}{\pi^2}\right]^2} \tag{3.7}$$

The variances are usually used to validate statistical significance for the estimates of δ and θ at some specific confidence level. For instance, standard deviation is easily computed from the variance (variance$^{0.5}$) and knowing that the one-sided 95% confidence interval is 1.96 of the standard deviation, the statistical significance of the estimates at the 5% level is simply established if the estimate ± 1.96 standard deviations does not include zero. Other confidence levels can be used, such as 90%, but 95% seems most common. Equations 3.3 through 3.7 are the rudiments of quantifying the safety effectiveness of any implemented treatment.

In order to estimate the value of π one needs to account for the effects of the regression to the mean phenomenon. This is discussed next.

3.1 Regression to the mean

If an entity is selected to be treated due to its high accident counts, the safety benefits after treatment will usually be overestimated (Hauer 1997). This is because accident counts on any entity oscillates around the true (long term) average of its accident count. Therefore, if an entity experiences high accident count it will likely decrease subsequently even if no treatment was implemented. This phenomenon is known as the regression-to-the-mean (RTM).

Several sources (Hauer 1997; Hauer and Persaud 1983; Persaud 2001, etc.) have explained this phenomenon in detail. An illustration of this phenomenon was developed for this study using data from California, HSIS database. Data from 1993-1997 at 1,463 rural intersections were used. The intersections were all rural four-legged, two-way-stop-controlled, with two-lane major and minor approaches.

Table 3.1 - Illustrating the Regression to the Mean Phenomenon

No. of Intersections with given No. of Accidents in 1994-1996	Accidents/ Intersection in 1994-1996	Accidents/ Year/ Intersection in 1994-1996	Accidents/ Year in 1994-1996 for Group (rounded)	Accidents in 1997 for Group	Accidents/ Intersection in 1997	Percent (%) Change
Column 1	**Column 2**	**Column 3**	**Column 4**	**Column 5**	**Column 6**	**Column 7**
468	0	0.00	0	86	0.18	**LR***
315	1	0.33	105	136	0.43	30
183	2	0.67	122	135	0.74	11
135	3	1.00	135	131	0.97	-3
92	4	1.33	123	98	1.07	-20
58	5	1.67	97	83	1.43	-14
49	6	2.00	98	74	1.51	-24
35	7	2.33	82	71	2.03	-13
128	≥ 8	3.78	484	56	0.44	-88

Sample Calculations:
Columns 1, 2 and 5 are inputs from the data.
Column 3: For second row, accidents/year = 1/3 since we are using three years of data.
Column 4: For second row, accidents/year = 315 x 0.33 = 105
Column 6: For second row, accidents / intersections in 1997 = 136/315 = 0.43
Column 7: Percent change (increase or decrease) between column 3 and 6.
* - LR = Large Reduction

The average accidents per intersection per year over the 3 years of data were approximately constant at 0.85. Yet, as Table 1 shows, accidents in individual groups changed. For example, the 315 intersections that recorded 1 accident during 1994-1996 (0.33 per year) averaged 0.43 accidents per intersection in 1997. Intersections that recorded 3 or more accidents in 1994 – 1996 experienced a reduction in accidents in

1997. Conversely, intersections that recorded 2 or less accidents during 1994 - 1996 experienced an increase in accident in 1997. These changes are nothing more than the effects of the RTM phenomenon which must be accounted for in quantifying the safety effects of a treatment. The ground breaking work by Hauer (1997) uses the Empirical Bayes (EB) procedure for doing so. The EB procedure is reviewed in the following section.

3.2 The Empirical Bayes (EB) Procedure

In section 3.1, it was made clear that two fundamental questions need to be answered to estimate the safety effect of a treatment at an entity. They are; what safety would have been had the treatment not been implemented, and what is safety after the treatment was implemented? The EB procedure was intentionally developed to estimate π by accounting for RTM. The second question is rather trivial – what safety was after treatment; the accident frequency after treatment is an unbiased estimate of safety after the treatment was implemented (Hauer 1997).

Hauer (1997) has proven that the most efficient means of accounting for RTM in estimating safety is by joining of two clues. These two clues are:

1. *Traits of the entity in question* – traits here are referring to roadway geometric elements such as number of lanes, median type (divided or undivided), rural or urban environment, etc. Safety is affected by the traits of an entity; for example, a rural signalized intersection will tend to have more severe accidents than urban signalized intersections because of their high design speeds. Hauer (1997) used the concept of a reference population when referring to entities that share the

same traits as the entity of interest. Safety of entities in the reference population, E{κ}, is a very useful measure in understanding what safety is at the entity of interest.

2. *Accident count (K) at the entity in question* – the accidents that took place on the entity of interest before the treatment was implemented.

The entire framework of the Empirical Bayes (EB) procedure is supported by the joining of these two clues. The same notations used by Hauer (1997) are used here and joining of the two clues resulted in Equation 3.8:

$$E\{\kappa \mid K\} = \alpha E\{\kappa\} + \{1-\alpha\}K \tag{3.8}$$

where E{κ} = expected accident frequency in the reference population

K = accident counts (history) at the entity of interest

α = is a weight between 0 and 1

If α is close to 1 then the safety in the reference population is close to the safety at the entity of interest. Conversely, if α is close to 0, then the safety, κ, of the entity of interest is close to the count of accidents recorded on it. Clearly the weight, α, in Equation 3.8 plays a pivotal role in the EB procedure and is calculated using Equation 3.9:

$$\alpha = \frac{1}{1+\frac{VAR\{\kappa\}}{E\{\kappa\}}} \tag{3.9}$$

where as the variance of the accident frequency, VAR{κ}, is estimated using Equation 3.10:

$$VAR\{\kappa \mid K\} = \{1-\alpha\}E\{\kappa \mid K\} \tag{3.10}$$

Therefore, one must be able to estimate E{κ} and VAR{κ} before the EB procedure can be performed. Hauer (1997) assumed that the accident count at the entity of interest, K, obeys the Poisson distribution while the κ's in the reference population follow a gamma probability distribution. By knowing the probability distributions of accidents in the reference population and treated sites, Hauer (1997) was able to prove that the probability density function of the κ's in the reference population is indeed gamma distributed by using Bayes theorem. The derivation is not shown here, only the results. The gamma distribution of the κ's in the reference population, assuming that the distribution of the accidents at the entity of interest is Poisson distributed, is given by Equation 3.11:

$$g\{\kappa \mid K\} = \frac{(1+a)^{K+b}\,\kappa^{K+b-1}e^{-\kappa(1+a)}}{\Gamma(K+b)} \tag{3.11}$$

where the parameters 'a' and 'b' are associated with the gamma distribution. From the properties of gamma distribution:

- The mean is the ratio of the parameter b to parameter a; that is $E\{\kappa\} = \frac{b}{a}$
- The variance is the ratio of the parameter b to the squared of parameter a, that is $VAR\{\kappa\} = \frac{b}{a^2}$

Therefore,

$$a = \frac{b}{E\{\kappa\}} \qquad \text{and} \qquad b = a^2 VAR\{\kappa\}$$

implying that

$$a = \frac{a^2 VAR\{\kappa\}}{E\{\kappa\}} \Rightarrow \frac{a}{a^2} = \frac{VAR\{\kappa\}}{E\{\kappa\}} \Rightarrow \frac{1}{a} = \frac{VAR\{\kappa\}}{E\{\kappa\}}$$

resulting in

$$a = \frac{E\{\kappa\}}{VAR\{\kappa\}} \tag{3.12}$$

Thus substituting for parameter 'a' parameter we can find parameter 'b'

$$b = \frac{(E\{\kappa\})^2}{(VAR\{\kappa\})^2} VAR\{\kappa\}$$

$$b = \frac{(E\{\kappa\})^2}{VAR\{\kappa\}} \tag{3.13}$$

Therefore, if parameter 'b' is known, we can then easily estimate E{κ} and VAR{κ}. The estimation of 'a' and 'b' hinges on the fact that E{κ} and VAR{κ} are known. Two methods are proposed (Hauer 1997) to calculate E{κ} and VAR{κ}. They are:

1. Method of Sample Moments
2. Multivariate Regression Method

More emphasis will be placed on the multivariate regression method in this thesis because it's the method that is widely accepted as being the better of the two.

Before moving on, some clarification is worthwhile with respect to Equations 3.8 and 3.9. Simplifying Equation 3.9 will result in the following:

$\alpha = \dfrac{1}{1 + \dfrac{VAR\{\kappa\}}{E\{\kappa\}}} = \dfrac{E\{\kappa\}}{E\{\kappa\} + VAR\{\kappa\}}$, but we know what VAR{κ} is equal to,

$VAR\{\kappa\} = \frac{(E\{\kappa\})^2}{b}$, therefore substituting for VAR{κ} will result in

$$\frac{E\{\kappa\}}{E\{\kappa\} + \frac{(E\{\kappa\})^2}{b}} = \frac{b}{b + E\{\kappa\}} = \alpha$$

Now $1 - \alpha = 1 - \frac{b}{b + E\{\kappa\}}$; simplifying we have $\frac{E\{\kappa\}}{b + E\{\kappa\}} = 1 - \alpha$

Therefore, Equation 3.8 when simplified resulted in Equation 3.14:

$$E\{\kappa \mid K\} = \frac{bE\{\kappa\}}{b + E\{\kappa\}} + \frac{E\{\kappa\}K}{b + E\{\kappa\}} = \frac{E\{\kappa\}(b + K)}{b + E\{\kappa\}} \qquad (3.14)$$

where E{κ|K} estimates the *number* of crashes

An important aspect of Equation 3.14 is the duration for which E{κ} pertains to. E{κ} *must* have the same duration as the before period accidents accounts, K. Usually E{κ} is estimated in crashes per year, and that estimate is simply multiplied by the number of years for which we have before period accident counts, K.

3.2.1 Multivariate Regression Method

As the name suggest, this method is simply using a fitted regression equation to estimate E{κ} and VAR{κ}. The models use accident frequency as the dependent variable and traits such as AADT, lane widths, lighting etc. as the independent variables. Fitting such a regression model falls in the domain of statistics and is discussed in more detail in Section 3.4 since it is very crucial to the validity of the results for this research.

The estimate of a multivariate model that is fitted to accident counts is E{κ}; this is the expected accident frequency in the reference population. The other estimate that is needed is the VAR{κ}. Based on previous work (Persaud et al. 2001, Persaud et al. 2003, McGee et al. 2003, etc.) the VAR{κ} is estimated by Equation 3.15

$$VAR\{\kappa\} = \frac{(E\{\kappa\})^2}{b} \tag{3.15}$$

The relationship is the same as Equation 3.13. Again, the parameter 'b' is associated with the gamma distribution. This parameter is usually estimated during the fitting of a multivariate regression model to the accident data. The method of maximum likelihood is usually used in the parameter estimation process. More details of the modeling approach are discussed in Section 3.4. The relationship shown in Equation 3.15 was confirmed when the estimates of VAR{κ} were plotted against the estimated E{κ} for a specific entity (Hauer 1997).

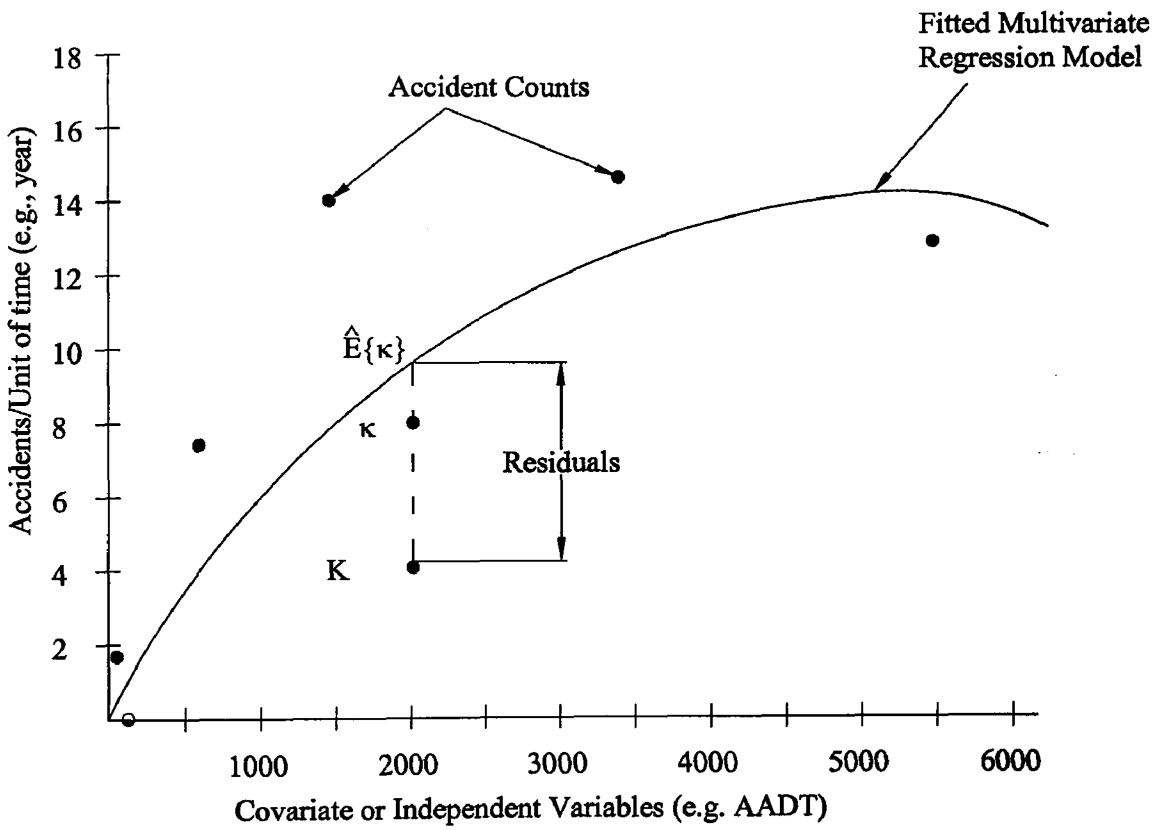


Figure 3.2 – Plot of a typical SPF (Hauer 1997)

Figure 3.2 illustrates the concept of the multivariate regression method and the estimate of VAR{κ}. Hauer (1997) suggested that the estimate of VAR{κ} for a single data point is as shown in equation 3.16

$$VAR\{\kappa\} = \left[\left(E\{\kappa\} - K \right)^2 - \ Estimate\,of\; E\{\kappa\} \right] \qquad (3.16)$$

Figure 3.2 depicts the residual for a single data point (accident count) on an entity of interest. The relationship of VAR{κ} that is commonly adapted is not actually 'always' the case (Hauer 1997). This equation is usually dependent on the data on which the multivariate regression model was fitted. However, this relationship was adopted in several sources (Council et al. 2005, Persaud et al. 2003, McGee et al. 2003 etc.) and is also used in this thesis.

The discussions so far have been based on how safety is quantified for an entity during the before period. In summary, if one needs to estimate the safety of an entity, the following steps would be followed:

- Estimate E{κ} and VAR{κ} from either the multivariate regression method
- Find the weight, α, using Equation 3.9, and
- Use Equation 3.8 to estimate the expected accident frequency on the entity of interest.

The EB procedure (Hauer et al. 2002) explained so far has been referred to as the abridged version since it only uses the average traffic flow and a few (2-3) years of accident counts. However, Hauer et al. (2002) has shown that by using longer before and after years of traffic and accident counts, the safety estimate is much more precise than when estimated by the shorter version. The full EB procedure is explained next.

In road safety analysis, more than one year of before and after period of accident data are usually collected for an entity. That is, in the before period one could have data on an entity for several years as $b_1, b_2, \ldots, b_Y$, where b_Y represents the last year of before period data. The same applies to the after period, that is, one could have data on an entity for several years as $a_1, a_2, \ldots, a_Z$ where a_Z is the last year of data in the after period. With yearly data on an entity 'i', the expected accident frequency $E\{\kappa_i\}$ will be different each year. The estimate of $E\{\kappa_i\}$ is usually derived from a multivariate regression model, thus one would have estimates of $\kappa_{i,b1}, \kappa_{i,b2}, \ldots, \kappa_{i,bY}$.

The multivariate regression model developed from the reference population is used to calculate $E\{\kappa_{i,b1}\}, E\{\kappa_{i,b2}\}, \ldots, E\{\kappa_{i,bY}\}$ using time (yearly) sensitive parameters. As will be explained in Section 3.4, a model is usually used to capture the systematic variations among the data (covariates) and link it to a dependent variable (expected accident frequency). Since each year will experience a different count of accidents, the average model with its same intercept term or multiplier cannot be used to estimate the $E\{\kappa_i\}$ for each year. The intercept is usually calibrated to capture the non-systematic variations in the data, such as driver's age, demography, traffic volume trends, education, accident reporting etc. The procedure on how the intercept is calibrated for each year is illustrated in (Harwood et al. 2002 and McGee et al. 2003).

The next question that must be answered is, how does the EB procedure account for the changes in expected accident frequency for each year? Hauer (1997) assumes that the change in the expected accident frequencies for each year can be used to capture the non-systematic variations in the data by using Equation 3.17:

$$C_{i,y} = \frac{\kappa_{i,y}}{\kappa_{i,b1}} \tag{3.17}$$

where $\kappa_{i,y}$ = prediction in year 'y'

$\kappa_{i,b1}$ = prediction in year '1'

The purpose of using $C_{i,y}$ in Equation 3.17 is to capture the changes at an entity over time. These changes are non-systematic, and cannot be quantified by the regression model. The best estimate is given by Equation 3.17. The reason is that, in a model equation, the accident counts are a function of independent variables (AADT, lane widths etc.) and by definition of Equation 3.17, $C_{i,y}$ is in turn a function of the independent variables.

Therefore, under the full empirical Bayes (EB) framework, the safety at some entity 'i' during the first year 'b_1' is *estimated* using Equation 3.18 (Hauer 1997):

$$\hat{\kappa}_{i,b_1} = \frac{\hat{b} + \sum_{y=1}^{Y} K_{i,y}}{\frac{\hat{b}}{E\{\kappa_{i,b_1}\}} + \sum_{y=1}^{Y} C_{i,y}} \tag{3.18}$$

and

$$VAR\{\hat{\kappa}_{i,b_1}\} = \frac{\hat{b} + \sum_{y=1}^{Y} K_{i,y}}{\left(\frac{\hat{b}}{E\{\kappa_{i,b_1}\}} + \sum_{y=1}^{Y} C_{i,y}\right)^2} = \frac{\kappa_{i,1}}{\frac{\hat{b}}{E\{\kappa_{i,b_1}\}} + \sum_{y=1}^{Y} C_{i,y}} \tag{3.19}$$

where $\hat{b}$ is a parameter estimated during the fitting of a regression model (the dispersion parameter). Just illustrated differently here for simplicity:

$\sum_{y=1}^{Y} K_{i,y}$ = sum of accident counts on the entity of interest 'i' during the before years '$b_1, b_2, \ldots, b_Y$'

$E\{\kappa_{i,b_1}\}$ = the estimate of the first year of safety at entity 'i" using the regression model.

The only difference between Equations 3.18 and 3.8 is the new parameter $C_{i,y}$. The concept of weight explained in the shorten EB procedure applies here also. However, the weight in the full EB procedure uses different expected accident frequencies for each year whereas the shortened version uses the same for each year. Simplifying Equation 3.18 will result in Equation 3.14 with the addition of the $C_{i,y}$ parameter

$$\hat{\kappa}_{i,b_1} = \frac{\hat{b} + \sum_{y=1}^{Y} K_{i,y}}{\frac{\hat{b}}{E\{\kappa_{i,b_1}\}} + \sum_{y=1}^{Y} C_{i,y}} = \frac{E\{\kappa_{i,b_1}\}\left(\hat{b} + \sum_{y=1}^{Y} K_{i,y}\right)}{\hat{b} + E\{\kappa_{i,b_1}\}\sum_{y=1}^{Y} C_{i,y}}$$

where $\dfrac{E\{\kappa_{i,b_1}\}\left(\hat{b}+\sum_{y=1}^{Y} K_{i,y}\right)}{\hat{b}+E\{\kappa_{i,b_1}\}}$ is exactly the same as Equation 3.8, only here we are taking into account $E\{\kappa_{i,y}\}$ and not simply $E\{\kappa_i\}$.

Knowing what the safety is in year '1' is only part of what is needed. The safety at the entity in later years b_2, b_3, ... b_Y and a_1, a_2 ..., a_Z is also needed. Manipulating Equation 3.17, one can estimate the safety of an entity each year after year '1'; that is,

$$\hat{\kappa}_{i,y_1} = \hat{\kappa}_{i,b_1} C_{i,y} \tag{3.20}$$

and

$$VAR\{\hat{\kappa}_{i,y_1}\} = VAR\{\hat{\kappa}_{i,b_1}\}(C_{i,y})^2 \tag{3.21}$$

In essence, Equation 3.17 is simply normalizing the regression estimate of $E\{\kappa_i\}$ from year 1 and thereby accounting for the non-systematic changes based on year 1 as the reference point. Therefore, using longer before and after periods does actually help to estimate safety more precisely since the shortened EB version does not account for these changes.

The first question posed earlier, that is, what would have been the safety of the entity in the after period had the treatment not been implemented, is now answered by the use of Equation 3.20.

So far, there has been discussion of the EB procedure for estimating the safety at an entity if a treatment was implemented. Another common method is the Naive approach which is discussed next. The Naïve method is usually not favoured by safety analysts at present but they can serve as a guide if accident data are not available to perform the EB procedure.

3.3 The Naïve Before-After Procedure

As in the EB method, we need to answer the same two fundamental questions in the Naive approach:

1. What would have been the safety of the entity in the after period had the treatment not been implemented, π?
2. What is the safety after the treatment was implemented, λ?

The basis of the naive method is that the before period accident counts, K_b, are used as an estimate of the expected accident in the after period, K_a, had the treatment not been implemented (Hauer 1997). For this method, another parameter must first be defined, namely the "ratio of duration" r_d:

$$r_d = \frac{\textit{After period duration at entity}}{\textit{Before period duratio at entity}} \tag{3.22}$$

The following formulae outline the approach of the Naive procedure:

$$\hat{\lambda} = K_a \tag{3.23}$$

$$VAR\{\hat{\lambda}\} = K_a \tag{3.24}$$

$$\hat{\pi} = r_d K_b \tag{3.25}$$

$$VAR\{\hat{\pi}\} = (r_d)^2 K_b \tag{3.26}$$

Accident counts are assumed to be Poisson distributed, implying that the mean and the variance are equal. With this assumption, K_a is equal to the mean and variance of the after period accident counts. More detail derivations of these equations are found in Hauer (1997).

Knowing λ and π of an entity, the safety effectiveness is found using either of the two methods discussed previously, namely:

1. Reduction in Expected Number of Crashes, δ
2. Index of Effectiveness, θ

The naive procedure does not account for changes in several factors such as traffic and accident trend, weather, driver's demography etc. So, in essence, the naive procedure simply assumes that had a treatment not been implemented, these factors will remain the same from the before to the after period. Obviously this is not the case. Based on several previous findings (Lord 2000; Hauer 1988; Persaud 2001; etc) traffic flow seems to always have a direct relationship with accidents. Accordingly, Hauer (1997) proposed a modification of the simple naive procedure whereby changes in traffic flow between the before and after periods are accounted for.

There are two ways in which traffic flows are accounted for in the naive procedure:

1. *linear or proportional assumption* - traffic flow and accident counts are considered to exhibit a proportional relationship with each other and

2. *Non-linear or non-proportional assumption* - traffic flow and accident counts are considered to follow a non-proportional relationship.

The same set of equations (Equations 3.22 through 3.26) is used in this modified naive approach. Traffic flow changes are accounted for by use of a variable introduced by Hauer (1997) referred to as the traffic flow correction factor, r_{tf}, and is defined as the ratio of the after period flow to the before period flow as shown in Equation 3.27:

$$r_{tf} = f\left(After\ flow\right) / f\left(Before\ flow\right) \qquad (3.27)$$

where:

f(after flow) = expected number of crashes with the after period flow (average after period if more than one year is used)

f(before flow) = expected number of crashes with the before period flow (average before period if more than one year is used).

Under the assumption that the traffic flow and expected accidents have a proportional relationship Equation 3.27 will result in Equation 3.28:

$$r_{tf} = \frac{Average\ after\ traffic\ flow, A_{average}}{Average\ before\ traffic\ flow, B_{average}} \qquad (3.28)$$

and

$$VAR\left\{r_{tf}\right\} = r_{tf}^{2}\left\{v^{2}\left(A_{average}\right) + v^{2}\left(B_{average}\right)\right\} \qquad (3.29)$$

where:

v = the coefficient of variation defined as the standard deviation of a variable divided by its mean.

When the relationship is not proportional (not linear), Equation 3.30 is used to find r_{tf}.

$$r_{tf} = \frac{f\left(Average\,after\,traffic\,flow, A_{average}\right)}{f\left(Average\,before\,traffic\,flow, B_{average}\right)} \tag{3.30}$$

The variance of r_{tf} is given by

$$VAR\left\{r_{tf}\right\} = r_{tf}^{\,2}\left\{\frac{c_A VAR\left(A_{average}\right)}{f^2\left(A_{average}\right)} + \frac{c_B VAR\left(B_{average}\right)}{f^2\left(B_{average}\right)}\right\} \tag{3.31}$$

where

c_A = derivative of the SPF with respect to the traffic flow at $A_{average}$

c_B = derivative of the SPF with respect to the traffic flow at $B_{average}$

Knowing what r_{tf} and $VAR(r_{tf})$ are, the modified form of Equation 3.25 is given by:

$$\hat{\pi} = r_d r_{tf} K_b \tag{3.32}$$

3.4 Accident Prediction Models

Development of safety performance functions (SPFs) is a critical component in road safety analysis, especially in empirical Bayes estimation. SPFs are simply statistical multivariate models developed from the theory of multivariate regression analysis, in which the dependent variable is the accident count and the independent variables are the traits at the entity. Traits refer to geometric, traffic and operational composition of an entity (lane width, traffic control, number of lanes, speed limit, traffic volumes etc,). This component of safety estimation lies in the domain of statistics in which SPFs are usually calibrated using generalized linear modeling (GLM). The rudiments of GLM are presented next.

The time honoured theory of linear multivariate regression analysis has been used extensively when one aims in developing some functional relationship, to relate a dependent variable to several independent variables or a single independent variable. The classic form of a multivariate model is given by Equation 3.33:

$$E\{y\} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{3.33}$$

where y is the dependent (outcome or response) variable, $x_1, x_2, \ldots, x_k$ are a set of independent (predictor or regressor) variables and β_0, $\beta_1, \beta_2, \ldots, \beta_k$ are unknown coefficients. However, there are situations where the relationship between the dependent and independent variables is not linear. The theory of non-linear regression analysis is the alternative choice with a typical model taking the following form:

$$y = \beta_1 e^{\beta_2 x} + \varepsilon \tag{3.34}$$

The reason Equation 3.34 is referred to as a non-linear model is that the parameter β_2 exhibits a non-linear form. Another approach in judging when a model is non-linear is when the partial derivatives are functions of the unknown parameters. Regardless of the model form used in the linear and non-linear multivariate regression analysis, the errors are always assumed to follow the normal distribution (Raymond et al 2002). With any regression analysis that follows the normal distribution the assumption of homoscedasticity is made. This assumption is one in which the variance of the error term is constant for each value of the independent variables.

The method of least squares or maximum likelihood (ML) is usually used to estimate the unknown parameters in Equations 3.33 and 3.34. The least squares approach has been very popular due to its simplicity; where estimates of the unknown parameters are achieved when the residual sum of squares (Equation 3.35) is minimized:

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3.35}$$

where y_i = observed dependent variable, $\hat{y}_i$= estimated dependent variable. The maximum likelihood approach on the other hand is a tad more mathematically inclined but is the only way one can estimate the unknown parameters for an SPF that takes the form of a GLM.

The first step in performing the maximum likelihood estimation is to have a likelihood function with a known probability distribution. The likelihood function is then the joint probability distribution of the observations. The normal distribution is mathematically defined as:

$$f(x)=\frac{1}{(2\pi)^{n/2}\sigma^n}e^{\frac{(x-\mu)}{(2\sigma^2)}} \tag{3.36}$$

where μ = mean of the observation and σ^2 = variance. From Equation 3.33, the mean μ = y since the equation is simply finding the expected value of y, (E{y}). The component of (x-μ) in Equation 3.36 is the error of the normal distribution (Raymond et al. 2002). Therefore Equation 3.33 will result in:

$$\varepsilon_i = y_i - \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \tag{3.37}$$

where i = 1, 2, …, n parameters and j = 1, 2, …, k coefficients for observation y_i.

Knowing that ε = (x-μ) in Equation 3.37, we can substitute into Equation 3.36:

$$f(x)=\frac{1}{(2\pi)^{n/2}\sigma^n}e^{\frac{(y_i-\beta_0-\sum_{j=1}^{k}\beta_j x_{ij})}{(2\sigma^2)}} \tag{3.38}$$

The likelihood function is then the joint probability distribution of Equation 3.38, giving rise to:

$$L(x,\beta,\sigma^2)=\prod_{i=1}^{n}\frac{1}{(2\pi)^{n/2}\sigma^n}e^{\frac{(y_i-\beta_0-\sum_{j=1}^{k}\beta_j x_{ij})}{(2\sigma^2)}} \tag{3.39}$$

From the likelihood function (Equation 3.39), values of β and σ^2 are iteratively substituted into the function until the maximum possible value attainable is achieved. The value of β that maximizes the likelihood function is in turn its estimate.

In the discussion so far, the assumption that the dependent variable follows the normal distribution was made. However, crashes do not follow the normal distribution, implying that the classic linear or non-linear regression, as commonly applied, cannot be used for developing SPFs. The distribution of crashes at an entity was found (Hauer 1997; Poch 1996; Lord 2005; etc) to follow the negative binomial distribution instead. When the error structure does not follow the normal distribution, a unique form of regression technique known as generalized linear models (GLMs) is adopted instead.

The use of GLM was developed so linear or non-linear regression models may be fitted by a very diverse set of distributions called the exponential family (Myers 2002, Dobson 2002; Myers 1990; and McCullah 1989). The exponential distributions are normal, binomial, Poisson, geometric, negative binomial, exponential, gamma, and inverse normal distributions. The choice of using the negative binomial distribution in modeling accidents has been shown to be the preferred approach in several sources (Hauer 1997; Hauer et al 1988; Poch 1996 etc).

The negative binomial distribution (NB) is defined mathematically as:

$$f\left(y_i, P\{x_i\}, \alpha\right) = \frac{\left(y_i + \alpha - 1\right)!}{\left(y_i\right)!\left(\alpha - 1\right)!} P\left(x_i\right)^{\alpha} \left[1 - P(x_i)\right]^{y_i} \tag{3.40}$$

where $P(x_i)$ = probability of x crashes at entity i; therefore $0 \leq P(x_i) \leq 1$

α = NB distribution parameter $\alpha > 0$; usually referred to as the dispersion parameter

y_i = number of accidents an entity is likely to experience; $y_i = 0, 1, 2 \ldots$

The mean is given by Equation 3.41:

$$E(y_i) = \mu_i = \frac{\alpha(1-P\{x_i\})}{P\{x_i\}} \tag{3.41}$$

and the variance is given by Equation 3.42:

$$VAR(y_i) = \mu_i\left(1+\frac{\mu_i}{\alpha}\right) \tag{3.42}$$

where the probability, $P\{x_i\}$, is assumed to follow the gamma distribution (Hauer 1997; Lord et al 2005; Miaou 1996 etc).

The GLM has three fundamental components (McCullah and Nelder 1989):

1. The *random component* – this component consists of the dependent variable with a specific error distribution. In our case the dependent variable is accident counts, E{accident} = μ, with the negative binomial error distribution
2. The *systematic component* – this component consists of the independent variables that will be used to develop a linear model. This model will serve as the predictor, η
3. The *link* component – this is a very critical component in GLM methodology because it links the random component to the systematic component. The link between μ and η is usually in the form of a function that depends on the error structure. For the negative binomial distribution, the link function is the $\log\left(\frac{\alpha+\mu_i}{\mu_i}\right)$.

The parameter estimation process in a GLM uses the maximum likelihood (ML) approach. In this study, since we know the distribution of the dependent variable, the ML method was adapted.

PROPERTY OF
RYERSON UNIVERSITY LIBRARY

The likelihood function for the negative binomial distribution (Equation 3.40) is given by Equation 3.43 and is solved by taking the log. Therefore the likelihood equation is

$$L(y_i) = \prod_{i=1}^{n} \frac{(y_i + \alpha - 1)!}{(y_i)!(\alpha - 1)!} P(x_i)^{\alpha} [1 - P(x_i)]^{y_i} \tag{3.43}$$

By taking the log of the likelihood function, Equation 3.43 is reduced in a much simpler form that can be easily solved. The log of the likelihood function is given by Equation 3.44:

$$\log(L) = \sum_{i=1}^{n} \left[\log(y_i + \alpha - 1)! - \ln y_i - \log(\alpha - 1)! + \alpha \log(P(x_i)) + y_i \log(\{1 - P(x_i)\}) \right] \tag{3.44}$$

The model parameters are found by iteratively substituting values for y_i, α and $P\{x_i\}$ in Equation 3.44 until the function is maximized.

3.4.1 Models Functional Form and Goodness-of-Fit-Measures

The functional forms of the models were developed after conducting exploratory analyses of the data. Use of the Integrate-differentiate (ID) method, developed by Hauer and Bamfo (1997), was the initial step in selecting a functional form. The procedure consisted of developing an Empirical Integral Function (EIF) by placing each potential variable (for each site) into a series of bins. The bins are sorted in an increasing order. The left boundary of a bin for a single site is located halfway between the current site and the preceding site. The right boundary of the bin is located halfway between the current site and the succeeding site. The height of the bin is the accident count on the site. The width of the bin is the difference of the right and left boundary at a site. The area of the bin (height x width) is the value of the EIF at that site. The EIF is then plotted and compared to some pre-established functions (power, gamma, polynomial etc) as shown in Figure 3.3.

If the EIF should have a similar shape to any of those in Figure 3.3, then the pertinent functional form is selected. This method has been applied in several studies (Lord 2000; Persaud et al. 2002) in selecting a functional form for their safety performance functions. Other functional forms (McGee et al. 2003; Lord 2002; Hauer 1988 etc) were also investigated in this project with the aim of using the best model form for the data.

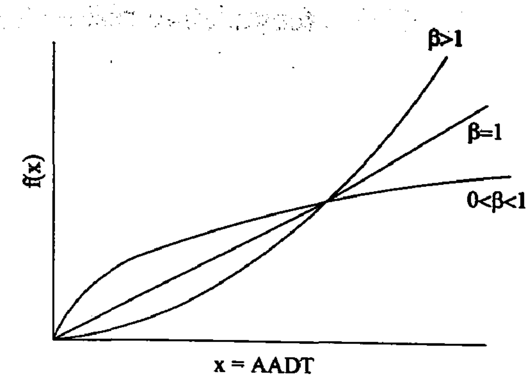
The functional form is by no means a simple selection from already built ones, even though those are helpful. In engineering practice, common sense plays a vital role. The developed model with its estimated parameters must adhere to common knowledge. For example if one were to develop an SPF for a roadway segment, the parameter estimated for the road length will have to be positive to adhere to common knowledge that a longer roadway segment is more likely to have more crashes than a shorter one. This type of reasoning was adopted in developing SPFs for this thesis.

Once parameters were estimated using the maximum likelihood approach, their statistical significance was checked. A 5% significance level was used to validate the statistical significance of each parameter in the models.

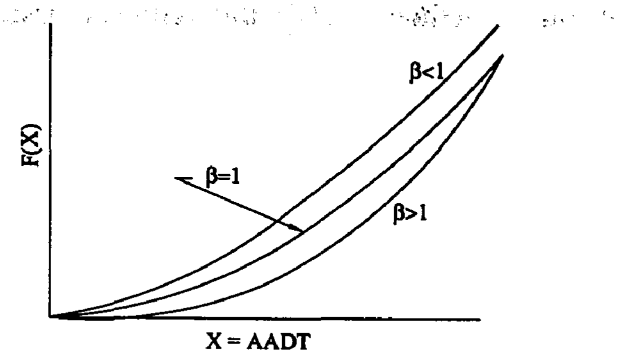
Two other statistics were used to evaluate the goodness-fit (GOF) statistics for models:

1. The Deviance/(n-p) – this measure is ratio of the deviance of the full model to the degrees of freedom, n-p. The value of this ratio serves as a gauge of the dispersion of the data.
2. The Pearson Chi-Statistic – this measure is the ratio of the Pearson Chi-Statistic to the degree of freedom.

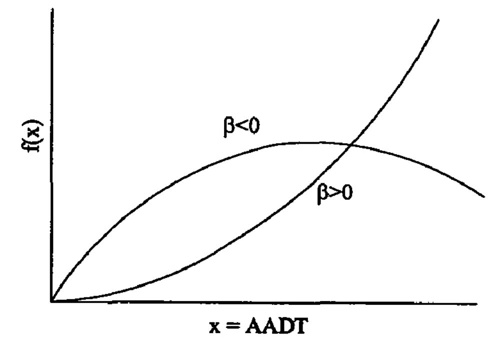
The underlying theory for these two statistics was developed by McCullah and Nelder (1989). In theory, both measures should be close to 1 if a model was to be considered adequate. However, safety analysts (Bauer and Harwood 1996) suggest that if the Pearson Chi-Statistic is somewhere in the range of 0.8 to 1.2, the models should still be considered as a well fitted model to the data.



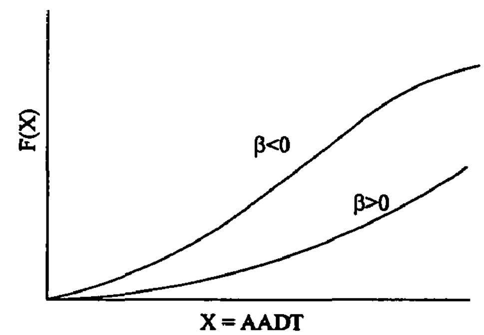
Power Function $f_1(x) = \alpha x^{\beta}$



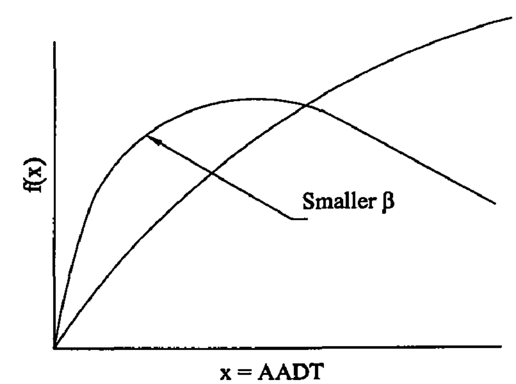
Power Function $F_1(X) = \left[\alpha/(\beta+1)\right] X^{\beta+1}$



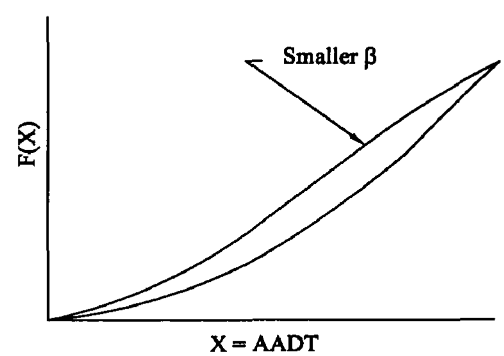
Polynomial Function $f_2(x) = \alpha x + \beta x^2$



Ploynomial Function $F_2(X) = (\alpha/2) X^2 + (\beta/3) X^3$



Gamma Function $f_3(x) = \alpha x e^{\beta x}$
This Gamma function is also the same as $\alpha x^{\beta_1} e^{\beta_2 x}$



Gamma Function $F_3(X) = \alpha \left[\dfrac{e^{\beta X} (\beta X - 1)}{\beta^2 + 1/\beta^2} \right]$

Figure 3.3 Plots of pre-established functions (Hauer and Bamfo, 1997)

Another method used in this report to assess GOF is the CURE (cumulative residuals) plot developed by Hauer and Bamfo (1997). The method consists of three plots:

1. The plot of the cumulative residuals (difference of the actual and predicted accident for each entity) against each variable separately. The cumulative residual is sorted in increasing order.
2. The $+2\sigma$ and -2σ plots. The parameter σ signifies the standard deviation

For a model to have a good fit under the CURE method, the plot in 1 (preceding point) should oscillate around the value of 0 and within the two standard deviation boundaries. For instance, Figures 3.4 and 3.5 illustrate the essence of the CURE plots. The cure plots developed in Figures 3.4 and 3.5 were from an SPF developed from California rural intersections (same data used in this thesis). The model form used for this illustration is $E\{\kappa\}=\alpha(\text{Major AADT})^{\beta1}(\text{Minor AADT})^{\beta2}$. Figure 3.4 illustrate the plot of the major AADT variable while Figure 3.5 illustrates the plot of the minor AADT variable.

Both plots depict a constant oscillation of the cumulative residuals around zero which was also bounded within the 2 standard deviations limits. Since the plot was bounded within the 2σ limits, the functional form and its estimated parameters are judged as being from a well fitted model. The main advantage of the CURE plot against conventional GOF measures (level of significance, Pearson Chi-Statistic etc) is that safety analysts can examine how well the model fits the independent variables along the entire range of its values. As in our case (Figure 3.4) it shows that the model does not estimate the expected accident frequency well when the major AADT is larger than 20,000. For the minor AADT variable, the model had the same problem of poor performance when the

AADT range is greater than 7,000. For a more detailed description of the method, see Hauer and Bamfo (1997).

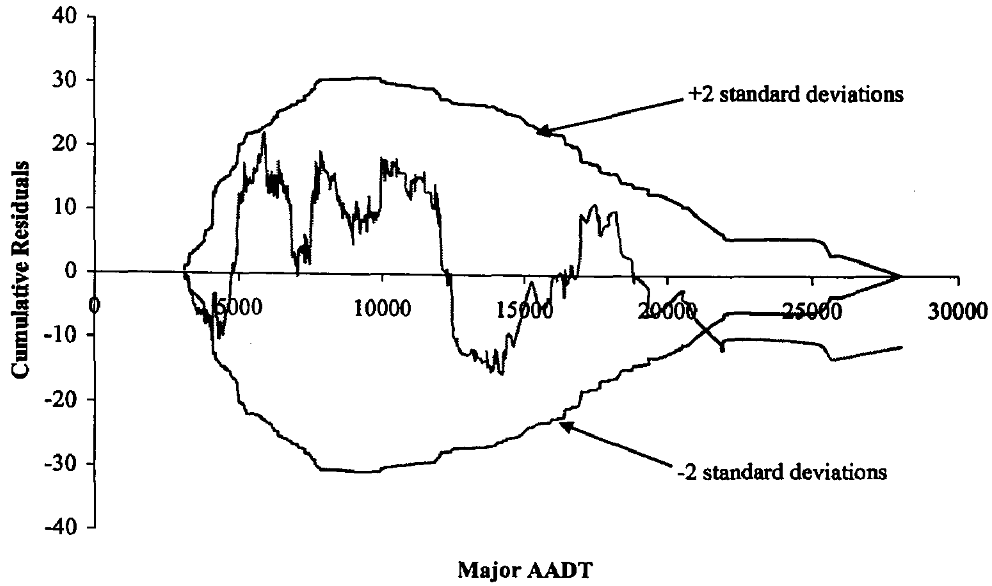


Figure 3.4 – Cumulative residuals for '4 leg with 2 lanes' stop controlled intersections (Major AADT as the parameter)

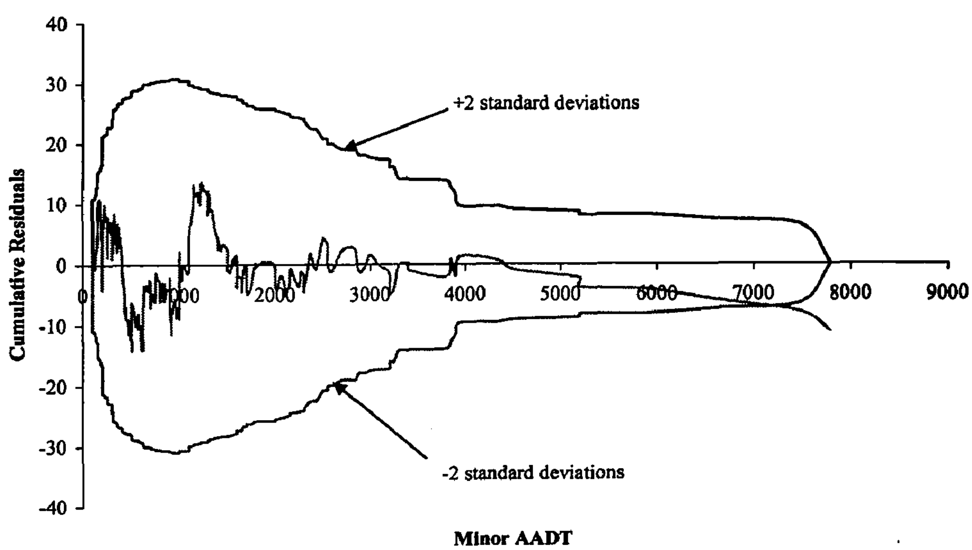


Figure 3.5 - Cumulative residuals for '4 leg with 2 lanes' stop controlled intersections (Minor AADT as the parameter)

CHAPTER 4: Database Used

The fundamental objective of this study was to quantify the safety effects of installing a traffic signal at rural intersections using the Empirical Bayes (EB) and conventional methodologies. To achieve this, appropriate treated sites were retrieved from the Highway Safety Information Systems (HSIS) database for two states, California and Minnesota.

4.1 California

There were 28 intersections with 10 years (1993 – 2002) of data, that were converted from stop to signal control in the California database and the treatment sites were disaggregated into three categories:

1. Three-legged with 2 lanes on the major approach (4 sites),
2. Four-legged with 2 lanes on the major approach (12 sites), and
3. Four-legged with 4 lanes on the major approach (10 sites).

This data is summarized in Tables 4.1, 4.2 and 4.3.

Table 4.1 – Summary of Treated Intersections with 3 legs and 2 lanes on major approach

Crash Types	Periods	All Severity			Injury Severity			PDO Severity		
		Mean	*Min*	*Max*	*Mean*	*Min*	*Max*	*Mean*	*Min*	*Max*
Total	Before	22.3	3.0	71.0	9.5	1.0	31.0	12.8	2.0	40.0
	After	22.8	3.0	65.0	7.0	0.0	21.0	15.5	2.0	44.0
Right Angle	Before	0.3	0.0	1.0	0.3	0.0	1.0	0.0	0.0	0.0
	After	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Left Turn	Before	13.8	1.0	50.0	6.5	0.0	24.0	7.3	1.0	26.0
	After	0.8	0.0	2.0	0.5	0.0	1.0	0.3	0.0	1.0
Rear End	Before	0.5	0.0	1.0	0.0	0.0	0.0	0.5	0.0	1.0
	After	0.8	0.0	2.0	0.5	0.0	1.0	0.3	0.0	1.0
		Mean			*Min*			*Max*		
Avg. Major AADT	Before	12,975			5,750			19,100		
	After	15,105			7,400			26,944		
Avg. Minor AADT	Before	5,613			201			10,300		
	After	5,640			201			10,300		
Number of Years	Before	3.5			1.0			5.0		
	After	5.5			4.0			8.0		

Table 4.2 – Summary of Treated Intersections with 4 legs and 2 lanes on major approach

Crash Types	Periods	All Severity			Injury Severity			PDO Severity		
		Mean	***Min***	***Max***	***Mean***	***Min***	***Max***	***Mean***	***Min***	***Max***
Total	Before	15.2	1.0	51.0	8.6	1.0	28.0	6.5	0.0	29.0
	After	13.2	2.0	34.0	4.2	1.0	9.0	9.0	1.0	26.0
Right Angle	Before	3.8	0.0	11.0	2.5	0.0	7.0	1.2	0.0	4.0
	After	0.9	0.0	3.0	0.3	0.0	2.0	0.6	0.0	3.0
Left Turn	Before	4.3	0.0	26.0	2.3	0.0	13.0	2.0	0.0	13.0
	After	3.1	0.0	10.0	0.8	0.0	5.0	2.3	0.0	8.0
Rear End	Before	0.9	0.0	3.0	0.3	0.0	2.0	0.6	0.0	3.0
	After	0.6	0.0	2.0	0.2	0.0	1.0	0.5	0.0	2.0
		Mean			***Min***			***Max***		
Avg. Major AADT	Before	10,344			7,400			18,738		
	After	11,204			7,763			21,700		
Avg. Minor AADT	Before	2,150			101			5,280		
	After	2,187			101			5,280		
Number of Years	Before	4.0			1.0			8.0		
	After	5.0			1.0			8.0		

Table 4.3 – Summary of Treated Intersections with 4 legs and 4 lanes on major approach

Crash Types	Periods	All Severity			Injury Severity			PDO Severity		
		Mean	***Min***	***Max***	***Mean***	***Min***	***Max***	***Mean***	***Min***	***Max***
Total	Before	18.2	4.0	38.0	9.0	2.0	20.0	9.2	2.0	28.0
	After	27.2	11.0	75.0	12.9	3.0	37.0	14.3	4.0	38.0
Right Angle	Before	5.2	0.0	14.0	3.1	0.0	9.0	2.1	0.0	8.0
	After	2.3	0.0	7.0	1.3	0.0	4.0	1.0	0.0	5.0
Left Turn	Before	5.1	0.0	13.0	3.0	0.0	9.0	2.1	0.0	5.0
	After	5.5	0.0	12.0	2.3	0.0	7.0	3.2	0.0	9.0
Rear End	Before	0.7	0.0	3.0	0.3	0.0	2.0	0.0	0.0	2.0
	After	2.1	0.0	8.0	1.4	0.0	5.0	1.0	0.0	3.0
		Mean			***Min***			***Max***		
Avg. Major AADT	Before	15,958			7,018			25,666		
	After	18,235			7,155			29,750		
Avg. Minor AADT	Before	2,716			600			9,700		
	After	2,791			600			9,646		
Number of Years	Before	3.4			1.0			6.0		
	After	5.7			3.0			8.0		

4.2 Minnesota

The Minnesota database consisted of 17 treated intersections with 12 years (1991 – 2002) of data and the sites were disaggregated into two categories:

1. Three legged intersections (2 sites), and
2. Four legged intersections (15 sites).

This data is summarized in Tables 4.4 and 4.5.

Table 4.4 – Summary of Treated Intersections with 3 legs

Crash Types	Periods	All Severity			Injury Severity			PDO Severity		
		Mean	*Min*	*Max*	*Mean*	*Min*	*Max*	*Mean*	*Min*	*Max*
Total	Before	30.0	0.0	60.0	16.0	0.0	32.0	14.0	0.0	28.0
	After	16.0	8.0	24.0	8.5	2.0	15.0	7.5	6.0	9.0
Right Angle	Before	9.0	0.0	18.0	4.0	0.0	8.0	5.0	0.0	10.0
	After	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Left Turn	Before	5.0	0.0	10.0	4.0	0.0	8.0	1.0	0.0	2.0
	After	1.0	0.0	2.0	1.0	0.0	2.0	0.0	0.0	0.0
Rear End	Before	8.5	0.0	17.0	5.0	0.0	10.0	3.5	0.0	7.0
	After	9.0	4.0	14.0	4.5	0.0	9.0	4.5	4.0	5.0
		Mean			*Min*			*Max*		
Avg. Major AADT	Before	18,361			18,223			18,498		
	After	17,278			17,065			17,491		
Avg. Minor AADT	Before	2,068			602			3,535		
	After	3,666			1,077			6,255		
Number of Years	Before	6.5			6.0			7.0		
	After	4.5			4.0			5.0		

Table 4.5 – Summary of Treated Intersections with 4 legs

Crash Types	Periods	All Severity			Injury Severity			PDO Severity		
		Mean	*Min*	*Max*	*Mean*	*Min*	*Max*	*Mean*	*Min*	*Max*
Total	Before	53.1	6.0	142.0	23.5	0.0	46.0	29.7	4.0	96.0
	After	60.7	6.0	202.0	24.7	0.0	88.0	36.1	6.0	114.0
Right Angle	Before	28.5	2.0	66.0	14.5	0.0	34.0	14.1	2.0	43.0
	After	16.2	2.0	62.0	8.5	0.0	30.0	7.7	0.0	32.0
Left Turn	Before	4.7	2.0	14.0	2.4	0.0	10.0	2.3	0.0	6.0
	After	7.3	0.0	62.0	2.7	0.0	28.0	4.7	0.0	34.0
Rear End	Before	9.8	0.0	30.0	4.2	0.0	14.0	5.6	0.0	22.0
	After	27.9	2.0	66.0	9.8	0.0	30.0	18.1	0.0	42.0
		Mean			*Min*			*Max*		
Avg. Major AADT	Before	13,739			3,261			29,926		
	After	17,614			3,327			38,179		
Avg. Minor AADT	Before	2,659			986			5,210		
	After	5,324			1,759			18,165		
Number of Years	Before	4.5			1.0			8.0		
	After	6.5			3.0			10.0		

The different crash types considered in this research were:

1. Total,
2. Right angle,
3. Left-turn, and
4. Rear-end.

The definition of the each crash types is illustrated in Figures 4.1 through 4.3. Other forms of rear-end crash definitions used in this thesis are illustrated in Appendix A.

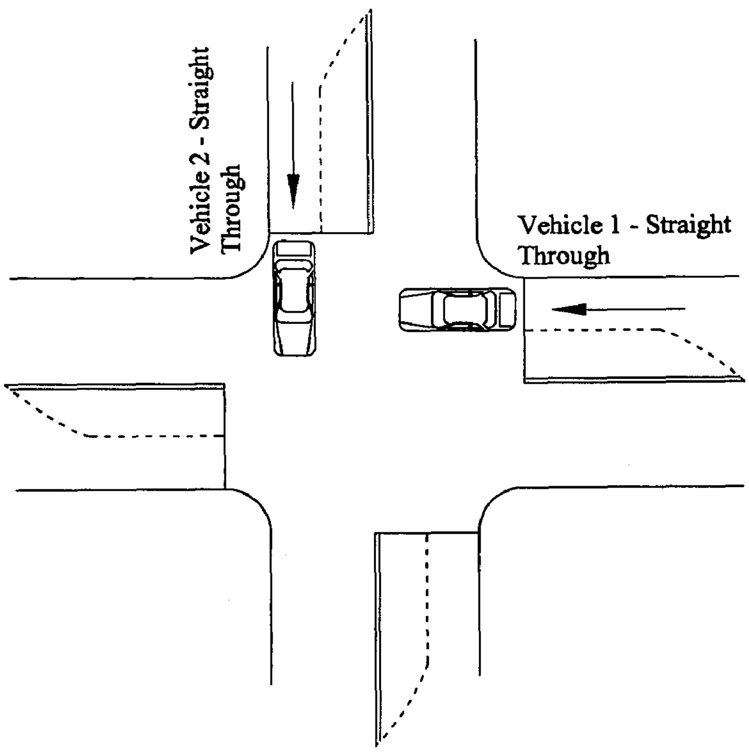


Figure 4.1 – Typical Right-Angle Cashes - Vehicles 1 and 2 both continue straight through the intersection after they were traveling perpendicular to each other

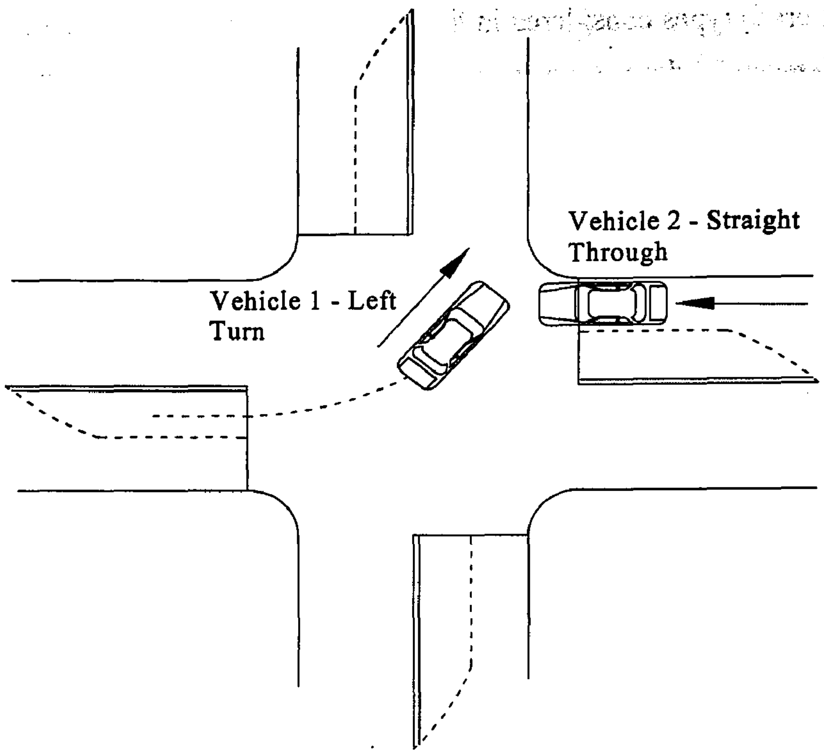


Figure 4.2 – Typical Left Turn Crashes - Vehicle 2 traveling straight through intersection while vehicle 1 turns left

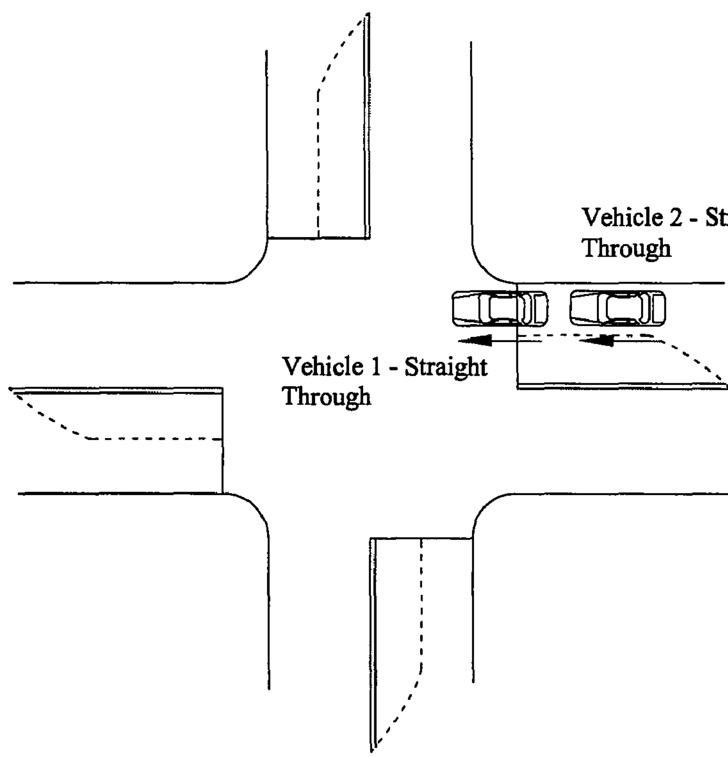


Figure 4.3 - Typical Rear End Crashes (Other forms of rear-end crashes illustrated in Appendix A)

The California and Minnesota database both followed the KABCO severity scale (Council et al. 2005). The KABCO severity scale is used by police officers at the scene of a crash to categorise injury severity as either K (killed), A (disabling injury), B (evident injury), C (possible injury) or O (no apparent injury). In this study, K, A, B and C severity were aggregated and referred to as injury severity due to very sparse K and A level severities in both data sets which resulted in poorly fitted regression models for the disaggregated severities.

4.3 Reference Populations

4.3.1 California

The following three reference populations were compiled for the treated intersections in the California database:

1. Three-legged with 2 lanes on the major approach (1405 sites),
2. Four-legged with 2 lanes on the major approach (726 sites), and
3. Four-legged with 4 lanes on the major approach (183 sites).

Tables 4.6, 4.7 and 4.8 summarize the basic information used in calibrating SPFs for the three reference groups.

Table 4.6 – Summary of Unconverted intersections with 3 legs and 2 lanes on the major approach

Crash Types	All Severity				Injury Severity				PDO Severity			
	Mean	*Min*	*Max*	*Sum*	*Mean*	*Min*	*Max*	*Sum*	*Mean*	*Min*	*Max*	*Sum*
Total	8.5	0	139	11,880	3.6	0	80	4,958	4.9	0	74	6,736
Right Angle	0.2	0	6	258	0.1	0	5	146	0.1	0	2	108
Left Turn	1.7	0	87	2,401	0.9	0	57	1,217	0.8	0	30	1,155
Rear End	0.6	0	10	6,736	0.3	0	5	362	0.4	0	5	509
	Mean				*Min*				*Max*			
Avg. Maj. AADT	9,019				2,950				31,450			
Avg. Min. AADT	554				100				10,001			

Table 4.7 – Summary of Unconverted intersections with 4 legs and 2 lanes on the major approach

Crash Types	All Severity				Injury Severity				PDO Severity			
	Mean	*Min*	*Max*	*Sum*	*Mean*	*Min*	*Max*	*Sum*	*Mean*	*Min*	*Max*	*Sum*
Total	13.9	0	94	10,314	6.4	0	49	4,685	75	0	46	5,476
Right Angle	3.3	0	47	2,428	2.0	0	29	1,427	1.3	0	19	976
Left Turn	2.3	0	31	1,719	1.2	0	21	895	1.1	0	17	809
Rear End	1.0	0	18	798	0.4	0	10	305	0.6	0	8	432
	Mean				*Min*				*Max*			
Avg. Maj. AADT	8,557				3,101				28,055			
Avg. Min. AADT	656				100				7,800			

Table 4.8 – Summary of Unconverted intersections with 4 legs and 4 lanes on the major approach

Crash Types	All Severity				Injury Severity				PDO Severity			
	Mean	*Min*	*Max*	*Sum*	*Mean*	*Min*	*Max*	*Sum*	*Mean*	*Min*	*Max*	*Sum*
Total	12.3	0	75	2,251	5.9	0	41	974	6.4	0	41	1,156
Right Angle	2.8	0	35	513	1.6	0	22	273	1.1	0	13	193
Left Turn	2.5	0	23	449	1.4	2	15	227	1.0	0	10	187
Rear End	1.0	0	8	176	0.4	0	4	63	0.6	0	6	105
	Mean				*Min*				*Max*			
Avg. Maj. AADT	12,441				3,087				30,500			
Avg. Min. AADT	12.3				0				75			

4.3.2 Minnesota

The following two reference populations were compiled for the treated intersections in the Minnesota database:

1. Three-legged intersections (522 sites), and
2. Four-legged intersections (736 sites).

These reference groups were not disaggregated by number of lanes on the major approach because they resulted in very small sample sizes which generated poorly fitted regression models. Tables 4.9 and 4.10 summarize the basic information used in calibrating SPFs for these two reference groups.

Table 4.9 – Summary of Unconverted Intersections with 3 legs

Crash Types	All Severity				Injury Severity				PDO Severity			
	Mean	*Min*	*Max*	*Sum*	*Mean*	*Min*	*Max*	*Sum*	*Mean*	*Min*	*Max*	*Sum*
Total	485	409	601	5,822	220	189	279	2,640	265	206	322	3,182
Right Angle	134	93	170	1,609	74	54	96	889	60	31	82	720
Left Turn	35	25	44	415	17	10	25	203	18	10	26	212
Rear End	136	109	200	1,633	60	38	73	723	76	59	134	910
	Mean				*Min*				*Max*			
Avg. Maj. AADT	6,710				1,165				32,645			
Avg. Min. AADT	989				196				12,750			

Table 4.10 – Summary of Unconverted Intersections with 4 legs

Crash Types	All Severity				Injury Severity				PDO Severity			
	Mean	*Min*	*Max*	*Sum*	*Mean*	*Min*	*Max*	*Sum*	*Mean*	*Min*	*Max*	*Sum*
Total	1,445	1,196	1,630	17,335	708	582	828	8,492	737	614	850	8,843
Right Angle	718	577	834	8,610	423	308	506	5,073	295	222	380	3,537
Left Turn	92	64	120	1,104	47	20	60	563	45	34	70	541
Rear End	281	219	364	3,375	123	76	149	1,472	159	96	215	1,903
	Mean				*Min*				*Max*			
Avg. Maj. AADT	5,538				1,173				31,074			
Avg. Min. AADT	902				194				18,774			

4.4 Exploratory Analyses of the Data

The reference populations for both California and Minnesota data were examined to reveal frequency distributions and possible functional forms before any form of regression analysis was attempted. Frequency plots were developed for total crashes after which the shape of the plots was used to attest possible probability distributions (Bauer and Hardwood 1998).

4.4.1 Frequency Distribution

4.4.1.1 California

The frequency distributions for the California reference groups were developed by plotting total number of crashes in the 10 year period against the number of corresponding intersections that experience such crashes. Resulting frequency distributions are shown in Figure 4.4, 4.5 and 4.6.

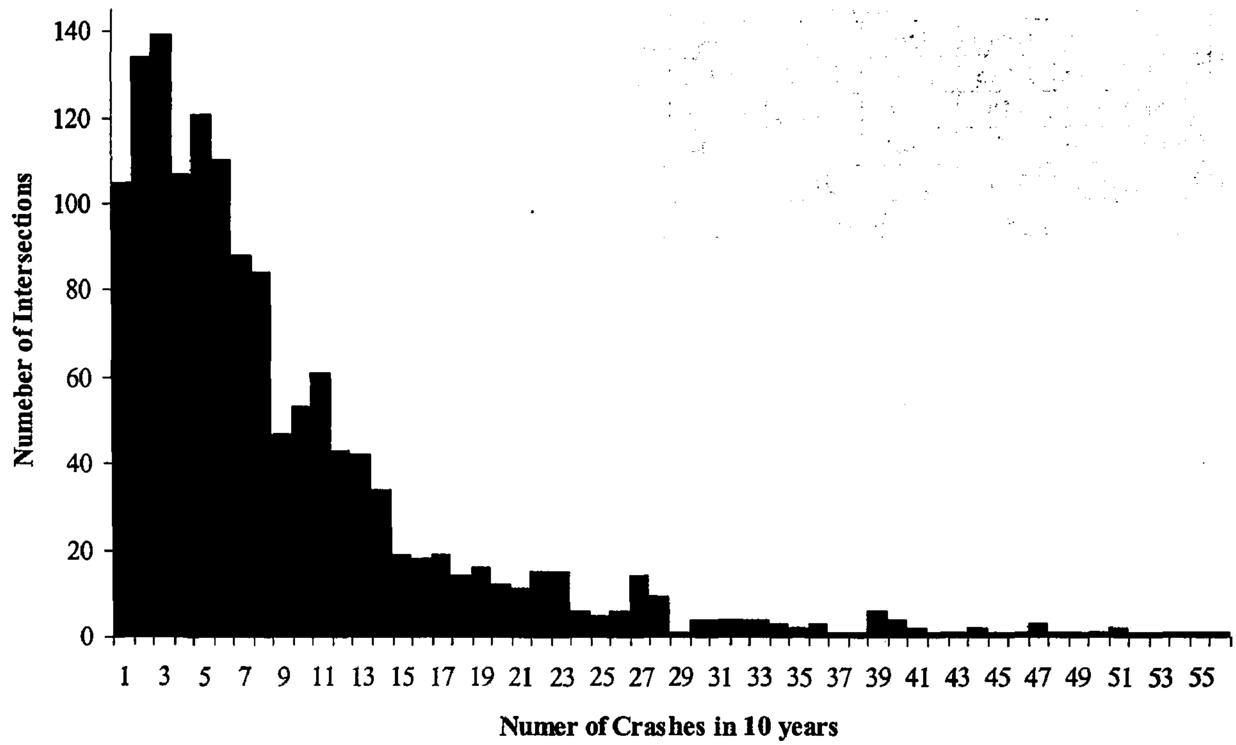


Figure 4.4 – Total Accident Frequency Distributions for three-legged intersections with 2 lanes on major approach

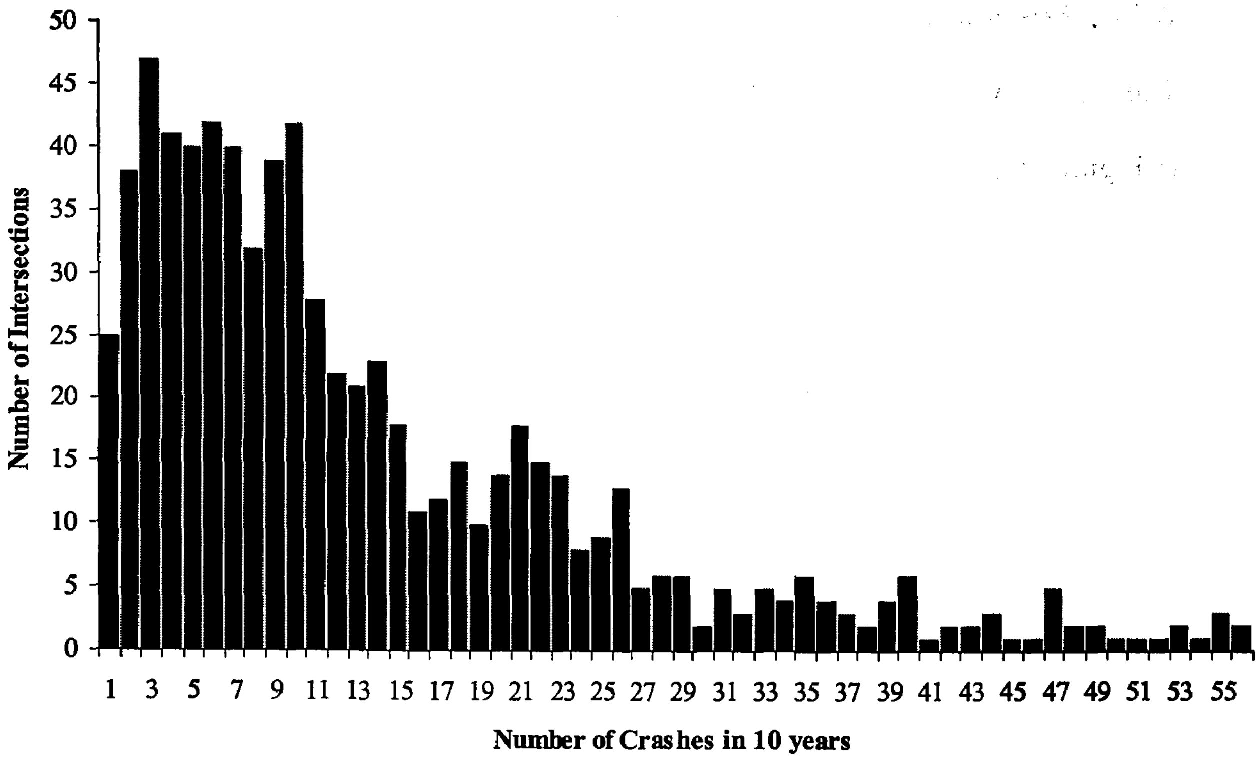


Figure 4.5 – Total Accident Frequency Distributions for four-legged intersections with 2 lanes on major approach

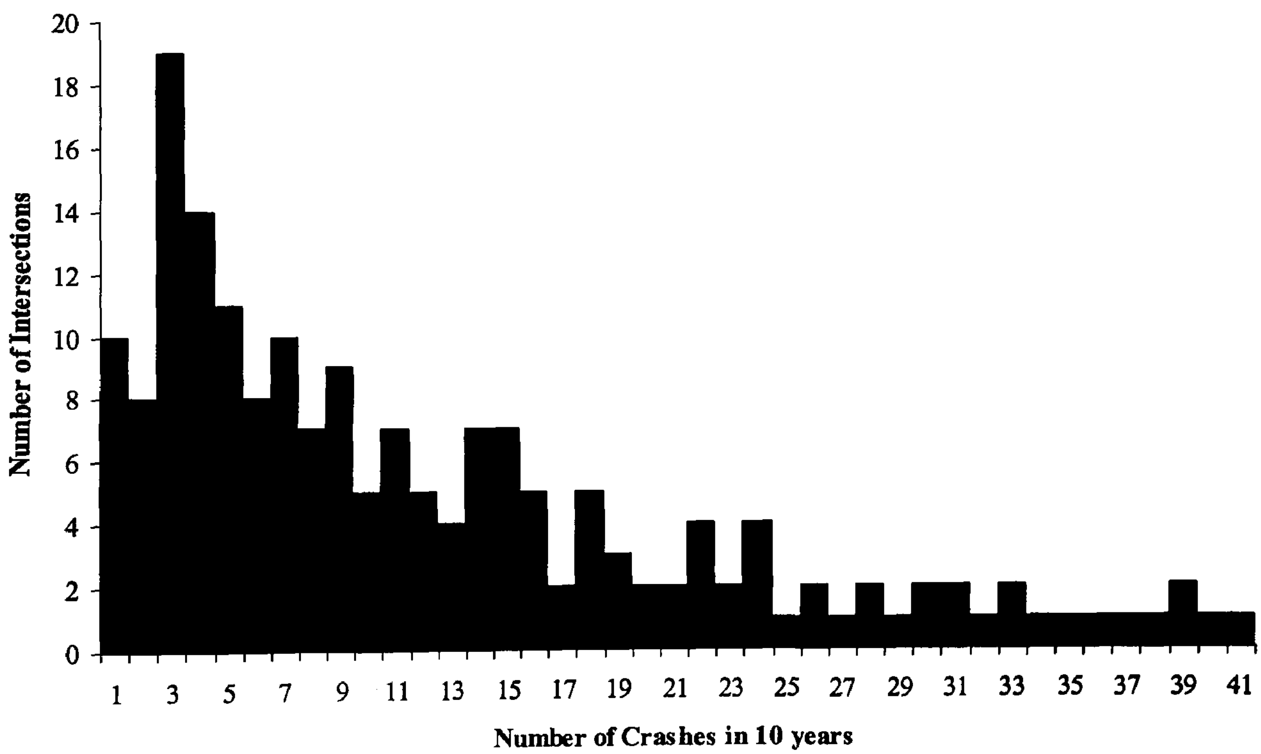


Figure 4.6– Total Accident Frequency Distributions for four-legged intersections with 4 lanes on major approach

4.4.1.2 Minnesota

The summary for the Minnesota frequency distributions using 12 years of data is shown in Figures 4.7 and 4.8.

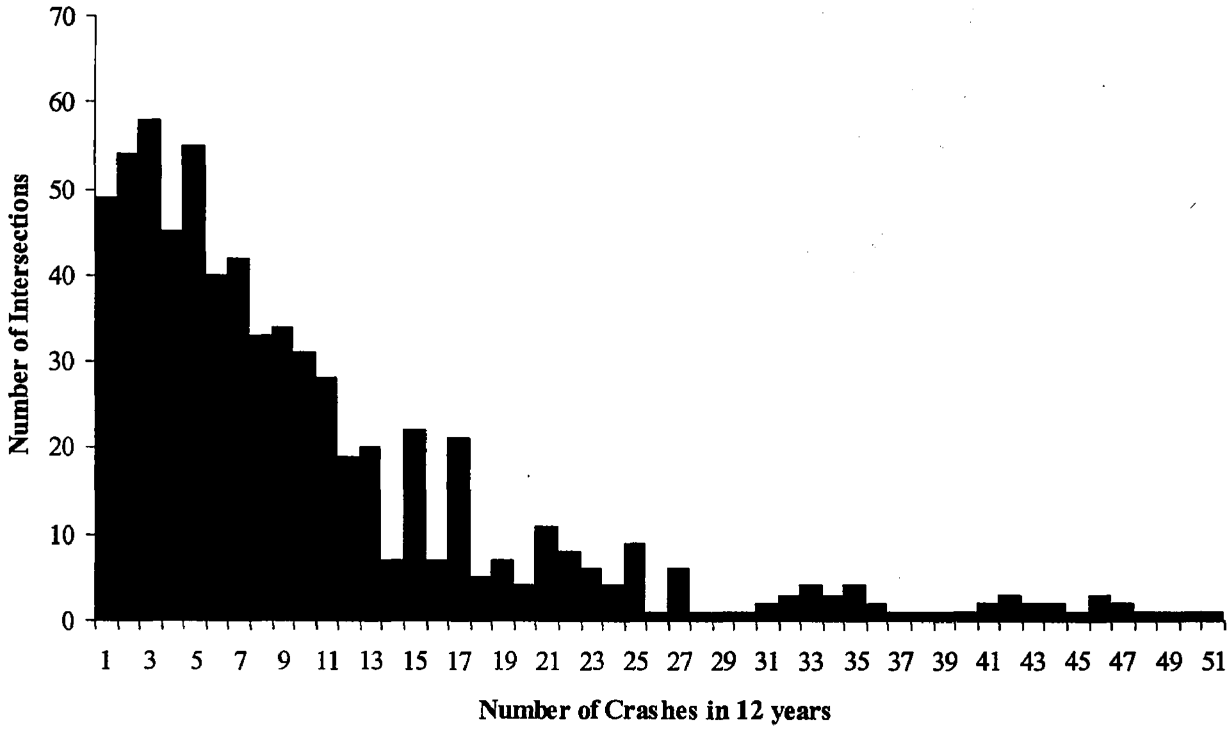


Figure 4.7 – Total Accident Frequency Distributions for 3 legged intersections

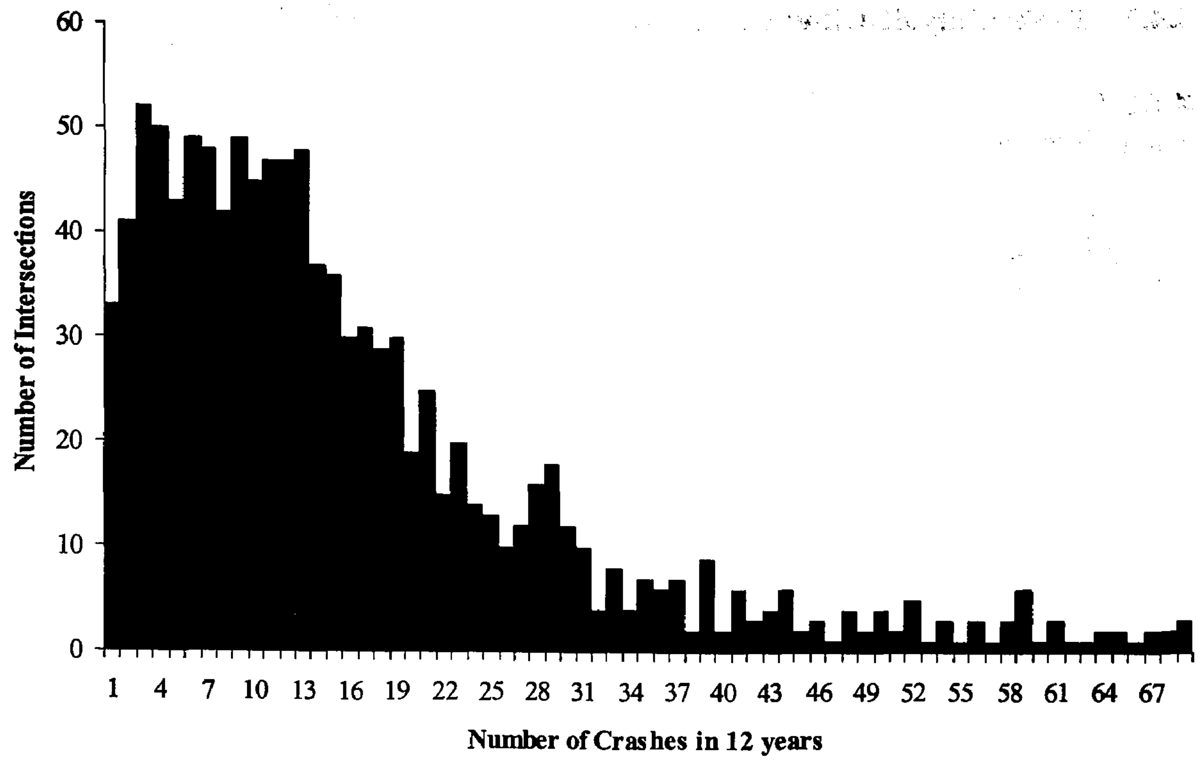


Figure 4.8 – Total Accident Frequency Distributions for 4 legged intersections

The shape of the frequency distributions in Figures 4.4 through 4.8 exhibits a left skewed distribution, indicative of a Poisson or negative binomial distribution, according to Bauer and Harwood (1998). Validating this assumption is only possible if the probability distribution is fitted to the data (reference population). For instance, using California's three-legged intersection reference group (Figure 4.4), it was assumed that the crashes follows both a Poisson and a negative binomial distribution and the probability distribution was estimated for illustration purposes.

4.4.2 Probability distribution

4.4.2.1 Assuming the Poisson distribution

The Poisson distribution is expressed mathematically as:

$$f(y_i, \mu_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \qquad (4.1)$$

where y_i = number of crashes an entity is likely to experience; $y_i = 0, 1, 2,$

μ_i = expected number of crashes at entity 'i' in a given time interval

The mean and variance of a Poisson distribution are same; that is $E(y_i) = \mu_i = VAR(y_i)$. If the expected accident is computed, the probability y_i of having a crash at entity 'i' can be estimated. Table 4.11 (Column 5) shows the distribution of the crashes in the reference population under the Poisson distribution. Column 6 of Table 4.11 gives the number of sites that would likely have the number of crashes specified in Column 2. If the reference population followed the Poisson distribution, the quantities in Columns 1 and 6 should be similar (relatively close) and in this case it is not true. The difference between the actual number of sites (Column 1) and the estimated number of sites (Column 6) is quite large, which implies that the crash counts in this particular reference group did not follow a Poisson distribution.

Another means of judging if a specific probability distribution is adequate for the data is to compare the plot of the specified probability distribution against the plot of frequency distribution. Figure 4.9 shows the Poisson probability distribution of the crashes in the reference population for three-legged intersections with two lanes on the major approach. The shape of Figure 4.4 (frequency distribution of the reference group) is not even close to that in Figure 4.9 (Poisson probability distribution of crashes). Therefore it would be safe to assume that the Poisson distribution is not adequate in describing the outcomes (crashes) for this reference population.

4.4.2.2 Assuming the Negative Binomial distribution

The negative binomial distribution is defined mathematically according to Equation 3.40 in Chapter 3. As in the preceding case, if the expected accident and variance is computed, the probability y_i of having a crash at entity 'i' can be estimated. Table 4.11 illustrates the calculations based on the negative binomial distribution. Clearly one can see that the NB distribution is a more reliable probability distribution of the accident counts since the calculated numbers of intersections (sites) in Column 3 are very close to the numbers in the actual data (Column 1). As with the Poisson case, comparing the frequency plot (Figure 4.4) to the probability distribution plot can also serve as an indicator of adequacy for the assumed probability distribution.

Comparing the shape of Figures 4.4 and 4.10, a distinct similarity emerges. This implies that the NB distribution is quite reasonable for describing the probability of a crash in this reference population.

Table 4.11 – Summary of Poisson and Negative Binomial Distribution

1	2	3	4	5	6
No. of Sites	Accident Counts, K	P(K), Under NB distribution	n(K) if all 1266 sites have gamma distributed means	P(K) under Poison Distribution	n(K) if all 1266 sites have identical means
170	0	0.113	143	0.002	2
169	1	0.112	142	0.012	15
155	2	0.103	131	0.036	46
93	3	0.092	117	0.077	97
79	4	0.081	103	0.121	153
69	5	0.071	90	0.152	192
59	6	0.061	78	0.160	202
57	7	0.053	67	0.144	182
51	8	0.046	58	0.113	143
54	9	0.039	50	0.079	100
43	10	0.034	43	0.050	63
35	11	0.029	36	0.029	36
25	12	0.025	31	0.015	19
30	13	0.021	27	0.007	9
23	14	0.018	23	0.003	4
15	15	0.015	19	0.001	2
14	16	0.013	16	0.001	1
20	17	0.011	14	0.000	0
12	18	0.009	12	0.000	0
15	19	0.008	10	0.000	0
18	20	0.007	9	0.000	0
15	21	0.006	7	0.000	0
13	22	0.005	6	0.000	0
7	23	0.004	5	0.000	0
9	24	0.004	4	0.000	0
12	25	0.003	4	0.000	0
4	26	0.003	3	0.000	0

For the Negative Binomial Distribution

Giving $\mu = 6.250$ *and variance,* $VAR = 39.90$*; equations 44 and 45 is equal to*

$$6.250 = \frac{\alpha(1-P\{x_i\})}{P\{x_i\}} \qquad and \qquad 39.90 = \mu\left(1+\frac{\mu}{\alpha}\right)$$

$\because$ *knowing what* μ *is we have* $39.90 = 6.250\left(1+\frac{6.250}{\alpha}\right) \Rightarrow 5.385\alpha = 6.250; \therefore \alpha = 6.250/5.385 = 1.1606.$

Thus, substituting α *into equation 44 we can estimate the probability of a success*

$$6.250 = \frac{1.1606(1-P\{x_i\})}{P\{x_i\}} \Rightarrow P\{x_i\} = 1/6.6006 = 0.1515$$

Therefore, the probability of having zero crashes $(y_i = 0)$ *at site 'i' is*

$$f(0, 0.1515, 1.1606) = \begin{pmatrix} 0+1.1606-1 \\ 1.1606-1 \end{pmatrix} 0.1515^{1.1606}[1-0.1515]^0 = \begin{pmatrix} 0.1606 \\ 0.1606 \end{pmatrix} 0.1515^{1.1606}[1] = 0.1515^{1.1606} = 0.112$$

For the Poisson Distribution

the mean and variane = 6.250; therefore

$$f(0; 6.250) = \frac{e^{-6.250}\,6.250^0}{0!} = 0.00193$$

The sample calculation in Table 4.11 illustrates the computation of Row 1 in Columns 3 and 5 with the answers not being exactly the same as Row 1 due to rounding off errors.

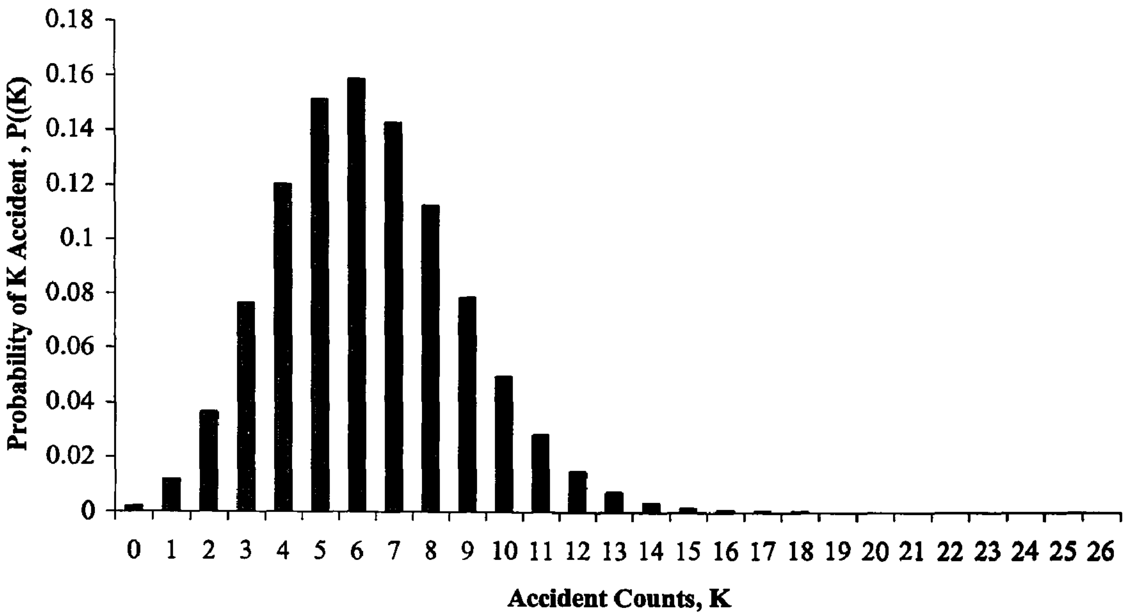


Figure 4.9 – Plot of the Poisson distribution for California's three-legged intersections

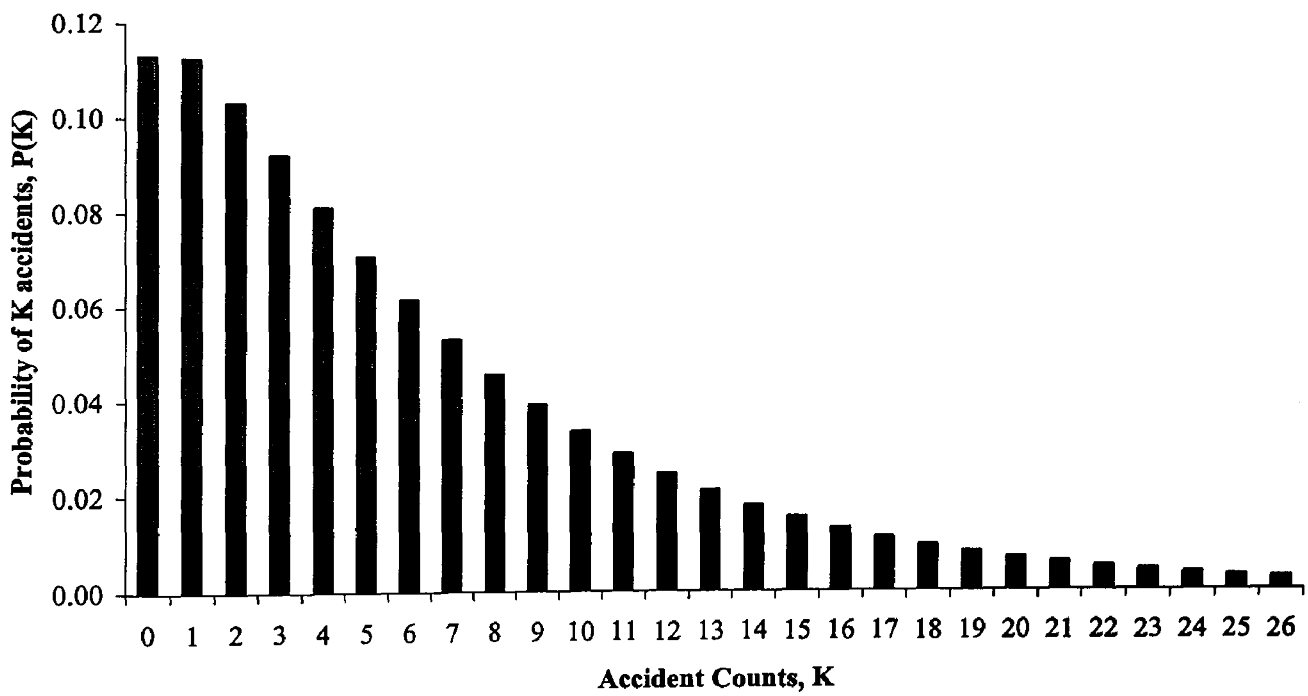


Figure 4.10 – Plot of the NB distribution for California's three-legged intersections.

4.4.3 Possible Functional forms of Regression Models using the ID method

After determining the probability distribution of the accident counts for the reference population, the next step in the exploratory analyses was regression analysis using the GLM procedures. The SAS GENMOD (SAS V8) procedure was used to develop models with various functional forms and the Integrate Differentiate (ID) method developed by Hauer and Bamfo (1997) was used to explore potential functional forms. Also, models developed from previous research (McGee et al. 2003; Lord 2002; Hauer 1988 etc) were also considered during this stage. Using the ID method, the EIF plots for all five of the reference groups (3 for California and 2 for Minnesota) were explored for various functional forms. For instance, using the California's four-legged intersection with 2 lanes on the major approach, the EIF plots were developed using major and minor AADT as the independent variables (Figures 4.11 and 4.12).

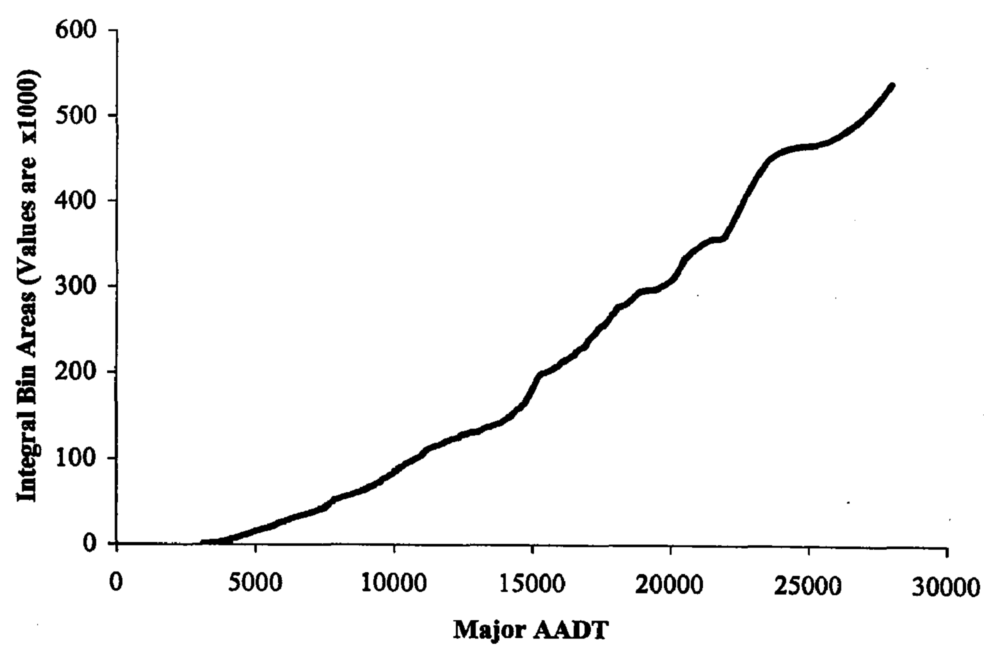


Figure 4.11 – EIF plot for Major AADT

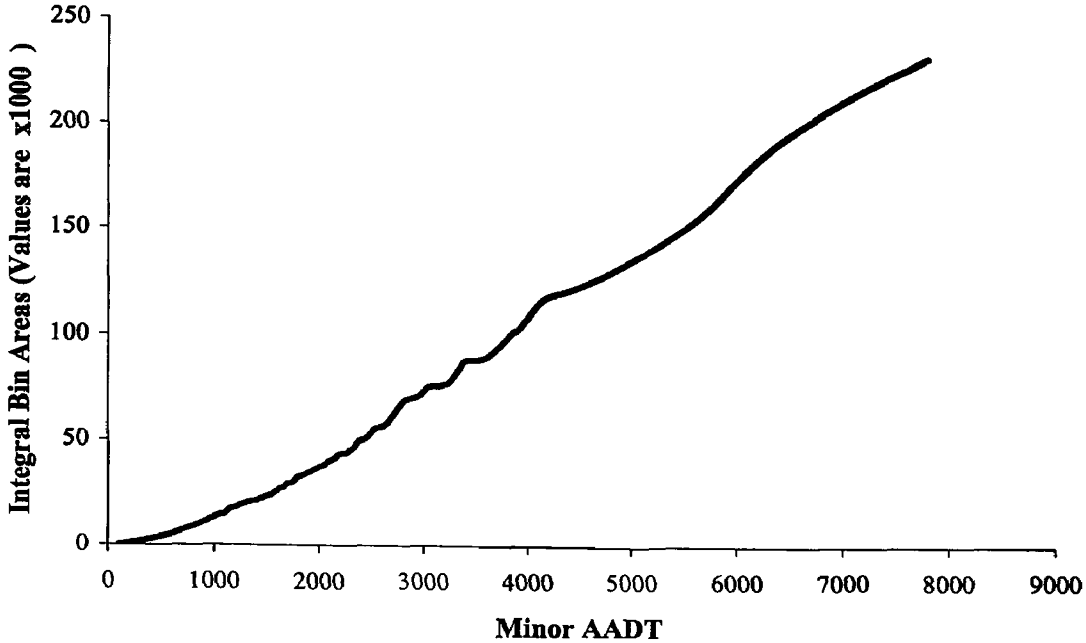


Figure 4.12 - EIF plot for Minor AADT

After the EIF plots were developed, they were compared to pre-established plots for specific functional forms, such as those in Figure 3.3. Because of the similarity of Figures 4.11 and 4.12 to the power function (left hand side) of Figure 3.3, the functional form selected is shown in Equation 4.2.

$$Accident\,/\,year = \alpha\left(Major\ AADT\right)^{\beta_1}\left(Minor\ AADT\right)^{\beta_2} \tag{4.2}$$

To summarize, the following two exploratory analyses were performed on each data set:

1. Frequency distributions were developed for each crash types, and
2. EIF plots on each data set were developed to select an initial functional form.

After performing these two exploratory analyses, the SPFs can then be calibrated for each data set and this calibration process is presented in the next chapter.

CHAPTER 5: Calibrated Safety Performance Functions

SPFs were developed using the SAS GENMOD (SAS V8) procedure for which the negative binomial error distribution was assumed. Explanatory variables considered included AADT, number of lanes, lane width, turning lanes (left or right), lighting, etc. However, after several attempts with various logical variables, only the AADT set of variables (Major and Minor AADT) was found to be statistically significant, which resulted in SPFs consisting of only AADT variables. Even so, opting to use AADT alone was not a simple task because several functional forms (with AADT alone) were possible and CURE plots proposed by Hauer and Bamfo (1997), which were introduced in Chapter 3 , had to be used to decide on the best form.

5.1 California SPFs

SPFs were developed for all severity levels (Total, Injury and PDO) and for each crash type defined in Chapter 4. Therefore, with three severity levels and four crash types (all, right-angle, left turn and rear end) 12 models were calibrated for each reference group, resulting in 36 models developed for the three California reference groups. Tables 5.1, 5.2, and 5.3 summarize the SPFs for the "3-legged intersections with 2 lanes on the major approach" reference group. Summaries for the SPFs of "4-legged intersections with 2 lanes on the major approach" are in Tables 5.4, 5.5 and 5.6, while Tables 5.7, 5.8 and 5.9 summarize the SPFs for "4-legged intersections with 4 lanes on the major approach".

Table 5.1 – Total SPFs for California Three-legged intersections with 2 lanes on Major Approach

	All	Right-Angle	Left-Turn	Rear-End
Model	2	2	2	1
LN(α) **(s.e.)**	-9.321 (0.408)	-10.660 (1.418)	-13.672 (0.716)	-11.446 (0.776)
β_1 **(s.e.)**	1.113 (0.046)	0.907 (0.156)	1.483 (0.080)	0.824 (0.084)
β_2 **(s.e.)**	0.325 (0.023)	0.523 (0.071)	0.571 (0.037)	0.212 (0.040)
Dispersion	1.771	0.455	0.839	1.253
χ^2/DOF	1.096	1.272	1.665	1.060
Dev./DOF	1.118	0.447	0.953	0.872

Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$ **Model 2:** $E\{\kappa\} = \alpha(Maj + Min)^{\beta_1}\left(\frac{Min}{Maj + Min}\right)^{\beta_2}$

Table 5.2 – Injury SPFs for California Three-legged Intersections with 2 lanes on Major Approach

	All	Right-Angle	Left-Turn	Rear-End
Model	2	2	2	1
LN(α) **(s.e.)**	-10.09 (0.468)	-11.304 (1.860)	-14.906 (0.875)	-10.748 (1.079)
β_1 **(s.e.)**	1.108 (0.053)	0.918 (0.206)	1.5450 (0.0969)	0.637 (0.116)
β_2 **(s.e.)**	0.351 (0.025)	0.525 (0.092)	0.589 (0.042)	0.228 (0.056)
Dispersion	1.749	0.297	0.772	0.994
χ^2/DOF	1.104	0.31	0.813	0.635
Dev./DOF	1.104	1.22	1.4920	0.982

Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$ **Model 2:** $E\{\kappa\} = \alpha(Maj + Min)^{\beta_1}\left(\frac{Min}{Maj + Min}\right)^{\beta_2}$

Table 5.3 – PDO SPFs for California Three-legged Intersections with 2 lanes on Major Approach

	All	Right-Angle	Left-Turn	Rear-End
Model	2	2	2	2
LN(α) **(s.e.)**	-10.027 (0.456)	-11.125 (0.782)	-13.853 (0.8679)	-13.194 (0.966)
β_1 **(s.e.)**	1.122 (0.051)	0.8751 (0.031)	1.426 (0.095)	1.119 (0.106)
β_2 **(s.e.)**	0.305 (0.024)	0.5573 (0.041)	0.573 (0.044)	0.114 (0.049)
Dispersion	1.651	1.395	0.760	1.227
χ^2/DOF	1.125	0.91	0.806	0.727
Dev./DOF	1.117	1.12	1.4058	1.090

Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$ **Model 2:** $E\{\kappa\} = \alpha(Maj + Min)^{\beta_1}\left(\frac{Min}{Maj + Min}\right)^{\beta_2}$

Table 5.4 – Total SPFs for California Four-legged Intersections with 2 lanes on Major Approach

	All	Right-Angle	Left-Turn	Rear-End
Model	1	1	1	2
LN(α) **(s.e.)**	-9.148 (0.554)	-11.029 (1.224)	-13.290 (0.918)	-10.870 (1.030)
β_1 **(s.e.)**	0.719 (0.060)	0.554 (0.129)	0.821 (0.095)	1.023 (0.114)
β_2 **(s.e.)**	0.481 (0.028)	0.781 (0.062)	0.763 (0.046)	0.252 (0.052)
Dispersion	2.070	0.515	1.099	1.255
χ^2/DOF	1.108	0.997	1.246	1.115
Dev./DOF	1.067	0.998	1.036	0.984

Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$ **Model 2:** $E\{\kappa\} = \alpha(Maj+Min)^{\beta_1}\left(\frac{Min}{Maj+Min}\right)^{\beta_2}$

Table 5.5 – Injury SPFs for California Four-legged Intersections with 2 lanes on Major Approach

	All	Right-Angle	Left-Turn	Rear-End
Model	1	1	2	2
LN(α) **(s.e.)**	-9.996 (0.660)	-11.696 (1.389)	-13.749 (1.098)	-11.336 (1.458)
β_1 **(s.e.)**	0.724 (0.070)	0.603 (0.146)	1.453 (0.123)	0.988 (0.161)
β_2 **(s.e.)**	0.480 (0.033)	0.738 (0.070)	0.572 (0.055)	0.285 (0.161)
Dispersion	1.723	0.453	0.995	0.921
χ^2/DOF	1.143	0.905	0.940	0.745
Dev./DOF	1.053	1.026	0.055	1.020

Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$ **Model 2:** $E\{\kappa\} = \alpha(Maj+Min)^{\beta_1}\left(\frac{Min}{Maj+Min}\right)^{\beta_2}$

Table 5.6 PDO SPFs for California Four-legged Intersections with 2 lanes on Major Approach

	All	Right-Angle	Left-Turn	Rear-End
Model	2	1	1	1
LN(α) **(s.e.)**	-10.014 (0.600)	-11.119 (1.333)	-14.182 (1.131)	-11.565 (1.178)
β_1 **(s.e.)**	1.198 (0.067)	0.416 (0.140)	0.812 (0.116)	0.763 (0.06)
β_2 **(s.e.)**	0.413 (0.030)	0.845 (0.067)	0.733 (0.055)	0.301 (0.01)
Dispersion	2.109	0.563	1.036	1.335
χ^2/DOF	1.108	0.858	0.901	0.873
Dev./DOF	1.066	1.024	1.124	1.060

Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$ **Model 2:** $E\{\kappa\} = \alpha(Maj+Min)^{\beta_1}\left(\frac{Min}{Maj+Min}\right)^{\beta_2}$

Table 5.7 – Total SPFs for California Four-legged Intersections with 4 lanes on Major Approach

	All	Right-Angle	Left-Turn	Rear-End
Model	1	2	2	1
LN(α) **(s.e.)**	-9.650 (1.191)	-12.790 (2.146)	-16.669 (2.302)	-12.758 (1.970)
β_1 **(s.e.)**	0.769 (0.116)	1.454 (0.247)	1.802 (0.265)	0.969 (0.197)
β_2 **(s.e.)**	0.426 (0.069)	0.707 (0.120)	0.559 (0.125)	0.211 (0.103)
Dispersion	1.551	0.614	0.656	1.749
χ^2/DOF	1.222	1.014	1.273	1.000
Dev./DOF	1.136	1.019	0.996	0.989

Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$ **Model 2:** $E\{\kappa\} = \alpha(Maj+Min)^{\beta_1}\left(\frac{Min}{Maj+Min}\right)^{\beta_2}$

Table 5.8 – Injury SPFs for California Four-legged Intersections with 4 lanes on Major Approach

	All	Right-Angle	Left-Turn	Rear-End
Model	2	1	2	3
LN(α) **(s.e.)**	-12.648 (1.555)	-14.506 (2.356)	-18.804 (3.148)	-14.250 (3.147)
β_1 **(s.e.)**	1.413 (0.179)	0.794 (0.223)	1.997 (0.365)	1.159 (0.161)
β_2 **(s.e.)**	0.457 (0.090)	0.833 (0.140)	0.643 (0.172)	
Dispersion	1.100	0.602	0.415	0.878
χ^2/DOF	1.150	0.918	0.809	0.697
Dev./DOF	1.187	1.047	1.157	0.969

Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$ **Model 2:** $E\{\kappa\} = \alpha(Maj+Min)^{\beta_1}\left(\frac{Min}{Maj+Min}\right)^{\beta_2}$

Model 3: $E\{\kappa\} = \alpha(Maj+Min)^{\beta_1}$

Table 5.9 PDO SPFs for California Four-legged Intersections with 4 lanes on Major Approach

	All	Right-Angle	Left-Turn	Rear-End
Model	2	1	1	1
LN(α) **(s.e.)**	-9.145 (1.230)	-10.466 (2.249)	-13.655 (2.226)	-12.803 (2.268)
β_1 **(s.e.)**	1.034 (0.140)	0.664 (0.216)	0.854 (0.214)	0.908 (0.225)
β_2 **(s.e.)**	0.325 (0.068)	0.433 (0.131)	0.540 (0.124)	0.230 (0.111)
Dispersion	1.850	0.713	1.023	2.013
χ^2/DOF	1.162	0.908	0.935	0.900
Dev./DOF	1.209	0.940	1.069	1.018

Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$ **Model 2:** $E\{\kappa\} = \alpha(Maj+Min)^{\beta_1}\left(\frac{Min}{Maj+Min}\right)^{\beta_2}$

5.2 Minnesota SPFs

The SPFs for Minnesota were developed using a similar approach as for the California SPFs in terms of severity levels (Total, Injury and PDO), and for each crash types defined in Chapter 4. Therefore, with three severity levels and four crash types (all, right-angle, left turn and rear end) 12 models were calibrated for each reference group. However, the Minnesota dataset consists of only 2 reference groups, implying that 24 models were developed in total. Tables 5.10, 5.11 and 5.12 summarize the SPFs for the "3-legged intersections" reference group, while Tables 5.13, 5.14 and 5.15 summarize the SPFs for the "4-legged intersections" reference group.

Table 5.10 – Total SPFs for Minnesota Three-legged Intersections

	All	Right-Angle	Left-Turn	Rear-End
Model	1	1	1	1
LN(α) **(s.e.)**	-8.699 (0.524)	-13.249 (0.891)	-11.955 (1.421)	-11.482 (0.976)
β_1 **(s.e.)**	0.496 (0.050)	0.722 (0.083)	0.292 (0.127)	0.7602 (0.092)
β_2 **(s.e.)**	0.624 (0.050)	0.816 (0.083)	0.983 (0.127)	0.521 (0.090)
Dispersion	1.760	0.725	0.332	0.597
χ^2/DOF	1.177	1.056	0.682	1.042
Dev./DOF	0.889	0.873	0.876	0.949
Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$				

Table 5.11 – Injury SPFs for Minnesota Three-legged Intersections

	All	Right-Angle	Left-Turn	Rear-End
Model	1	1	2	1
LN(α) **(s.e.)**	-8.841 (0.626)	-13.689 (1.117)	-12.557 (1.927)	-12.020 (1.179)
β_1 **(s.e.)**	0.488 (0.060)	0.779 (0.106)	1.247 (0.242)	0.742 (0.111)
β_2 **(s.e.)**	0.553 (0.061)	0.720 (0.102)	0.948 (0.190)	0.501 (0.113)
Dispersion	1.265	0.457	0.207	0.439
χ^2/DOF	1.190	0.875	0.466	0.856
Dev./DOF	0.901	0.844	0.880	0.962
Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$			**Model 2:** $E\{\kappa\} = \alpha(Maj+Min)^{\beta_1}\left(\frac{Min}{Maj+Min}\right)^{\beta_2}$	

Table 5.12 – PDO SPFs for Minnesota Three-legged Intersections

	All	Right-Angle	Left-Turn	Rear-End
Model	1	1	2	1
LN(α) **(s.e.)**	-9.695 (0.624)	-14.223 (1.183)	-12.356 (1.966)	-12.074 (1.146)
β_1 **(s.e.)**	0.503 (0.058)	0.637 (0.108)	1.218 (0.247)	0.751 (0.107)
β_2 **(s.e.)**	0.687 (0.058)	0.949 (0.112)	0.901 (0.195)	0.535 (0.101)
Dispersion	1.437	0.489	0.172	0.496
χ^2/DOF	1.200	0.845	0.444	0.919
Dev./DOF	0.869	0.885	0.839	0.896

Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$ **Model 2:** $E\{\kappa\} = \alpha(Maj+Min)^{\beta_1}\left(\frac{Min}{Maj+Min}\right)^{\beta_2}$

Table 5.13 – TOTAL SPFs for Minnesota Four-legged Intersections

	All	Right-Angle	Left-Turn	Rear-End
Model	1	1	1	1
LN(α) **(s.e.)**	-8.850 (0.362)	-9.961 (0.518)	-13.814 (0.886)	-12.577 (0.616)
β_1 **(s.e.)**	0.5661 (0.039)	0.462 (0.054)	0.725 (0.091)	0.909 (0.0671)
β_2 **(s.e.)**	0.698 (0.041)	0.891 (0.060)	0.815 (0.096)	0.559 (0.065)
Dispersion	2.040	1.021	0.481	0.852
χ^2/DOF	1.140	1.186	0.839	1.103
Dev./DOF	1.0695	0.935	1.360	1.300

Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$

Table 5.14 – Injury SPFs for Minnesota Four-legged Intersections

	All	Right-Angle	Left-Turn	Rear-End
Model	1	1	1	1
LN(α) **(s.e.)**	-8.840 (0.423)	-9.811 (0.575)	-13.052 (1.149)	-12.442 (0.799)
β_1 **(s.e.)**	0.484 (0.045)	0.414 (0.060)	0.563 (0.124)	0.889 (0.090)
β_2 **(s.e.)**	0.650 (0.049)	0.854 (0.068)	0.809 (0.128)	0.446 (0.086)
Dispersion	1.536	0.841	0.263	0.519
χ^2/DOF	1.183	1.160	0.607	0.919
Dev./DOF	1.005	0.932	1.017	0.982

Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$

Table 5.15 – PDO SPFs for Minnesota Four-legged Intersections

	All	Right-Angle	Left-Turn	Rear-End
Model	1	1	1	1
LN(α) **(s.e.)**	-10.189 (0.405)	-11.786 (0.627)	-16.140 (1.290)	-13.792 (0.732)
β_1 **(s.e.)**	0.644 (0.043)	0.534 (0.064)	0.892 (0.124)	0.911 (0.077)
β_2 **(s.e.)**	0.695 (0.045)	0.934 (0.070)	0.838 (0.135)	0.647 (0.647)
Dispersion	1.784	0.819	0.272	0.677
χ^2/DOF	1.179	1.115	0.577	0.985
Dev./DOF	1.128	0.891	1.285	1.254
Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$				

5.3 Goodness-of-Fit of the SPFs

During the calibration of the SPFs, the goodness-of-fit measures discussed in Chapter 3 were all assessed to ensure the models' adequacy. For most of the SPFs the Pearson Chi-Statistic divided by its degree of freedom (χ^2/DOF) and the deviance divided its degree of freedom (Dev/DOF) for each of the models were close to 1, indicating good fit to their respective data (MaCullagh and Nelder 1989) or within the range of 0.8 - 1.2 deemed acceptable by Bauer and Harwood (1996). However, this ratio was not within the acceptable range for a few models simply because the sample sizes for those crash types were relatively small. Another method of assessing the goodness-of-fit of the SPFs is by the dispersion parameter. The variance of the expected accident frequency at any particular site in a reference group is computed using the following equation:

$$VAR\{\kappa\} = \frac{(E\{\kappa\})^2}{b} \tag{5.1}$$

where parameter 'b' is the inverse of the dispersion parameter from the SAS GENMOD procedure. Therefore, the larger the dispersion parameter is, the smaller will be the variance of the expected accident frequency and hence a better model. Using this goodness of fit criterion, models with the largest possible dispersion parameter (equation 5.1) were selected for this study.

The CURE plots for each model were also developed to supplement the decision in selecting the best 'fitting' SPF. Figures 5.1 to 5.6 are the CURE plots of total crashes for all severitie for California. The complete set of CURE plots developed for all SPFs is shown in Appendix B.

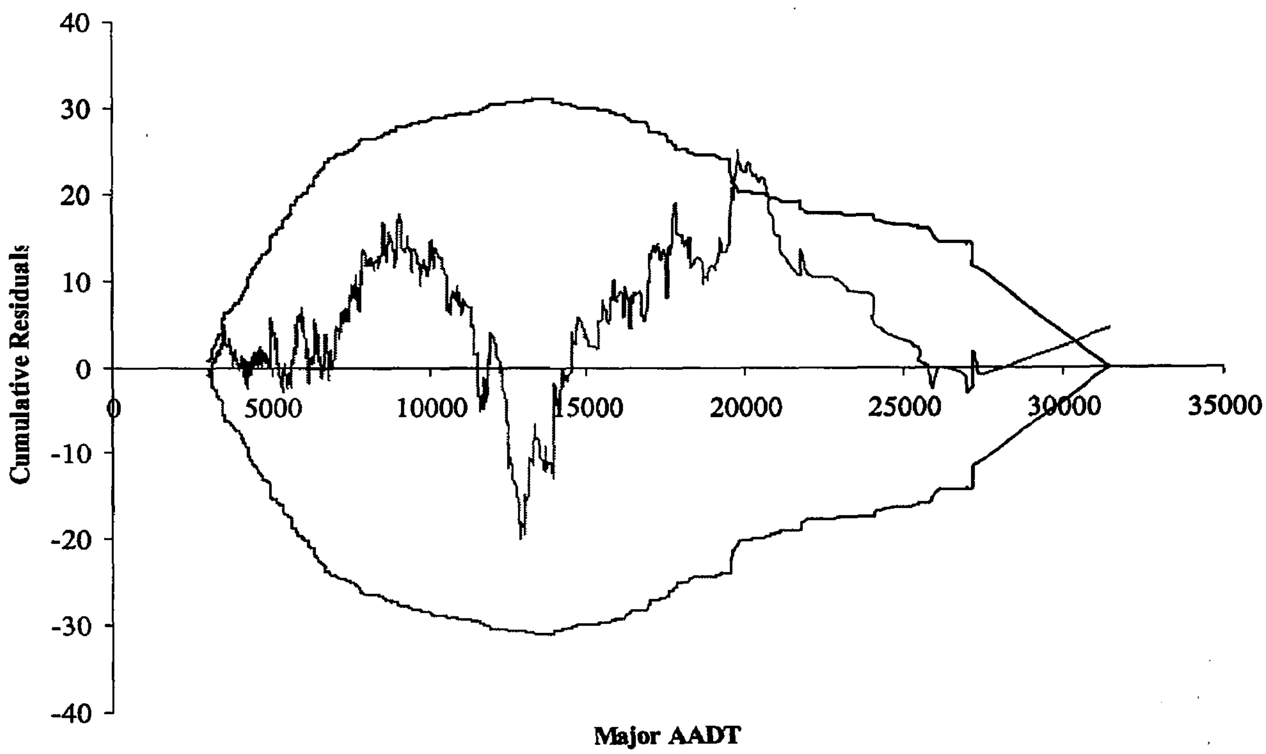


Figure 5.1 – California: CURE plot for '3-legged Intersections with 2 lanes on Major Approach' (Major AADT as the parameter)

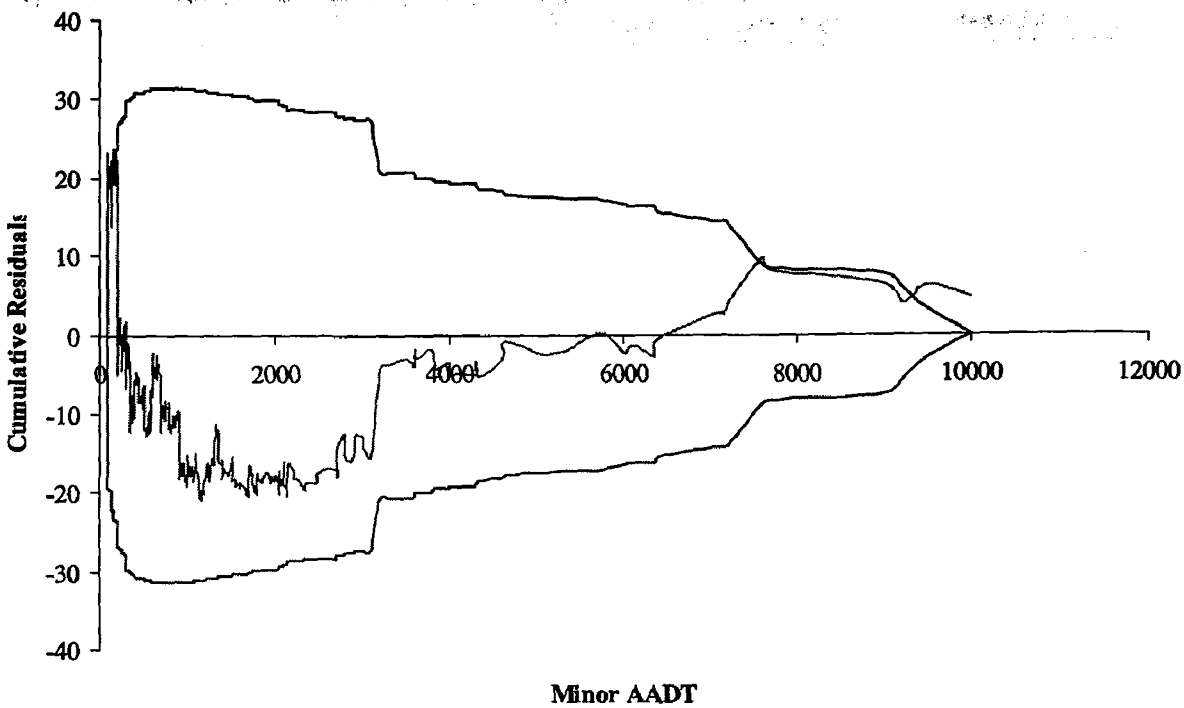


Figure 5.2 – California: CURE plot for '3-legged Intersections with 2 lanes on Major Approach' (Minor AADT as the parameter)

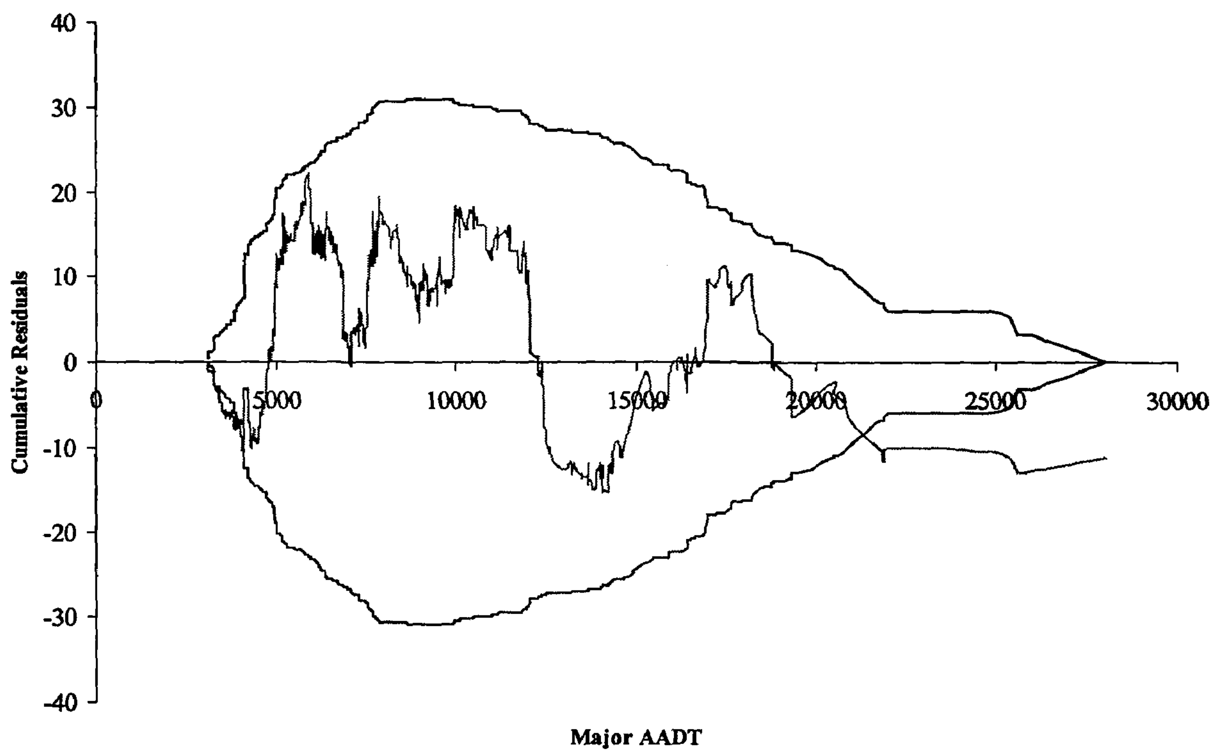


Figure 5.3 – California: CURE plot for '4-legged Intersections with 2 lanes on Major Approach' (Major AADT as the parameter)

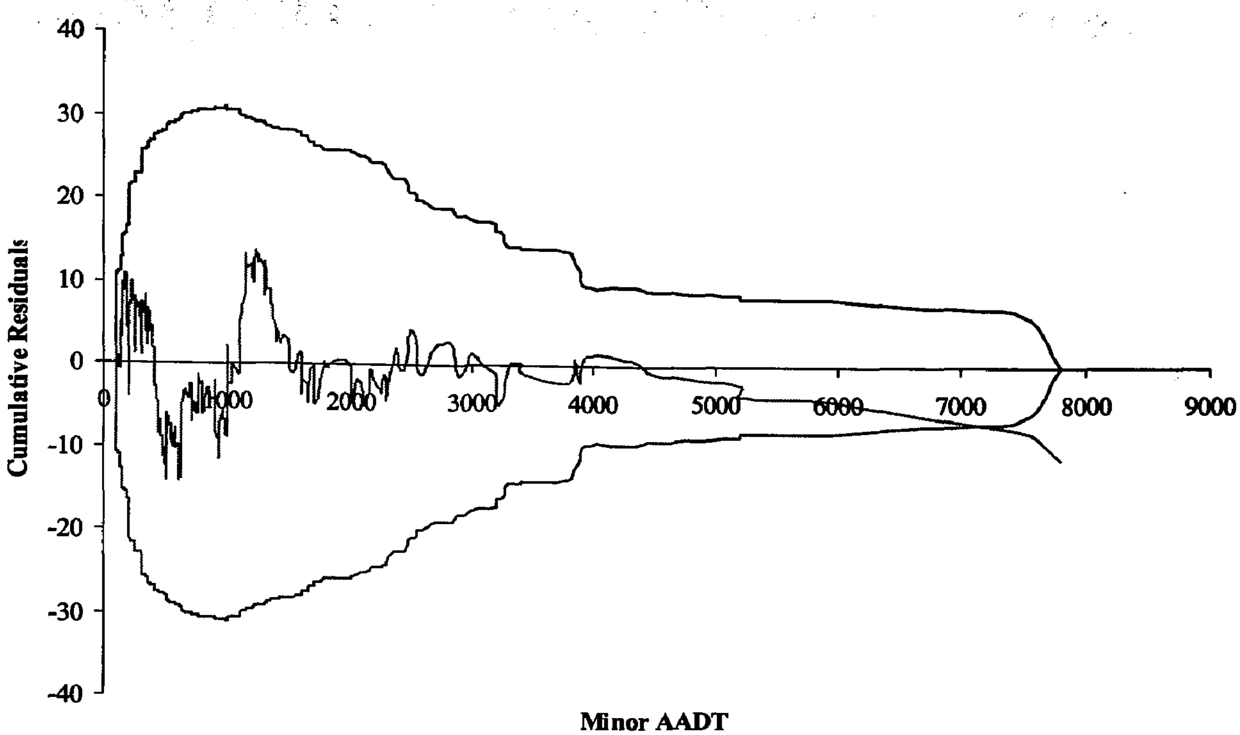


Figure 5.4 - California: CURE plot for '4-legged Intersections with 2 lanes on Major Approach' (Minor AADT as the parameter)

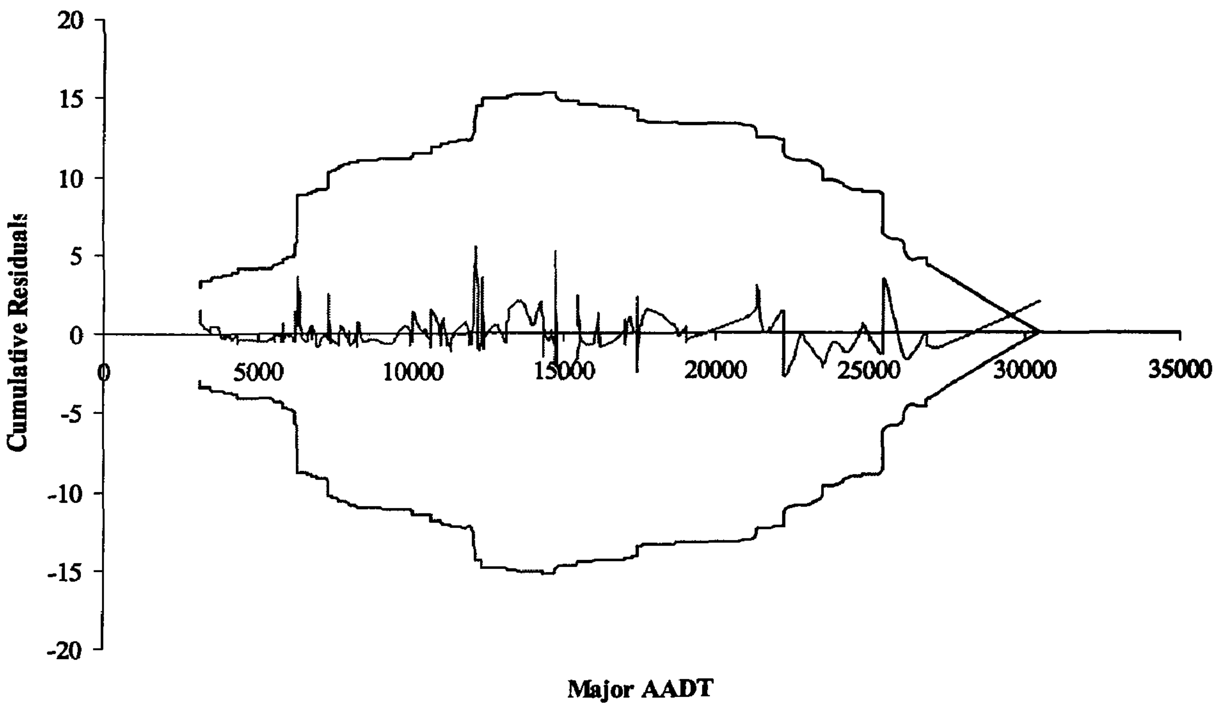


Figure 5.5 – California: CURE plot for '4-legged Intersections with 4 lanes on Major Approach' (Major AADT as the parameter)

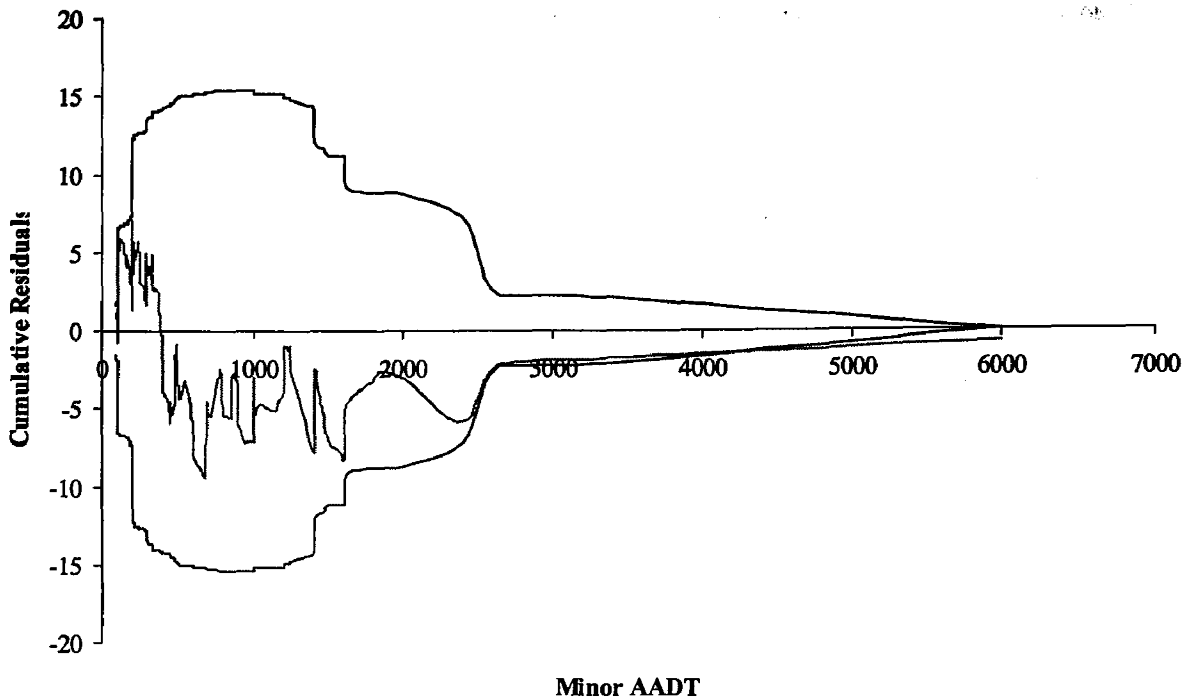


Figure 5.6 – California: CURE plot for '4-legged Intersections with 4 lanes on Major Approach' (Minor AADT as the parameter)

The models for total crashes in the California reference populations seem to have a good fit of the data. The CURE plots for each of the independent variables (major and minor AADT) seem to oscillate around zero (0) and finish around zero (0). This implies (Hauer and Bamfo 1997) that the model using the major and minor AADT as independent variables fit the data well. Cumulative residuals plots did cross the minus 2 standard deviation boundaries in Figures 5.4 and 5.6, indicating some inadequacy in the model to estimate at large major and minor AADTs. However, there could be a simpler explanation than inadequacy of the model; there are only 8 sites out of 743 with major AADT larger than 20,000 and only 2 sites with minor AADT larger than 7000, and, therefore, the problem may lie with the small sample size for these AADT ranges.

These calibrated SPFs were then used in the safety estimation procedures which are presented next.

CHAPTER 6: Safety Effects Estimation

The safety effects of installing traffic signals at rural intersections were quantified by using both the conventional and EB procedures with the treated intersections considered as a composite entity (Hauer 1997). The concept of a composite entity is quite simple. In Chapter 3 two fundamental questions needed to be answered before one can quantify the safety effects of a particular treatment:

1. What is the safety of the entity in the after period had the treatment ***not*** been implemented , π? and
2. What is the safety after the treatment was implemented λ?

The safety effect estimation methodologies introduced in Chapter 3 (conventional and EB) used π and λ to only quantifying safety effects at a single entity. However, more than one treated entity is usually used before inferences can be made on a particular treatment. Therefore, in the composite entity, π and λ are aggregated over the entire set treated sites, that is:

$$\lambda_{sum} = \sum \lambda_1, \lambda_2, \cdots \lambda_n \tag{6.1}$$

$$\pi_{sum} = \sum \pi_1, \pi_2, \cdots \pi_n \tag{6.2}$$

where the number of sites in the composite entity is n. The variances of λ_{sum} and π_{sum} in the composite entity are estimated similarly:

$$VAR(\lambda_{sum}) = \sum VAR(\lambda_1), VAR(\lambda_2), \cdots VAR(\lambda_n) \tag{6.3}$$

$$VAR(\pi_{sum}) = \sum VAR(\pi_1), VAR(\pi_2), \cdots, VAR(\pi_n) \tag{6.4}$$

The rudiments of both conventional the EB methodology remains unchanged when using a composite entity, because the composite entity uses the end products of these methods, that is π and λ.

Therefore, using treated intersections from California and Minnesota as one composite entity, the following safety effects were quantified using both the conventional and EB methodology.

6.1 Results for the Conventional Methods

To recap, the conventional methods applied in this study were:

1. The simple Before-After Naive,
2. The Before-After Naive procedure with the correction for traffic flow, and
3. The Before-After Naive procedure using the SPFs.

Simple before-After Naïve

The safety effects for all severity levels (Total, Injury, and PDO) for the simple before and after Naïve method on the composite entity of California and Minnesota are given in Table 6.1

Table 6.1 –Simple Before-After Naïve for ALL Severity (Negative indicate increase in crash) at California and Minnesota as a Composite Entity

	TOTAL Severity			
	All	Right Angle	Left Turn	Rear End
Naive estimate of crashes expected in the after period without signals ,π (s.e)	1974 (86.9)	847 (53.8)	456 (51.4)	238 (26.6)
Accident Counts after treatment was implemented λ (s.e)	1487 (38.6)	281 (16.8)	214 (14.6)	468 (21.6)
Estimate of the change in crash frequency, δ (s.e)	487 (95.1)	566 (56.4)	242 (53.5)	-230 (34.2)
Index of Effectives ,θ (s.e)	*0.752 (0.04)*	*0.333 (0.06)*	*0.463 (0.11)*	*1.939 (0.14)*
Percent Change in crashes	*24.8*	*67*	*53.7*	*-93.9*
	INJURY Severity			
Naive estimate of crashes expected in the after period without signals ,π (s.e)	983 (61.4)	458 (38.7)	250 (38.5)	114 (18.9)
Accident Counts after treatment was implemented λ (s.e)	601 (24.5)	145 (12.0)	78 (8.8)	174 (13.2)
Estimate of the change in crash frequency, δ (s.e)	382 (66.1)	313 (40.6)	172 (39.5)	-60 (23.1)
Index of Effectives ,θ (s.e)	*0.609 (0.06)*	*0.315 (0.09)*	*0.304 (0.15)*	*1.491 (0.19)*
Percent Change in crashes	*39.1*	*68.5*	*69.6*	*-49.1*
	PDO Severity			
Naive estimate of crashes expected in the after period without signals ,π (s.e)	991 (61.5)	389 (37.4)	206 (34.1)	125 (18.6)
Accident Counts after treatment was implemented λ (s.e)	886 (29.8)	136 (11.7)	136 (11.7)	294 (17.1)
Estimate of the change in crash frequency, δ (s.e)	105 (68.3)	253 (39.1)	70 (36.1)	-169 (25.3)
Index of Effectives ,θ (s.e)	*0.891 (0.07)*	*0.346 (0.10)*	*0.643 (0.17)*	*2.304 (0.20)*
Percent Change in crashes	*10.9*	*65.4*	*35.7*	*-130.4*

Naïve Before-After with Traffic Flow Correction

The safety effects for all severity levels (Total, Injury, and PDO) for the "Naïve before-after with Traffic Flow Correction" method on the composite entity of California and Minnesota are given in Table 6.2.

Table 6.2 –Naïve Before-After with Traffic Flow Correction for ALL Severity (Negative indicate increase in crash) at California and Minnesota as a Composite Entity

	TOTAL Severity			
	All	Right Angle	Left Turn	Rear End
Naive estimate of crashes expected in the after period without signals ,π (s.e)	2799 (147.9)	1217 (90.6)	687 (94.9)	355 (51.3)
Accident Counts after treatment was implemented λ (s.e)	1487 (38.6)	281 (16.8)	214 (14.6)	468 (21.6)
Estimate of the change in crash frequency, δ (s.e)	1312 (152.8)	936 (92.2)	473 (96.0)	-113 (55.7)
Index of Effectives ,θ (s.e)	*0.530* *(0.03)*	*0.230* *(0.00)*	*0.306* *(0.04)*	*1.290* *(0.19)*
Percent Change in crashes	*47*	*77*	*69.4*	*-29*
	INJURY Severity			
Naive estimate of crashes expected in the after period without signals ,π (s.e)	1430 (111.4)	659 (67.5)	406 (76.8)	171 (36.5)
Accident Counts after treatment was implemented λ (s.e)	601 (24.5)	145 (12.0)	78 (8.8)	174 (13.2)
Estimate of the change in crash frequency, δ (s.e)	829 (114.0)	514 (68.5)	328 (77.3)	-3 (38.8)
Index of Effectives ,θ (s.e)	*0.418* *(0.03)*	*0.218* *(0.03)*	*0.186* *(0.04)*	*0.974* *(0.21)*
Percent Change in crashes	*58.2*	*78.2*	*81.4*	*2.6*
	PDO Severity			
Naive estimate of crashes expected in the after period without signals ,π (s.e)	1369 (97.3)	558 (60.5)	281 (55.7)	185 (36.1)
Accident Counts after treatment was implemented λ (s.e)	886 (29.8)	136 (11.7)	136 (11.7)	294 (17.1)
Estimate of the change in crash frequency, δ (s.e)	483 (101.7)	422 (61.6)	145 (56.9)	-110 (39.9)
Index of Effectives ,θ (s.e)	*0.644* *(0.05)*	*0.241* *(0.03)*	*0.465* *(0.09)*	*1.535* *(0.30)*
Percent Change in crashes	*35.6*	*75.9*	*53.5*	*-53.5*

Naïve Before-After with SPF

The safety effects for all severity levels (Total, Injury, and PDO) for the Naïve before and after with SPF method on the composite entity of California and Minnesota are given in Table 6.3.

Table 6.3 –Naïve Before-After with SPF for ALL Severity (Negative indicate increase in crash) at California and Minnesota as a Composite Entity

	TOTAL Severity			
	All	Right Angle	Left Turn	Rear End
Naive estimate of crashes expected in the after period without signals ,π (s.e)	3255 (303.8)	2033 (282.0)	1468 (281.4)	484 (87.2)
Accident Counts after treatment was implemented λ (s.e)	957 (30.9)	281 (16.8)	214 (14.6)	468 (21.6)
Estimate of the change in crash frequency, δ (s.e)	2298 (305.4)	1752 (282.5)	1254 (281.8)	16 (89.8)
Index of Effectives ,θ (s.e)	*0.291* *(0.03)*	*0.136* *(0.00)*	*0.141* *(0.03)*	*0.936* *(0.17)*
Percent Change in crashes	*70.9*	*86.4*	*85.9*	*6.4*
	INJURY Severity			
Naive estimate of crashes expected in the after period without signals ,π (s.e)	2111 (242.9)	1075 (198.7)	928 (229.9)	209 (48.8)
Accident Counts after treatment was implemented λ (s.e)	601 (24.5)	145 (12.0)	78 (8.8)	174 (13.2)
Estimate of the change in crash frequency, δ (s.e)	1510 (244.1)	930 (199.1)	850 (230.1)	35 (50.5)
Index of Effectives ,θ (s.e)	*0.281* *(0.03)*	*0.13* *(0.03)*	*0.079* *(0.00)*	*0.789* *(0.18)*
Percent Change in crashes	*71.9*	*87*	*92.1*	*21.1*
	PDO Severity			
Naive estimate of crashes expected in the after period without signals ,π (s.e)	1812 (189.3)	945 (199.4)	533 (162.9)	272 (74.4)
Accident Counts after treatment was implemented λ (s.e)	886 (29.8)	136 (11.7)	136 (11.7)	294 (17.1)
Estimate of the change in crash frequency, δ (s.e)	926 (191.7)	809 (199.7)	397 (163.3)	-22 (76.3)
Index of Effectives ,θ (s.e)	*0.484* *(0.05)*	*0.138* *(0.03)*	*0.233* *(0.07)*	*1.005* *(0.26)*
Percent Change in crashes	*51.6*	*86.2*	*76.7*	*-0.5*

The conventional methods reveal that after a rural stop controlled intersection is converted to a signalized intersection, there will be a reduction in 'all', 'right-angle' and 'left-turn' crash types for all severity levels (Total, Injury, and PDO). Rear-end crash effects on the other hand were mixed. For the simple Naïve procedure, rear-end crashes were estimated to increase for all severity levels and for the "Naïve with 'traffic flow' correction procedure", rear-end crashes only increased for 'total' and 'PDO' severity after a signal was installed. The direction of the safety estimate for rear-end crashes was also mixed when the Naïve with 'SPF' procedure was used; however, only the 'PDO' severity level showed a very small increase.

Another noticeable trend in the results is visible if the safety effects for each conventional method, for each crash type and severity level, are compared to each other. For 'all', 'right-angle' and 'left-turn' crash types, the "Naive with SPF" gives the highest gain in estimating safety, followed by the "Naive with traffic correction" and the simple Naïve methods. The same trend is visible for rear-end crashes. This trend is attributable to the underlying assumptions for each methodology.

The simple Naïve procedure assumes that the before period crash counts is the expected crash count in the after period had a signal not been installed. The problem with this method is that one cannot attribute the change in the crash frequencies to the treatment alone. Other factors such as regression-to-the-mean, traffic volume changes and accident trend could have contributed to the change, since all of these factors affect safety (Persaud 2001). Therefore, the estimated percent change in crashes given in Table 6.1

cannot entirely be attributed to installing signals alone as implied in the simple Naïve procedure.

The Naïve with 'traffic flow' and 'SPF' is somewhat more precise (Hauer 1997) in that traffic flow during the before and after period is accounted for. However, the "Naïve with traffic flow correction" simply uses the ratio of the average after period AADT to before period AADT in accounting for traffic flow in its estimate of safety in after period had the signal not been installed. On the other hand, the "Naïve with SPF" uses the same ratio but here the ratio is the SPF estimate of the expected crashes in the after period to that for the before period, not necessarily assuming a linear crash-traffic relationship as the "Naïve with traffic flow correction" method does.

The trend in the safety effects among the Naïve procedures expressed previously can be attributed to the effect of the AADT since the treated entities exhibits an increasing trend in AADT from the before to the after period (Tables 4.1 through 4.5, Chapter 4). A higher AADT in the after period implies that the estimate of expected crashes in the after period had the treatment not been implemented, π, will be larger. This is simply because the ratio of traffic flows is larger than 1 and since this ratio serves as a multiplicative correction for AADT (Hauer 1997), π will in turn be larger than in the simple Naïve procedure. Therefore, with a larger value for π, the index of effectiveness (Equation 3.5, Chapter 3) will be smaller than in the simple Naïve procedure.

One other interesting aspect is the difference between the safety effects for the "Naïve with 'traffic flow' correction" and the "Naïve with 'SPF'". A larger safety gain was estimated using the Naïve with 'SPF' procedure simply because the ratio of the expected crashes in the after to before period is quantified by a non-linear crash-traffic flow relationship, i.e., by a non-linear nature of SPF (Figure 3.1, depicts a typical example.). Under this assumption, with larger AADTs in the after period with respect to the before period, the ratio of expected crashes before to after will be larger than the ratio under the linear crash-traffic flow assumption of the "Naïve with traffic flow correction" method. However, because the relationship between crash frequency and AADT is non-linear, this ratio will in turn be larger than in the "Naïve with traffic correction".

Assuming that the "Naïve with SPF" procedure gives a better estimate of the safety effects of installing a signal, there are still several factors that are not considered such as regression-to-the-mean and trends in the crashes due to a variety of factors. Therefore, the safety effects are still not without its bias under the three conventional approaches. However, with the aid of the empirical Bayes (EB) procedure (which accounts for all those factors), the bias in the safety effects would be eliminated. The EB results are presented next.

6.2 Results from the empirical Bayes method
(Comparison with the results from the other methods)

The safety effects for all severity levels (Total, Injury, and PDO) using the Empirical Bayes procedure on the composite entity of California and Minnesota is given in Table 6.4. The results reveal that by installing a signal, 'all', 'right-angle', and 'left-turn' crashes will be reduced, but there will be an increase in rear-end crashes for all severity levels. A summary of the percent change in crashes for each method is given in table 6.5.

Table 6.4 –Empirical Bayes ALL Severity (Negative indicate increase in crash) at California and Minnesota as a Composite Entity

	TOTAL Severity			
	All	Right Angle	Left Turn	Rear End
EB estimate of crashes expected in the after period without signals ,π (s.e)	2386 (160.7)	1083 (132.0)	524 (106.6)	245 (42.8)
Accident Counts after treatment was implemented λ (s.e)	1487 (38.6)	281 (16.8)	214 (14.6)	468 (21.6)
Estimate of the change in crash frequency, δ (s.e)	899 (165.3)	802 (133.1)	310 (107.6)	-223 (47.9)
Index of Effectives ,θ (s.e)	0.620 (0.04)	0.256 (0.03)	0.392 (0.08)	1.855 (0.33)
Percent Change in crashes	38	74.4	60.8	-85.5
	INJURY Severity			
EB estimate of crashes expected in the after period without signals ,π (s.e)	1195 (119.4)	644 (94.7)	304 (72.4)	100 (27.3)
Accident Counts after treatment was implemented λ (s.e)	601 (24.5)	145 (12.0)	78 (8.8)	174 (13.2)
Estimate of the change in crash frequency, δ (s.e)	594 (121.9)	499 (95.4)	226 (72.9)	-74 (30.3)
Index of Effectives ,θ (s.e)	0.498 (0.05)	0.22 (0.03)	0.243 (0.06)	1.623 (0.44)
Percent Change in crashes	50.2	78	75.7	-62.3
	PDO Severity			
EB estimate of crashes expected in the after period without signals ,π (s.e)	1043 (76.7)	409 (87.2)	217 (50.1)	169 (34.9)
Accident Counts after treatment was implemented λ (s.e)	886 (29.8)	136 (11.7)	136 (11.7)	294 (17.1)
Estimate of the change in crash frequency, δ (s.e)	157 (82.3)	273 (88.0)	81 (51.4)	-125 (38.9)
Index of Effectives ,θ (s.e)	0.845 (0.07)	0.318 (0.07)	0.596 (0.14)	1.669 (0.35)
Percent Change in crashes	15.5	68.2	40.4	-66.9

These directional effects are in complete agreement with those from the simple Naive procedure. However, the effects for the other two forms of the Naïve procedure and the EB are only in agreement for total, right-angle and left-turn crashes for all severity levels. For rear-end crash effects, the other two forms of the Naïve procedure show a reduction in rear-end crashes for injury crashes after a signal is installed, thus yielding mixed results when compared to the EB effects.

In terms of percentage change in crashes the EB results showed a higher safety gain when compared to the simple Naïve estimate, but a smaller gain than the Naïve with 'traffic correction' and 'SPF' procedures. These differences in percentage changes and patterns can be directly attributed to the assumptions governing each procedure.

Table 6.5 – Summary of safety effects from each method

Safety Method	Percent change in crashes (negative implies increase)											
	Total				Injury				PDO			
	All	RA	LT	RE	ALL	RA	LT	RE	ALL	RA	LT	RE
Simple Naïve	24.8	67	53.7	-93.9	39.1	68.5	69.6	-49.1	10.9	65.4	35.7	-130.4
Naïve - traffic flow	47	77	69.4	-29	58.2	78.2	81.4	2.6	35.6	75.9	53.5	-53.5
Naïve - SPF	70.9	86.4	85.9	6.4	71.9	87	92.1	21.1	51.6	86.2	76.7	-0.5
EB	38	74.4	60.8	-85.5	50.2	78	75.7	-62.3	15.5	68.2	40.4	-66.9

The larger safety effect for the EB compared to the simple Naïve can be attributed to the fact that the simple Naïve procedure fails to account for the safety impact of increase in AADT in the after period compared to the before period, while the EB does. The other two forms of the Naïve procedure (traffic flow and SPF) improved on this shortcoming of the simple Naive procedure by accounting for AADT. The question is why these other two Naïve procedures estimated a larger safety gain than the EB procedure, since by accounting for AADT, one would have expected that the percentage change in crashes would have been in closer agreement to the EB estimate. This question can be answered by the noting that the Naïve procedures, including these two, all fail to account for regression to the mean, which would cause them to overestimated safety gains.

6.3 Disaggregate EB estimates

The composite entity of California and Minnesota consisted of 45 treated intersections comprising of 39 four-legged and 6 three-legged intersections. By evaluating groups of similar variables, useful inferences can be made about the safety effects after a signal is installed. Tables 6.6, 6.7 and 6.8 give the EB estimates of the safety effect by disaggregating the composite entity into smaller composite entities. Three legged-intersections were not disaggregated in smaller groups because of its small sample size.

Based on the disaggregate results for the 4-legged intersections after a signal is installed, total crashes for right-angle and left-turn types for total severity (See Table 6.6) would be reduced while rear-end crashes increase. This pattern is similar to the aggregate EB effect. However, there was a mixture of safety effects based on the different levels at which the variables were disaggregated. For example, under *total* severity for all ranges of AADT tested, a reduction in the total crashes always existed, but for injury crashes with AADT < 10,000 and PDO crashes with AADT > 20,000, an increase in crashes was observed after a signal was installed. This increase in injury crashes for AADT < 10,000 is in conformance with findings by Persaud et al. (1997) where traffic signals were removed at low ranges of AADT to reverse this undesirable safety effect. Therefore signalizing intersections with AADT < 10,000 might have an adverse effect on injury accidents.

Table 6.6 –Empirical Bayes Disaggregate Estimates for Safety Effects – 4 leg TOTAL CRASHES

Smaller Composite Entity (# of intersections)	TOTAL		RIGHT-TURN		LEFT-TURN		REAR-END	
	θ	VAR(θ)	θ	VAR(θ)	θ	VAR(θ)	θ	VAR(θ)
AADT ≤ 10000 (7)	0.66	0.01	0.51	0.02	0.64	0.05	0.92	0.13
10,000< AADT ≤ 15000 (15)	0.50	0.00	0.22	0.00	0.33	0.01	1.72	0.41
15,000< AADT ≤ 12000 (10)	0.59	0.00	0.18	0.00	0.52	0.02	1.49	0.07
AADT > 20000 (7)	0.95	0.01	0.41	0.01	0.52	0.02	2.87	0.69
Total (all severity) Crashes/year ≥ 5 (22)	0.57	0.00	0.26	0.00	0.37	0.01	1.96	0.15
Total (all severity) Crashes/year < 5 (17)	0.91	0.01	0.25	0.01	0.77	0.03	0.84	0.07
Major approach lanes = 2 (21)	0.62	0.00	0.23	0.00	0.67	0.02	1.63	0.08
Major approach lanes = 4 (18)	0.61	0.00	0.27	0.00	0.36	0.01	1.92	0.26
California (24)	0.79	0.00	0.21	0.00	0.65	0.01	1.27	0.10
Minnesota (15)	0.55	0.00	0.27	0.00	0.32	0.01	1.92	0.15
All (Minnesota and California) (39)	0.61	0.00	0.26	0.00	0.44	0.01	1.87	0.12

Table 6.7 –Empirical Bayes Disaggregate Estimates for Safety Effects – 4 leg INJURY CRASHES

Smaller Composite Entity (# of intersections	TOTAL		RIGHT-TURN		LEFT-TURN		REAR-END	
	θ	VAR(θ)	θ	VAR(θ)	θ	VAR(θ)	θ	VAR(θ)
AADT ≤ 10000 (7)	1.09	0.08	0.94	0.11	0.45	0.08	0.82	0.22
10,000< AADT ≤ 15000 (15)	0.34	0.00	0.15	0.00	0.20	0.01	1.17	0.40
15,000< AADT ≤ 12000 (10)	0.43	0.00	0.12	0.00	0.26	0.01	1.23	0.11
AADT > 20000 (7)	0.88	0.02	0.42	0.01	0.29	0.01	2.64	1.86
Total (all severity) Crashes/year ≥ 5 (22)	0.40	0.00	0.17	0.00	0.22	0.01	1.36	0.22
Total (all severity) Crashes/year < 5 (17)	0.68	0.01	0.35	0.00	0.30	0.01	2.21	0.78
Total Injury Crashes/year ≥ 5 (11)	0.40	0.00	0.17	0.00	0.22	0.01	1.37	0.22
Total Injury Crashes/year < 5 (28)	0.68	0.01	0.35	0.00	0.30	0.01	2.21	0.78
Major approach lanes = 2 (21)	0.46	0.00	0.21	0.00	0.38	0.00	0.91	0.07
Major approach lanes = 4 (18)	0.47	0.01	0.22	0.00	0.22	0.01	1.84	0.56
California (24)	0.62	0.01	0.16	0.00	0.40	0.01	1.74	0.34
Minnesota (15)	0.44	0.00	0.23	0.00	0.18	0.00	1.57	0.24
All (Minnesota and California) (39)	0.49	0.00	0.22	0.00	0.25	0.00	1.60	0.21

An interesting finding was observed in the safety effects between California and Minnesota when considering PDO crashes (See Table 6.8) from the disaggregate analysis. When taking California alone into consideration, the results revealed that there will be an increase in PDO crashes after a signal was installed while Minnesota experienced a reduction for this crash severity type. This difference in safety effect could be related to how PDO crashes are considered by the two states. PDO crashes reports are usually based on dollar value of damages sustained and most likely California and Minnesota could have been using different threshold values for reporting these crashes. This issue is also exhibited for rear-end PDO severity crashes for which California is estimating a reduction in crashes in contrast to the increase for Minnesota.

Table 6.8 –Empirical Bayes Disaggregate Estimates for Safety Effects– 4 leg PDO CRASHES

Smaller Composite Entity (# of intersections	TOTAL		RIGHT-TURN		LEFT-TURN		REAR-END	
	θ	VAR(θ)	θ	VAR(θ)	θ	VAR(θ)	θ	VAR(θ)
AADT ≤ 10000 (7)	0.62	0.02	0.24	0.01	0.75	0.08	0.69	0.08
10,000< AADT ≤ 15000 (15)	0.81	0.02	0.30	0.01	0.54	0.05	1.62	0.49
15,000< AADT ≤ 12000 (10)	0.86	0.01	0.28	0.00	0.65	0.04	1.66	0.17
AADT > 20000 (7)	1.08	0.03	0.38	0.02	1.31	0.25	1.90	0.33
Total (all severity) Crashes/year ≥ 5 (22)	0.80	0.01	0.27	0.01	0.47	0.04	1.76	0.35
Total (all severity) Crashes/year < 5 (17)	0.93	0.01	0.39	0.01	0.88	0.00	1.52	0.11
Total PDO crashes / year ≥ 5 (19)	0.80	0.01	0.27	0.01	0.46	0.04	1.76	0.35
Total PDO crashes / year <5 (30)	0.93	0.01	0.39	0.01	0.93	0.03	1.52	0.11
Major approach lanes = 2 (21)	0.86	0.01	0.28	0.00	0.78	0.03	2.08	0.17
Major approach lanes = 4 (18)	0.86	0.01	0.33	0.01	0.61	0.06	1.55	0.22
California (24)	1.05	0.01	0.40	0.01	1.07	0.05	0.91	0.08
Minnesota (15)	0.79	0.01	0.31	0.01	0.50	0.04	1.80	0.19
All (Minnesota and California) (39)	0.87	0.01	0.32	0.01	0.70	0.04	1.73	0.15

6.4 Empirical Bayes with Proportional SPFs

The need for and functionality of SPFs is a very important component in an EB study. However, SPFs are sometimes impossible to calibrate due to small sample sizes that usually lead to poorly fitted SPFs, or simply because it can be extremely time consuming. In such situations, another form of SPFs, referred to as proportional SPFs, is usually developed (Persaud et al. 2005). This method simply applies a factor (proportion) to the SPF for total accidents to estimate the accidents for a specific crash type. For instance, assuming that the total accident SPF for a particular intersection is of the form:

$$Total\ Accident\,/\,year = \alpha\left(Total\ Entering\ AADT\right)^{\beta} \tag{6.5}$$

To illustrate, suppose one needs to calibrate an SPF for left-turning accidents, but the actual dataset is insufficient for calibrating a statistically significant model; therefore under the proportional SPF approach, a factor (proportion of total accidents that are left-turn accidents) is applied to Equation 6.5:

$$Left-Turn\ \ Accident\,/\,year = \left(FACTOR\right)\alpha\left(Total\ Entering\ AADT\right)^{\beta} \tag{6.6}$$

To assess the validity of this approximation to calibrating SPFs, this method was used to develop secondary Injury and PDO severity crash type SPFs for the California and Minnesota reference groups. The total severity SPFs were used, analogous to Equation 6.5. These SPFs were then applied to the EB procedure to test the difference between the results so obtained and the EB results with SAS directly calibrated SPFs for the specific crash type. Dispersion parameters for the proportion SPFs were estimated by using a maximum likelihood software procedure. A summary of the California proportional

SPFs for Injury and PDO severity is given in Tables 6.9, 6.10 and 6.11. Tables 6.12 and 6.13 give the summary for the Minnesota proportion SPFs for Injury and PDO.

Table 6.9 – Proportion Injury and PDO SPFs for California Three-legged with 2 lanes on Major Approach

	All	Right-Angle	Left-Turn	Rear-End
Model	2	2	2	1
Injury LN(α)	-10.134	-11.206	-14.345	-12.460
PDO LN(α)	-9.880	-11.528	-14.403	-12.044
β_1 (s.e.)	1.112 (0.046)	0.907 (0.156)	1.482 (0.080)	0.824 (0.084)
β_2 (s.e.)	0.324 (0.023)	0.522 (0.071)	0.571 (0.037)	0.212 (0.040)
Dispersion Injury	1.74	0.31	0.78	0.97
Dispersion PDO	1.65	1.54	0.77	1.22
Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$		**Model 2:** $E\{\kappa\} = \alpha(Maj + Min)^{\beta_1}\left(\frac{Min}{Maj + Min}\right)^{\beta_2}$		

Table 6.10– Proportion Injury and PDO SPFs for California Four-legged with 2 lanes on Major Approach

	All	Right-Angle	Left-Turn	Rear-End
Model	1	1	1	2
Injury LN(α)	-9.925	-11.541	-13.949	-11.761
PDO LN(α)	-9.781	-11.929	-14.053	-11.419
β_1 (s.e.)	0.719 (0.060)	0.554 (0.129)	0.821 (0.095)	1.023 (0.114)
β_2 (s.e.)	0.481 (0.028)	0.781 (0.062)	0.763 (0.046)	0.252 (0.052)
Dispersion Injury	1.72	0.45	0.97	0.88
Dispersion PDO	2.1	0.56	1.03	1.33
Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$		**Model 2:** $E\{\kappa\} = \alpha(Maj + Min)^{\beta_1}\left(\frac{Min}{Maj + Min}\right)^{\beta_2}$		

Table 6.11 – Proportion Injury and PDO SPF for California Four-legged with 4 lanes on Major Approach

	All	Right-Angle	Left-Turn	Rear-End
Model	1	2	2	1
Injury LN(α)	-10.379	-13.268	-17.208	-13.720
PDO LN(α)	-10.316	-13.768	-17.545	-13.286
β_1 (s.e.)	0.769 (0.116)	1.454 (0.247)	1.802 (0.265)	0.969 (0.197)
β_2 (s.e.)	0.426 (0.069)	0.707 (0.120)	0.559 (0.125)	0.211 (0.103)
Dispersion Injury	0.99	0.54	0.42	0.94
Dispersion PDO	1.81	0.71	0.97	1.99

Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$ **Model 2:** $E\{\kappa\} = \alpha(Maj+Min)^{\beta_1}\left(\frac{Min}{Maj+Min}\right)^{\beta_2}$

Table 6.12 – Proportion Injury and PDO SPF for Minnesota Three-legged Intersections (Minnesota)

	All	Right-Angle	Left-Turn	Rear-End
Model	1	1	1	1
Injury LN(α)	-9.3958	-13.844	-12.652	-12.296
PDO LN(α)	-9.2107	-14.052	-12.646	-12.068
β_1 (s.e.)	0.496 (0.050)	0.722 (0.083)	0.292 (0.127)	0.760 (0.092)
β_2 (s.e.)	0.624 (0.050)	0.816 (0.083)	0.983 (0.127)	0.521 (0.090)
Dispersion Injury	1.26	0.46	0.21	0.45
Dispersion PDO	1.43	0.49	0.17	0.5

Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$

Table 6.13 – Proportion Injury and PDO SPF for Minnesota Four-legged Intersections (Minnesota)

	All	Right-Angle	Left-Turn	Rear-End
Model	1	1	1	1
Injury LN(α)	-9.562	-10.486	-14.479	-13.277
PDO LN(α)	-9.525	-10.859	-14.537	-13.011
β_1 (s.e.)	0.566 (0.039)	0.462 (0.054)	0.725 (0.091)	0.909 (0.067)
β_2 (s.e.)	0.698 (0.041)	0.891 (0.060)	0.815 (0.096)	0.559 (0.065)
Dispersion Injury	1.52	0.84	0.26	0.51
Dispersion PDO	1.78	0.82	0.28	0.68

Model 1: $E\{\kappa\} = \alpha(Maj)^{\beta_1}(Min)^{\beta_2}$

The results for the EB analysis using the proportional SPFs are given in Table 6.14. Comparing these result with those of the EB study using SAS calibrated SPFs (Table 6.4), small differences between the percent changes in crashes were observed in general. The only marked difference is for the PDO and Injury estimates for rear-end crashes. For this particular dataset, it can, therefore, be concluded that in general valid EB results can be obtained by using simpler proportional SPFs. This is a useful conclusion in the light of the difficulties that are typical in obtaining sufficient reference group data for the direct calibration of SPFs.

Table 6.14 – Empirical Bayes safety effects for Injury and PDO crashes using proportional SPFs (Negative indicate increase in crash)

	INJURY Severity			
	All	Right Angle	Left Turn	Rear End
EB estimate of crashes expected in the after period without signals ,π (s.e)	1190 (119.1)	656 (99.6)	302 (71.4)	86 (37.1)
Accident Counts after treatment was implemented λ (s.e)	601 (24.5)	145 (12.0)	78 8.8	174 13.2
Estimate of the change in crash frequency, δ (s.e)	589 (121.6)	511 (100.3)	224 (71.9)	-88 (39.4)
Index of Effectives ,θ (s.e)	0.500 (0.05)	0.216 (0.04)	0.244 (0.06)	1.702 (0.68)
Percent Change in crashes	50.0	78.4	75.6	-70.2
	PDO Severity			
EB estimate of crashes expected in the after period without signals ,π (s.e)	1150 (98.9)	412 (82.4)	218 (48.1)	149 (39.2)
Accident Counts after treatment was implemented λ (s.e)	886 (29.8)	136 (11.7)	136 (11.7)	294 (17.1)
Estimate of the change in crash frequency, δ (s.e)	264 (103.3)	276 (83.2)	82 (49.5)	-145 (42.8)
Index of Effectives ,θ (s.e)	0.765 (0.07)	0.317 (0.07)	0.596 (0.14)	1.840 (0.48)
Percent Change in crashes	23.5	68.3	40.4	-84.0

CHAPTER 7: Economic Analysis

Based on the results of the safety effects analysis, no definitive claim can be made about the overall safety effect for the implementation of traffic signals since the results show that this treatment will reduce right-angle accidents, but increase rear-end crashes. The question is whether the increase in rear-end crashes will negate the benefits gained from the reduction of the right-angle crashes. This question can be addressed by an examination of the economic costs for the two crash types, base on severity levels (injury and non-injury). Council et al. (2005) did a similar economic analysis after they found that rear end crashes will tend to increase, and right angle accidents decrease, after red light cameras are installed at a signalized intersection. The same procedure was adopted in this study.

To apply the procedure, the economic cost per crash was needed for the right-angle, rear-end and "other" crash types. The crash type 'other' is simply the difference between the total crash and the sum of the right-angle and left-turn crashes. Crash costs were developed in the U.S. for each of the KABCO severity level (Council, Eduard, Miller and Persaud 2005) for rural and urban intersections by control type. The distinction between urban and rural crash cost was made solely on the posted speed. Intersections with main road posted speeds higher than 50mph (80 km/h) were considered rural whereas those with speed limits lower than 45 mph (72 km/h) were considered as urban.

Council et al. (2005) developed two types of crash cost - human capital costs and comprehensive costs. The human capital cost is simply the actual dollar value incurred for each severity, whereas the comprehensive costs is the sum of the human capital cost and cost for loss of quality of life for the two severities (K+A+B+C, and O). This study used the comprehensive cost for the economic analysis, following Council et al. Section 7.1 gives a synopsis of the procedure.

7.1 Methodology for Economic Analysis

The EB calculations were done for each crash types (right-angle, rear-end and other). Crash types were further sub-divided into two severity groups, injury (K+A+B+C) and PDO (O). Then, following Council et al. (2005) the EB crash estimates were used with the following costs to perform the analysis:

$\Lambda_{Cost\,A}$ = cost of crashes that occurred after the treatment was installed. This cost is simply the count of crashes in the after period multiplied by the corresponding unit comprehensive crash cost for the crash type and severity at rural *signalized* intersections.

$VAR\left(\Lambda_{Cost\,A}\right)$ = variance of the crash costs in the after period. Each unit crash cost developed by Council et al. 2005) has a corresponding standard error. The variance is simply computed as the square of the standard error multiplied by the count of after period crashes.

$\Pi_{Cost\,A}$ = cost of crashes in the after period had there been *no* treatment. The estimation of this cost accounts for RTM, traffic volume trend and other factors. This parameter is the product of EB estimate of π (the crashes expected in the after period without treatment) and the corresponding unit comprehensive cost for the crash type and severity at rural *stop controlled* intersections.

$VAR\left(\Pi_{Cost\,A}\right)$ = variance of the crashes in the after period had no signal been installed.

Similar to above, the standard error of the unit crash cost was used to calculate the variance, which was then multiplied by EB estimate of π (the crashes expected in the after period without treatment).

The parameters listed above were aggregated over the treatment sites for both severity groups, (K+A+B+C, and O or, simply, injury and PDO), giving an aggregated economic cost with the change in crash cost given by Equation 7.1:

$$\Phi_{cost} = \Pi_{cost\,A} - \Lambda_{cost\,A} \tag{7.1}$$

The variance of the change in crash cost is given by Equation 7.2:

$$VAR\left(\Phi_{cost}\right) = VAR\left(\Pi_{cost\,A}\right) + VAR\left(\Lambda_{cost\,A}\right) \tag{7.2}$$

The index of effectiveness with respect to crash cost is given by Equation 7.3:

$$\theta_{cost} = \frac{\dfrac{\Lambda_{Cost\,A}}{\Phi_{Cost\,A}}}{\left[1 + \dfrac{VAR\left(\Pi_{Cost\,A}\right)}{\left(\Pi_{Cost\,A}\right)^2}\right]} \tag{7.3}$$

The variance of the index of effectiveness is given by Equation 7.4:

$$VAR\left(\theta_{cost}\right) = \frac{\left(\theta_{cost}\right)^2 \left[\dfrac{VAR\left(\Lambda_{Cost\,A}\right)}{\left(\Lambda_{Cost\,A}\right)^2} + \dfrac{VAR\left(\Phi_{Cost\,A}\right)}{\left(\Phi_{Cost\,A}\right)^2}\right]}{\left[1 + \dfrac{VAR\left(\Pi_{Cost\,A}\right)}{\left(\Pi_{Cost\,A}\right)^2}\right]^2} \tag{7.4}$$

7.2 Estimates of Economic Effects of Installing Traffic Signals

Table 7.1 gives the estimate of the unit comprehensive crash cost for each crash type and severity. Unit crash costs for "other" is the same as the crash cost for right-angle accidents due to recommendation made by Council et al. (2005).

Table 7.1 – Unit comprehensive crash costs used in the economic analysis

Control Type	Severity Level	Right Angle Cost (s.e)	Rear End Cost (s.e.)	Other Cost (s.e.)
Signal	Injury (s.e)	\$126,878 \$9,619	\$52,276 \$13,794	\$126,878 \$9,619
	PDO (s.e)	\$8,544 \$1,294	\$5,901 \$1,802	\$8,544 \$1,294
Stop Sign	Injury (s.e)	\$199,788 \$27,768	\$34,563 \$12,854	\$199,788 \$27,768
	PDO (s.e)	\$5,444 \$1,265	\$3,788 \$978	\$5,444 \$1,265
s.e. – standard error				

The EB estimates of the economic effect of crashes for each crash type are shown in Tables 7.2 and 7.3. As noted earlier, the objective for this economic assessment was to see if the increase in rear-end accidents after a signal was installed will offset the benefits gained from the reduction of right-angle accidents. If all accidents are considered (injury and non-injury) the results (Table 7.2) show a positive aggregate economic benefit of \$155,883,978 which represents a 69 percent reduction in cost over the 45 treated sites. The results for the injury crashes only (Table 7.3) show a benefit of \$157,280,562, representing a 71 percent reduction in cost. Note the benefit is aggregated over the 45 sites with a total of 253 years of data in the after period. To get a better perspective on the reduction in crash cost it would be prudent to estimate the benefit per site-year. These results are:

1. If Injury and PDO accidents are considered together, the reduction in accident costs is (155,883,978)/(253) = $616,142 per site-year.
2. If PDO crashes are not considered, the reduction in accident costs is (157,280,562)/ (253) = $621,662 per site.

Table 7.2– EB economic effects for composite entity with injury and on-injury (PDO) crashes

	Right Angle	Rear-End	Other	All Crashes
EB estimate of crash costs without Signals (s.e)	$130,908,188 $705,181	$4,446,765 $12,935	$90,594,897 $583,634	$225,949,850 $915,465
Cost of crashes recorded after Signals were installed (s.e)	$19,559,294 $116,807	$10,830,918 $182,899	$39,675,660 $163,877	$70,065,872 $271,940
Index of Effectives ,θ (s.e)	0.149 0.001	2.436 0.042	0.438 0.003	0.310 0.002
Percent reduction in Crash Cost (Negative implies an increase)	85	-144	56	*69*
Aggregated Crash Cost decrease	$111,348,894	-$6,384,153	$50,919,237	*$155,883,978*

Table 7.3– EB economic effects for composite entity with injury crashes only

	Right Angle	Rear-End	Other	All Crashes
EB estimate of crash costs without Signals (s.e)	$128,679,455 $704,717	$3,806,961 $2,400	$88,067,076 $582,997	$220,553,492 $914,613
Cost of crashes recorded after Signals were installed (s.e)	$18,397,310 $115,828	$9,096,024 $181,955	$35,779,596 $161,530	$63,272,930 $269,474
Index of Effectives ,θ (s.e)	0.143 0.001	2.389 0.056	0.406 0.003	0.287 0.002
Percent reduction in Crash Cost (Negative implies an increase)	86	-139	59	*71*
Aggregated Crash Cost decrease	$110,282,145	-$5,289,063	$52,287,480	*$157,280,562*

The economic analysis shows a positive gain in cost even after considering the increase in rear end crashes after a signal was installed, so it was useful for this purpose alone. The aggregation of economic costs also allowed for an evaluation of the change in crash costs for various disaggregated groups such as those based on before period AADT, on crash levels, and on intersection geometry. For instance, the disaggregated EB economic effects for the 4-legged intersections are shown in Tables 7.4. The results for the 3-legged intersections are based on a small sample size and were not disaggregated.

The disaggregated economic analysis for the 4-legged intersections revealed positive economic effects for all classes of intersections with small standard errors. That is 'θ' was less than 1 for all disaggregated levels implying a reduction in total crash cost for each level. Therefore, the following inferences were possible:

1. A net gain in economic benefit was achieved for the entire range of AADT used after a signal is installed. However, there is an apparent pattern in the percent *decrease* in crash cost at various AADT levels. Intersections with AADT less than 10,000 experienced the smallest economic gain but an increase in the economic gain was observed for AADTs within a range of 10,000 to 20,000. However, the economic gain then subsides for AADT greater than 20,000.
2. A large economic gain was observed after signalization for total and injury severity crash frequencies larger than 5 per year during the before period. PDO crashes were not considered in this analysis because of inconsistent reporting of these crashes types.

3. Major approaches with 2 or 4 lanes experience approximately the same economic gain.

4. Intersections from California and Minnesota exhibited substantial economic gain as both separate entities and together (as one composite entity). An interesting observation is the difference between the percent decrease in crash cost for "injury and PDO" and injury alone; the difference for Minnesota is quite small compared to that for California. This anomaly can be attributed to the difference in reporting PDO crashes in each of the States. It is believed that California tends to severely underreport PDO crashes.

Table 7.4-Disaggregate EB Economic Effects for Composite Entity with 4-legged Intersections: 39 Sites

Intersection Class (number of sites)	Combined Injury and PDO Severity				Only Injury Severity			
	θ	VAR(θ)	% decrease in crash cost	Std. error	θ	VAR(θ)	% decrease in crash cost	Std. error
AADT ≤ 10000 (7)	0.74	2.65E-4	26	0.016	0.73	2.84E-4	27	0.017
10,000< AADT ≤ 15000 (15)	0.21	3.30E-6	79	0.002	0.19	3.13E-5	81	0.006
15,000< AADT ≤ 12000 (10)	0.26	1.13E-5	74	0.003	0.24	1.08E-5	76	0.003
AADT > 20000 (7)	0.54	4.89E-5	46	0.007	0.51	4.75E-5	49	0.007
Total (all severity) Crashes/year ≥ 5 (22)	0.28	3.23E-6	72	0.002	0.26	3.12E-6	74	0.002
Total (all severity) Crashes/year < 5 (17)	0.45	4.27E-5	55	0.007	0.41	3.95E-5	59	0.006
Total Injury Crashes/year ≥ 5 (11)	0.24	3.39E-6	76	0.002	0.22	3.26E-5	78	0.006
Total Injury Crashes/year < 5 (28)	0.43	1.61E-5	57	0.004	0.40	1.55E-5	60	0.004
Major approach lanes = 2 (21)	0.31	1.02E-5	69	0.003	0.28	9.60E-6	72	0.003
Major approach lanes = 4 (18)	0.30	4.39E-6	70	0.002	0.28	4.23E-6	72	0.002
California (24)	0.42	1.66E-5	58	0.004	0.39	1.59E-5	61	0.004
Minnesota (15)	0.26	3.81E-6	74	0.002	0.24	3.68E-6	76	0.002
All (Minnesota and California) (39)	0.30	3.07E-6	70	0.002	0.28	2.94E-6	72	0.002

CHAPTER 8: Conclusions

This study was primarily focused on estimating the safety effects after traffic signals are installed at stop controlled rural intersections. Use was made of both the conventional and the Empirical Bayes (EB) procedures. The conventional procedures consisted of the "simple before-after Naïve", "Naive with traffic flow correction", and "Naive with safety performance functions" methods. Data from 45 treated intersections located in California and Minnesota, in the US (28 from California and 17 from Minnesota) were used in this research to execute various analyses.

Both the conventional and the EB methods revealed a reduction in right-angle and left turn crashes and an increase in rear-end crashes after signals were installed. However, as expected the percent changes in crashes were quite different for each method, which is mainly due differences in underlying assumptions. The results are generally substantiated by similar findings revealed by a comprehensive literature review. Thus, it can be concluded that safety analysts should expect similar changes in crashes to those found whenever the implementation of traffic signals at stop controlled rural intersections is contemplated.

Results obtained from the EB procedure showed that "all", right angle and left turn crashes of all severities combined were reduced by 38%, 74% and 60% respectively, while an 85% increase was observed in rear-end crashes.

Clearly the percent changes in crashes estimated using the EB method should not be used alone in justifying the need for a signal since the increase in rear-end crashes has the potential to offset benefits obtained from a reduction in right-angle and left-turn crashes after a signal was installed. An economic analysis performed on accident types and severities after treatment (See Chapter 7) addressed the issue of whether the increase in rear-end crashes will offset the safety gained from the reduction of right-angle accidents. The results obtained from this analysis (Tables 7.3 and 7.4) show that even allowing for the increase in rear-end crashes after signals were installed, the overall economic effect achieved was quite positive.

In terms of dollar value, the economic analysis of the 45 treated sites shows that, after installation of the signals, the aggregated crash costs when considering both injury and non-injury crashes were reduced by $155,883,978 site year. This works out to an approximate reduction of $616,142 per site year. The reduction for injury crash costs alone amounted to about $157,280,562 per site year, which is an approximate reduction of $621,662 per site-year. Therefore, it can be concluded that after signalizing stop controlled rural intersections a definite increase in the overall safety at these intersections could be expected. This benefit is likely to outweigh by far the capital and maintenance costs of installing a traffic signal. It should be noted, however, that these benefits cannot be extrapolated to signalizing *any* intersection, since the installations evaluated in this study were most likely "warranted".

References

Bauer, K. M and Harwood, W.D (1997). "Statistical Models of Accidents on Interchange Ramps and Speed Change-Lanes" Report No. FHWA-RD-97-106, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C.

Bauer, K. M and Harwood, W.D (1998). "Statistical Models of At-Grade Intersection Accidents – Addendum" Report No. FHWA-RD-99-094, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C.

Council, M. F et al. (2005). "Economic Analysis of the Safety Effects of Red Light Camera Programs and the Identification of Factors Associated with the Greatest Benefits" Transportation Research Board, Washington, D.C. CD-ROM paper.

Council, M. F et al. (2005). "Safety Evaluation of Red Light Cameras" Report No. FHWA-HRT-05-048, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C.

Council, M.F., Zaloshnja, E., Miller, T. and Persaud, B. (2005), "Crash Cost Estimates by Maximum Police-Reported Injury Severity within Selected Crash Geometries" Report No. FHWA-HRT-OR-051, Federal Highway Administration, U.S. Department of Transportation, Washington D.C.

Datta, K.T. and Dutta, U. (1990). "Traffic Signal Installation and Accident Experience", ITE Journal, Washington, D.C., pp. 39-42.

Datta, K.T. (1991). "Head-on, Left-turn Accidents at Intersections with Newly Installed Traffic Signals", Transportation Research Record No. 1318, TRB, The National Academies, Washington, D.C.,pp. 58-63.

Dobsob, J.A. (2002). "An Introduction to Generalized Linear Models" Second Edition, Chapman and Hall/CRC.

Hanna, J.T. ,Flynn, T.E., and Tyler, W.L. (1976). "Characteristics of Intersection Accidents in Rural Municipalities", Transportation Research Record No. 601, TRB, The national Academies, Washington, D.C., pp. 79-82.

Harwood, D. W., F.M. Council, E. Hauer, W.E. Huges, and A. Vogt, (2000). "Prediction of the Expected Safety Performance of Rural Two-lane Highways." Vol. FHWA-RD-99-207, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C.

Hauer E., Harwood W.D., Council M.F., and Grffith S.M., (2002). " Estimating Safety by the Empirical Bayes Method: A Tutorial", Transportation Research Record No. 1784, TRB, The National Academies, Washington, D.C., pp. 126-131.

Hauer, E. (1997). "Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety". Pergamon Press, Elseviser Science Ltd., Oxford, U.K.

Hauer, E. and Bamfo, J. (1997). "Two Tools for Finding What Function Links the Dependent Variable to the Explanatory Variables", ICTCT, Lund, Sweden.

Hauer, E. and Persaud, B., (1983). "A common bias in before and after accident comparisons and its elimination". Transportation Research Record 905, 164-174. Transportation Research Board, Washington, D.C.

Hauer, E., C.N. Jerry, and Lovell, J. (1988). "Estimation of Safety at Signalized Intersections". Transportation Research Record 1185, 48-61. Transportation Research Board, Washington, D.C.

King, G. F. and Goldblatt, R. B. (1975). " Relationship of Accident Patterns to Type of Intersection Control". Transportation Research Record No. 540, TRB, The National Academies, Washington, D.C., pp. 1-12.

Lord, D., Wasgington, P. Simon, and Ivan, N.J. (2005). "Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory", Accident Prevention and Analysis, 35, 35-46.

Lord, D., and Persaud, N.B.. (2000). "Accident Prediction Models with and without Trend". Transportation Research Record 1717, 102-108. Transportation Research Board, Washington, D.C.

Manual on Uniform Traffic Control Devices, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., (2003).

McCullagh. P and Nelder. A. J (1989) "Generalised Linear Models", Chapman and Hall/CRC.

McGee, H., Taori S., and Persaud, N. B, (2003) "Crash Experience Warrant for Traffic Signals: NCHRP Report 491", Transportation Research Board, Washington, D.C.

Miaou, S. P. (1996). "Measuring the Goodnes-of-Fit of Accident Predicton Models", Publication No. FHWA-96-040, Federal Highway Administration, Washington, D.C.

MTO 2002 "Ministry of Transportation of Ontario" web site
http://www.mto.gov.on.ca/english/safety/orsar/orsar02/

Myers H.R, Montgomery, C.D., and Vining G. G. (2002). "Generalized Linear Models: With Applications in Engineering and the Sciences", A Wiley-Interscience Publication, John Wiley and Sons, INC.

Myers. H.R., (1990). "Classical and Modern Regression with Applications: Second Edition", Duxbury Classic Series, Thomson Learning Academic Recourse Centre, Duxbury.

Persaud, B., (1988). "Do Traffic Signals Affect Safety? Some Methodological Issues" Transportation Research Record No. 1183, TRB, The National Academies, Washington, D.C., pp. 37-47.

Persaud, B., et al. (2005). "A Multi-jurisdictional Safety Evaluation of Red Light Cameras", Transportation Research Board, Washington, D.C. CD-ROM paper.

Persaud, B., Hauer, E., Retting. R, Vallurupalli, R and Mucsi, K. (1997). "Crash Reduction Related to Traffic Signal Removal in Philadelphia", Accident Analysis and Prevention, Vol. 29, No.6., pp. 803-810.

Persaud, N. B, McGee, H., Lyon, Craig., and Lord, D., (2003). "Development of a Procedure for Estimating the Expected Safety Effects of a Contemplated Traffic Signal Installation", Transportation Research Record No. 1840, Transportation Research Board, Washington, D.C.

Persaud, N. B, R. Retting, P. Garder, and D. Lord, (2001) "Safety Effect of Roundabout Conversions in the U.S.: Empirical Bayes Observation Before-After Study," Transportation Research Record No. 1757, TRB, The National Academies, Washington, D.C., pp. 1-8.

Persaud, N.B., (2001). "Statistical Method in Highway Safety Analysis: A Synthesis of Highway Practice". NCHRP Synthesis 295. Transportation Research Board, Washington, D.C.

Persaud, N.B, Lord, D and Palmisano, J. (2002) "Calibration and Transferability of Accident Prediction Models for Urban Intersections", Transportation Research Record, No. 1784, Washington. D.C. p.p. 57-64.

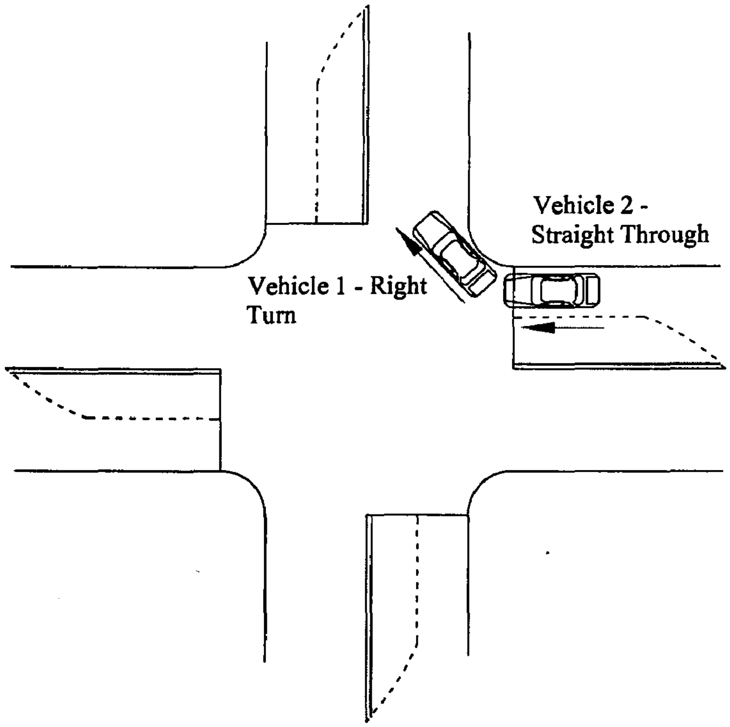
Persaud, N.B. (2001). "Statistical Methods in Highway Safety Analysis: A systhesis of Highway Practice." NCHRP Synthesis 295, Transportation Research Board, National Academy Press, Washington, D.C.

Poch, M. and Mannering, F. (1996). " Negative Binomial Analysis of Intersection-Accident Frequencies". American Society of Civil Engineering, Journal of Transportation Engineering, Vol. 122, No.2.

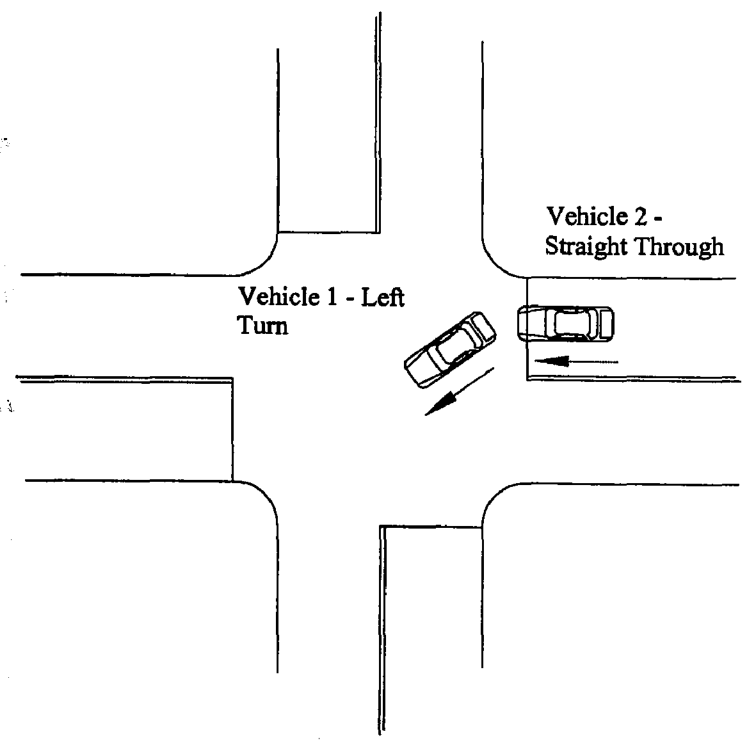
SAS Institute Inc., SAS/STAT® Software: The GENMOD Procedure, Release 8.00, http://gsbwww.uchicago.edu/computing/research/SASManual/stat/chap29/index.htm

Appendices

Appendix A – Alternative Rear-End Crash Types



Vehicle 1 turns right while vehicle 2 heads straight through the intersection

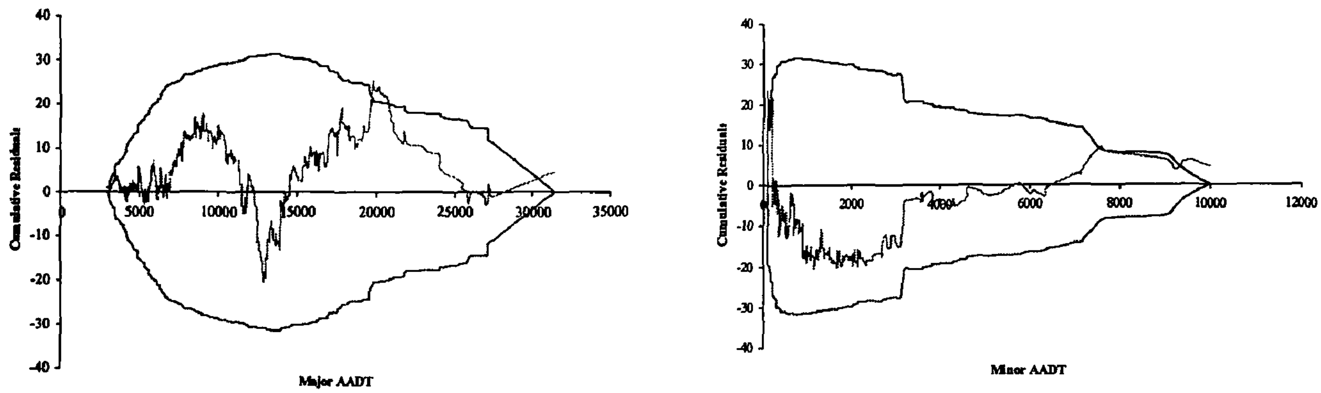


Vehicle 1 turns left while vehicle 2 heads straight through the intersection

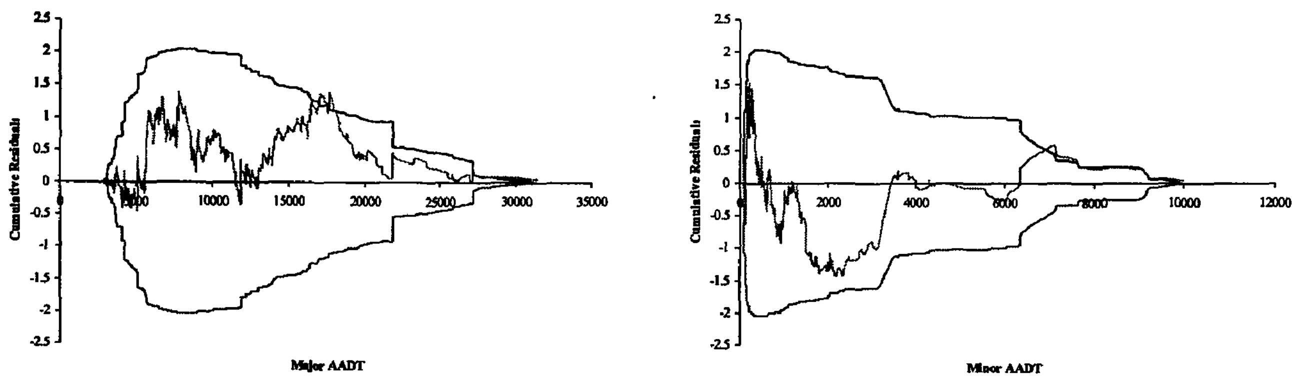
Appendix B - Cumulative Residual (CURE) plots for SPFs in Chapter 5

CURE PLOTS for SPFs calibrated from the California Reference Groups

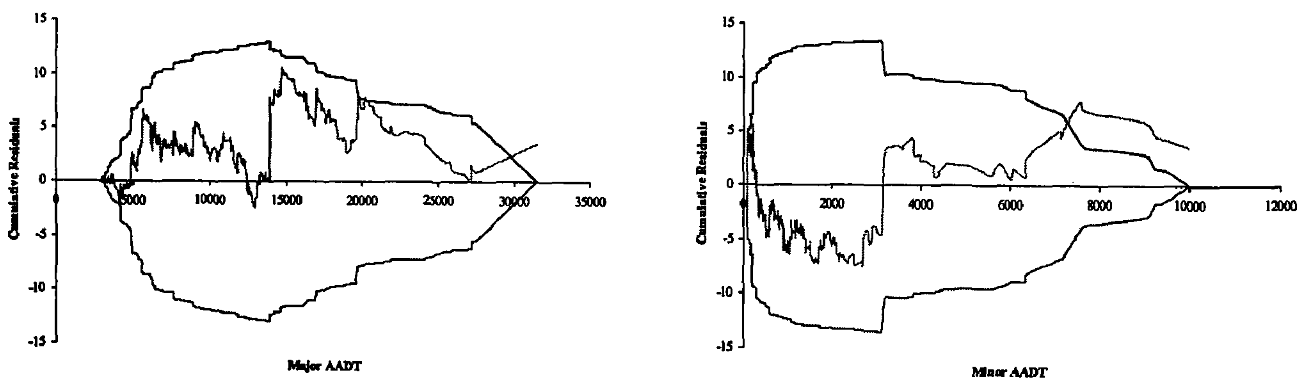
Total SPFs – 3-legged intersections with 2 lanes on major approach



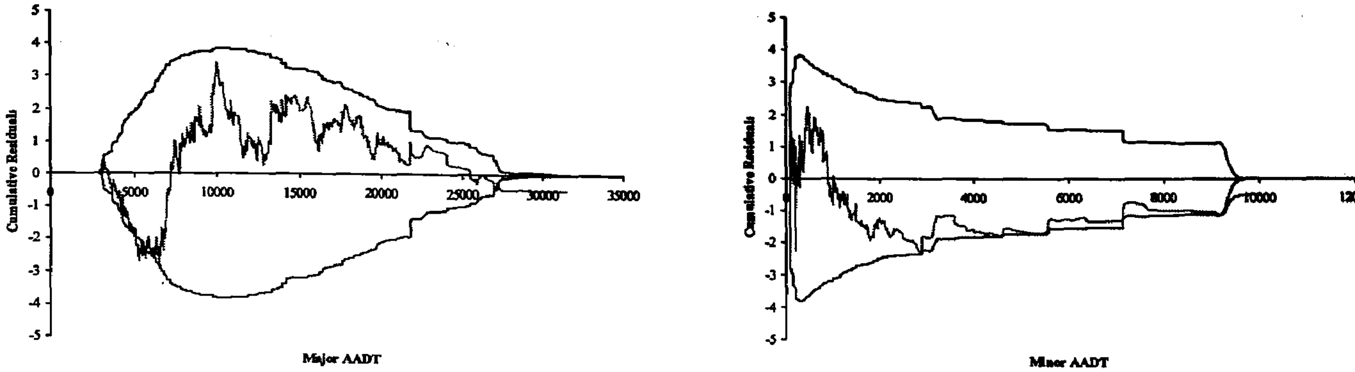
All



Right Angle

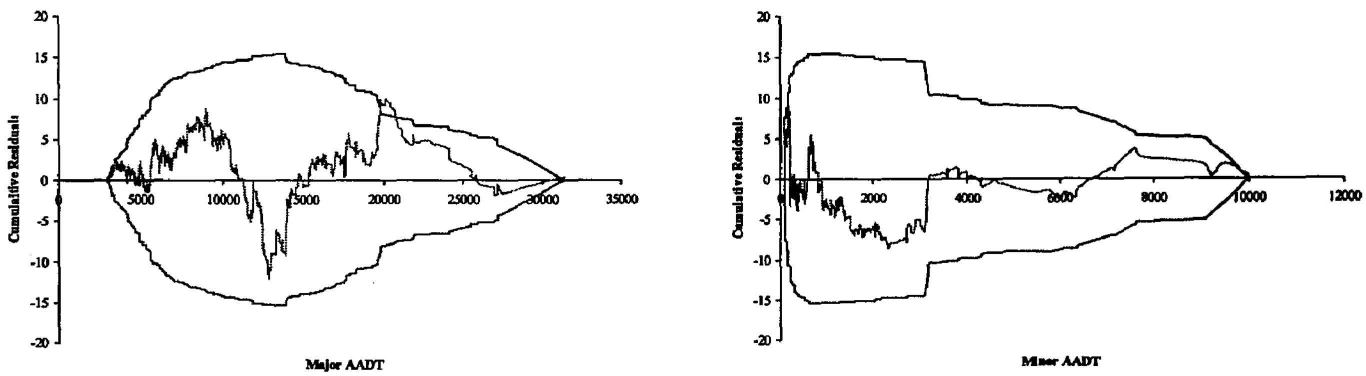


Left Turn

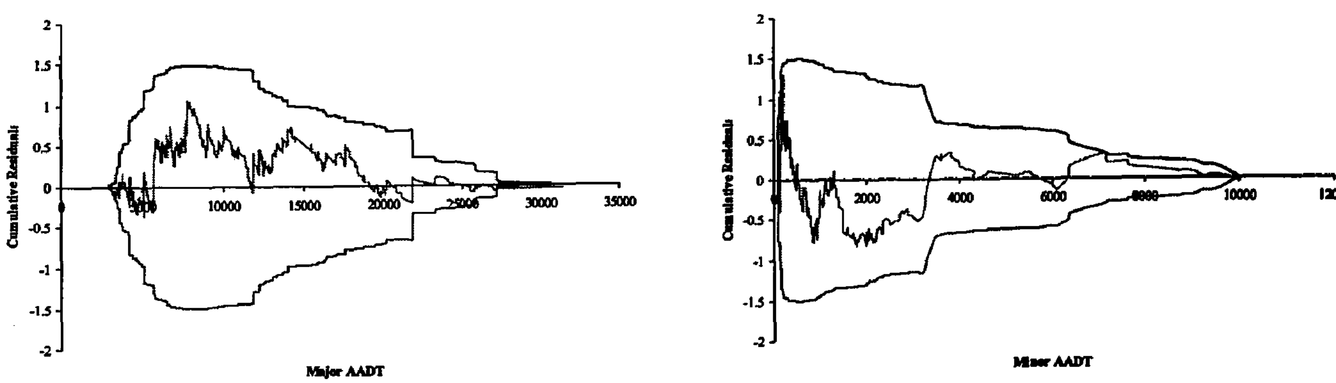


Rear End

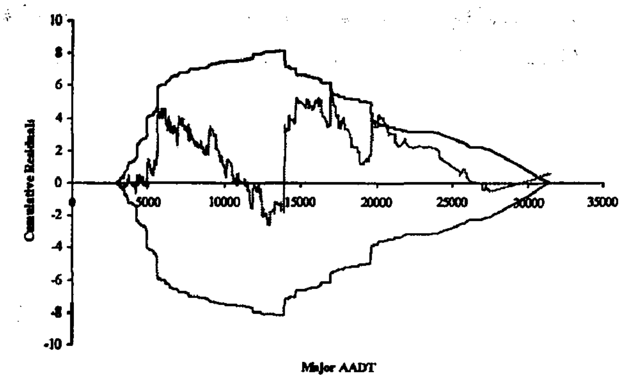
Injury SPFs – 3-legged intersections with 2 lanes on major approach

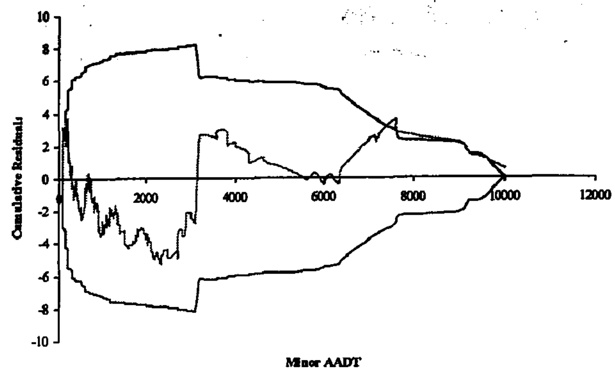


All

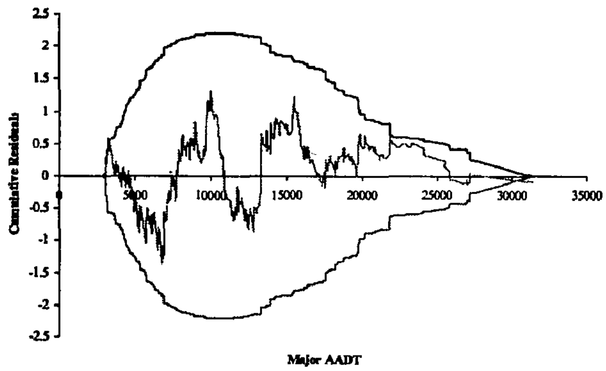


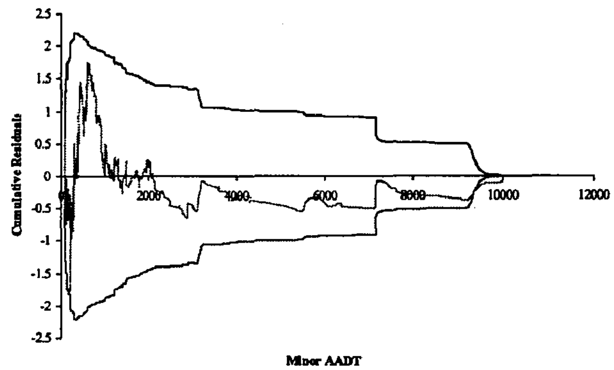
Right Angle





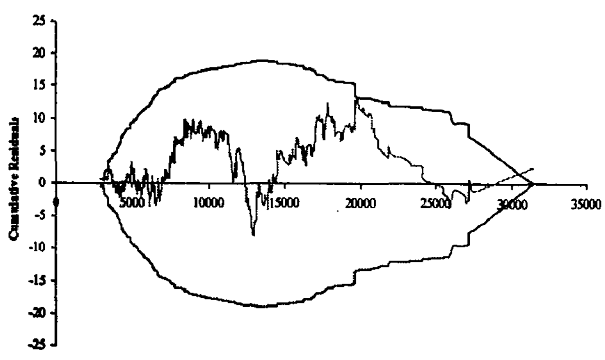
Left Turn

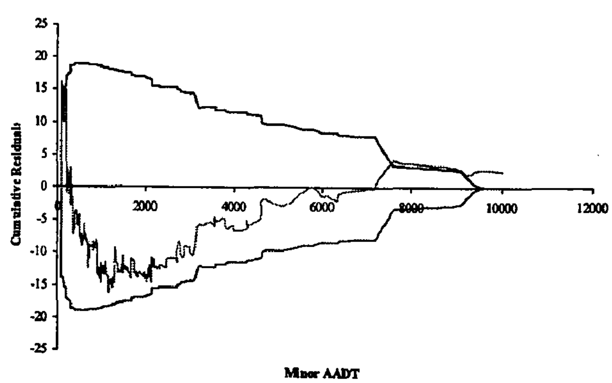




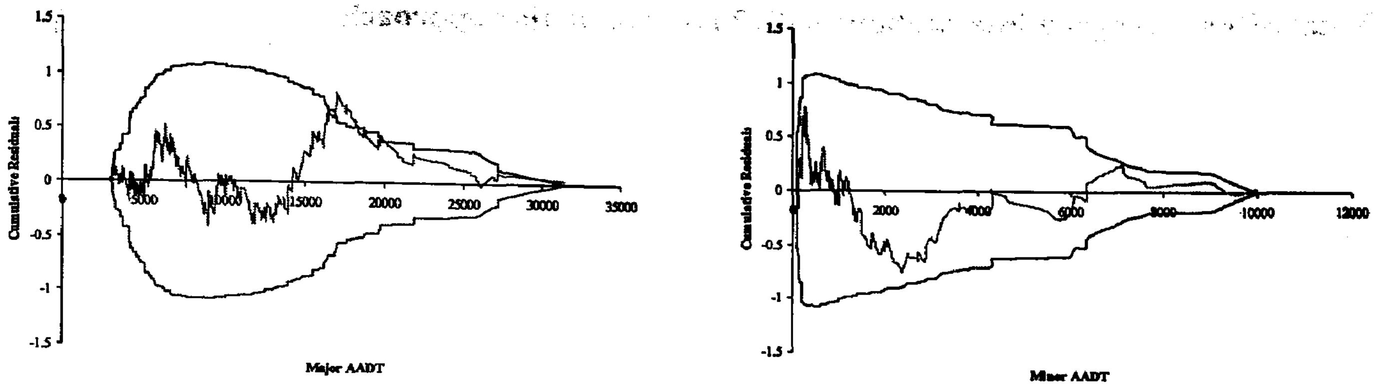
Rear End

PDO SPFs – 3-legged intersections with 2 lanes on major approach

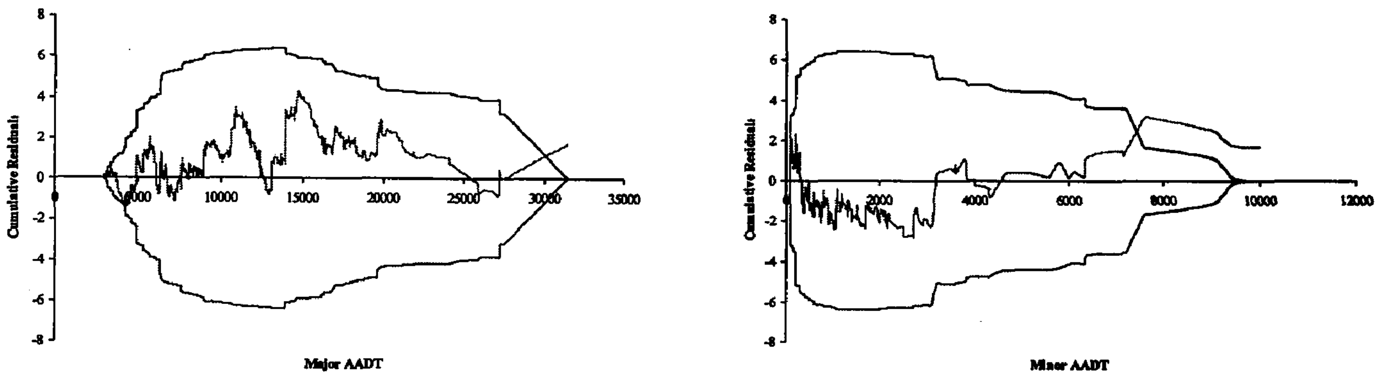




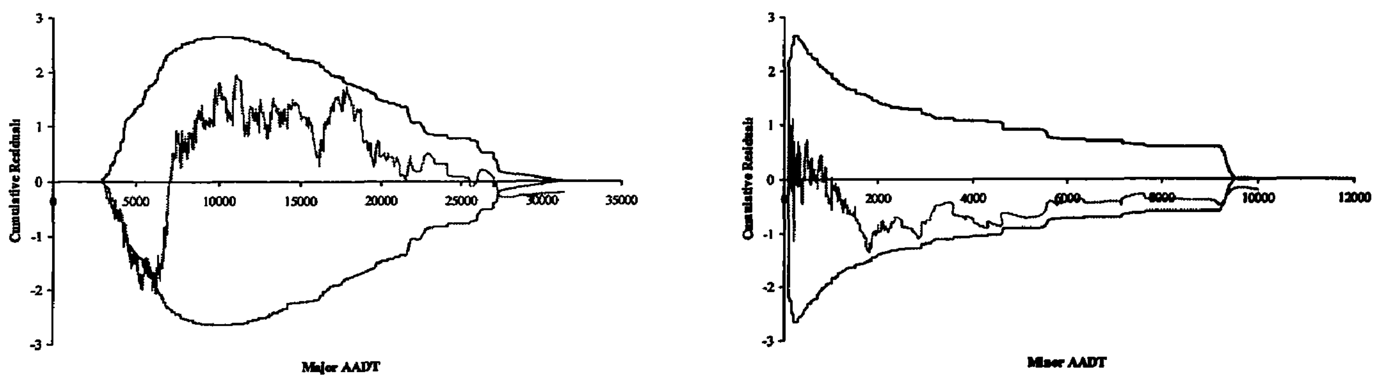
All



Right Angle

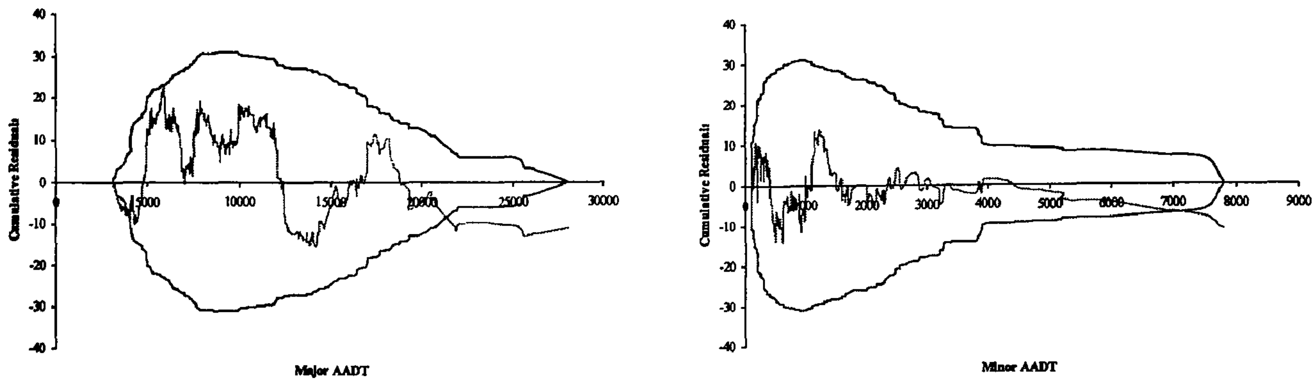


Left Turn

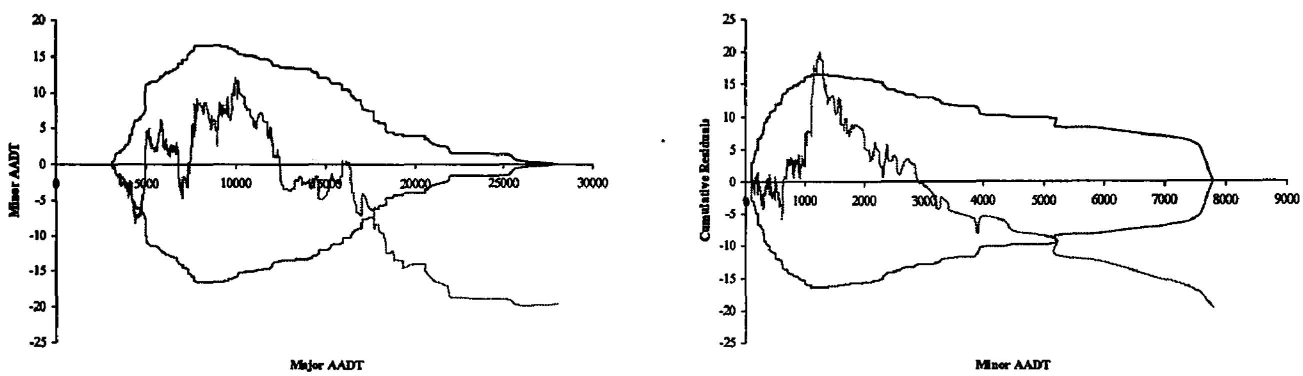


Rear End

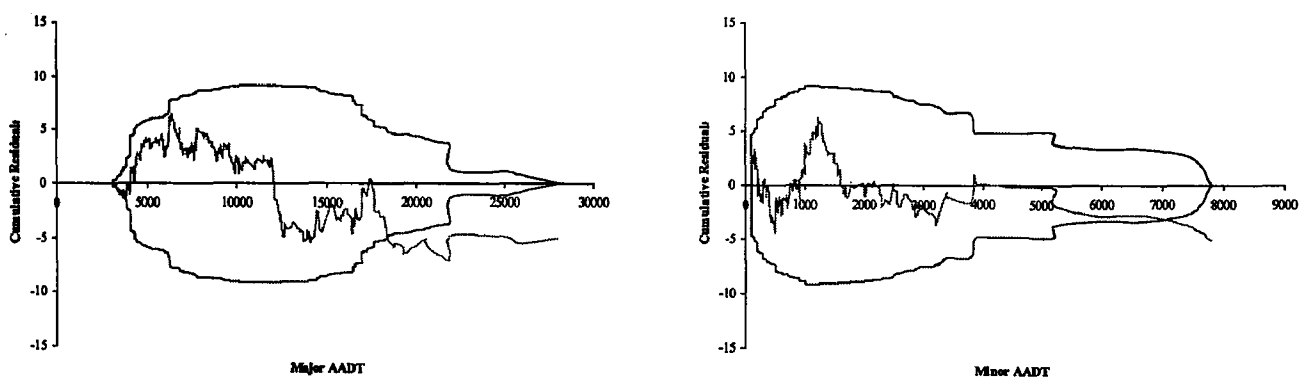
Total SPFs – 4-legged intersections with 2 lanes on major approach



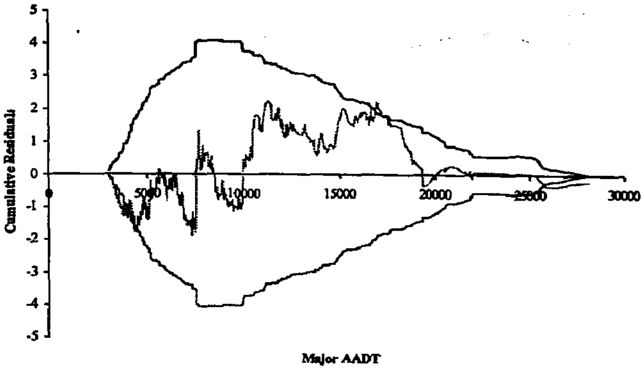
All

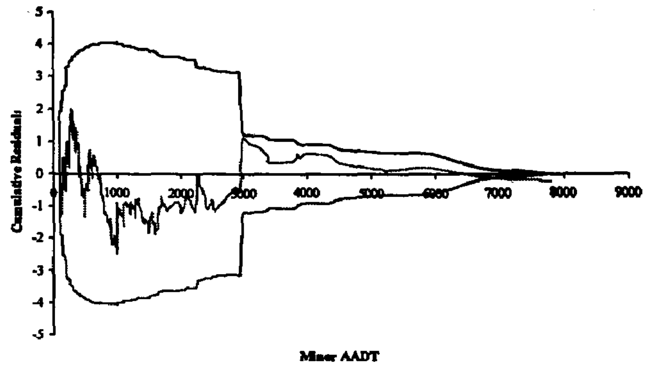


Right Angle



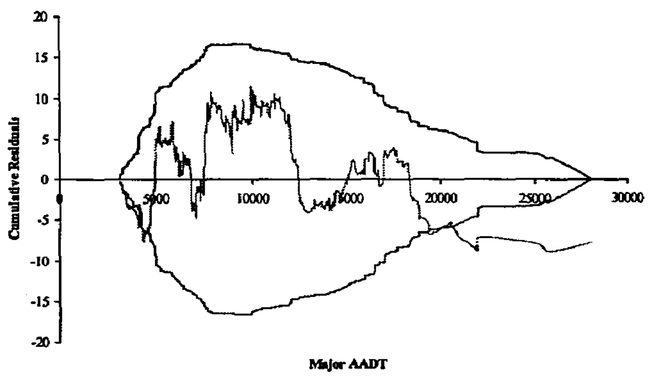
Left Turn

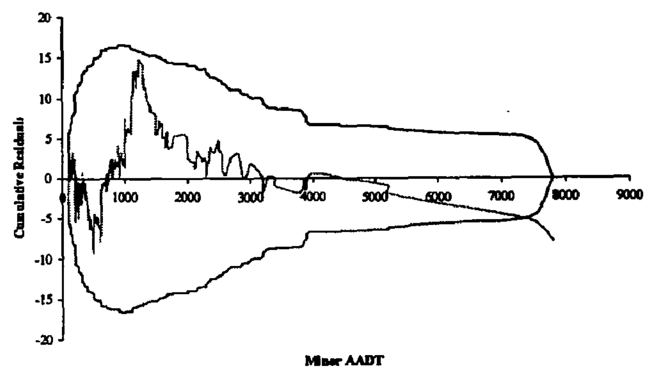




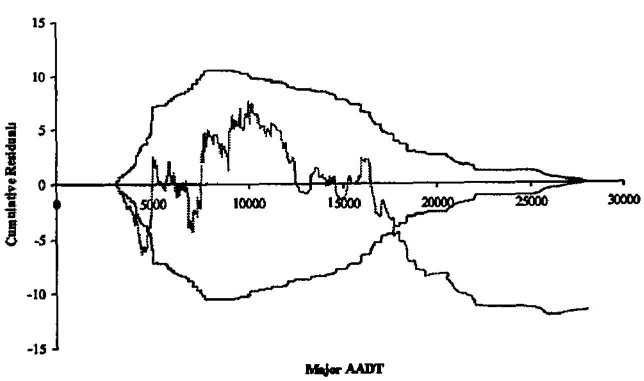
Rear End

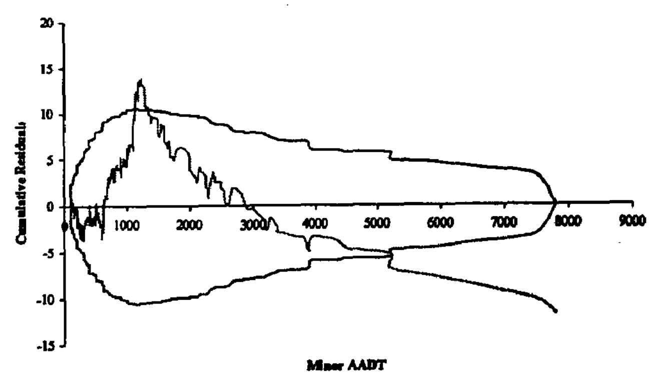
Injury SPFs – 4-legged intersections with 2 lanes on major approach



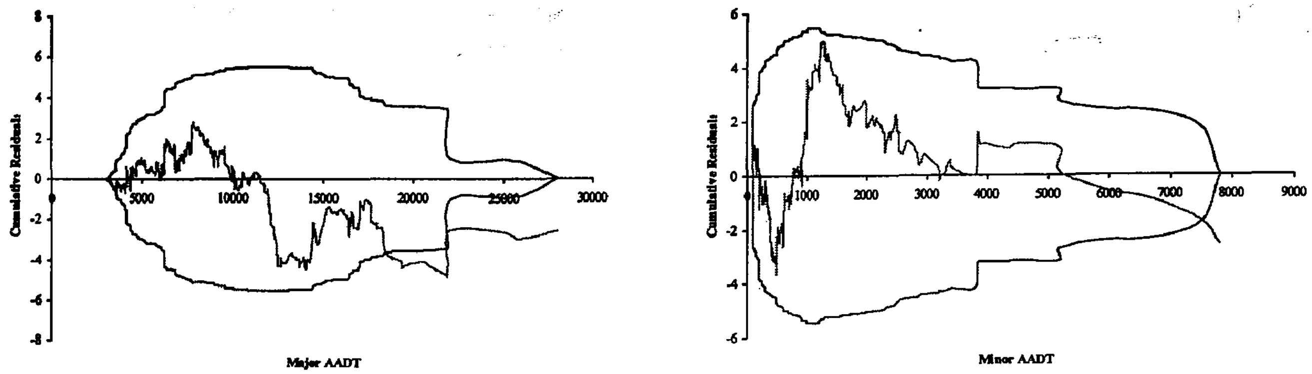


All

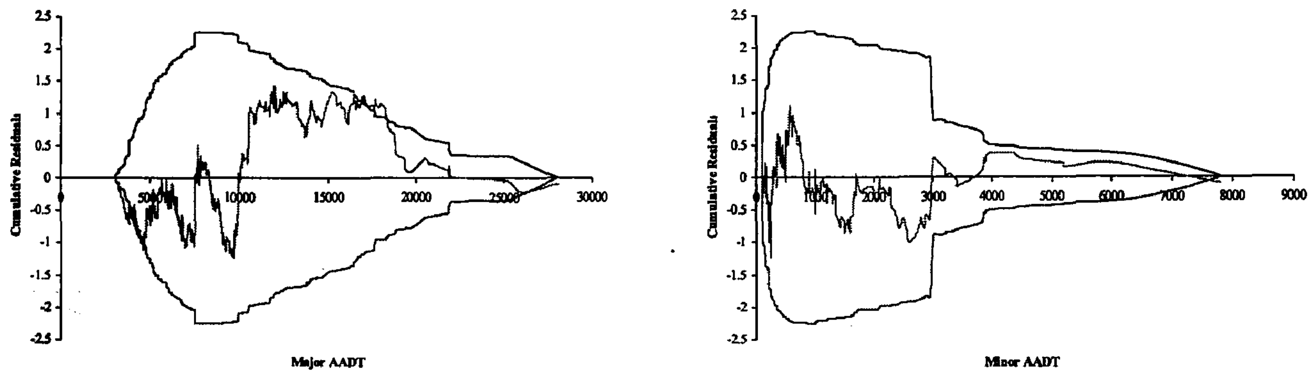




Right Angle

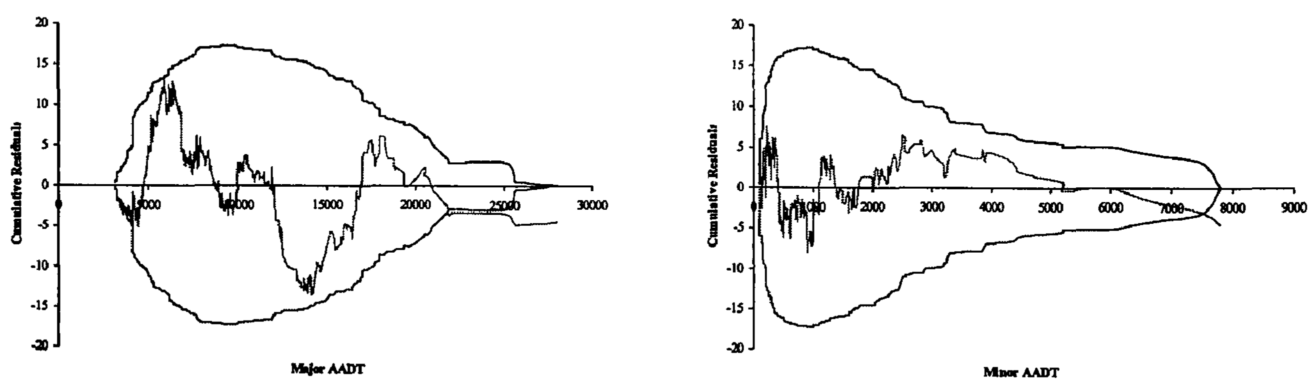


Left Turn

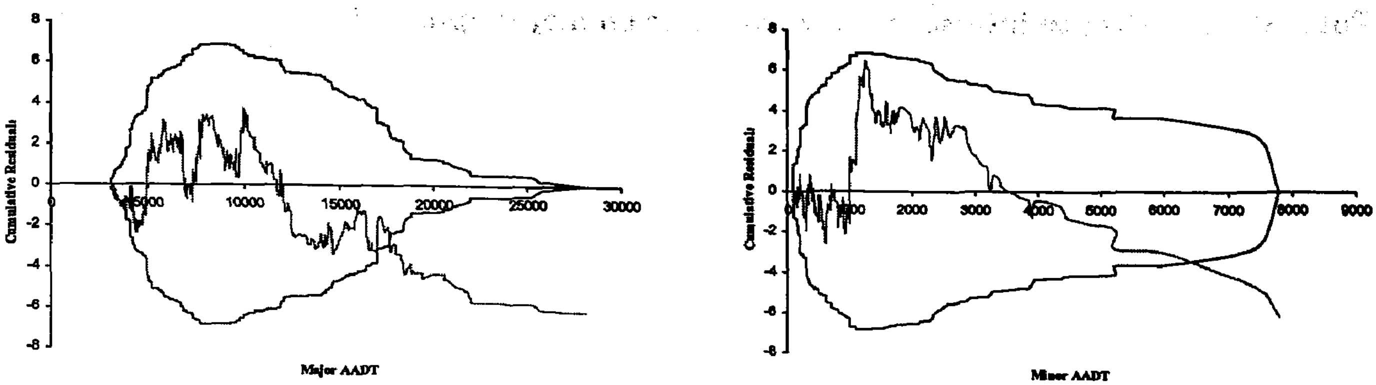


Rear End

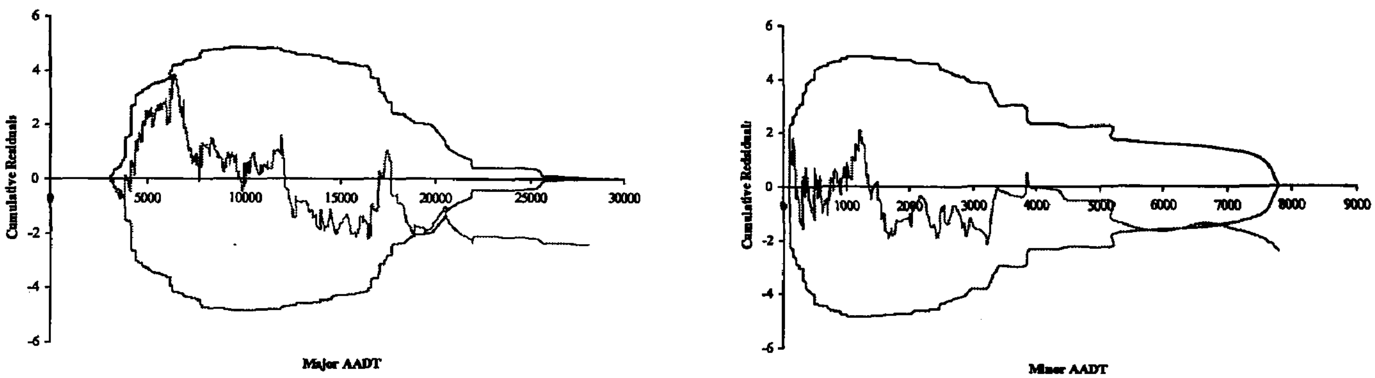
PDO SPFs – 4-legged intersections with 2 lanes on major approach



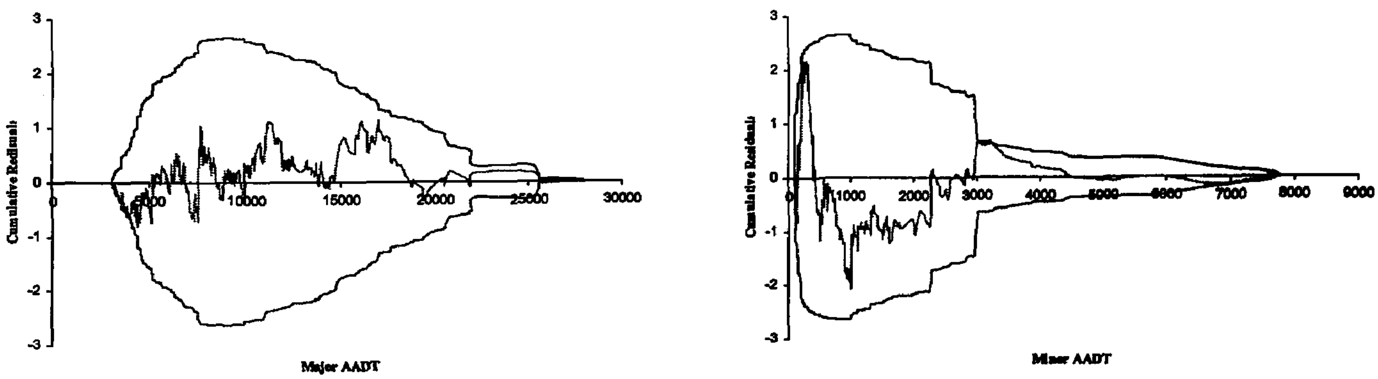
All



Right Angle

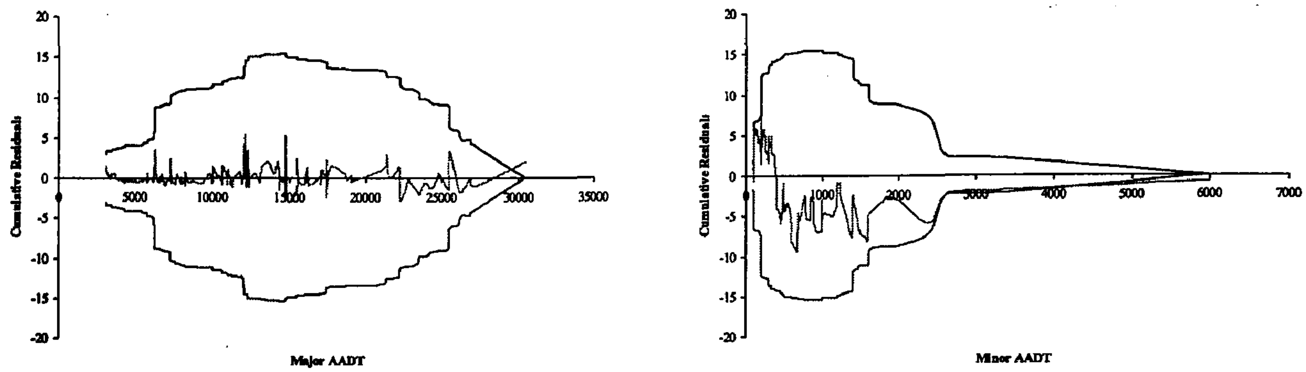


Left Turn

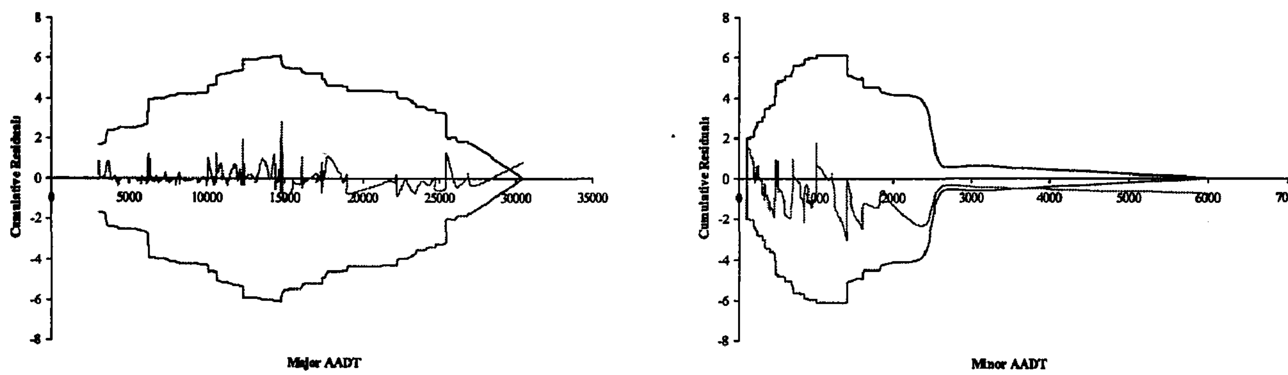


Rear End

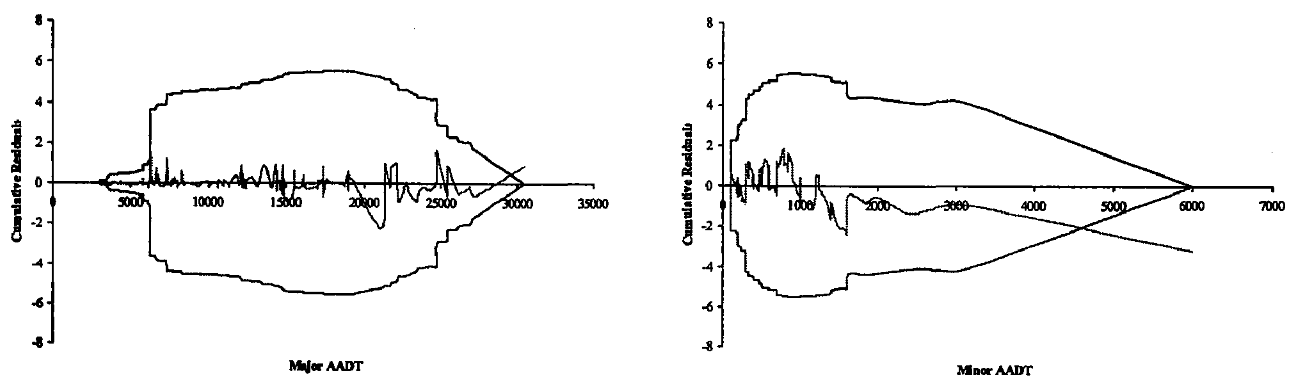
Total SPFs – 4-legged intersections with 4 lanes on major approach



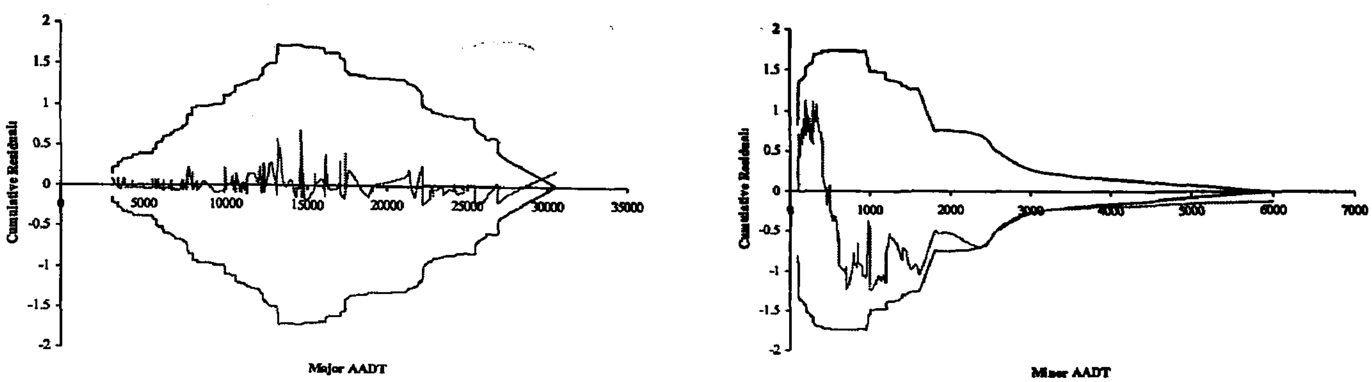
All



Right Angle

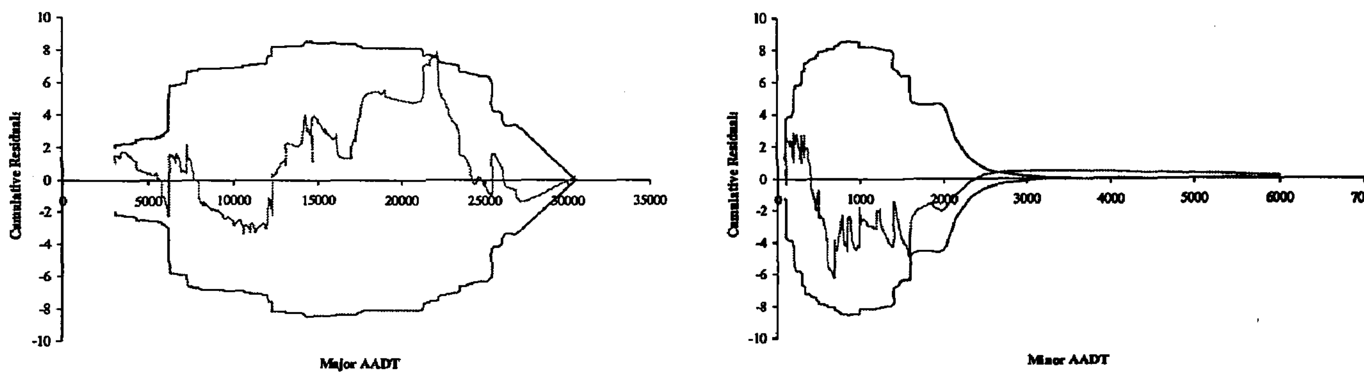


Left Turn

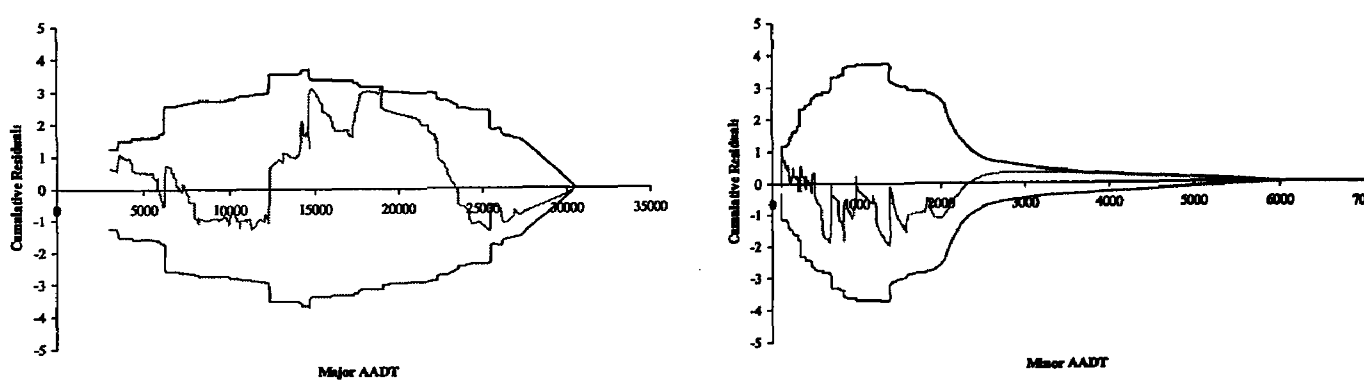


Rear End

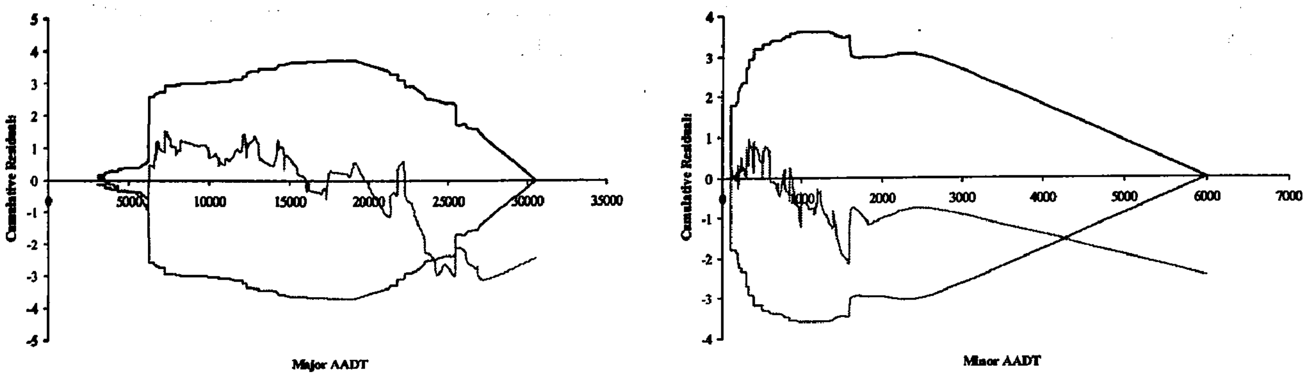
Injury SPFs – 4-legged intersections with 4 lanes on major approach



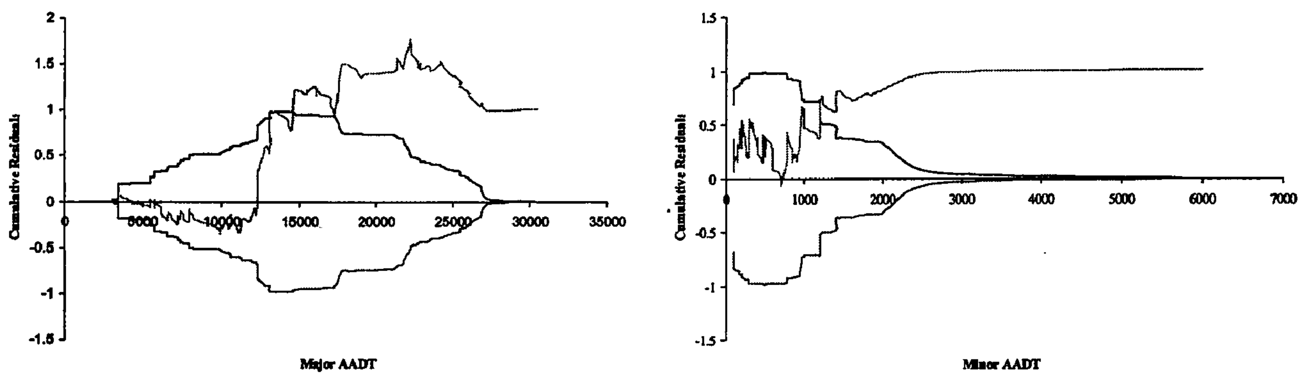
All



Right Angle

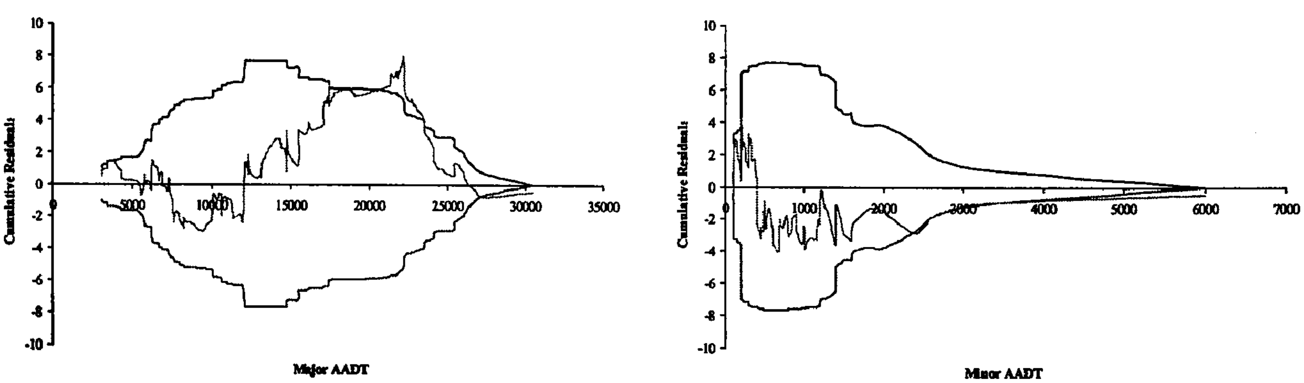


Left Turn

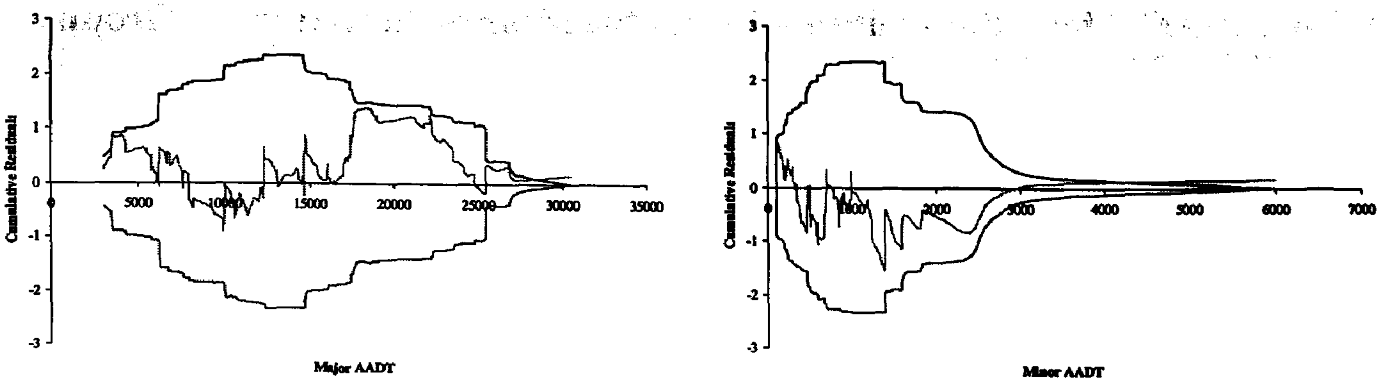


Rear End

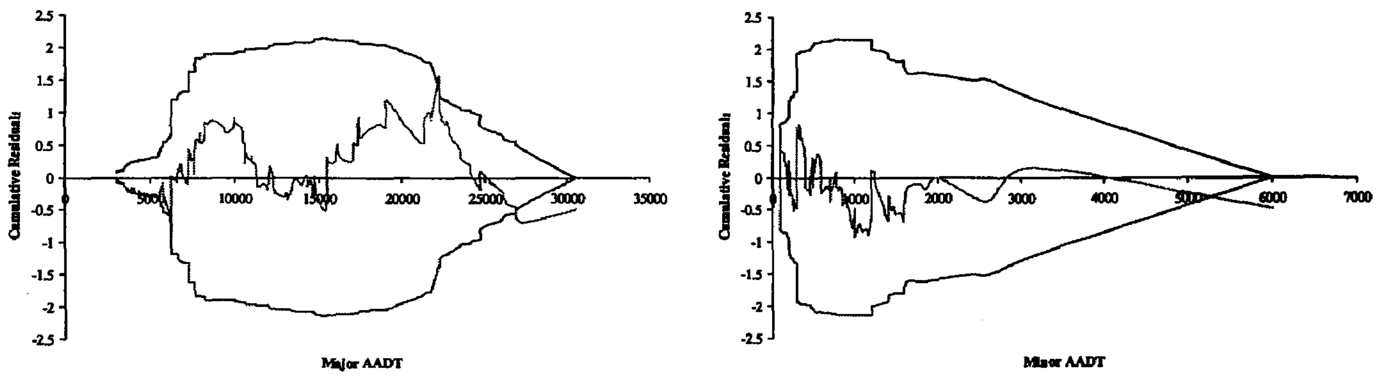
PDO SPFs – 4-legged intersections with 4 lanes on major approach



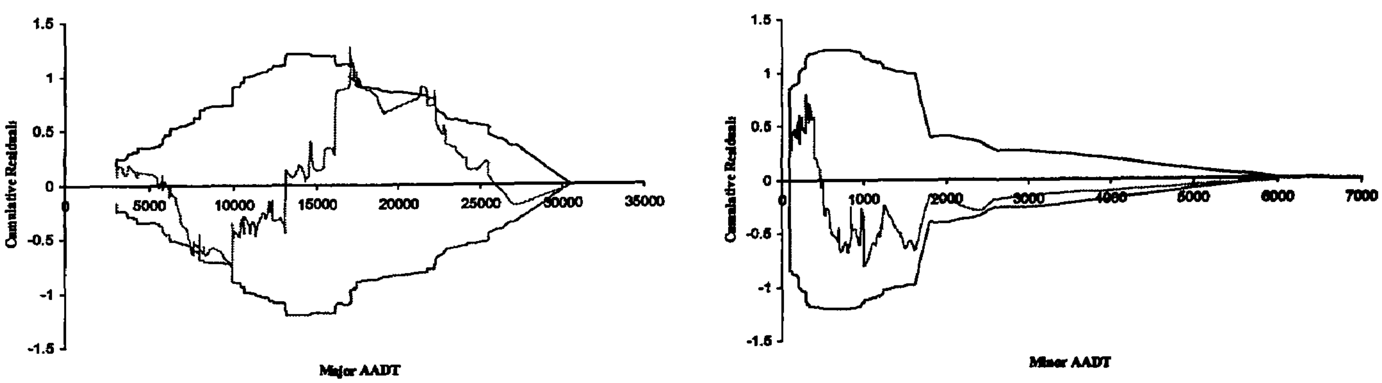
All



Right Angle



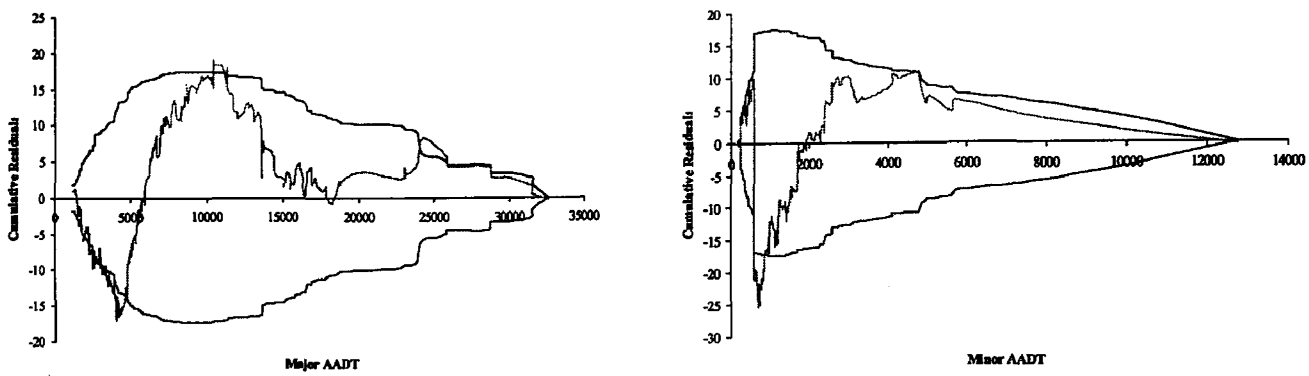
Left Turn



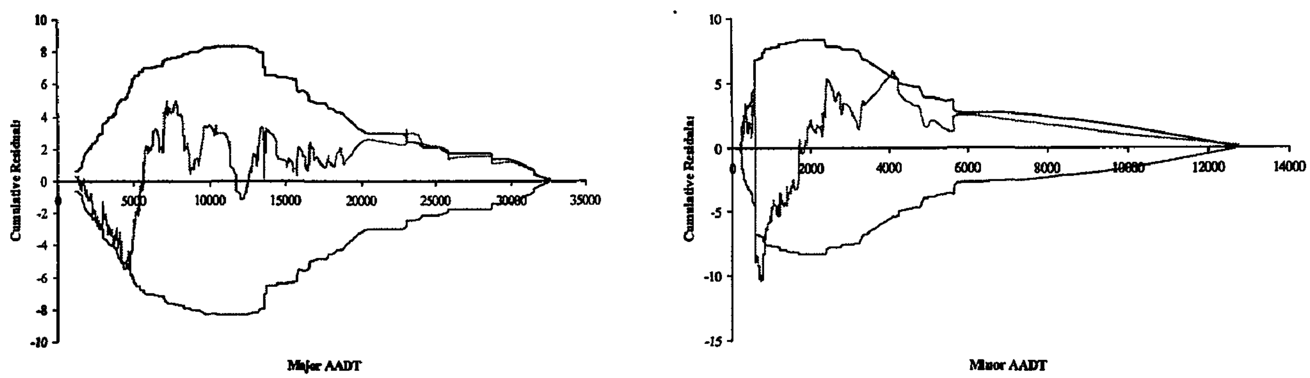
Rear End

CURE PLOTS for SPFs calibrated from the Minnesota Reference Groups

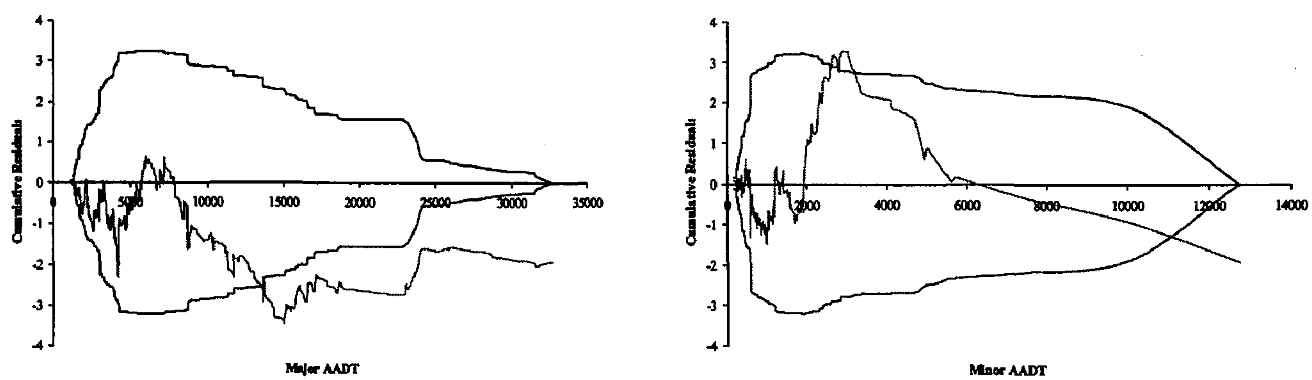
Total SPFs – 3-legged intersections



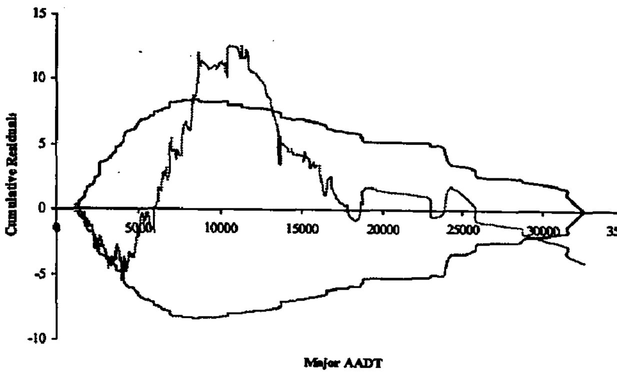
All

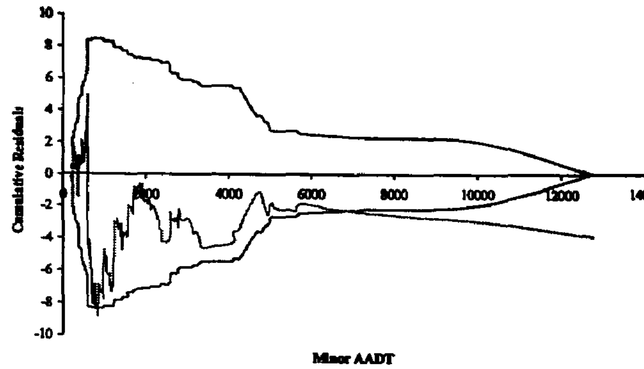


Right Angle



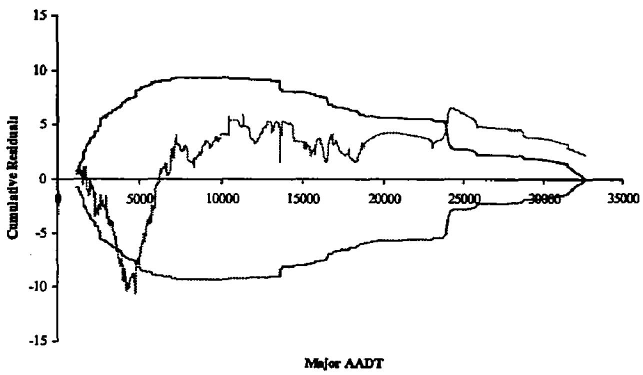
Left Turn

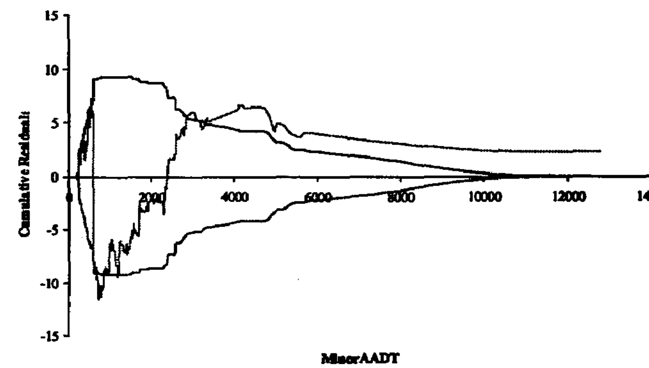




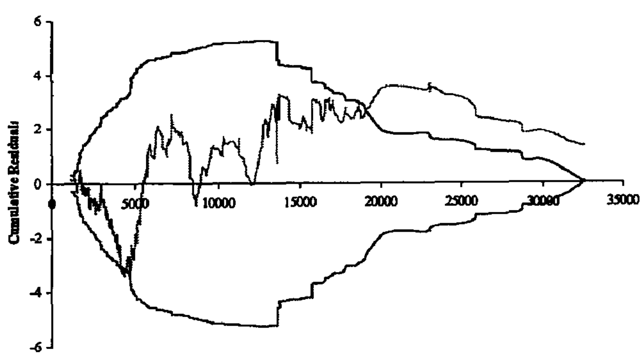
Rear End

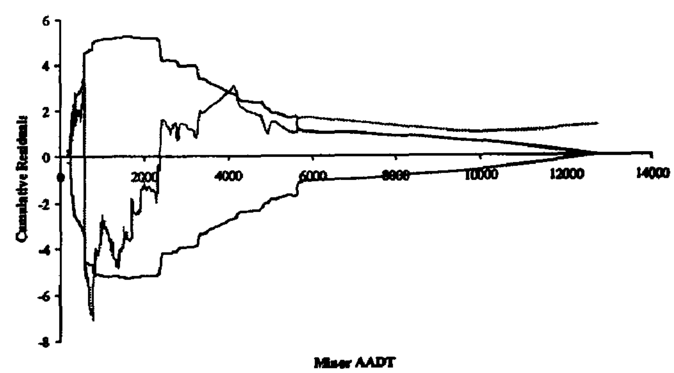
Injury SPFs – 3-legged intersections



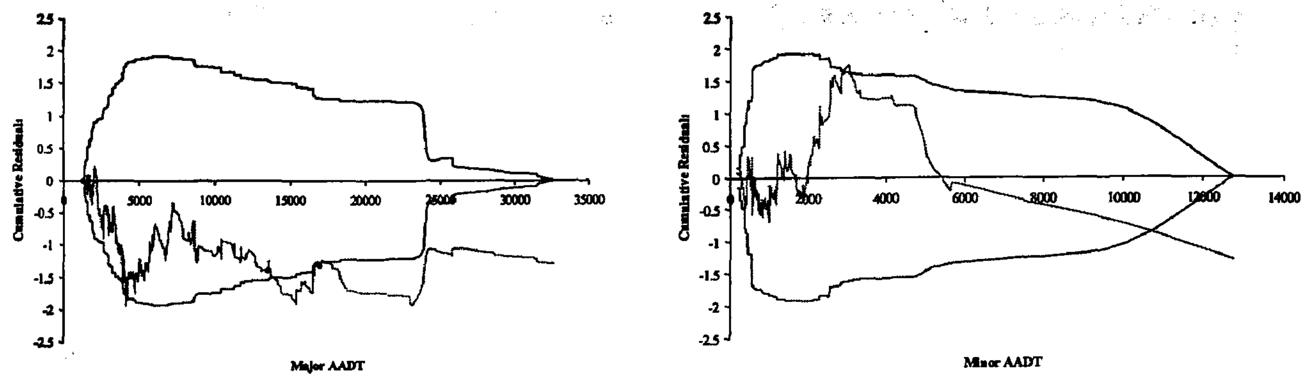


All

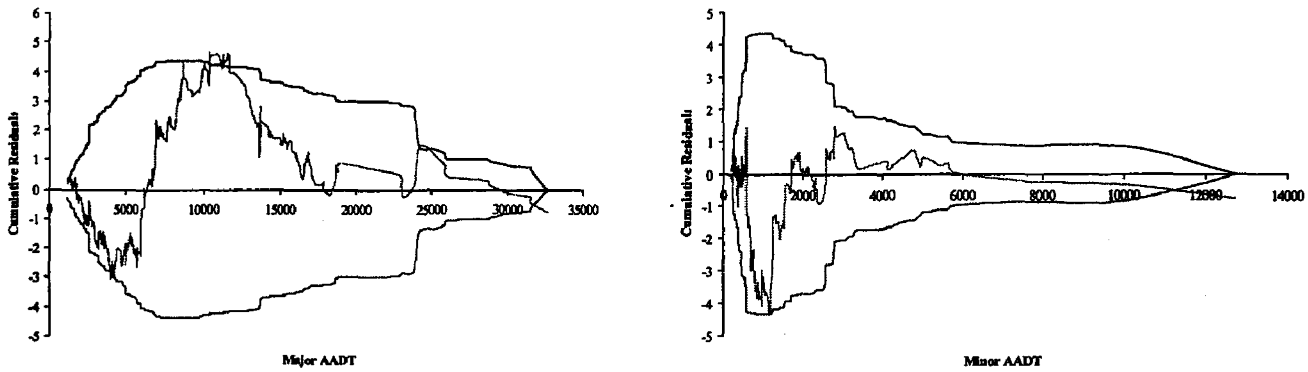




Right Angle

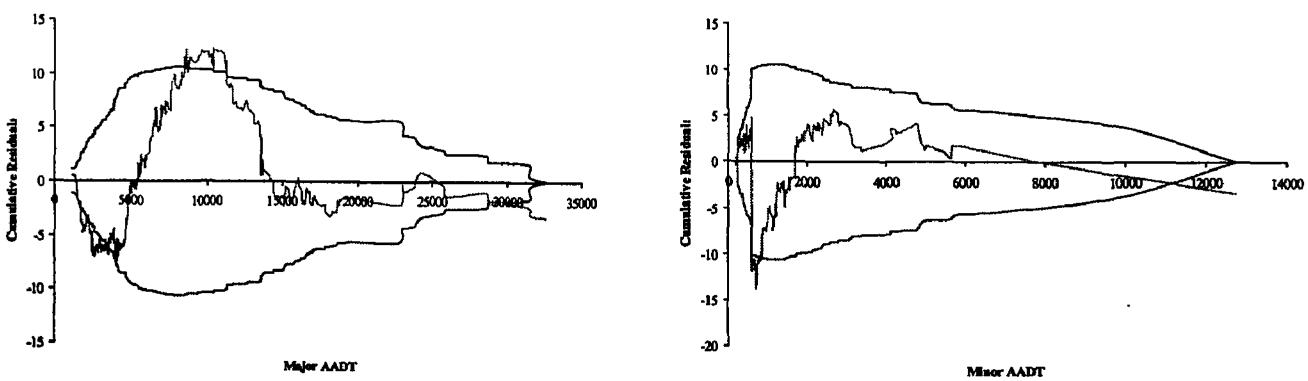


Left Turn

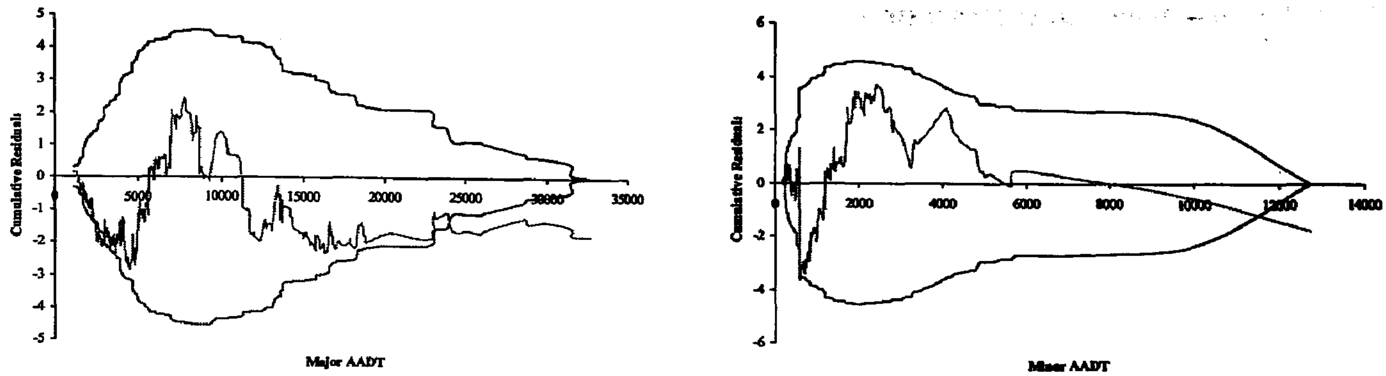


Rear End

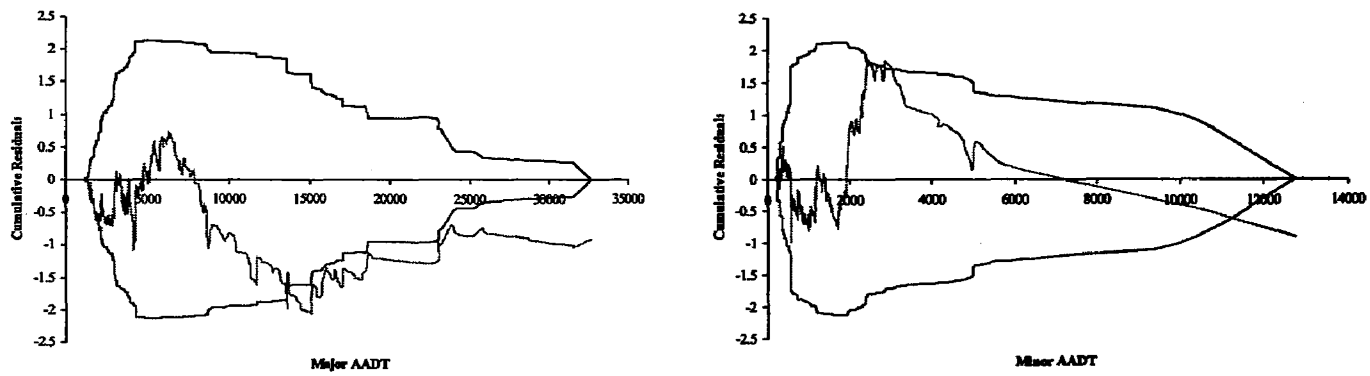
PDO SPFs – 3-legged intersections



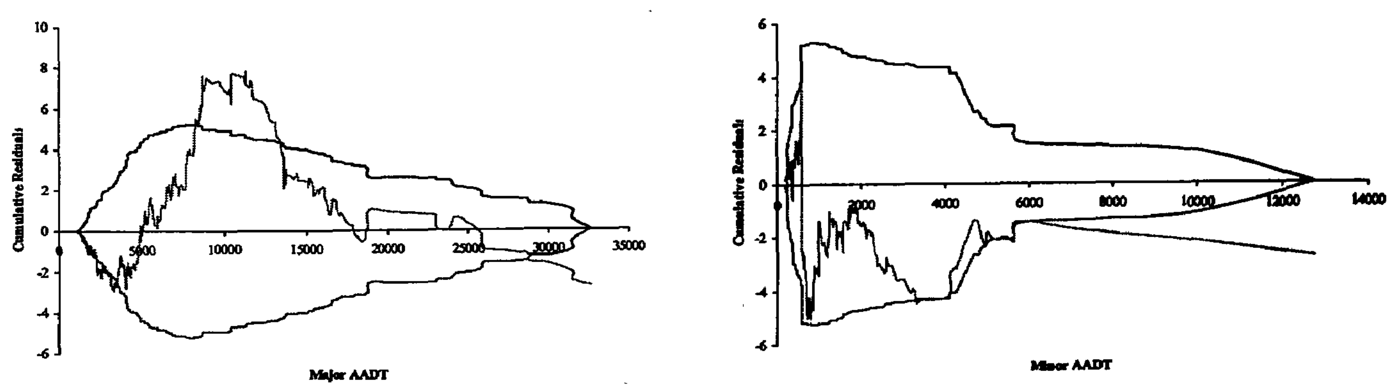
All



Right Angle

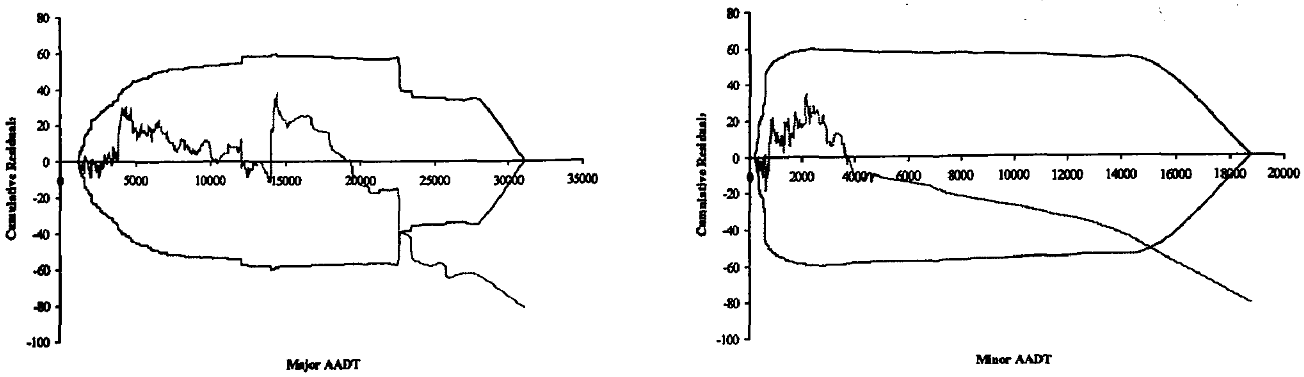


Left Turn

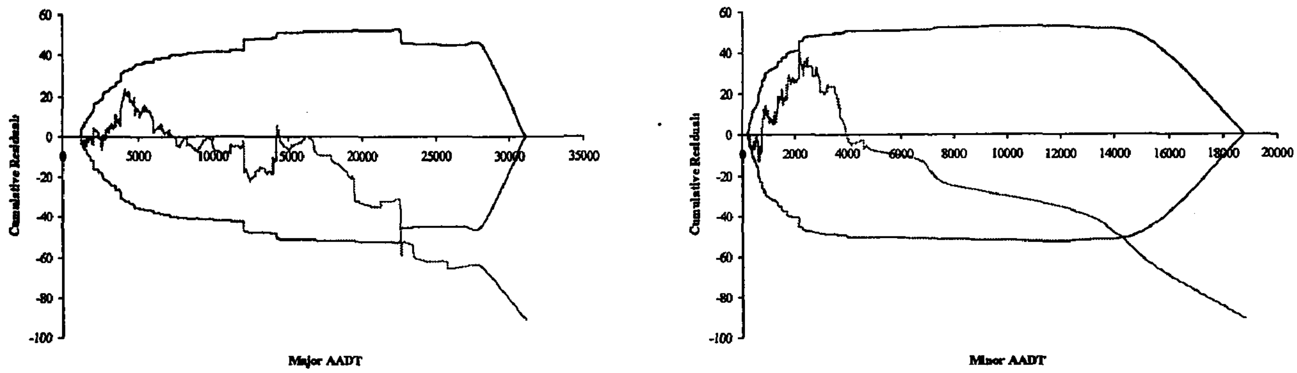


Rear End

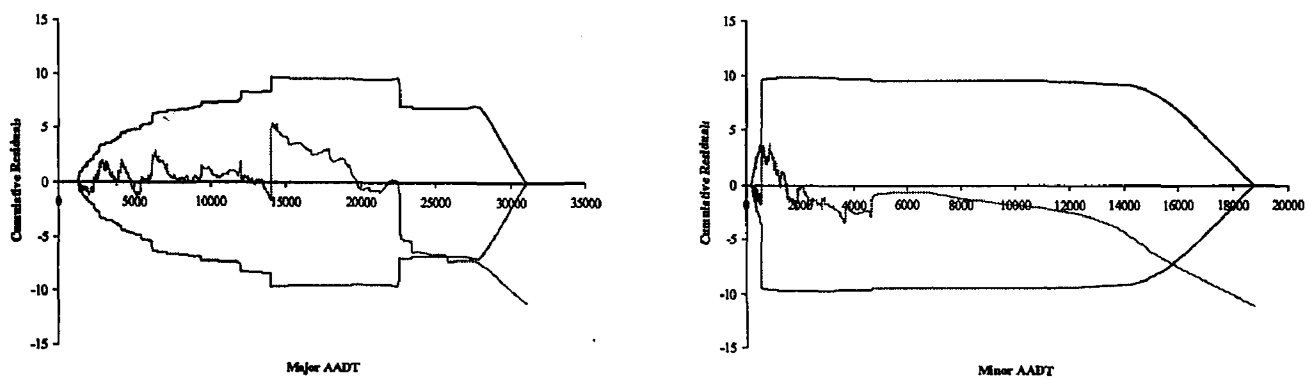
Total SPFs – 4-legged intersections



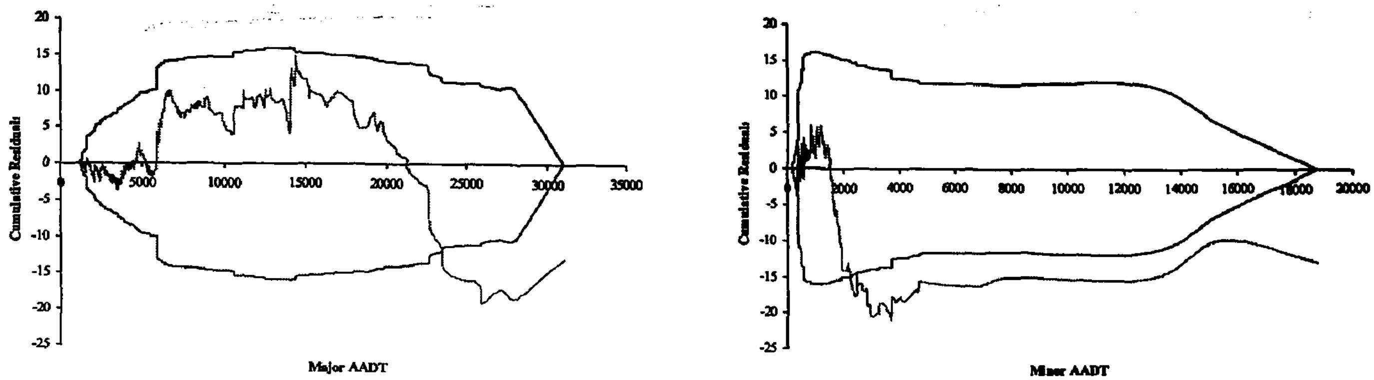
All



Right Angle

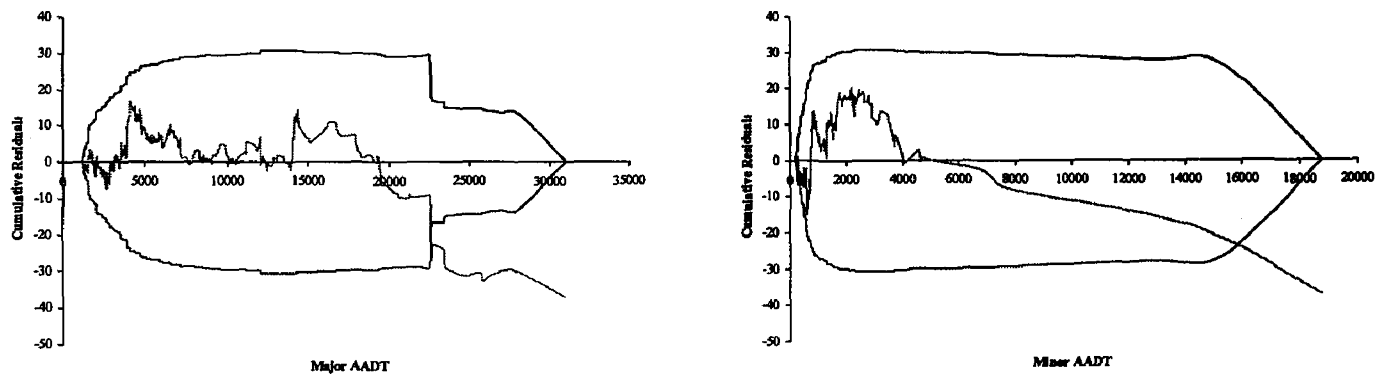


Left Turn

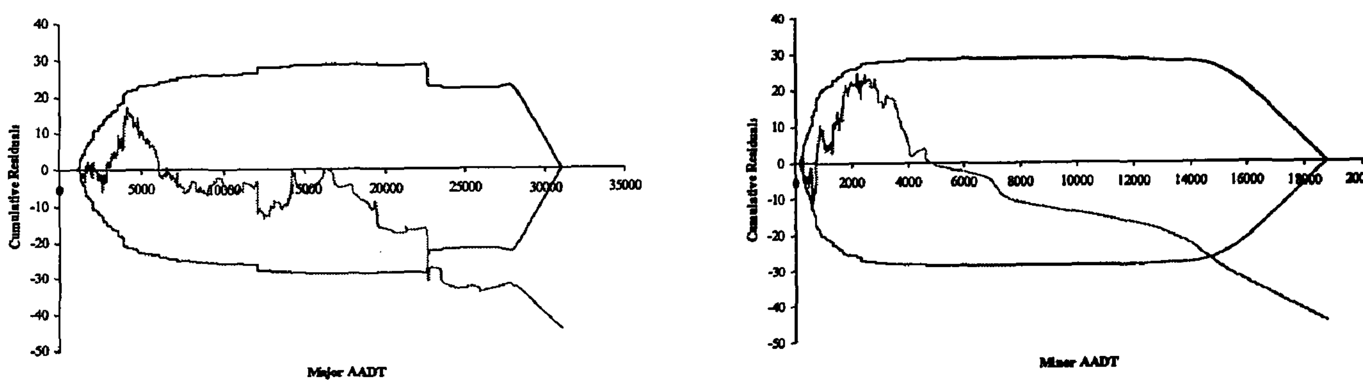


Rear End

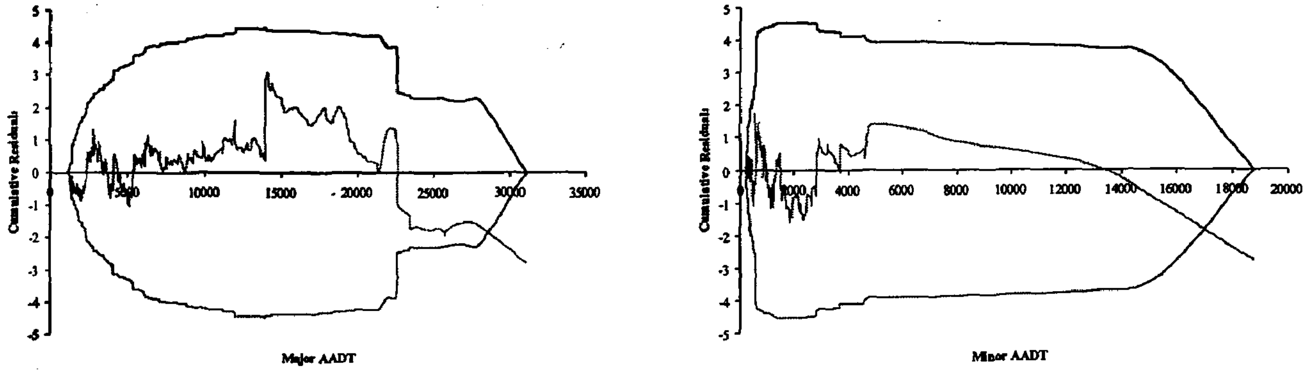
Injury SPFs – 4-legged intersections



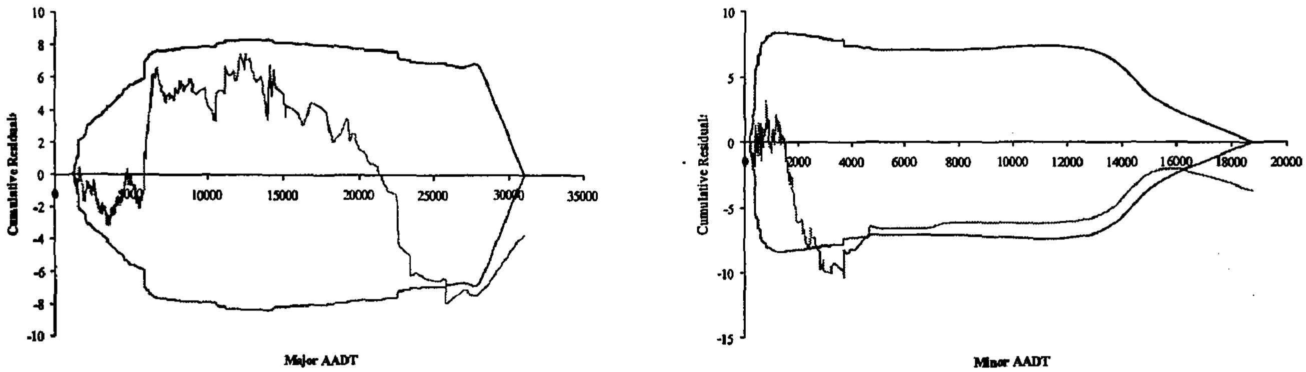
All



Right Angle

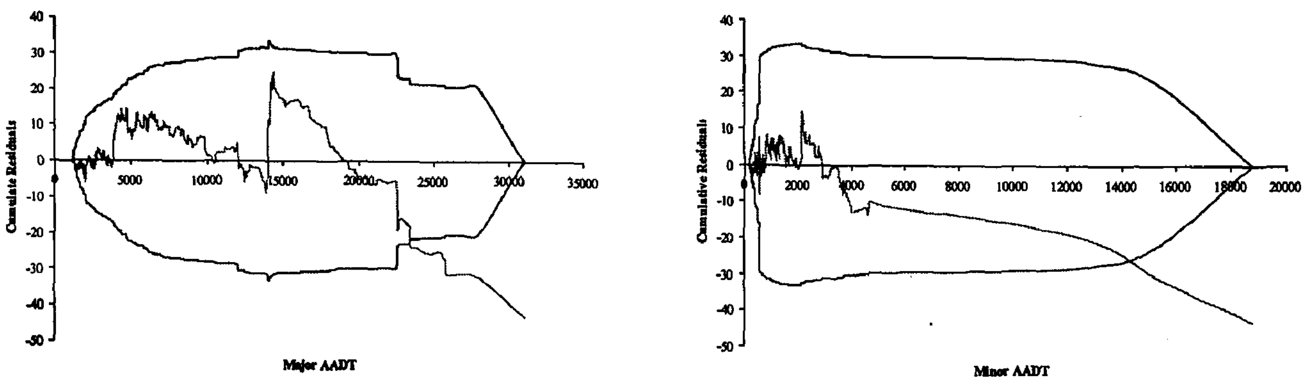


Left Turn

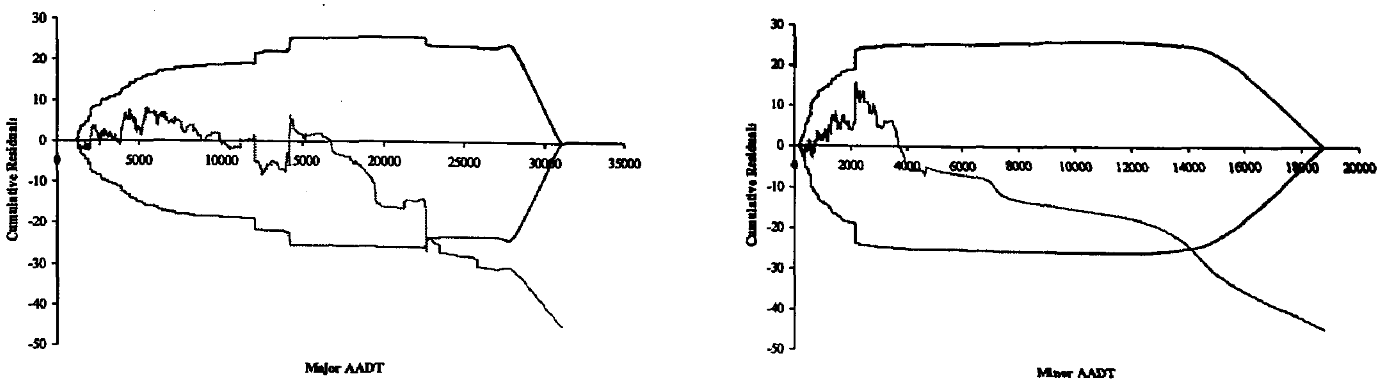


Rear End

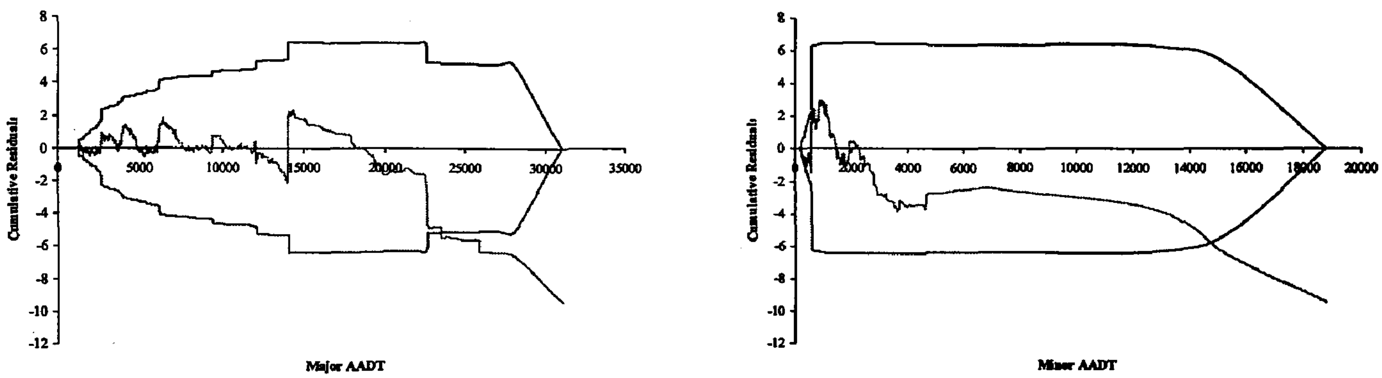
PDO SPFs – 4-legged intersections



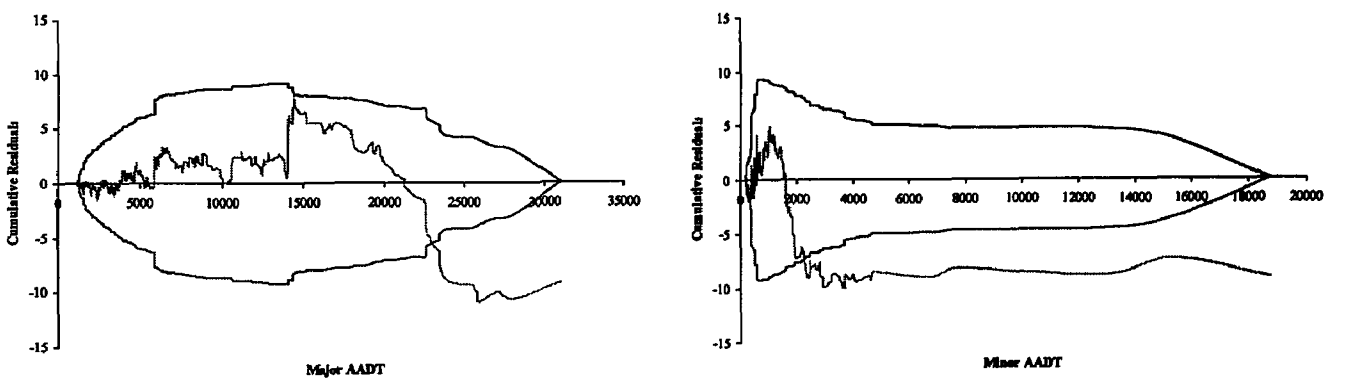
All



Right Angle



Left Turn



Rear End