Spring 2015

A probabilistic model of benefit-cost analysis for highway construction projects

Yue He
Purdue University

Graduate School Form 30
Updated 1/15/2015

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Yue He

Entitled
A PROBABILISTIC MODEL OF BENEFIT-COST ANALYSIS FOR HIGHWAY CONSTRUCTION PROJECTS

For the degree of Master of Science in Building Construction Management

Is approved by the final examining committee:

Yi Jiang
Chair

Bryan J Hubbard

James L Jenkins

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University's "Policy of Integrity in Research" and the use of copyright material.

Approved by Major Professor(s): Yi Jiang

Approved by: Bryan J Hubbard 4/8/2015

Head of the Departmental Graduate Program Date

A PROBABILISTIC MODEL OF BENEFIT-COST ANALYSIS FOR HIGHWAY CONSTRUCTION PROJECTS

A Thesis

Submitted to the Faculty

of

Purdue University

by

Yue He

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science in Buliding Construction Management

May 2015

Purdue University

West Lafayette, Indiana

For my mother, Li Xiaoqiong

ACKNOWLEDGEMENTS

I would like to give my greatest gratitude to Professor Yi Jiang for his supports, advices, and encouragements in guiding me not only to finish this work but also to complete my entire graduate study. I would also like to thank Professor Bryan Hubbard and Professor James Jenkins for their timely feedbacks and suggestions.

TABLE OF CONTENTS

Page

LIST OF TABLES ... vi
LIST OF FIGURES ... vii
ABSTRACT ... ix
CHAPTER 1. INTRODUCTION ... 1
1.1 Background ... 1
1.2 Significance ... 2
1.3 Statement of Purpose ... 3
1.4 Research Question ... 3
1.5 Assumptions ... 4
1.6 Limitations ... 4
1.7 Delimitations ... 5
1.8 Definition of Key Terms ... 5
1.9 Chapter Summary ... 7
CHAPTER 2. LITERATURE REVIEW ... 8
2.1 Literature on BCA ... 8
2.2 Literature on Deterministic BCA Approach ... 13
2.3 Literature on Probabilistic BCA Approach ... 14
2.3.1 The Determination of Distribution ... 15
2.3.2 Simulation Methods ... 19
2.3.2.1 Monte Carlo Simulation ... 20
2.3.2.2 Latin Hypercube Simulation ... 21
2.4 Chapter Summary ... 23
CHAPTER 3. METHODOLOGY ... 24

3.1 Framework 24
3.2 Data Sources 28
3.3 Chapter Summary 29
CHAPTER 4. TRAFFIC VOLUME DATA ANALYSIS 30
4.1 Analysis of Annual Daily Traffic Data 30
4.2 Data Simulation 45
4.3 Chapter Summary 51
CHAPTER 5. PROBABILISTIC BENEFIT-COST ANALYSIS 52
5.1 Build the Analysis Algorithm 52
5.1.1 Agency costs 52
5.1.2 User Benefits 55
5.1.2.1 Travel Time Saving 55
5.1.2.2 Vehicle Operation Cost Saving 60
5.1.2.3 Safety Saving 64
5.2 Build the Probabilistic Model 66
5.3 Model Outcomes Analysis 68
5.3.1 Travel Time Savings 70
5.3.2 Vehicle Operation Costs Savings 74
5.3.3 Safety Savings 79
5.3.4 Net Present Value (NPV) 84
5.4 Chapter summary 90
CHAPTER 6. CONCLUSIONS AND RECOMMENDATIONS 91
LIST OF REFERENCES 96

LIST OF TABLES

Table Page

Table 4.1 *Simulated ADT Data of I-70 (Part)* 48

Table 5.1 *Construction Costs per Mile in Indiana* 53

Table 5.2 *Maintenance Cost for Flexible Pavements (1984 Dollars)* 54

Table 5.3 *Hourly Traffic Distribution in Indiana (2004)* 57

Table 5.4 *Hourly Values of Travel Time Savings (2009 U.S. $ per person-hour)* 58

Table 5.5 *Value of Travel Time in Indiana (in 2003 dollar)* 59

Table 5.6 *Unit Gas Costs* 61

Table 5.7 *Average Truck Gross Vehicle Weight Indiana (Unit: 1000 lbs.)*.......... 61

Table 5.8 *Equations for Fuel and Oil Consumption (Unit: gal/1,000 miles)*....... 62

Table 5.9 *Fatality and Injury Rates per Million VMT*... 64

Table 5.10 *Crash Reduction Factors*.. 65

Table 5.11 *Estimated Accident Costs in Indiana Highways*................................. 65

Table 5.12 *Basic Project Information from Cal-B/C*... 68

Table 5.13 *Travel Time Saving Values in 20 Years (Part)* 70

Table 5.14 *Vehicle Operation Costs Saving in 20 Years (Part)* 75

Table 5.15 *Safety Saving in 20 Years (Part)*.. 80

Table 5.16 *NPV Values in 20 Years (Part)*.. 86

LIST OF FIGURES

Figure Page

Figure 3.1 Probabilistic Framework of User Benefits 25

Figure 3.2 Probabilistic Framework of Agency Cost and Net Present Values 26

Figure 4.1 Parts of Daily Traffic Data 31

Figure 4.2 Cullen and Frey Graph of I-64 with Outliers 33

Figure 4.3 Cullen and Frey Graph of I-64 without Outliers 35

Figure 4.4 Cullen and Frey Graph of I-70 with Outliers 36

Figure 4.5 Cullen and Frey Graph of I-70 without Outliers 37

Figure 4.6 Cullen and Frey Graph of I-80/90 with Outliers 38

Figure 4.7 Cullen and Frey Graph of I-80/90 without Outliers 40

Figure 4.8 ADT Data Distribution for I-64 with Outliers 41

Figure 4.9 ADT Data Distribution for I-70 with Outliers 42

Figure 4.10 ADT Data Distribution for I-80/90 with Outliers 42

Figure 4.11 ADT Data Distribution for I-64 without Outliers 43

Figure 4.12 ADT Data Distribution for I-70 without Outliers 44

Figure 4.13 ADT Data Distribution for I-80/90 without Outliers 44

Figure 4.14 The Process of the Monte Carlo Simulation of I-70 46

Figure 4.15 Simulated Daily Traffic Data Distribution Plot for I-64 49

Figure 4.16 Simulated Daily Traffic Data Distribution Plot for I-70 49

Figure 4.17 Simulated Daily Traffic Data Distribution Plot for I-80/90............. 50

Figure 5.1 Process of Proposed Probabilistic Model... 67

Figure 5.2 Simulated Daily Traffic Distribution of Project in Cal-B/C............... 69

Figure 5.3 Present TTS in Year 1.. 71

Figure 5.4 Present TTS in Year 20.. 72

Figure 5.5 Present TTS of 20 Years.. 73

Figure 5.6 Present Values of Vehicle Operation Cost Saving in Year 1............. 76

Figure 5.7 Present Values of Vehicle Operation Cost Saving In Year 20 77

Figure 5.8 Present Values of Vehicle Operation Cost Saving In 20 Years......... 78

Figure 5.9 Present Values of Safety Saving in Year 1 81

Figure 5.10 Present Values of Safety Saving in Year 20 82

Figure 5.11 Present Values of Safety Saving in 20 Years................................... 83

Figure 5.12 Interface of Rstudio.. 85

Figure 5.13 NPV for 20 Years.. 87

Figure 5.14 NPV for 25 Years.. 88

Figure 5.15 NPV for 30 Years.. 89

ABSTRACT

He, Yue. M.S.B.C.M., Purdue University, May 2015. A Probabilistic Model of Benefit-Cost Analysis for Highway Construction Projects. Major Professor: Yi Jiang.

Deterministic method occupies the predominant position in economic analysis. However, by using a single value to estimate the project value, deterministic method would hide the risk under it. Analysts may make wrong judgments by looking at a value without knowing the likelihood of its occurrence and the consequences of its occurrence. A probabilistic model was developed in this study to address this problem. It was found in this study that the traffic flows on Indiana highways follow logistic distributions rather than normal distributions as generally believed. The application of the probabilistic model was illustrated through a case study. The results indicate that the net present values (NPV) of benefit and cost do not follow normal distributions. Instead, they show asymmetric patterns with large skews toward right. The probabilistic model developed in this study provides a confidence range rather than a single value for decision makers to make rational decisions.

CHAPTER 1. INTRODUCTION

This chapter outlines basic information of the study. Starting from the background of economic analysis, this chapter states the importance of performing probabilistic economic analysis. The research questions of this study are then identified. A main purpose and sub-goals of this study are proposed to answer research questions. Assumptions, limitations, and delimitations are discussed later in this chapter. Key terms are defined at the end.

1.1 Background

Economic analysis has been extensively used for highway investment decisions for decades. It is the most comprehensive methodology to quantify the benefits and costs of a project over multiple years (FHWA, 2003). The economic analysis could recommend the most cost-efficient choice among the same beneficial alternatives and could suggest the most profitable choice among alternatives in the same investment level. Several measures used in economic analysis such as internal return rate (IRR) and benefits and costs ratio (BCR) can help decision makers to understand the nature of this project and better control the balance between benefits and costs. The process of performing economic analysis also provides good documentations for interpreting investment decisions in the future (FHWA, 2003).

Deterministic economic analysis is the prevailing analysis approach in past decades. This simple and efficient method uses a single value to represent a input variable. By using a timeline of cash inflows and outflows, analysts can evaluate the life-cycle values of a project easily and directly.

However, the deterministic method ignores potential risk underneath the “best-fit” value of input variables. The notion of uncertainty (Lawson, 1985) in economic analysis was introduced in the 1980s but was not widely applied to highway projects until 1998, in which FHWA released an interim technical bulletin called the *Life-Cycle Cost Analysis in Pavement Design*. The technical bulletin first introduced probabilistic methods in life-cycle cost analysis (LCCA) for highway projects. After that, probabilistic economic analysis method started to arouse decision makers’ interests.

1.2 Significance

Over the past decade, many researchers have studied the application of probabilistic analysis in highway construction. However, most works examined the probabilistic nature of agency costs and their outcomes. Very few of them discussed user benefits and costs. Of course, the impacts of user benefits and costs were ignored in these probabilistic economic analysis studies. However, an economic analysis would not be a comprehensive and persuasive one if it only focused on agency costs. User benefits play significant roles in economic analysis outcomes and should be emphasized more. This study starts from the probabilistic nature of traffic volume and mainly examines user benefits and costs with risk and uncertainty in benefit cost analysis (BCA). By using deterministic analysis of agency costs, this study illustrates the probabilistic nature of user benefits and their impacts to project net present values.

Besides, this study examines the daily traffic data of interstates in Indiana. The analysis outcomes reveal that the distributions of traffic data, travel time savings, vehicle operation cost savings, safety savings, and project net present values are different from general presumptions in deterministic approach.

1.3 Statement of Purpose

This study intends to develop a probabilistic economic analysis methodology for highway construction projects. To achieve this purpose, this study has set up four sub-goals:

First, this study would examine the nature of traffic flow and determine the distribution of daily traffic data.

Second, this study would perform data simulation to reveal future risk associated with traffic volume uncertainty.

Third, this study would establish the relationship between traffic volume and user benefits and costs.

Fourth, this study would show the impacts of uncertain traffic volume on the project net present values.

1.4 Research Question

This study will answer the following questions:

1. How to build a probabilistic model addressing benefits and costs for highway construction projects?
2. What are the impacts of probabilistic approaches on highway construction economic analysis?

1.5 Assumptions

This study included the following assumptions:

1. The daily traffic data used in this study reflects real highway traffic flow.
2. The traffic flow in highways varies from the lower limit (0) to the upper limit (X_{max}).
3. The traffic volume grows year by year. The selected annual growth rate could reflect the growth trend.
4. Highway projects discussed in this study were under normal operations without experiencing unknown irresistible forces such as earthquakes.
5. The selected discount rate could reflect present values of money during the analysis period.
6. The maintenance costs and rehabilitation costs data are discrete and could not form distributions.

1.6 Limitations

This study included the following limitations:

1. Because of the data source, this study was limited to highway projects in Indiana State.

2. This study only focused on highway projects. Bridge projects were excluded in the discussion.
3. Among agency costs, small acts of maintenance were not considered in this study because their economic impacts were negligible.
4. This study only focused on agency costs and user benefits and costs. Non-user benefits and costs were out the scope of this study.

1.7 Delimitations

This study included the following delimitations:

1. The availability of daily traffic data from the Indiana Department of Transportation (INDOT).
2. This study intends to develop a methodology for probabilistic economic analysis. The data used in this study is the daily traffic data in the year of 2004, which may not reflect current (2015) traffic flow.

1.8 Definition of Key Terms

Average annual daily traffic – "The total yearly volume divided by the number of days in the year, commonly abbreviated as AADT" (AASHTO, 2010, p. 15).

Maintenance cost – "A subset of Operating Cost relating to keeping a highway and its appurtenances in serviceable condition" (AASHTO, 2010, p. 17).

Present value – "It is the present amount that is equivalent to specified amounts of money or time in different time periods, at a given discount rate. Two related considerations underlie the need for computing PV: (1) the fact that money has a time value of capital cost, due to its productiveness and scarcity (see the Definition of Discount rate), and (2) the need in an economy study for comparing or summing outlays or savings of money or time in different time periods" (AASHTO, 2010, p. 19).

Primary non-user benefit – "An impact that falls on non-users and occurs as a direct consequence of a performance feature of the project. For example, emissions reductions that result from improved traffic flow may constitute a non-user benefit" (AASHTO, 2010, p. 19).

Project Alternatives – "Any variations to the basic project plan that (1) involve significantly different costs, (2) result in significantly different levels of service or demand, or (3) incorporate different route locations or other distinctive design features" (AASHTO, 2010, p. 19).

Rehabilitation – "Rebuilding or restoring an existing facility that is under disrepair or not up to standards" (AASHTO, 2010, p. 19).

User costs – "Costs incurred by highway users traveling on the facility and the excess costs incurred by those who cannot use the facility because of either agency or self-imposed detour requirements. User costs typically are an aggregation of three separate components: Vehicle Operating Costs (VOC), Crash Costs, and User Delay Costs" (Wall & Smith, 1998, p. 3).

1.9 Chapter Summary

This chapter states the background, significance, and research purpose of this study. In addition, assumptions, limitations, and delimitations are discussed in this chapter for further clarifications.

CHAPTER 2. LITERATURE REVIEW

This chapter provides an overview of literature specifically related to the highway project economic analysis including the basic differences between life-cycle cost analysis (LCCA) and benefit cost analysis (BCA), commonly used deterministic model in BCA, and possible probabilistic ways to conduct economic analysis.

2.1 Literature on BCA

Life-cycle cost analysis (LCCA) and benefit cost analysis (BCA) are two widely used economic evaluation methods adopted in the decision-making stage of pavement projects. As the first systematic way to evaluate highway projects, the LCCA method can help find the least cost alternative among projects. LCCA assumes benefits among projects are the same. In other words, this method considers each alternative will generate the same economic benefits (FHWA, 2003). By comparing the cost differential among projects, it would be able to select the lowest-cost project. This method is simple to use and very efficient in making an investment decision. However, one violation to the LCCA assumption is that the generated benefits among projects would not stay the same in reality. Different pavement design, the environment, and geography lead to various project benefits. The outcome of LCCA might be too theoretical and deviate from real results.

The other economic analysis, the BCA method, is more comprehensive and practical by considering both costs and benefits during the life cycle of a project (FHWA 2003). BCA method compares the changes caused by a potential action that might improve the status of a project (Jiang et al, 2013; FHWA, 2003). The major difference between LCCA and BCA is that the former one evaluates only the cost among two or more projects while the latter considers the status of a project before and after a change. The BCA method is also applicable for projects comparisons. It could be used to select the most beneficial alternative with a limited budget (FHWA, 2003). This study will dynamically analyze the economic impact of traffic volume change. Because traffic volume strongly affects user benefits, this study will adopt the BCA method.

Net present value (NPV) is one of the basic measurements in economic analysis. NPV is the discounting mathematical summed value of future benefits and costs in each year (Wall & Smith, 1998). The NPV formula shows as below:

$$NPV = Construction\ Cost + \sum_{t=1}^{N}\left[\frac{A_t}{(1+r)^t}\right]$$

Where:

t = the *t*th year

NPV = the summed discount present value of t years

A_t = the benefits (+) and costs (-) of the *t*th year

r = discount rate

The benefits and costs here represent agency costs and user benefits/costs. Agency costs happened in the whole life of a highway project and are incurred by agency. According to Wall and Smith (1998), agency costs typically consist of initial construction, rehabilitation costs, major maintenance costs, and salvage value (Li et al., 2008; Wall & Smith, 1998). In most analyses, routine maintenance costs are generally ignored in NPV calculations because their costs are small and impacts are minor. Except initial constructions that incur in the base year, rehabilitation costs and maintenance costs happen in the life cycle of the project. Salvage value is the value of a project at the end of project service life. User benefits and costs are defined as benefits and costs incurring along the overall life cycle of highway projects by highway users. User benefits/costs sometimes are categorized into two situations: normal operation and work zone (Li et al., 2008; Wall & Smith, 1998). Under each category, it is mainly divided into traffic time saving (TTS), vehicle operating costs (VOC), and safety saving (SS) (AASHTO 2003).

Once the benefits and costs are decided, the decision maker needs to select whether to use constant dollars or nominal dollars (Tighe, 2001). A constant dollar is also called a base year dollar or real dollar. It means the purchasing power of each dollar stays the same in the future as it is in the base year (FHWA 2003). However, a nominal dollar reflects the impact of inflation and its purchasing power varies year by year (Tighe, 2001; FHWA 2003). If a decision maker chooses to use a nominal dollar, it is essential, although difficult, to predict inflation, especially in a long future. The inflation rate is constantly changing by various uncontrollable factors. If the predicted inflation is not accurate, the predicted outcome will largely deviate from the true value. Given this situation, most decision makers would like to use the constant dollar with a time value in economic analysis (FHWA 2003).

The time value of money is also considered the opportunity cost of money. It is equal to the benefits people might potentially gain in other opportunities. For example, suppose a person had 100 dollars and saved it in the bank for three years. This person would receive interest, plus his/her capital after three years. If this person did not save this money into the bank, his/her loss would be the interest, which is called the time value of the original 100 dollars. The formula of the time value of money is indicated as follows:

$$F = (1 + r)^t P$$

$$P = \left[\frac{1}{(1 + r)^t}\right] F$$

Where:

P = present value

F= future value

r= discount rate

t= t years

As we can see from the above formula, one essential element in the calculations of present value (P) or future value (F) is the discount rate. The selection of discount rate is important on economic analysis because it switches future costs and benefits to present values (Jiang et al., 2013). Different discount rates will result in different project values, and of course, different investment decisions. In most situations, a conservative discount rate is approximately equal to the interest rate charged by a regional Federal Reserve Bank to commercial banks on loans (Jiang et al., 2013). Agencies sometimes use the interest rate charged by the bank when borrowing money (Tighe, 2001). The discount rate is also determined by the financial conditions of a region. For example, the Loan Prime Rate (LPR) in China, is about 5.73 percent for one-year loans between commercial banks (Bank of China, 2014). It is common, and safe, to use an eight percent or ten percent discount rate in the economic analysis of projects in China. This research will use the traffic data from the Indiana Department of Transportation (INDOT). A four percent discount rate in commonly used in INDOT bid documents (Phillips & Vancleave, 2007).

The last part should be included into the economic analysis is the analysis period. Basically, the analysis period should be able to cover all construction and rehabilitation activities related to a specific pavement project (Wall & Smith, 1998). The FHWA (1996) recommend at least a 35-year analysis period for all pavement projects. However, the old designs might need to change to adapt to new traffic growth and new transportation technology. A shorter period is also acceptable nowadays. In some economic analyses for Indiana highways, a 20-year analysis period is used (FHWA, 2009). In this study, a comparison of different analysis periods (20/25/30 years) will be conducted to reveal the best fitted design in response to various traffic flow.

For BCA analysis, the other useful measure is the benefit-cost ratio (BCR). The BCR is the ratio of benefits to costs. In the formula of BCR, the numerator is the present value of benefits and the denominator is the present value of the costs (FHWA, 2003). In the comparison of projects with a limited budget, the higher BCR value might represent the better choice.

Other measures, such as internal rate of return (IRR) and sensitivity analysis could also assist in further selection of analysis results.

2.2 Literature on Deterministic BCA Approach

Most decision makers choose to use deterministic approach in current economic analyses because of its simplicity. They could simply estimate a single value for each input variable to perform the analysis (Tighe, 2001). In BCA, for example, the monetary value of the travel time of a person could be assigned as half of the average wage of his/her by USDOT (2012).

Many tools are available to help make decisions in BCA process. In 1977, AASHTO issued a book called the *Manual on User Benefit Analysis of Highway and Bus-Transit Improvements*. This book introduced calculations and steps of user benefits analysis in highway and bus-transit improvement projects. It is referred to as the foundation of many other BCA models. The MicroBENCOST model was developed by the Texas Transportation Institute (FHWA, 2003). Later, several states had released their own BCA tools such as the California Life-Cycle Benefit/Cost Analysis Model (Cal-B/C) (1999).

Another influential model is the Highway Economic Requirement System (HERS) developed by FHWA (2000). The HERS is a computer-based model originally designed for economic analysis of national highways. A state version (HERS-ST) was developed later and was more popular in state-level BCA. It could help to select the most appropriate investments among up to six choices by the desirable benefit level or to find the most beneficial choice in a given investment level (FWHA, 2003). This model and the MicroBencost model is the basement of a more sophisticated model, the StatBENCOST model, developed by National Cooperative Highway Research Program (NCHRP).

These deterministic models mainly adopted several representative values such as 25 percent quartile, mean, median, and 75 percent quartile, to analyze the benefits and costs of projects. Using these discrete values might make decision makers lose sight of uncertainty. The ignorance of risks or uncertainty would lead to incompletion of information and errors in of analysis results.

2.3 Literature on Probabilistic BCA Approach

Probabilistic research on highway economic analysis first started in 1998 when FHWA incorporated risk analysis in its technical bulletin of the *Life-Cycle Cost Analysis in Pavement Design.* Then the notion, risk or uncertainty, came to decision makers minds. People started to realize the importance of probabilistic analysis in LCCA and BCA processes because most input variables are not certain (Wall & Smith, 1998). The probabilistic method describes variables in distribution, which incorporates the probability of occurrence and specific spread. After examining the spread of input variables, a distribution pattern could be determined and assigned. Researchers could conduct a simulation based on the distribution to predict future values and the likelihood of future outcomes. The probabilistic approach exposed the uncertainty or risk that was hidden in deterministic methods to decision makers. They could straightly face uncertainty and weigh it based on their acceptance levels to risk rather than avoid it like in the past. Decision makers could also reduce the uncertainty by adopting control methods using solid analysis method (FHWA, 2003).

2.3.1 The Determination of Distribution

In the probabilistic process of economic analysis, one critical step is to determine the distributions of input variables. It assists the prediction of future value and decides the accuracy of prediction.

Among various distributions of input variables, researchers prefer to see a normal distribution so they could easily examine fitness, perform predictions, and run programming. It is more convenient to study a normal distribution than any other distributions using current computational methods. Besides, based on the Central Limit Theorem (CLT), if the sample size is large enough, the mean of independent random variables will be approximately normally distributed (Rice, 1995). Most construction variables are commonly considered normally distributed. Unfortunately, real data is never perfectly bell-shaped. Several researchers have worked on this field and have provided meaningful distributions of construction variables.

Tighe (2001) proposed a study using mathematical methods and a comparison to similar financial variables to carefully examine two construction variables, material cost and layer thickness. In Tighe's work (2001), she considered the pavement variables follow lognormal or gamma distributions because pavement variables do not have upper limits but are bounded by non-negative value. This reveals the distributions of input variables are possibly right-skewed rather than symmetric.

To confirm her hypothesis, she input material cost and layer thickness variables in both lognormal distribution and normal distribution and ran a Monte Carlo simulation to get both construction initial cost distributions. After adding all the cost in the whole life cycle, the construction cost distribution should approach a normal distribution (Tighe, 1999) so the results can be compared by χ^2, an indicator of goodness of fit. Tighe (2001) compared the total life cycle cost distributions using lognormal distribution and normal distribution for input variables. She found the former has a much lower χ^2 (231.4) than the latter (429.6), which turns to an increase of life-cycle cost about $62,000 per kilometer. Tighe's research (2001) illustrated the importance of examining independent variables, that is, failure to identify the true nature of input variables may result in inaccuracy of economic analysis and largely increase the total cost.

Tighe's research (2001) introduced a brand new perspective in a probabilistic analysis area to view input pavement variables. Normal distribution is no longer the only distribution people could use in construction economic analysis. Several studies are encouraged by Tighe's work.

Li and Madanu (2008) considered Beta distribution to be a better way to define pavement factors, including flexible pavement cost, rigid pavement cost and concrete bridge cost. Compared with the single skewness and kurtosis of lognormal distribution, Beta distribution can provide a variety as left-skewed, right-skewed, and symmetric. Besides, the Beta distribution has a boundary defined by two of its four parameters, the lower limit (l) and upper limit (u), which is more reasonable because the main variable discussed in this research, traffic volume, obviously has an upper bound. The formula of Beta distribution is:

$$f(x) = \frac{1}{B(\alpha,\beta)} \frac{(x-l)^{\alpha-1}(u-x)^{\beta-1}}{(u-l)^{\alpha+\beta-1}}, \qquad l \leq x \leq u$$

Where:

l = the lower limit

u = the upper limit

α and β: define the shape

The Beta function:

$$\mathrm{B}\,(\alpha, \beta) = \int_0^1 t^{\alpha-1}\,(1-t)^{\beta-1} dt, \qquad (\alpha > 0, \beta > 0)$$

The Beta function can help normalize the distribution to make sure the area between *l* and u is equal to 1.

Another major contribution of Li and Madanu's research (2008) is that they proposed a framework to address the certainty, risk and uncertainty of input variables. For certain input variables with fixed values, a traditional deterministic economic analysis is applicable for decision making; for variables involving risk concern, a probabilistic economic analysis would be a better solution; for input risk variables without a meaningful statistical distribution, an approach of combining life cycle cost analysis with Shackle's model is used. Shackle's model (1949) used the "Potential Surprise Function" (Cantillo, 2010, p18), which involved an expected outcome (X_E), a standardized focus loss (X_{SFL}) and a standardized focus gain (X_{SFG}) to measure uncertainty of variables.

$$X_E = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} X_i}{M \times N}$$

X_E: expected outcome

X_i: a simulated outcome with an input variable

N: the number of iterations in each simulation

M: the times of simulation runs

$$X_{SFL} = \left| \frac{\sum_{m=1}^{M}\sum_{n=1}^{N_r} X_i}{M \times N_r} - X_E \right|$$

X_{SFL}: the standardized focus loss

Nr: a simulated outcome in the *r*th run

$$X_{SFG} = \left| \frac{\sum_{m=1}^{M}(\sum_{n=1}^{N} X_i - \sum_{n=1}^{N_r} X_i)}{M \times (N - N_r)} - X_E \right|$$

X_{SFG}: the standardized focus gain

Li and Madanu's study (2008) simplified Shackle's model (1949) and provided formula applicable for economic analysis. For input variables (X):

$$X = \begin{cases} X_E & X_{SFL} \leq \Delta X \\ \dfrac{X_E \pm X_{SFL}}{\left[1 \pm \dfrac{\Delta X}{X_E}\right]} & Otherwise \end{cases}$$

Where:

X: input variables that could not form a meaningful statistical distribution

ΔX: assumed acceptable loss of a decision-maker

The acceptable loss, ΔX, in the above formula serves to control the uncertainty. Different decision makers acceptance levels of uncertainty correspond to different ΔXs and different ΔXs deal with different uncertainty of input values.

After determining the distribution of input variables, it sometimes could conduct some transformations that normalize the distribution function and simplify the statistical analysis. For example, if X follow lognormal distribution, the log (X) will follow normal distribution. Then a transformation on input variables X ($x_1, x_2, x_3, \ldots, x_n$) could be made as:

$$X_{new} = \log(X)$$

However, in Beta distribution, there is no such transformation but an approximation. If α and β are large enough, the Beta distribution could approximately equal to normal distribution. The approximation might violate the nature of input variables, so in this study an approximation would not be adopted.

It is not easy, but crucial, to find a good distribution for input variables. There are also other choices, such as Poisson distribution. The correct choice would hugely contribute to prediction of true cost and benefits. This study will examine the distribution of traffic volume very carefully.

2.3.2 Simulation Methods

A simulation method could predict future value and unknown values based on known information. It uses random generated numbers to choose a set of input values and then calculates results by using these selected input values. The results are discrete because the selected number may not cover all possible values. Almost all simulation techniques require certain times of iteration for the above process in order to generate enough numbers that can cover as many results as possible (Wall & Smith, 1998). There are two prevailing sampling methods: the Monte Carlo simulation and Latin Hypercube simulation.

2.3.2.1 Monte Carlo Simulation

The Monte Carlo method is one of the most important sampling techniques in modern statistics used to generate observations from known distributions (Hogg, McKean, & Craig, 2012). Wall and Smith (1998) introduced the Monte Carlo method in its Interim Technical Bulletin as a major sampling method for decision makers. In the Monte Carlo process, X is the input variable with a known distribution of:

$$f(x) = (x_1, x_2, x_3, \dots, x_n), \qquad -\infty < x < +\infty$$

In this situation, *f*(x) is the Probability Density Function (PDF). By integrating *f*(x), the Cumulative Distribution Function (CDF) of X would be:

$$F(x) = \int_{-\infty}^{x} f(t)\, dt, \qquad -\infty < x < +\infty, 0 \leq \text{y} \leq 1$$

The CDF is strictly monotone increasing and is bounded by [0, 1] (Hayter, 2012). We inverse the above cumulative distribution function and get

$$F^{-1}(u) = (u_1, u_2, u_3, \dots, u_n), \qquad 0 \leq u \leq 1$$

Once the function is inversed, the next step is to generate a series of random numbers between 0 and 1 and plug them into the inversed function. There are various ways to generate random numbers. The oldest and most convenient way is to use the Random Digits Table (Vohra, 2006). Besides that, computational methods are more commonly used to generate data for more complex needs. One example is using the SAS software with '=rand' function.

After the generation of many random numbers (u_i), they can be plugged into the $F^{-1}(u)$ function to calculate corresponding X_g values, where X_g means generated X value. For example, suppose that a value of $u_i=0.5$ is generated, this value will then be applied in the $F^{-1}(u)$ function, and a $X_{0.5}$ would be predicted and sampled (Wall & Smith, 1998). In this study, the sampled $X_{0.5}$ will be used in another function, which shows the relationship between X (Traffic volume) and Y_1 (Travel time savings), and Y_2 (Vehicle operating costs) and Y_3 (Safety savings). The sampled X_i can predict Y_i and form a part of Y_i distribution. Then, an iteration about 100 times or larger of the sampling process can help form a statistical meaningful distribution of Y_i. The distribution can tell much more information than a single value, including the probability of the occurrence of certain values, the crest value and the deviations.

2.3.2.2 Latin Hypercube Simulation

The Latin Hypercube sampling method is an improvement to the Monte Carlo technique. When conducting the Monte Carlo sampling, the number of iterations has a huge impact on the outcomes. Because the CDF is a strictly monotonic distributed function, its slope is not constant. In its distribution plot, it is clear to see that some slope is more vertical and some is more horizontal. For those X_is in the more vertical slope, the possibility of being sampled is higher than those in the more horizontal slope. Therefore, the possibility of X_i being sampled is uneven when conducting a random sampling. The likelihood of clustering would be increased if the number of iterations is low (Wall & Smith, 1998). However, increasing iterations would cost much more time and efforts. One statistical way to avoid the uneven random sampling is to use a stratified sampling method. The Latin Hypercube technique is one of the stratified sampling methods (Wall & Smith, 1998).

Compared with the Monte Carlo technique, the Latin Hypercube sampling method is more efficient and more comprehensive (Wall & Smith, 1998). It can get similar results while saving many iterations. The Latin Hypercube sampling divides the series number between 0 and 1 into several sections evenly. In each section, it is sampled within a certain amount, say 30, of X_is. Once a section is sampled with 30 X_is, it would not be sampled again. This method ensures every section is sampled and eliminates the difference in probability between variables in a flatter slope and those in a steeper slope. It is much faster to generate distribution. Wall and Smith (1998) compared the outcomes of the Latin Hypercube method and Monte Carlo method by histogram. When both performing 100 iterations, it clearly shows that the graph of the Latin Hypercube method is normally distributed with legible means and standard deviations while the graph of the Monte Carlo method is still ambiguous and the distribution type is hard to determine.

2.4 Chapter Summary

This chapter discussed literature related to highway project economic analysis, including the basic principle of BCA, deterministic approach to BCA, and probabilistic approach to BCA. The basic principle of the BCA section also talked about the common places and differences between BCA and LCCA. It also reviewed measures commonly used in economic analysis, necessary information for economic analysis, and major steps to perform the analysis. Then, deterministic tools of BCA, such as HERS-ST, are introduced in the following section. In the probabilistic approach section, this study reviews ways to determine the distribution of variables and simulation techniques. These two parts plus the full process of the BCA analysis constitute the most meaningful part of the probabilistic study, which will be taken later in this study.

CHAPTER 3. METHODOLOGY

This study aims at developing a probabilistic life-cycle cost analysis model for highway construction projects. This study is a quantitative study and uses the Monte Carlo simulation techniques for data analysis. This chapter outlines the framework and data sources of this study.

3.1 Framework

This study is trying to build a probabilistic analysis model using traffic volume to analyze the project life-cycle benefits and costs. The model can dynamically analyze the life-cycle benefits and costs of a pavement project with the consideration of risk. Two major parts are included in this model: agency costs and user benefits/costs. Agency costs are further divided into construction costs, maintenance costs, and rehabilitation costs. User benefits are broken into travel time savings (TTS), vehicle operating costs savings (VOCS), and safety savings (SS). Traffic volume is the major explanatory variable in this model.

The first step of this analysis is to determine essential information for LCCA analysis, which includes the selection of constant/nominal dollar, the length of the analysis period, the value of the discount rate and traffic annual growth rate. Then a major framework of this study is developed in Figure 3.1 and Figure 3.2.

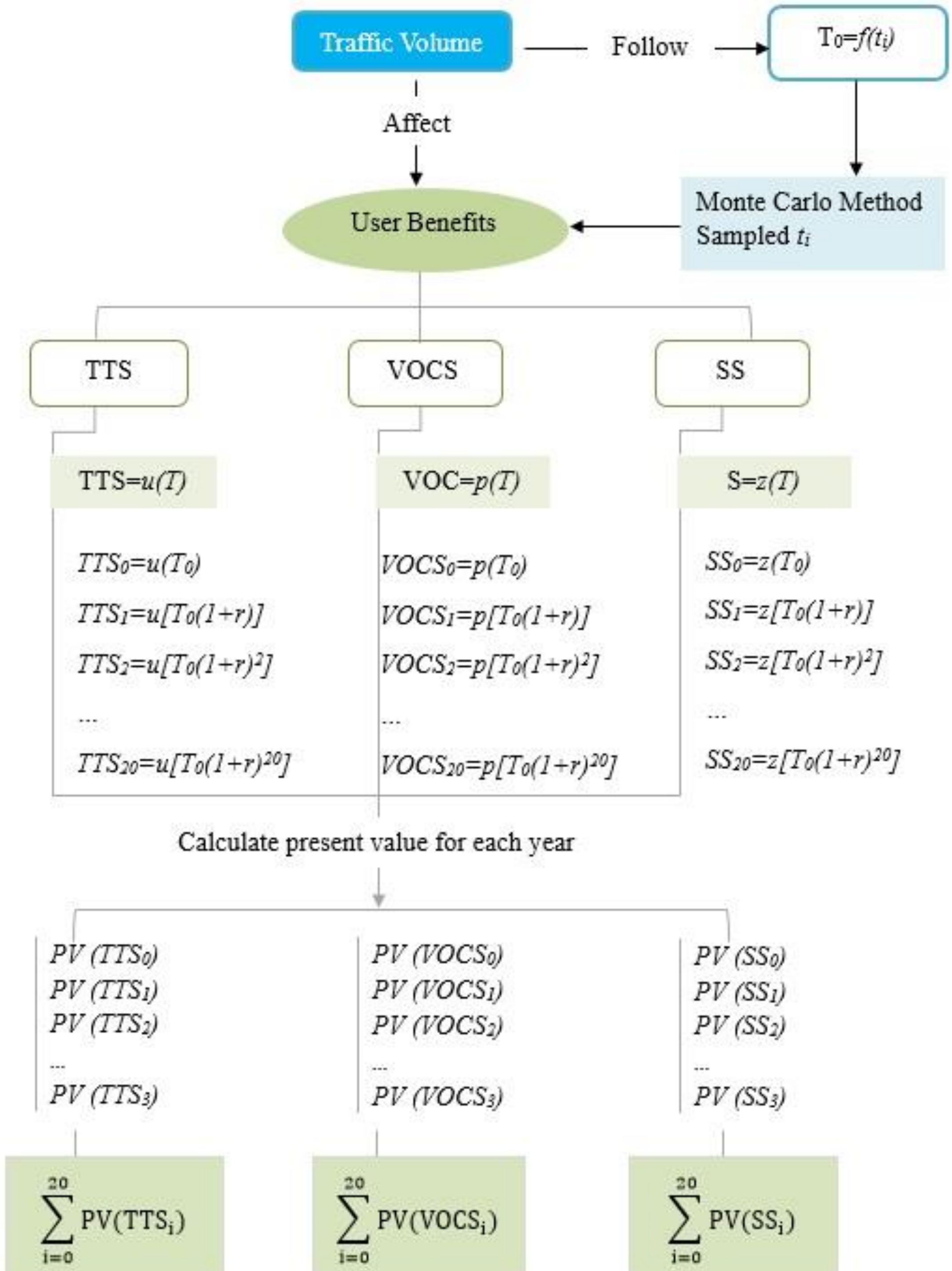

Figure 3.1 Probabilistic Framework of User Benefits

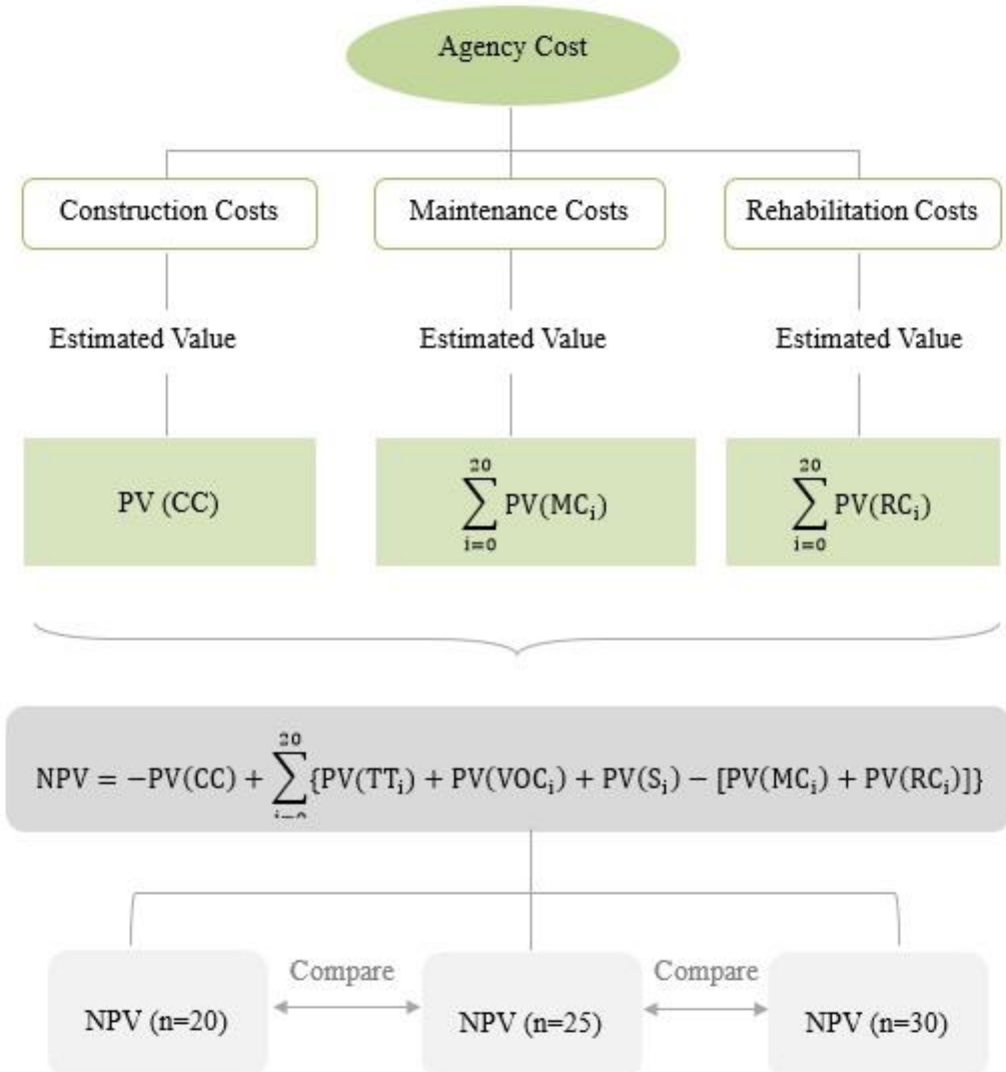

Figure 3.2 Probabilistic Framework of Agency Cost and Net Present Values

The traffic volume in this research serves as the primary explanatory variable. A careful examination of traffic flow distribution in the base year is performed. Because the traffic data is only a part of the whole population, the Monte Carlo simulation technique is adopted to generate more traffic volume observations. Traffic flow observations are then simulated and sampled. The sampled t_i will be used to calculate travel time savings (TTS), vehicle operating cost savings (VOCS), and safety savings (SS).

The relationships between traffic volume and TTS/VOCS/SS are established based on the model developed by Jiang et al. (2013). The results will be listed in each year for 20 years. Since the traffic flow is theoretically increasing each year, an annual growth rate is applied according to previous traffic data. The traffic volume of the $(i+1)^{th}$ year would be equal to the i^{th} year traffic volume plus traffic growth ($i \times r$, where r is the annual traffic growth rate). After the calculations of all three user benefits, TT, VOCS and SS, in each year are completed, and present values (PV) functions for all three phrases will then be computed. It is necessary to note that user benefits might become user costs according to a specific road type. Negative values are used in these formulas to represent user costs.

The other part of this model is agency costs, including construction costs, maintenance costs, and rehabilitation costs. Construction costs incur in the based year and are estimated on previous construction data. Maintenance costs in this study refer to annual routine maintenance. Small maintenance costs are excluded in this study due to its minor economic impact. Rehabilitations occur when the existing facility is in disrepair and is necessary to rebuild it. Maintenance costs and rehabilitation costs are also calculated based on historical data. The cost distributions data is divided into three categories: Interstate, US, and state road.

$$\mathrm{NPV} = -\mathrm{V(CC)} + \sum_{\mathrm{i=0}}^{\mathrm{n}} \{\mathrm{PV(TTS_i)} + \mathrm{PV(VOC_i)} + \mathrm{PV(SS_i)} - [\mathrm{PV(MC_i)} + \mathrm{PV(RC_i)}]\}$$

Where:

V = value

PV = present value

i = year

CC = construction costs

TTS_i = travel time saving in the *i*th year

VOC_i = vehicle operating costs in the *i*th year

SS_i = safety saving in the *i*th year

MC_i = maintenance costs in the *i*th year

RC_i = rehabilitation costs in the *i*th year

Comparisons of different analysis periods (20/25/30 years) will be conducted at the end of this analysis.

3.2 Data Sources

One primary piece of data used in this study is the statewide daily traffic flow in 2004. The data sets contain daily traffic volumes in most of the interstates, such as I-64, I-465 and I-70 of Indiana. Total data for each interstate is 365 in response to 365 days of a year. The data sets come from the Indiana Department of Transportation (INDOT). Other empirical data, such as discount rate, annual traffic growth, and rehabilitation costs, will be obtained legally through other DOT websites.

3.3 Chapter Summary

This chapter proposed a framework of a probabilistic model that would be used in this study. This model contains two major parts, user benefits and agency costs. For user benefits, traffic volume is the uncertain variable that has an influence on travel time savings, vehicle operating cost savings, and safety savings. By determining the distribution of traffic volume data, the Monte Carlo method is adopted to sample observations. Then calculations would be performed to compute TTS, VOCS and SS according to corresponding equations. For agency costs, except construction costs incurred in the base year, maintenance costs and rehabilitation costs only happen when needed. These costs adopt estimated values. After all the above information is computed and calculated, a net present value (NPV) function, which is equal to the total benefits minus total costs in each year, could be synthesized. At the end of this study, it will compare NPV distributions in different analysis periods (25/25/30 years) and analyze its distribution pattern.

This chapter also discussed the data resources of this study. These databases primarily come from the INDOT research division and other DOT websites.

CHAPTER 4. TRAFFIC VOLUME DATA ANALYSIS

Traffic volume has significant influences on highway user benefits and is the basis of the probabilistic model developed in this study. This chapter studied the annual daily traffic (ADT) data, discussed the distribution of ADT data, and performed data simulation to predict possible values and the likelihood of future outcomes.

4.1 Analysis of Annual Daily Traffic Data

Data used in this study is the statewide annual daily traffic (ADT) data in 2004 from the Indiana Department of Transportation (INDOT). The data sets contain daily traffic volume in most of the Interstates such as I-64, I-465 and I-70 of Indiana. Total data for each Interstate is 365 in response to 365 days of a year. Figure 4.1 is a screenshot of the data.

Rstudio is used in this study to fit and analyze ADT data. Rstudio is the software that provides an integrated development environment (IDE) for R language. R language is extensively used in statistical computing and graphical techniques. Compared with other statistical packages, such as SAS or SPSS, R has more advantages (Burns, 2006). It has effective and excellent data handling ability. The distribution and storage of data is easy in R. Besides, users can create and download packages online. This means R has the ability to meet all kinds of specialized statistical requirements.

		2 lanes	4	4	4	4	4	6	4	6	4	4	3	3	3	3
		I-64	I-164	I-70	I-74	I-80/90	I-65	I-94	I-64	I-65	I-80/94	I-80/94	I-465	I-465	I-465	I-465
		Dale	Evanssouth	Richmond	Wharrison	Sbendwest	Lafayette	Michcity	Newalbany	Southport	Garyeb	Bormanwb	dyEast NB	IdyEast NB	ıdyEast SB	ndyEast SB LT
		620	650	370	530	732	110	430	540	340	400	401	350	351	352	353
	1	4,054	14,142	22,937	19,369	16,848	26,777	25,300	17,100	48,434	50,206	43,765	15,704	19,826	16,235	19,015
	2	6,182	23,586	32,337	29,044	23,423	39,136	36,640	27,698	82,272	72,573	68,098	24,680	38,213	29,196	34,178
	3	5,772	20,299	33,964	27,452	25,705	41,010	36,679	23,287	74,540	69,782	63,827	21,703	32,046	23,124	28,693
	4	5,041	13,743	29,788	22,180	18,242	31,917	27,170	19,582	58,553	43,115	41,912	18,012	24,895	19,699	20,529
	5	5,475	21,316	29,345	26,045	19,010	32,973	28,382	27,554	80,272	68,991	62,944	26,964	41,688	34,121	36,286
	6	5,680	22,502	31,045	26,459	20,207	34,494	25,558	28,396	84,380	71,386	65,166	29,162	43,503	36,223	37,362
	7	5,847	22,797	32,368	27,087	20,903	36,039	25,969	28,835	85,914	73,088	69,389	29,612	44,373	35,990	38,605
	8	5,882	23,034	32,059	26,250	21,007	35,925	28,526	28,836	86,303	74,695	71,551	28,799	45,689	36,025	38,845
	9	6,518	-	33,670	28,625	21,401	39,244	31,619	29,332	90,110	77,054	73,232	30,431	47,232	37,592	40,052
	10	5,447	20,587	28,791	24,026	18,647	31,233	24,256	22,945	65,610	61,870	55,195	22,241	30,821	25,334	29,003
	11	4,665	15,417	26,300	20,539	19,152	31,323	26,894	18,747	52,808	55,930	51,418	17,715	23,351	19,521	21,703
	12	5,625	21,546	28,777	25,102	19,174	33,091	27,767	27,931	82,050	71,443	68,191	28,928	43,425	36,310	37,144
	13	6,039	23,404	32,031	27,148	20,978	35,293	28,365	28,625	87,039	76,201	72,306	22,782	45,434	37,111	39,371
	14	5,925	23,355	31,864	27,122	20,818	35,043	26,742	29,151	87,432	75,635	71,532	21,307	45,678	37,475	39,001
	15	6,262	24,011	33,466	27,572	21,526	37,316	29,594	29,969	89,326	75,842	73,730	13,951	46,775	38,399	38,972
Jan	16	7,149	28,849	37,478	32,198	24,055	44,617	37,280	32,732	96,443	82,082	77,828	5,790	49,806	40,041	42,723
	17	4,896	21,523	20,790	19,219	17,269	27,241	22,799	20,889	54,489	55,650	51,030	7,616	25,112	20,849	24,129
	18	4,457	15,873	25,192	20,340	17,846	28,171	28,905	18,270	50,322	54,745	50,446	7,225	22,639	18,775	20,951
	19	6,051	24,390	31,529	27,981	19,498	37,597	33,176	27,849	80,840	72,470	68,060	19,638	42,157	33,134	36,755
	20	5,791	26,214	31,687	26,849	20,394	35,401	28,155	28,382	86,485	75,888	71,741	24,352	45,327	38,300	38,675
	21	5,746	26,444	32,243	27,141	20,866	35,435	27,667	29,010	87,637	75,114	71,957	17,010	45,157	37,856	38,956
	22	6,265	26,905	32,905	27,249	21,208	36,323	27,405	29,447	88,239	75,523	72,573	22,854	45,590	37,423	39,243
	23	6,600	28,853	32,919	29,599	19,485	39,723	30,836	30,712	91,867	70,452	68,124	23,352	47,173	38,525	41,386
	24	5,235	22,999	28,784	25,156	18,353	31,732	24,143	22,972	69,593	57,602	52,202	19,241	31,814	26,831	30,138
	25	1,812	4,639	18,541	12,533	18,404	24,348	27,299	6,261	34,092	51,958	49,075	12,089	16,143	14,607	16,493
	26	5,067	20,262	23,438	18,501	16,661	27,930	25,729	3,443	65,912	67,557	62,837	22,333	36,264	30,319	30,821
	27	4,478	17,520	25,662	20,649	16,335	24,498	21,025	24,092	59,905	58,093	52,223	13,666	31,764	27,293	27,679
	28	5,574	22,167	31,699	25,586	20,738	33,430	25,948	27,655	83,467	72,513	62,934	26,794	42,528	36,466	36,671
	29	5,778	22,079	31,359	25,784	20,918	33,713	28,762	28,535	83,634	67,358	65,572	28,426	42,627	35,658	36,575
	30	6,094	20,935	33,656	27,669	21,920	38,766	32,180	26,345	89,156	73,089	59,232	30,393	45,171	36,788	39,332
	31	4,714	20,157	26,818	22,870	18,011	30,489	24,311	20,777	63,285	59,576	54,942	23,008	28,961	25,557	27,485
	1	3,965	14,589	23,824	19,609	16,133	27,547	25,494	17,275	52,029	53,512	49,182	16,473	23,196	19,331	20,898
	2	5,171	21,434	27,677	24,569	17,273	30,364	25,806	27,106	78,601	67,447	63,637	27,294	40,454	36,078	34,862
	3	5,558	21,970	30,466	25,715	1,475	31,483	26,901	28,024	83,016	72,026	67,183	30,342	43,035	36,250	37,537
	4	6,232	23,615	32,802	27,879	14,316	36,400	28,637	29,896	89,329	76,468	70,770	31,404	45,494	38,149	39,087
	5	4,987	18,918	30,179	25,060	21,149	33,217	29,538	26,526	79,125	66,657	59,244	6,908	41,291	35,816	34,677

Figure 4.1 Parts of Daily Traffic Data

The R package used to fit and analyze data distribution in this study is called "fitdistrplus." This package could maximize the possibility of quantile fitting, moment fitting, and goodness-of-fit (Muller, Dutang, Pouillot, and Denis, 2015). Three methods, the Cullen and Frey Graph, Akaike information criterion (AIC), and probability density function (PDF) plot, were used to examine data and determine data distribution for each data set.

The Cullen and Frey (1999) Graph is also called the skewness-kurtosis graph. It provides the choice of a best fit for an unknown distribution according to skewness level and kurtosis. It uses predefined distributions, such as normal, uniform, exponential, logistic, beta, lognormal, and gamma, to perform a moment fitting. It can also provide the maximum likelihood of data and assess the goodness-of-fit. In Cullen and Frey Graph, the x axis is the square of skewness and the y axis is kurtosis. The input data model is shown as a solid circle in this study. Different symbols represent different distribution types. If the skewness and kurtosis of the observation circle and the known distribution symbol is similar, it means the observation model and the known model may have similar, even the same, distribution.

Akaike information criterion (AIC) is also included in this package to help in model selection. AIC uses a set of candidate models to represent the unknown model. Because it uses one model to represent the other, information loss is unavoidable. If the information loss between the unknown model and one candidate model is the least, the candidate model has the highest probability to match the unknown model. For example, suppose model *k* is the unknown model. Candidate models *a, b, and c* are used to match model *k*. Their corresponding AIC values are AIC_a, AIC_b, AIC_c, and AIC_k. The information loss of matching is calculated as ($AIC_a - AIC_k$), ($AIC_b - AIC_k$), or ($AIC_c - AIC_k$). If the AIC value of model *a* is the smallest, they would have the least information loss and the best fit. Therefore, a candidate model with the lowest AIC value means it may match the unknown model best.

Three data sets, I-64, I-70, and I-80/90, are presented as examples to clarify the process and results of data analysis. Figure 4.2, Figure 4.4, and Figure 4.6 showed the Cullen and Frey Graph correspondingly for I-64, I-70, and I-80/90. Figure 4.3, Figure 4.5, and Figure 4.7 showed the Cullen and Frey Graph correspondingly for I-64, I-70, and I-80/90 after deleting outliers.

Figure 4.2 clearly shows the distribution of observation has a skewness of one. Uniform, normal, or logistic distribution allows zero skewness, but it is too early to exclude these distributions because the presence of outliers may account for the skew. The kurtosis of observation is close to lognormal and logistic distribution. Even though the distribution may be affected by outliers, it is still necessary to have a first look at its goodness-of-fit. AIC function could assist in further judgments by providing quantitative values. The smaller the value, the better they match. AIC results are shown as follows, and the lowest AIC value is 6358.941, which indicated logistic distribution is the best match for ADT data of I-64.

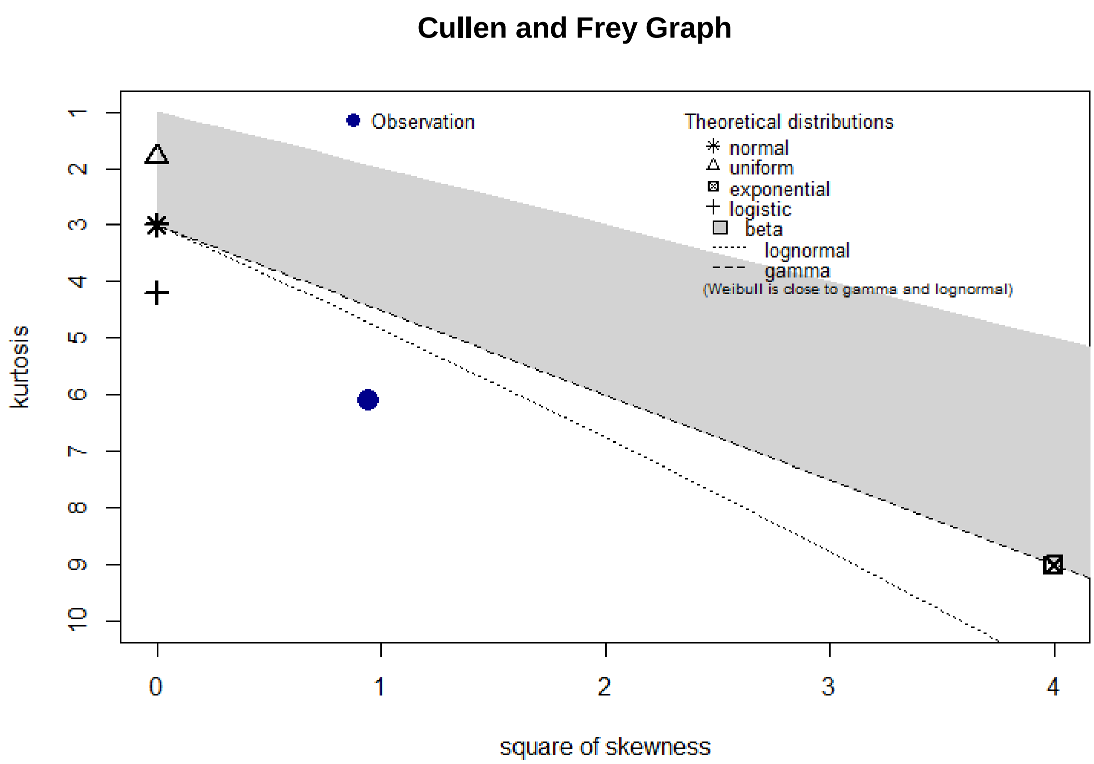

Figure 4.2 Cullen and Frey Graph of I-64 with Outliers

```
> fitdist(as.numeric(s),"norm",method="mme",gof="CvM")$aic
[1] 6395.635
> fitdist(as.numeric(s),"lnorm",method="mme",gof="CvM")$aic
[1] 6842.449
> fitdist(as.numeric(s),"gamma",method="mme",gof="CvM")$aic
[1] 6589.044
> fitdist(as.numeric(s),"logis",method="mme",gof="CvM")$aic
[1] 6358.941
> fitdist(as.numeric(s),"exp",method="mme",gof="CvM")$aic
[1] 7196.824
```

The input data of I-64 is original data, which may include outliers strongly affecting data distribution. Therefore, it is necessary to take out outliers and examine the changes in data distribution. Outliers in this study are defined as observations away from Q3 or Q1 larger than three times the difference between Q3 and Q1.

$$\text{Outliers} = \text{observations} > Q3 + 3 \times (\mathrm{Q3} - \mathrm{Q1})\ ;$$

$$\text{or; observations} < Q1 - 3 \times (Q3 - Q1)$$

Where:

Q3: 75 percent quartile of observations

Q1: 15 percent quartile of observations

An R function was used to remove these outliers. Figure 4.2 shows the Cullen and Frey Graph without outliers of daily traffic data of I-64.

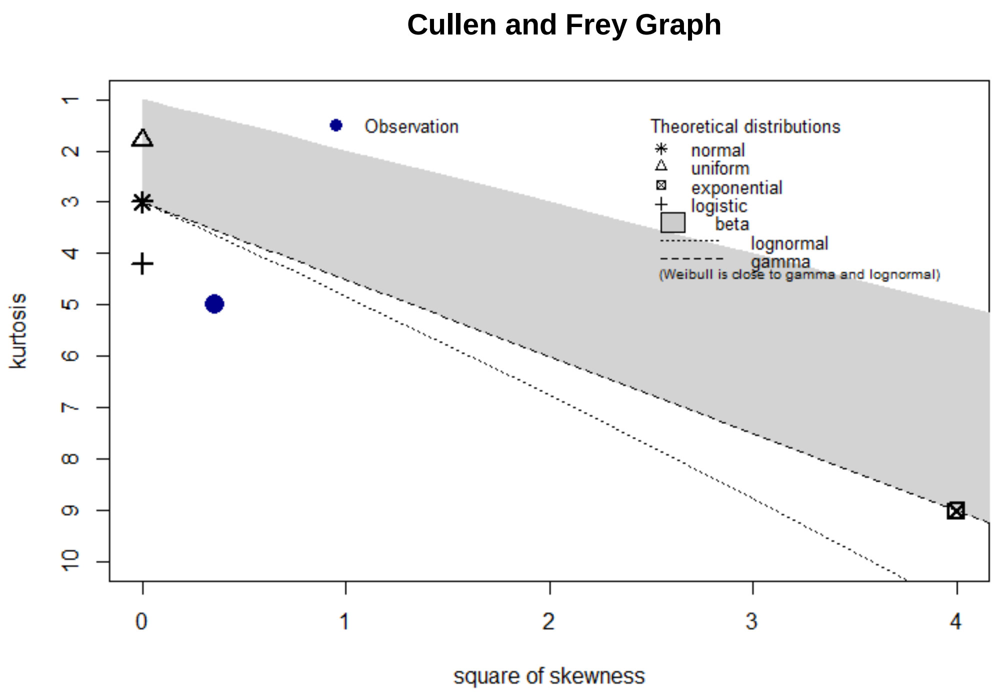

Figure 4.3 Cullen and Frey Graph of I-64 without Outliers

After deleting potential outliers, the skewness of observations has been eliminated to a large extent. Figure 4.3 indicates the input model has a very minor skewness and its kurtosis is close to the logistic symbol. It is better to estimate the input model following logistic distribution. AIC gives similar results. AIC values are shown as follows and the lowest AIC value is 6268.295, which also indicated logistic distribution is the best match for I-64 after deleting outliers.

```
> fitdist(as.numeric(s),"norm",method="mme",gof="CvM")$aic
[1] 6290.747
> fitdist(as.numeric(s),"lnorm",method="mme",gof="CvM")$aic
[1] 6474.135
> fitdist(as.numeric(s),"gamma",method="mme",gof="CvM")$aic
[1] 6378.753
> fitdist(as.numeric(s),"logis",method="mme",gof="CvM")$aic
[1] 6268.295
> fitdist(as.numeric(s),"exp",method="mme",gof="CvM")$aic
[1] 7143.285
```

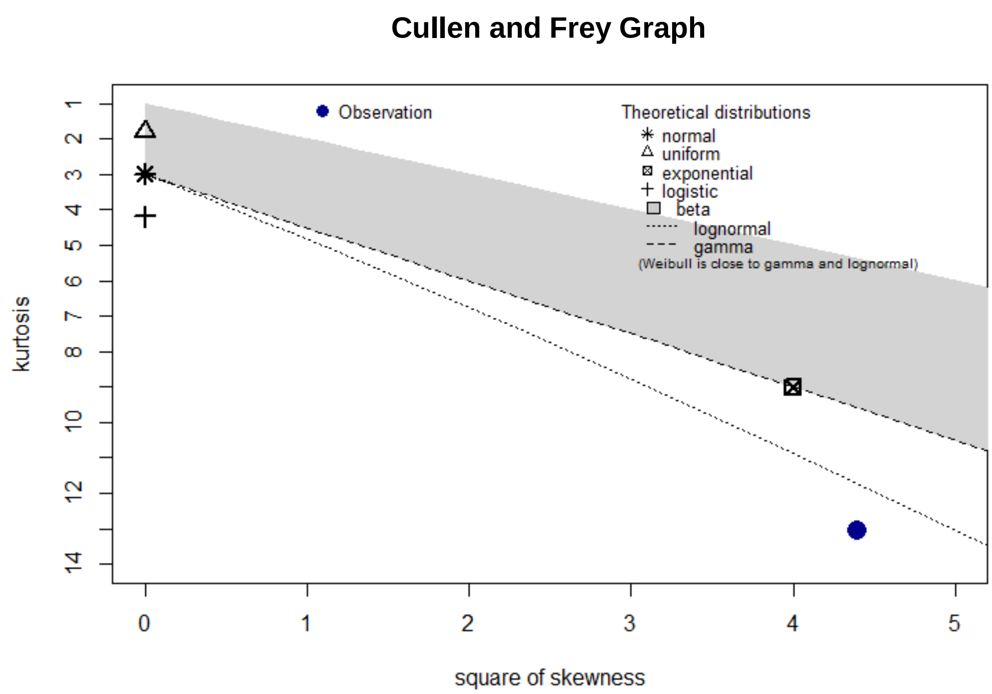

Figure 4.4 Cullen and Frey Graph of I-70 with Outliers

Figure 4.4 indicates the observation model has a strong skewness, which makes it look far away from all distributions except lognormal distribution. The situation could be improved after deleting outliers. AIC values are given as follows. The lowest AIC value is 5560.045 of logistic distribution followed by 5632.53 of normal distribution. Logistic distribution would also be considered the best match even with the effect of outliers.

```
> fitdist(as.numeric(s),"norm",method="mme",gof="CvM")$aic
[1] 5632.53
> fitdist(as.numeric(s),"lnorm",method="mme",gof="CvM")$aic
[1] 7914.375
> fitdist(as.numeric(s),"gamma",method="mme",gof="CvM")$aic
[1] 6309.687
> fitdist(as.numeric(s),"logis",method="mme",gof="CvM")$aic
[1] 5560.045
> fitdist(as.numeric(s),"exp",method="mme",gof="CvM")$aic
[1] 6313.66
```

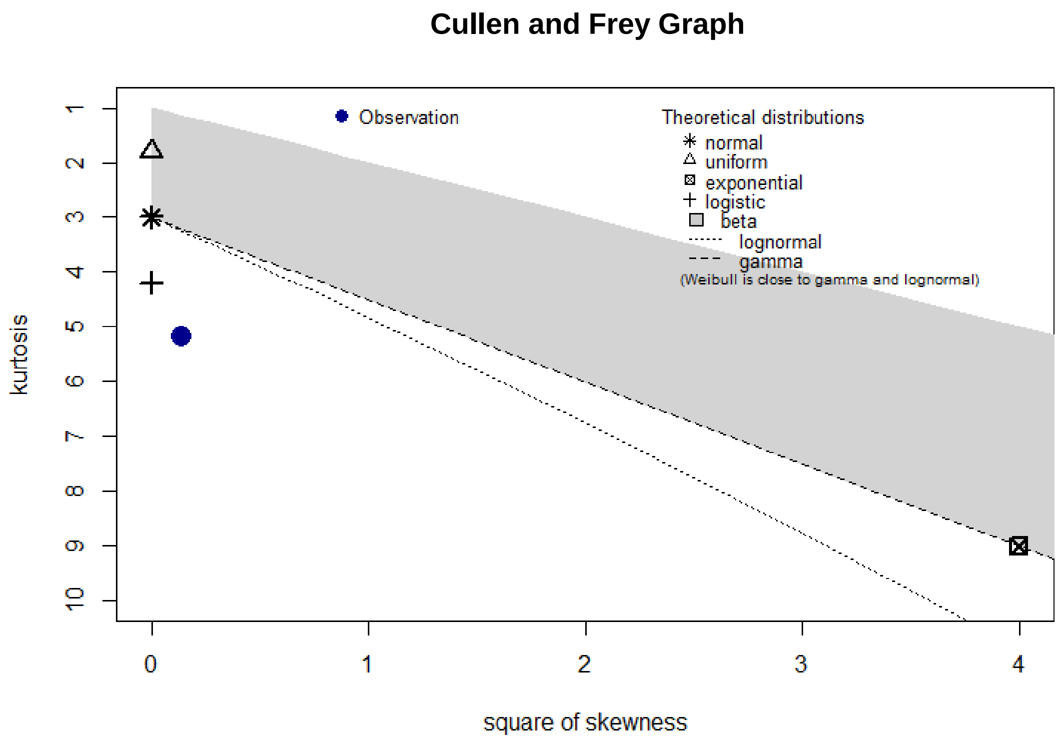

Figure 4.5 Cullen and Frey Graph of I-70 without Outliers

Figure 4.5 shows the skewness-kurtosis graph after deleting outliers. The skewness is approximately 0. It does not have a zero skewness that may be affected by the definition of outliers in this study. To avoid deleting of unnecessary data, this study defines outliers as three times that of the inter-quartile range (IQR), rather than 1.5 times in common studies because the skewness of most Interstates could be eliminated under this definition. Both skewness and kurtosis of the observations model shows it is very close to logistic distribution.

AIC values are as follows and the lowest AIC value is 5316.947, which also indicated logistic distribution is the best match for I-70 after deleting outliers.

```
> fitdist(as.numeric(s),"norm",method="mme",gof="CvM")$aic
[1] 5337.51
> fitdist(as.numeric(s),"lnorm",method="mme",gof="CvM")$aic
[1] 5376.428
> fitdist(as.numeric(s),"gamma",method="mme",gof="CvM")$aic
[1] 5358.512
> fitdist(as.numeric(s),"logis",method="mme",gof="CvM")$aic
[1] 5316.947
> fitdist(as.numeric(s),"exp",method="mme",gof="CvM")$aic
[1] 6185.023
```

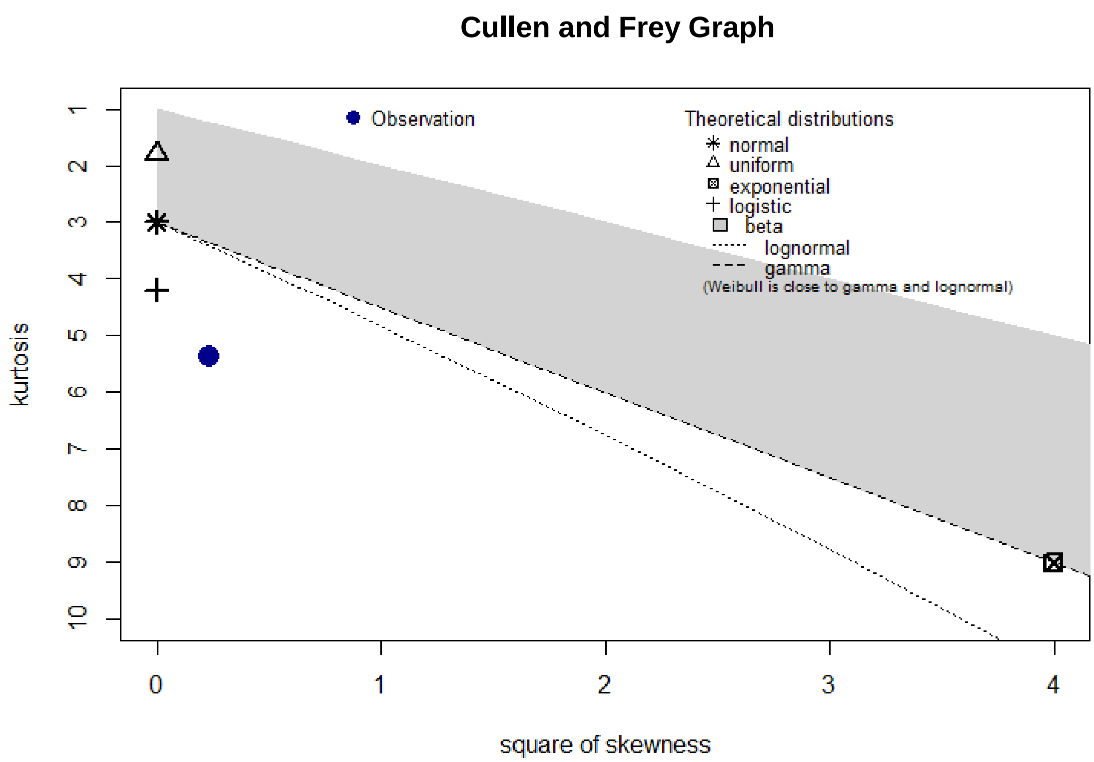

Figure 4.6 Cullen and Frey Graph of I-80/90 with Outliers

The observation of I-80/90 in Figure 4.6 has a minor skewness and it might be a normal, uniform, and logistic distribution. The kurtosis of observation model is close to logistic distribution. AIC provides similar outcomes. The lowest AIC value is 7469.152 that indicated logistic distribution is the best match for the input model with outliers.

```
> fitdist(as.numeric(s),"norm",method="mme",gof="CvM")$aic
[1] 7485.981
> fitdist(as.numeric(s),"lnorm",method="mme",gof="CvM")$aic
[1] 8306.177
> fitdist(as.numeric(s),"gamma",method="mme",gof="CvM")$aic
[1] 7760.13
> fitdist(as.numeric(s),"logis",method="mme",gof="CvM")$aic
[1] 7469.152
> fitdist(as.numeric(s),"exp",method="mme",gof="CvM")$aic
[1] 8268.321
```

After deleting potential outliers, the skewness of observations has been eliminated. From Figure 4.7, it clearly showed the kurtosis of observation model is close to the normal and logistic symbol. It is safe to say the distribution of input data is either normal or logistic distributed. AIC could help the judgment. AIC values are presented as follows and the lowest AIC value is from a normal model (7368.568) followed by a logistic model (7373.749). This indicates normal distribution might be the best match for I-80/90. However, these two values are close in regards to the loss of information. From the Cullen and Frey Graph, the unknown model shows an excess kurtosis. Therefore, it is reasonable to consider the unknown model follows normal distribution.

```
> fitdist(as.numeric(s),"norm",method="mme",gof="CvM")$aic
[1] 7368.564
> fitdist(as.numeric(s),"lnorm",method="mme",gof="CvM")$aic
[1] 7530.35
> fitdist(as.numeric(s),"gamma",method="mme",gof="CvM")$aic
[1] 7423.213
> fitdist(as.numeric(s),"logis",method="mme",gof="CvM")$aic
[1] 7373.749
> fitdist(as.numeric(s),"exp",method="mme",gof="CvM")$aic
[1] 8206.929
```

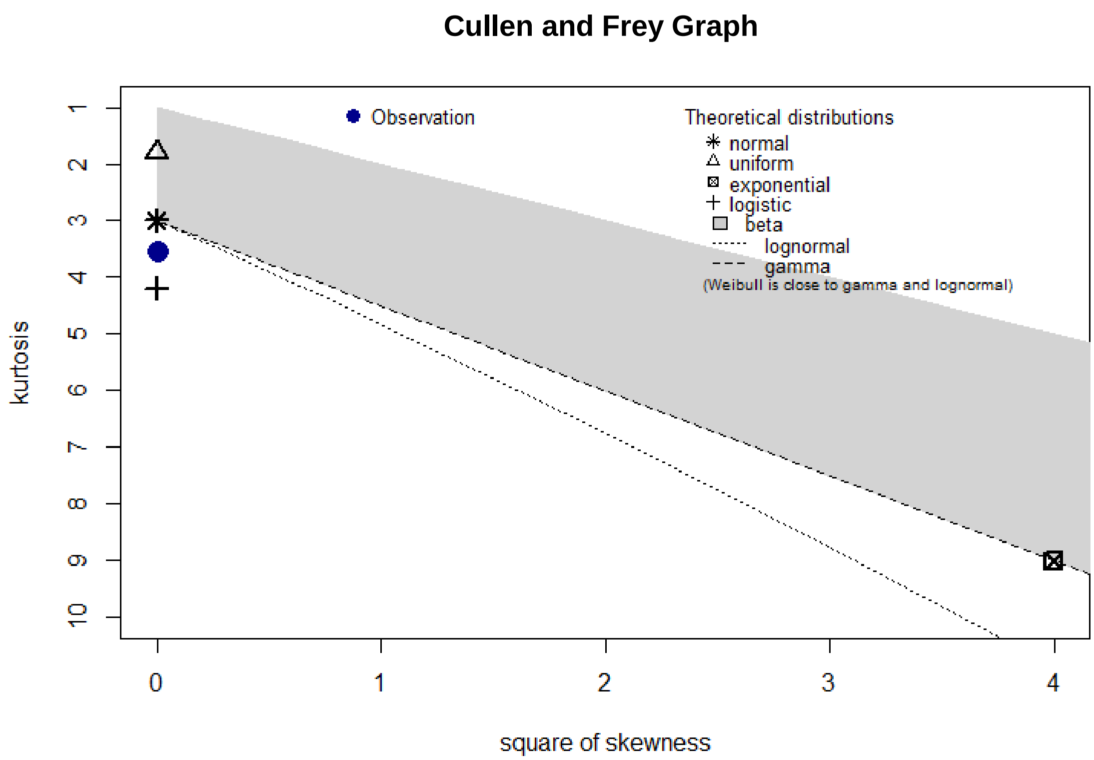

Figure 4.7 Cullen and Frey Graph of I-80/90 without Outliers

In the Cullen and Fray Graph for the original data without deleting outliers, all three data sets show clear skewness and kurtosis, a conflict to the assumption of normal and logistic distribution as they allow zero skewness. However, after deleting potential outliers, the skewness of the observation model has been almost removed, but it still shows an excess kurtosis. Normal distribution has an excess kurtosis of 0, while logistic distribution has an excess kurtosis of 1.2. Hence the logistic distribution would be a better match than normal distribution for input data model. Also, based on the AIC, the logistic model is the best candidate with the smallest AIC values compared with all other candidate models in both with and without outlier scenarios.

The probability density function (PDF) plot provides another method for determining the distribution of data. The PDF plot of I-64, I-70, and I-80/90 are shown as below to visually examine the data distribution. Figure 4.8, Figure 4.9, and Figure 4.10 are distribution plots of ADT data with outliers. Figure 4.11, Figure 4.12 and Figure 4.13 are distribution plots of corresponding roads after the elimination of outliers.

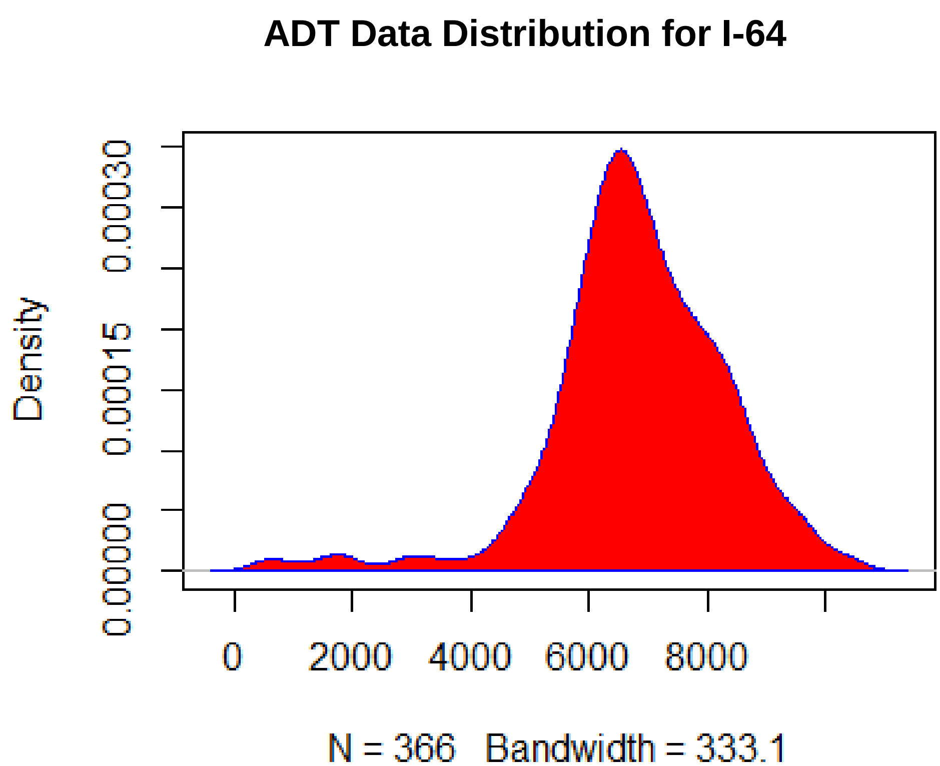

Figure 4.8 ADT Data Distribution for I-64 with Outliers

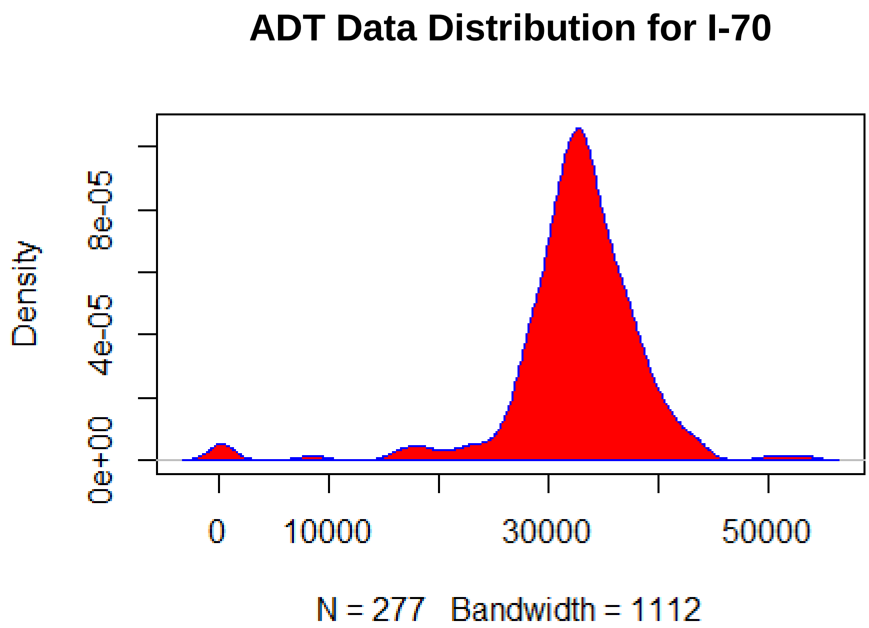

Figure 4.9 ADT Data Distribution for I-70 with Outliers

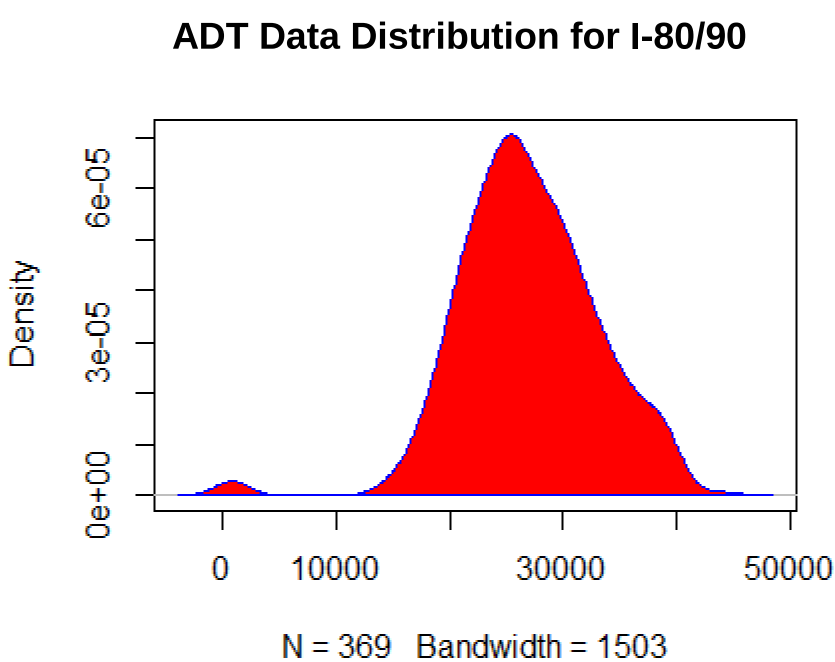

Figure 4.10 ADT Data Distribution for I-80/90 with Outliers

The graphs of Figure 4.8, Figure 4.9, and Figure 4.10 show the probability density plots of each data set with outliers. From all three plots, a clear fat-tail shows in each plot. Extreme values in the tails could be accounted for the cause of the skew. Each plot also indicates strong left skewness and a distinct peaked top.

A closer look was taken to eliminate the effect of outliers. After truncating those fat tails, each density plot shows a roughly symmetric pattern. Figure 4.11 to Figure 4.13 present the symmetric pattern of daily traffic distribution for I-64, I-70, and I-80/90.

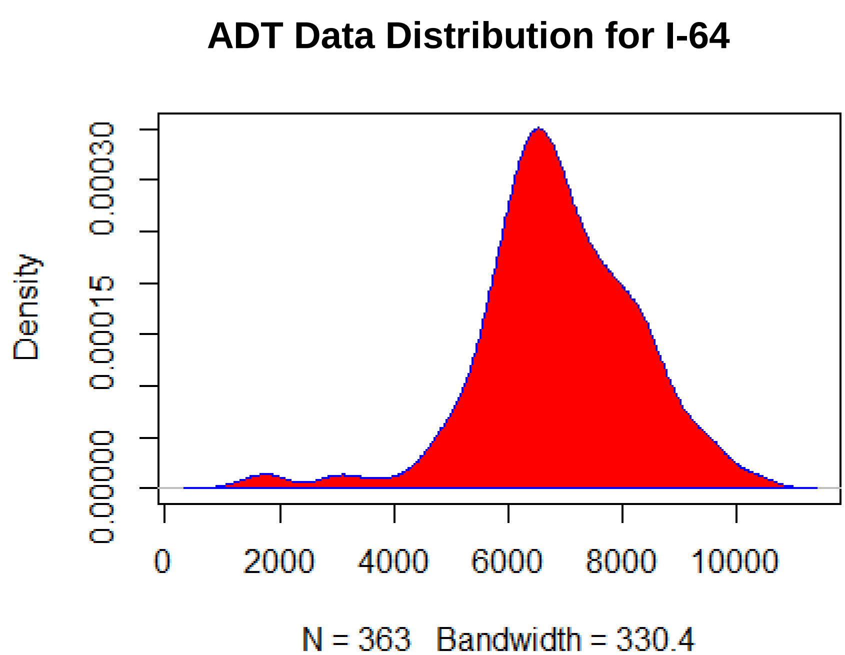

Figure 4.11 ADT Data Distribution for I-64 without Outliers

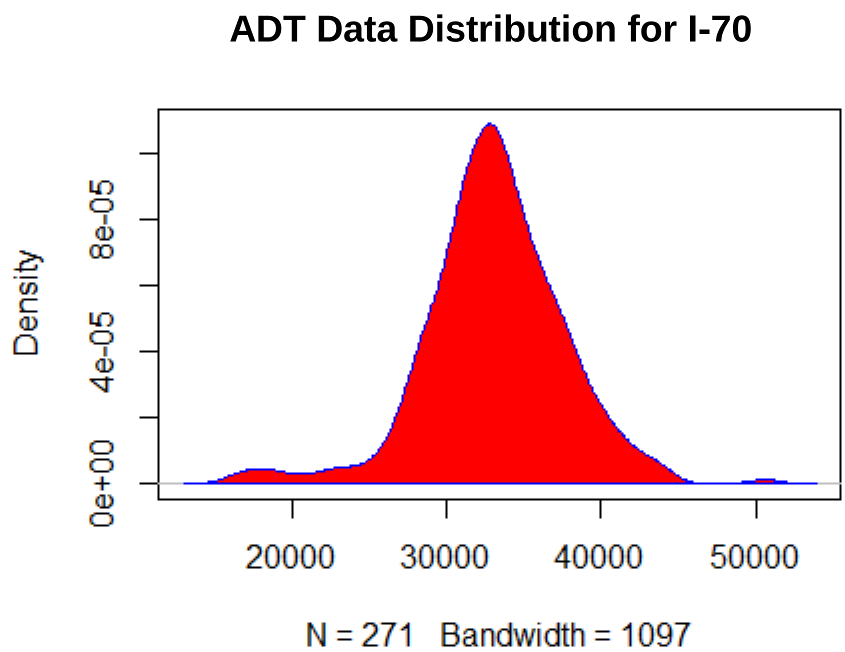

Figure 4.12 ADT Data Distribution for I-70 without Outliers

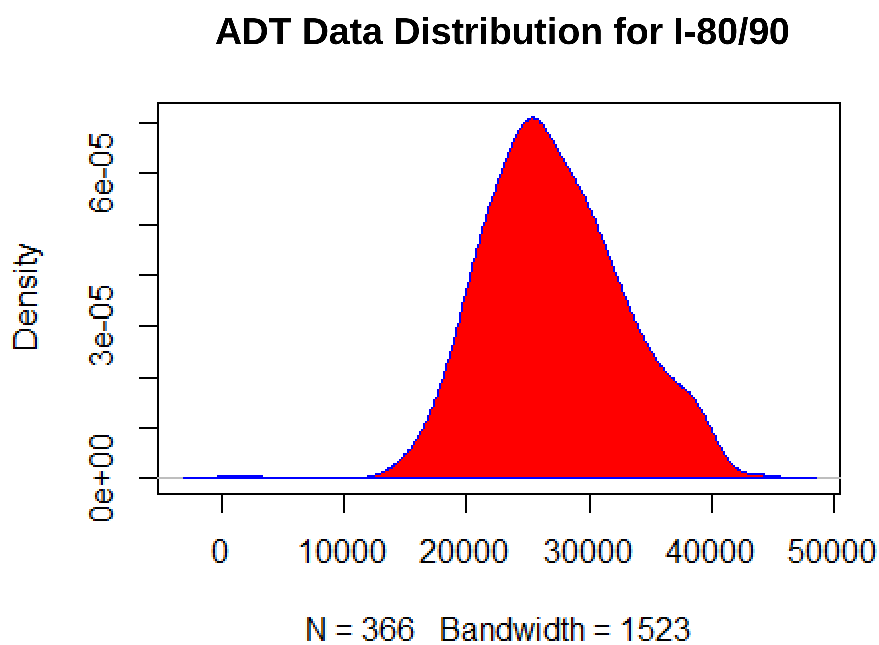

Figure 4.13 ADT Data Distribution for I-80/90 without Outliers

The above plots have a roughly symmetric distribution without obvious skewness. The mean value of daily traffic in I-64 is about 6,500, about 33,000 in I-70, and about 32,500 in I-80/90. Their distributions have distinct peaks around the means and have heavier tails in both sides compared to a normal distribution. The PDF plots proved the analysis results of the Cullen and Frey Graph, and AIC: logistic distribution is the best match for ADT data in Interstates of Indiana.

4.2 Data Simulation

The Monte Carlo Simulation method was adopted to generate random numbers based on given distribution patterns. The simulation process was also accomplished in R programming. It contained a couple of steps as follows:

Step 1: From the analysis of Section 4.1, logistic distribution is the known distribution for the input variable, ADT data, in of each road. The following code was used to estimate the location (mean) and scale (deviation) of input data:

```
t <-
fitdist(as.numeric(s),"logis",method="mme",gof="CvM")$estimate
```

Step 2: Integrating the PDF of logistic function to CDF. CDF is uniformly distributed and has a lower limit of 0 and higher limit of 1. In R, it is not necessary to conduct this step because R could provide the quantile value of each observation by adding a “q” command before the command of distribution type in step 4.

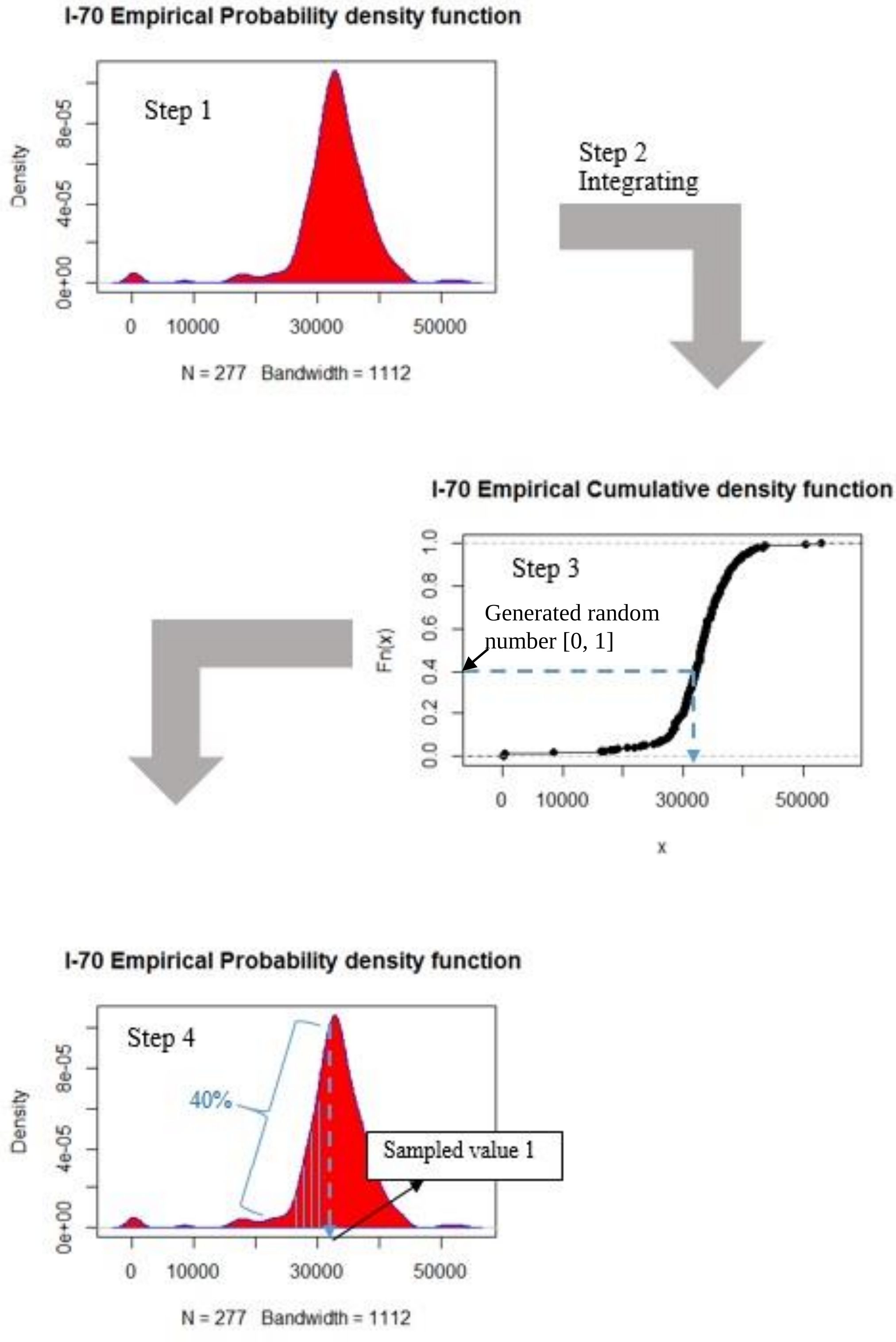

Figure 4.14 The Process of the Monte Carlo Simulation of I-70

Step 3: Generated random numbers from 0 to 1 and repeated the same process ten thousand times by using:

```
es <- runif(10000, min=0, max=1)
```

Step 4: Applied generated numbers to the CDF and the simulated ADT values were the outcomes of the CDF function.

```
p <- qlogis(es,location=t[1],scale=t[2])
```

The process of the Monte Carlo simulation is also presented in Figure 4.14 using I-70 as an example. Figure 4.14 only illustrates the simulation process for one iteration. As the number of iterations increases, the number of the sampled value would also be increased. The Monte Carlo simulation method could be applied to any road when the road traffic data distribution is determined. Table 4.1 shows part of the simulated daily traffic values of I-70. All these values were generated based on the logistic characters of the ADT data in I-70.

PDF plots can help to have a better look at the data. Figure 4.15, Figure 4.16, and Figure 4.17 are simulated distribution PDF plots for I-64, I-70, and I-80/90. This simulated data is constituted by ten thousand simulated observations based on the ADT data distribution of each road.

Table 4.1 *Simulated ADT Data of I-70 (Part)*

[1]	38839.48	32874.44	31470.80	21626.24	32344.92	30944.94
[7]	34443.51	33206.05	35450.83	36316.02	30298.42	30124.71
[13]	35524.60	30382.62	40198.16	34015.01	24484.23	35397.76
[19]	34050.03	36626.60	28899.74	29724.79	36604.22	34305.96
[25]	36730.91	36053.44	37584.19	33921.29	23886.34	38707.78
[31]	33731.75	34221.05	29206.72	25351.09	40054.66	26528.31
[37]	32953.22	32566.98	34811.89	32068.45	25856.29	31892.43
[43]	26857.66	36361.58	35936.45	35026.34	35785.29	24436.26
[49]	29012.25	27405.96	41404.26	31005.17	28267.16	28798.12
[55]	27099.77	34203.17	37901.50	38234.42	32606.42	31335.98
[61]	39719.14	31987.07	31748.61	30606.78	28616.13	33568.95
[67]	30188.32	41372.38	32291.70	30693.95	30588.64	31918.61
[73]	31610.21	32263.76	36701.02	29537.99	28224.37	32080.36
[79]	36465.67	31177.58	35197.13	37976.84	33744.86	36388.25
[85]	31174.79	22394.18	35982.88	33651.92	44185.13	34115.19
[91]	36147.97	23893.53	33858.29	37634.94	27125.39	33211.95
[97]	33467.37	28018.49	37267.30	27892.41	33058.46	32351.95
[103]	32867.00	34881.16	32692.27	25798.13	31447.24	26147.75
[109]	35282.90	32208.23	38476.40	30204.84	37748.06	41264.15
[115]	36844.50	32614.67	34484.66	30152.91	35914.80	33694.51
[121]	32587.24	28648.49	52377.39	37655.90	33772.64	24499.83
[127]	36766.07	34776.58	33293.48	32546.99	38289.11	35590.44
[133]	34580.65	35439.90	35356.18	23348.53	29999.77	32432.05
[139]	24842.32	38641.56	34597.14	35730.50	36490.22	38390.87
[145]	25141.46	28482.83	29756.72	36399.86	32030.85	38015.32
[151]	30206.64	30226.89	32631.01	35992.73	22427.54	40059.02
[157]	33204.20	31551.80	30716.77	40388.73	24520.46	33340.06
[163]	33861.71	35595.62	33701.75	39367.53	36394.93	33875.50
[169]	32785.76	30489.04	33438.66	36338.01	36655.08	43038.07
[175]	36585.78	36547.50	37094.85	43053.50	35557.57	35075.05
[181]	27616.24	33428.08	35832.26	38590.48	28629.49	29546.29
[187]	32500.42	31647.76	35908.75	33201.56	30638.23	34281.56
[193]	33563.58	32280.95	39256.81	28843.43	33454.99	33532.04
[199]	31965.78	33096.10	36135.29	30570.39	43692.30	33254.33
[205]	39468.23	28623.06	36187.19	33717.97	30956.18	33469.81
[211]	32935.64	28112.65	42328.71	35997.57	37037.91	29207.50
[217]	27054.49	29274.21	39894.46	33020.93	30730.13	36323.63
[223]	32736.87	36415.55	30493.45	32715.40	38414.54	33458.32
[229]	37649.91	34030.45	32986.20	34252.26	32645.75	35081.43
[235]	40858.61	32715.73	33439.08	33796.04	35026.46	36026.61
[241]	35405.55	26606.94	31412.67	31809.95	32537.24	31528.41
[247]	38672.39	30354.77	32927.18	34655.90	36339.42	31670.68
[253]	37231.16	34056.63	31498.39	26423.43	36269.83	35942.11
[259]	30545.62	31955.45	29678.91	30946.86	33708.37	31488.75
[265]	41185.44	36843.47	34251.81	36665.67	34209.74	35234.43
[271]	31248.06	29635.73	33125.23	27090.28	33153.71	32550.82
[277]	29310.93	30910.70	41677.66	32824.48	29795.72	32559.15
[283]	33816.76	38119.35	28012.23	38286.35	33948.07	25729.77
[289]	34907.56	37061.95	34146.45	33846.34	30944.04	33181.61
[295]	36497.26	28147.64	33985.52	29684.48	31586.89	32591.86
[301]	32435.25	37887.40	40173.40	28793.58	41008.71	35602.87
[307]	33820.79	32414.73	31531.31	35393.45	31547.59	31524.71
[313]	31760.11	34995.21	32249.97	33111.65	30107.27	33207.46
[319]	30581.49	33029.24	33182.68	34517.73	34493.59	35262.49
[325]	32177.10	29260.61	27879.48	31972.13	28780.05	33043.06
[331]	27080.85	37605.32	32447.25	35680.18	33033.15	47917.91
[337]	26795.31	31122.28	38885.81	32741.74	31056.32	32366.64
[343]	33637.37	38008.11	34832.21	34535.73	26280.64	36091.71

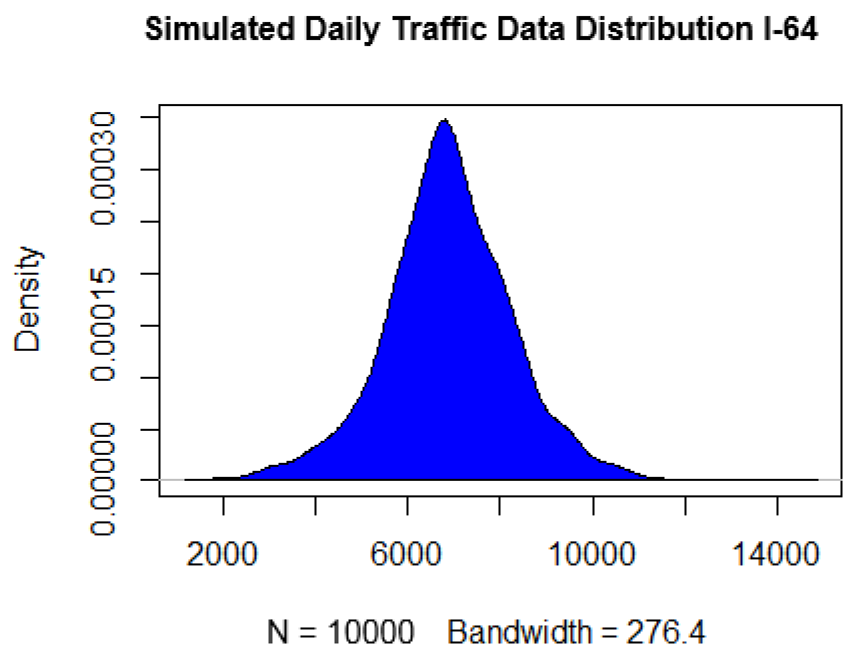

Figure 4.15 Simulated Daily Traffic Data Distribution Plot for I-64

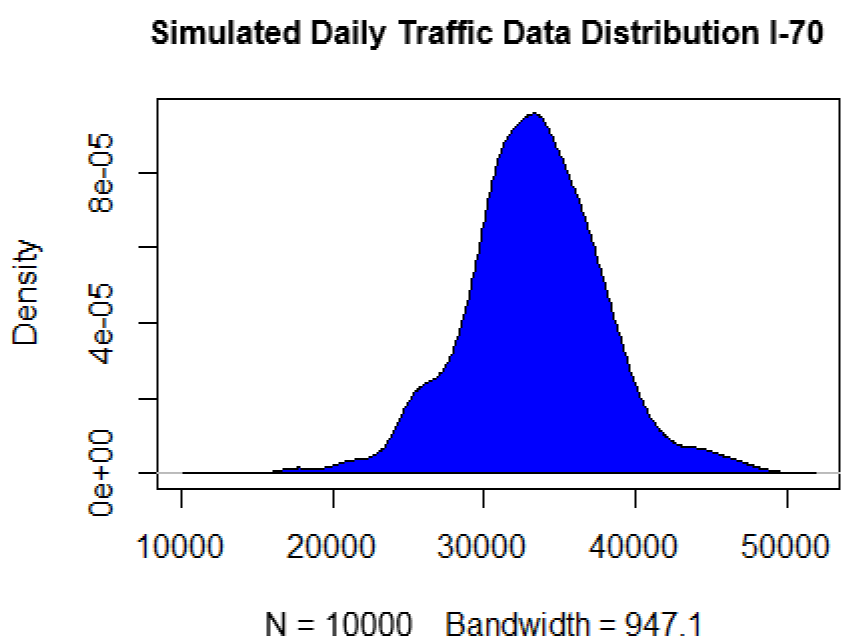

Figure 4.16 Simulated Daily Traffic Data Distribution Plot for I-70

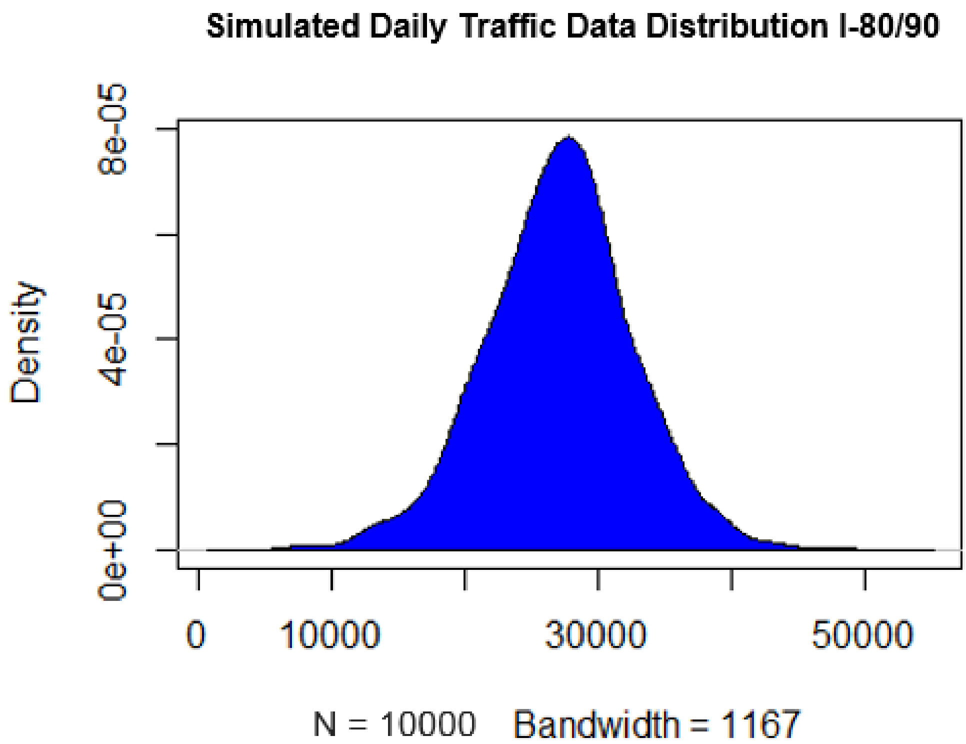

Figure 4.17 Simulated Daily Traffic Data Distribution Plot for I-80/90

It is necessary to mention that the simulation results always change when running the program every time. The above plots are only the simulated results of one time programming. However, all this simulated data shows similar patterns at each running and similar to their sample ADT data. All three figures above illustrate symmetric patterns without obvious skewness. The mean and deviation values of simulated data are close to their sample values. These simulated plots also demonstrate logistic features, and all of them have clear peaked tops and heavy tails on both sides.

4.3 Chapter Summary

This chapter analyzed the state-wide ADT data in Indiana. Three interstate roads, I-64, I-70, and I-80/90, were selected in this chapter to present the analysis results. By using the Cullen and Frey Graph, AIC and PDF plot to determine the ADT data distribution, all three roads show a symmetric distribution and a peaked top without obvious skewness. Logistic distribution was proved to be the best match for the ADT data in Indiana interstates. Monte Carlo simulation was then performed to generate random numbers to sample data for each road. Ten thousand iterations for each road were applied in the simulation process. The simulated values show similar mean, kurtosis, and deviation to the original ADT data.

CHAPTER 5. PROBABILISTIC BENEFIT-COST ANALYSIS

To further examine the impacts of daily traffic changes to highway facility benefits and costs, this chapter applied the simulated daily traffic data in the analysis model to study project benefits and costs with the consideration of risk.

5.1 Build the Analysis Algorithm

To build the probabilistic model of benefit-cost analysis, the first step was to understand the relationships between traffic volume changes and highway user benefits and costs. Therefore, it is necessary to figure out factors affecting highway costs and savings. The determinations of highway user costs and savings were discussed in the following sections.

5.1.1 Agency costs

Three agency costs were mainly discussed in this study: initial costs, maintenances costs and rehabilitations costs.

Initial costs account for over eighty percent of the total costs of a project. It includes the costs of management, engineering, construction costs, site condition improvement, and tax (WSDOT, 2004).

Table 5.1 *Construction Costs per Mile in Indiana*

(Adapted from: Jiang et al., 2013)

Project Type	Road Type	No. of Projects	Mean
Added Travel Lanes	Interstate	20	$14,097,614.90
	US	12	$6,440,628.67
	State Road	15	$5,502,277.11
HMA Overlay, Functional	Interstate	15	$1,069,419.81
	US	12	$839,291.16
	State Road	13	$470,455.63
		29	$363,398.26
HMA Overlay, Preventive Maintenance	Interstate	36	$728,156.54
		15	$529,642.89
	US	52	$436,967.89
		28	$278,111.37
	State Road	102	$395,904.04
		50	$240,850.16
		19	$183,510.52
Pavement Replacement	Interstate	4	$9,255,303.93
	US	12	$3,957,611.24
	State Road	12	$3,372,094.54
Road Reconstruction (3R/4R Standards)	US	6	$2,547,004.92
	State Road	23	$4,701,593.40
Road Rehabilitation (3R/4R Standards)	Interstate	4	$3,862,214.96
	US	9	$2,398,782.89
	State Road	10	$2,311,765.27
Surface Treatment, Microsurface	US	16	$120,952.23
	State Road	14	$138,907.71
Surface Treatment, Thin HMA Overlay	State Road	9	$112,172.96
Surface Treatment, Ultrathin Bonded Wearing Course	US	10	$185,574.56
	State Road	20	$129,462.80

Unit of initial costs is cost per lane mile, which mostly analyzes construction costs due to the complexity of determining issues specifically related to a project, such as right of way, site conditions, and environmental conditions (WSDOT, 2004). The estimation of initial costs needs sufficient project construction cost data in previous years and is out the scope of this study. A National Highway Construction Costs Index was provided by

FHWA in 2007 to show the trend of construction costs. Besides, several states provided state-level highway construction cost reports to assist planners in performing highway economic analyses in their early stages. Table 5.1 shows part of the construction costs per mile in Indiana adapted from Jiang et al. (2009).

Maintenance costs in this study refer to routine annual maintenance. Other small maintenance costs are excluded in this study due to its minor economic impacts. Routine maintenance costs obtained from HERS (2005) for flexible pavements are shown in Table 5.2.

Table 5.2 *Maintenance Cost for Flexible Pavements (1984 Dollars)*
(Adapted from: HERS, 2005)

Final PSI	Maintenance Cost Between PSI Levels ($/lane mile)	Cumulative Cost ($/lane mile)
Low SN/traffic: (SN=2.16)		
4.0	221.57	221.57
3.5	767.03	988.60
3.0	1,314.95	2,302.55
2.5	1,859.47	4,163.02
2.0	2,413.74	6,576.76
1.5	2,957.34	9,534.10
Medium SN/traffic: (SN=3.60)		
4.0	339.10	339.10
3.5	1,174.05	1,513.15
3.0	2,012.72	3,525.87
2.5	2,845.76	6,371.63
2.0	3,604.98	10,066.61
1.5	4,526.45	14,593.06
High SN/traffic: (SN=5.04)		
4.0	456.63	456.63
3.5	1,581.05	2,037.38
3.0	2,710.50	4,748.18
2.5	3,832.04	8,580.22
2.0	4,976.21	13,556.43
1.5	6,095.55	19,651.98

Rehabilitations occur when the existing facility is in disrepair and it is necessary to rebuild it. More than one rehabilitation activities should be included in an economic analysis period (Zhao, 2012). Rehabilitation costs could be obtained from previous data. Transportation departments of some states also provided recommended rehabilitation costs to assist planers in early planning stages when they do not have enough information of the project maintenance and rehabilitations. In most situations, project-specific maintenance and rehabilitation costs are decided according to the performance condition of a highway facility (Zhao, 2012).

5.1.2 User Benefits

This study mainly discussed three aspects of user benefits after a highway improvement: travel time savings, vehicle operation costs savings, and safety savings. The simulated daily traffic data from chapter 4 would be used in saving calculations to dynamically analyze benefits of a project before and after an improvement. Therefore, an essential step was to show how traffic volume changes affect project costs and benefits. The following sections discussed the relationships between traffic volume and the three aspects of user benefits.

5.1.2.1 Travel Time Saving

Most people are concerned about the time spent on the road. A major purpose of building a new road, adding lanes, or conducting road maintenance is to relieve or solve congestion, which is the major cause of many public irritations (FHWA, 2003).

Travel time cost of a certain type of vehicle per vehicle mile (TTCST) is defined in HERS (2005) as the average value of time (TTVAL) divided by the average effective speed (AES) of the type of vehicle. An average travel time cost for a specific type of road is the weighted value of the travel time costs of all vehicle types. Weights could be obtained by daily traffic distribution data from Weigh in Motion (WIM) and could vary from place to place. The average vehicle occupancy (AVO) is also an important index in estimating travel time costs. The National Household Travel Survey (NHTS) (2009) calculated AVO for work purposes as 1.13, shopping as 1.78, personal business as 1.84, recreational purpose as 2.20, and all purposes as 1.67. Free flow speed is the speed with no congestion and other adverse impacts. With the increase of vehicles in the same lane, free flow speed will decrease. In this study, the free flow speed for an eight-lane road with 1600 passenger cars per hour per lane (pcphpl) is set as 65 km/h after built. This study considered all purposes travel and used a weighted value of time of business trips and personal trips to simplify the analysis.

Hourly traffic distribution (HTD) is the percentage of each vehicle type in an hour (Jiang et al., 2013). The data is obtained from Weigh in Motion (2004) in Indiana and includes eight state roads, seven U.S. highways, and eighteen Interstates. Table 5.3 illustrated the HTD data in 24 hour.

Table 5.3 *Hourly Traffic Distribution in Indiana (2004)*

(Adapted from: Jiang et al, 2013)

	Average Interstate				Average Multilane				Average Two-lane			
Hour	All	Auto	S-U	Comb.	All	Auto	S-U	Comb.	All	Auto	S-U	Comb.
0->1	1.48%	0.96%	0.09%	0.43%	1.11%	0.89%	0.05%	0.18%	0.97%	0.68%	0.06%	0.23%
1->2	1.14%	0.68%	0.07%	0.39%	0.68%	0.49%	0.03%	0.16%	0.67%	0.42%	0.04%	0.20%
2->3	1.00%	0.55%	0.06%	0.39%	0.55%	0.36%	0.03%	0.16%	0.63%	0.39%	0.04%	0.20%
3->4	1.06%	0.58%	0.07%	0.41%	0.61%	0.39%	0.04%	0.18%	0.69%	0.38%	0.04%	0.26%
4->5	1.39%	0.84%	0.10%	0.46%	1.02%	0.74%	0.07%	0.21%	1.22%	0.78%	0.11%	0.34%
5->6	2.38%	1.70%	0.17%	0.51%	2.31%	1.89%	0.16%	0.25%	2.39%	1.70%	0.22%	0.47%
6->7	3.92%	3.11%	0.27%	0.54%	4.42%	3.80%	0.31%	0.31%	3.68%	2.71%	0.40%	0.57%
7->8	4.88%	3.96%	0.34%	0.58%	5.81%	5.06%	0.37%	0.38%	5.19%	4.00%	0.53%	0.66%
8->9	4.82%	3.79%	0.37%	0.66%	5.40%	4.57%	0.40%	0.43%	5.17%	3.85%	0.53%	0.79%
9->10	4.88%	3.76%	0.38%	0.73%	5.05%	4.18%	0.40%	0.46%	5.56%	4.17%	0.50%	0.89%
10->11	5.18%	3.99%	0.40%	0.79%	5.28%	4.39%	0.40%	0.49%	5.91%	4.45%	0.51%	0.95%
11->12	5.48%	4.26%	0.42%	0.81%	5.67%	4.77%	0.40%	0.49%	6.46%	4.91%	0.59%	0.96%
12->13	5.88%	4.47%	0.50%	0.92%	5.95%	5.06%	0.41%	0.49%	6.40%	4.93%	0.56%	0.91%
13->14	6.08%	4.69%	0.48%	0.90%	6.02%	5.12%	0.42%	0.47%	6.38%	4.96%	0.57%	0.85%
14->15	6.41%	5.00%	0.52%	0.89%	6.40%	5.50%	0.45%	0.45%	6.64%	5.22%	0.62%	0.80%
15->16	6.89%	5.53%	0.52%	0.85%	7.32%	6.43%	0.47%	0.42%	7.47%	6.03%	0.71%	0.73%
16->17	7.12%	5.84%	0.50%	0.78%	7.73%	6.91%	0.43%	0.39%	7.76%	6.45%	0.67%	0.65%
17->18	6.74%	5.55%	0.42%	0.77%	7.63%	6.91%	0.37%	0.35%	7.03%	5.81%	0.64%	0.57%
18->19	5.81%	4.55%	0.40%	0.86%	6.11%	5.52%	0.27%	0.31%	5.52%	4.55%	0.47%	0.50%
19->20	4.75%	3.66%	0.32%	0.77%	4.48%	4.00%	0.19%	0.28%	4.33%	3.52%	0.36%	0.45%
20->21	4.01%	3.03%	0.26%	0.72%	3.58%	3.16%	0.15%	0.26%	3.57%	2.89%	0.28%	0.40%
21->22	3.50%	2.55%	0.20%	0.74%	2.96%	2.60%	0.12%	0.24%	2.88%	2.30%	0.21%	0.37%
22->23	2.85%	1.97%	0.18%	0.69%	2.26%	1.95%	0.09%	0.22%	2.07%	1.64%	0.14%	0.29%
23->24	2.37%	1.50%	0.20%	0.66%	1.66%	1.40%	0.07%	0.19%	1.39%	1.03%	0.10%	0.26%
Total	100.00%	76.51%	7.24%	16.25%	100.00%	86.12%	6.09%	7.79%	100.00%	77.77%	8.91%	13.32%

Office of the Secretary of Transportation (OST) (2011) defined the value of travel time saving in three aspects. First, saved travel time could be used for other productive fields and bring benefits to travelers and society. Second, the cost of the same amount of time for more enjoyable activities. Third, the cost of the same amount of time for other enjoyable parts of the trip. Based on the above principles, OST (2011) has analyzed the factors, such as trip purpose, personal characteristics, and hourly income, and recommend the hourly monetary values of travel time saving. The results are shown in Table 5.4.

Table 5.4 *Hourly Values of Travel Time Savings (2009 U.S. $ per person-hour)* (Adapted from: Office of the Secretary of Transportation, 2011)

Category	Surface Modes (Except High-Speed Rail)
Local Travel	
Personal	$12.00
Business	$12.50
All purposes	$12.50
Intercity Travel	
Personal	$16.70
Business	$22.90
All purposes	$18.00

RealCost (2004) suggested VOT for passenger cars is $10 to $13, single-unit trucks is $17-$20, and combination trucks is $21 to $24. Cal-B/C (2007) suggested the value of travel time for the automobile is $11.6 and for trucks is $28.7. The value of travel time in Indiana (Gkritza, et al., 2006) is shown in Table 5.5.

Table 5.5 *Value of Travel Time in Indiana (in 2003 dollar)*

(Adapted from Gkritza, et al., 2006)

Type of Vehicle	Value of Travel Time ($ per person hour)
Automobile	$20.57
Single-Unit Truck	$24.46
Combination Truck	$29.55

Based on above information, travel time cost could be calculated as total travel time, which is determined by facility length and average speed, multiplied by value of time. Detailed equations are as followed:

$$\mathrm{TT} = \frac{\mathrm{D}}{\mathrm{AS}}$$

$$\mathrm{AS} = \frac{\mathrm{FFS}}{1 + 0.15 \times \left(\frac{\mathrm{ADT} \times \mathrm{HTD}}{\mathrm{Pcphpl} \times \mathrm{Lanes}}\right)^4}$$

$$\mathrm{HourlyTTS} = (\mathrm{TT_o} \times \mathrm{AADT} \times \mathrm{HTD} \times \mathrm{AVO} - \mathrm{TT_n} \times \mathrm{AADT} \times \mathrm{HTD} \times \mathrm{AVO}) \times \mathrm{VOT}$$

$$\text{Yearly Travel Time Saving} = 365 \times \sum_{i=1}^{24} \mathit{Hourly\ TTS}$$

Where:

ADT = annual daily traffic

AVO = average vehicle occupancy

AS = average speed

D = facility length (unit: mile)

FFS = free flow speed

HTD = hourly traffic distribution

Lanes = number of traffic lanes

n = new scenario after improvements

o = old scenario

pcphpl = passenger car per hour per lane

TTS = travel time saving

VOT = value of time

ADT is expecting a two percent increase year by year. To perform economic analysis, travel time cost was calculated on a yearly basis and was discounted to present values in base year.

5.1.2.2 Vehicle Operation Cost Saving

Vehicle operation cost (VOC) is the expenses associated with operating or owning a vehicle. Gas, maintenance, and tires are the main components of vehicle operation costs. Generally, gas cost is seventy percent of the overall vehicle operation costs (Gkritza, Labi, & Sinha, 2006). Thus, multiplying gas costs by 1.43 could be the total vehicle operation costs. Gas costs for a vehicle type are the results of unit gas cost times the gas consumption rate. According to current market prices, the recommended unit value of gas is presented in Table 5.6. Fuel consumption rate can be calculated when the average speed of a type of vehicle is known (McFarland et al., 1993) (shown in Table 5.8). It is noticeable that the gross vehicle weight (GVW) of trucks has an impact on its fuel consumption rate. Islam (2003) estimated GVW for northern, central, and southern Indiana (shown in Table 5.7).

Table 5.6 *Unit Gas Costs*

(Adapted from: Jiang et al., 2013)

Vehicle Type	Fuel ($/gal)
Automobiles	3.847
Trucks	4.132
Combination Trucks	4.132

Table 5.7 *Average Truck Gross Vehicle Weight Indiana (Unit: 1000 lbs.)*

(Adapted from: Islam, 2003)

Northern Areas		
Facility type	Single-unit trucks	Combination trucks
Rural interstates	12.6	50.6
Urban interstates	9.7	46.0
Urban other freeways & expressways	9.7	46.0
Urban other principal	9.7	46.0
Central Areas		
Rural interstates	10.5	56.6
Urban interstates	10.8	85.8
Urban other freeways & expressways	10.8	85.8
Urban other principal	14.3	49.1
Southern Areas		
Rural interstates	22.5	58.5
Urban interstates	11.2	47.8
Urban other freeways & expressways	11.2	47.8
Urban other principal	11.2	47.8

Table 5.8 *Equations for Fuel and Oil Consumption (Unit: gal/1,000 miles) (Adapted from: McFarland et al., 1993; Gkritza, Labi, & Sinha, 2006)*

Vehicle Type	Equation
Automobiles	Fuel = 65.46896 – 1.47217 × Speed + 0.02127 × Speed^2
Single-Unit Trucks	Log (Fuel) = 5.57605 + 0.00012 × Speed^2 –0.4656 × Log (Speed) + 0.29271 × Log (GVW)
Combination Trucks	Log (Fuel) = 5.57605 + 0.00012 × Speed^2 – 0.4656 × Log (Speed) + 0.29271 × Log (GVW)

Based on above information, vehicle operation cost savings for each class could be calculated by multiplying hourly vehicle-mile travelled and fuel unit cost and fuel consumption. Detailed equations are as followed:

$$\mathrm{AS} = \frac{\mathrm{FFS}}{1 + 0.15 \times \left(\frac{\mathrm{ADT} \times \mathrm{HTD}}{\mathrm{Pcphpl} \times \mathrm{Lanes}}\right)^4}$$

$$\begin{aligned}\text{Unit Fuel Consumption(Trucks)} \\ &= \exp[5.57605 + 0.00012 \times \mathrm{AS}^2 - 0.4656 \times \mathrm{Log}\,(\mathrm{AS}) + 0.29271 \\ &\times \mathrm{Log}\,(\mathrm{GVW})]\end{aligned}$$

$$\begin{aligned}\text{Unit Fuel Consumption(Automobile)} \\ &= 65.46896 - 1.47217 \times \mathrm{AS} + 0.02127 \times \mathrm{AS}^2\end{aligned}$$

$$\text{Hourly VOCS} = 1.43 \times \text{ADT} \times \text{HTD} \times \text{D} \times \text{Unit Fuel Cost} \times (\text{Unit Fuel Consumption}_o - \text{Unit Fuel Consumption}_n)$$

$$\text{Yearly VOCS} = 365 \times \sum_{i=1}^{24} Hourly\ VOC_i$$

Where:

ADT = annual daily traffic

AS = average speed

D = facility length (unit: mile)

exp = exponential function

FFS = free flow speed

GVW = Gross Vehicle Weight

HTD = hourly traffic distribution

Lanes = number of traffic lanes

n = new scenario after improvements

o = old scenario

pcphpl = passenger car per hour per lane

VOC = vehicle operation costs

VOCS = vehicle operation costs savings

Present values of VOCS in each year will then be calculated in order to perform economic analysis.

5.1.2.3 Safety Saving

Reducing traffic accident rates is always the highest goal of highway improvements. Poor road conditions, congestion, and bad road design could be a possible cause of traffic accidents. It is not easy to assign monetary value to traffic accidents because it sometimes involves human property and life loss (FHWA, 2003). Accident rates are applied based on facility type. Table 5.9 is the table of fatality and injury rates per million vehicle-mile travelled in Indiana. After the implementation of an improvement, the accident rates may decrease. In this study, the crash reduction factor (CRF) is adopted to adjust the accident rate in a new scenario. Table 5.10 is part of the CRF table obtained from Jiang et al. (2013) showing the reduction rate of an implementation. A CRF value of 0.75 in Table 5.10 means that after this specific action, the crash rate in this facility will be reduced by 25 percent Table 5.11 listed the estimated accident costs in Indiana highways (Jiang et al., 2013).

Table 5.9 *Fatality and Injury Rates per Million VMT*
(Adapted from Jiang et al., 2013)

Road Type	Fatal	Injury	PDO
Rural Area			
Interstate	0.0119	0.2430	0.4000
Multilane Highway	0.0158	0.8100	0.6700
Two Lane Highway	0.0240	1.1690	1.0100
Urban Area			
Interstate	0.0120	0.3310	0.6100
Multilane Highway	0.0180	1.8430	2.4650
Two Lane Highway	0.0203	2.4710	4.0533

Table 5.10 *Crash Reduction Factors*

(Adapted from Jiang et al., 2013)

Improvement	Facility	CRF Total	CRF I/F	CRF PDO
Road widening	Rural interstate	0.74	0.75	0.70
Road widening	Urban interstate	0.70	0.75	0.74
Road widening	Rural multilane/two-lane	0.40	0.50	0.30
Road widening	Urban multilane/two-lane	0.30	0.50	0.40
Median construction	Rural facilities	N/A	0.13	0.00
Median construction	Urban facilities	N/A	0.11	0.00
Interchange construction	Rural/Urban facilities	N/A	0.87	0.74
New road construction	Rural interstate	0.34	0.52	0.30
New road construction	Urban interstate	0.24	0.13	0.25
New road construction	Rural multilane/two-lane	0.13	0.10	0.15
New road construction	Urban multilane/two-lane	0.33	0.18	0.34

Table 5.11 *Estimated Accident Costs in Indiana Highways*

(Adapted from Jiang et al., 2013)

Facility Type	Injury/Fatal	Property Damage Only
Rural Interstate	$78,717	$6,822
Rural Other Principal	$81,866	$6,822
Rural Minor Arterial	$81,866	$6,822
Rural Major Collector	$81,866	$6,822
Rural Minor Collector	$81,866	$6,822
Rural Local	$59,300	$6,822
Urban Interstates	$54,577	$6,822
Urban Other Freeways & Expressways	$50,379	$6,822
Urban Other Principal	$50,379	$6,822
Urban Minor Arterial	$50,379	$6,822
Urban Minor Collector	$50,379	$6,822
Urban Local	$44,606	$6,822

Based on the above information, safety saving for each class could be calculated as follows:

$$Yearly\ Safety\ Saving = (365 \times \mathrm{ADT} \times \mathrm{D})/1{,}000{,}000 \times \mathrm{Accident\ Rate} \times \mathrm{CRF} \times \mathrm{Accident\ Costs}$$

Present values of SS in each year will then be calculated in order to perform an economic analysis.

5.2 Build the Probabilistic Model

An important part of building this probabilistic model is to apply traffic volume data in user benefits equations. The following steps illustrate the process of establishing this model:

Step 1: Simulating ten thousand daily traffic data according to the logistic distribution of a certain type of road.

Step 2: Input ten thousand data into TTS, VOCS, and SS equations to get ten thousand results for each equation (in base year).

Step 3: Because traffic volume will increase in each year (2% in this study), a loop in R is needed to calculate the values of TTS, VOCS, and SS for the future 20, 25, and 30 years.

Step 4: Discount the equation values to present values and sum all present values for 20, 25, or 30 years.

Figure 5.1 also demonstrates the process of building this model.

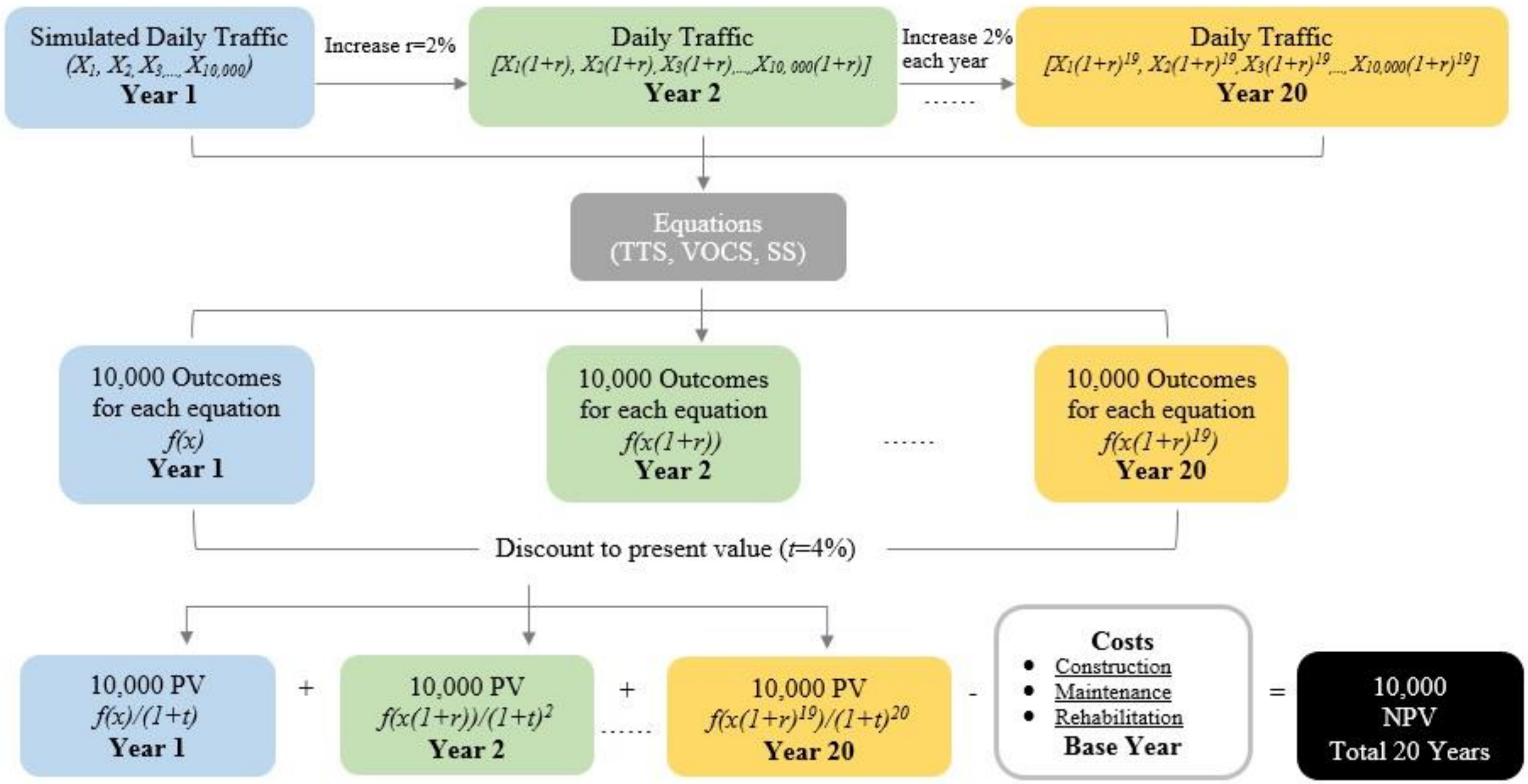

Figure 5.1 Process of Proposed Probabilistic Model

5.3 Model Outcomes Analysis

A case project from Cal-B/C User Guide (2009) is adopted in this study in order to obtain other necessary information, such as facility length, and to check the accuracy of the proposed probabilistic model. Table 5.12 shows the basic information of this project.

The information in this case is used in deterministic economic analysis in Cal-B/C. It used a single value of the average daily traffic data to calculate benefit-cost analysis. The average daily traffic in this case is 234,000. The data simulation is performed based on the traffic volume of this selected segment of road and the logistic distribution discussed in Chapter 4. The value used in this deterministic case is slightly smaller than the mean value of the ten thousand simulated daily traffic values. Figure 5.2 shows the distribution of the simulated daily traffic data.

Table 5.12 *Basic Project Information from Cal-B/C*

Project Information	
Project Type	Adding two lanes
Road way type	Freeway
Traffic lanes	8(non-build) to 10(build)
Analysis period	20 years
Base year	2010
Facility length	3.9 mile
Discount rate	4%
Annual traffic growth rate	2%
Construction costs (Base Year)	$99 million
Maintenance costs (every 5 years)	10% of base year construction costs
Rehabilitation costs	$1,500,000 per lane mile
Free-flow speed (Build=non-build):	65 mph
AADT	234,000
AVO	1.3
VOT	$18

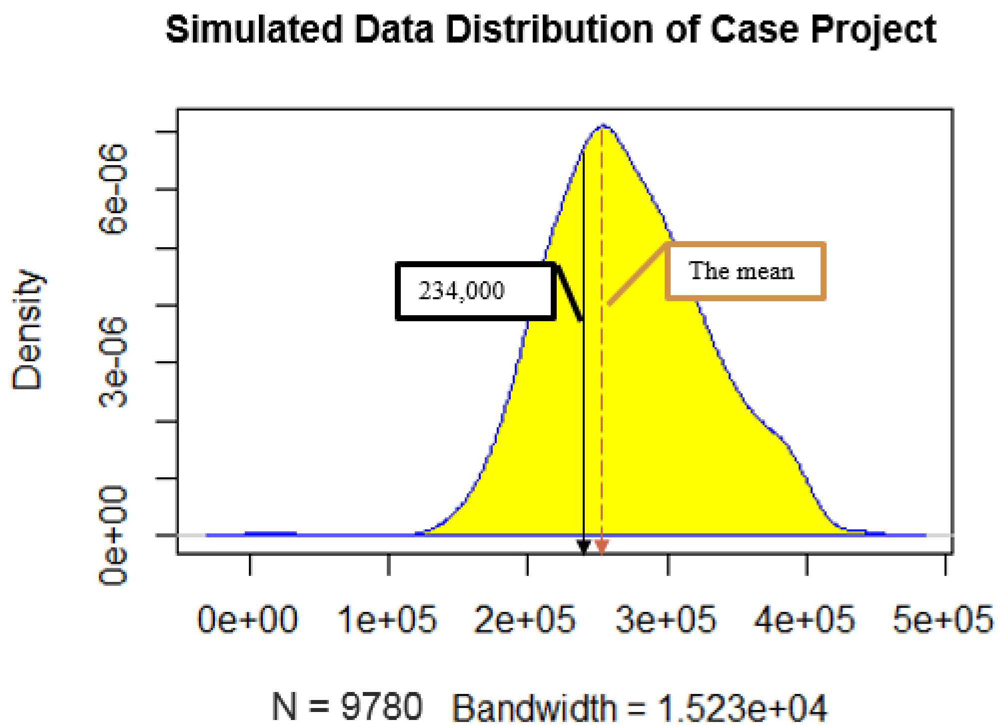

Figure 5.2 Simulated Daily Traffic Distribution of Project in Cal-B/C

Figure 5.2 is the PDF plot of the simulated daily traffic data of the case study. The x-axis is the daily traffic volume and the y-axis is the corresponding density. The mean value in this plot is 250,000 and the ADT value in this case is slightly smaller than the mean. This would result in the deterministic NPV located in a lower percentile of the probabilistic distribution. The distribution covers the single AADT of this project. The value is not largely deviated from the distribution.

The following sections discussed travel time savings, vehicle operation cost savings, and safety savings using the simulated daily traffic data.

5.3.1 Travel Time Savings

Table 5.13 *Travel Time Saving Values in 20 Years (Part)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	3.56E+06	3.93E+06	4.34E+06	4.79E+06	5.29E+06	5.84E+06	6.45E+06	7.12E+06	7.86E+06	8.68E+06	9.59E+06	1.06E+07	1.17E+07	1.29E+07	1.42E+07	1.57E+07	1.74E+07	1.92E+07	2.12E+07	2.34E+07
2	2.47E+07	2.73E+07	3.01E+07	3.33E+07	3.67E+07	4.06E+07	4.48E+07	4.94E+07	5.46E+07	6.03E+07	6.65E+07	7.35E+07	8.11E+07	8.95E+07	9.89E+07	1.09E+08	1.21E+08	1.33E+08	1.47E+08	1.62E+08
3	5.85E+06	6.46E+06	7.13E+06	7.87E+06	8.69E+06	9.60E+06	1.06E+07	1.17E+07	1.29E+07	1.43E+07	1.57E+07	1.74E+07	1.92E+07	2.12E+07	2.34E+07	2.58E+07	2.85E+07	3.15E+07	3.48E+07	3.84E+07
4	6.49E+06	7.17E+06	7.91E+06	8.74E+06	9.65E+06	1.07E+07	1.18E+07	1.30E+07	1.43E+07	1.58E+07	1.75E+07	1.93E+07	2.13E+07	2.35E+07	2.60E+07	2.87E+07	3.17E+07	3.49E+07	3.86E+07	4.26E+07
5	5.65E+06	6.23E+06	6.88E+06	7.60E+06	8.39E+06	9.26E+06	1.02E+07	1.13E+07	1.25E+07	1.38E+07	1.52E+07	1.68E+07	1.85E+07	2.05E+07	2.26E+07	2.49E+07	2.75E+07	3.04E+07	3.36E+07	3.71E+07
6	1.40E+07	1.54E+07	1.71E+07	1.88E+07	2.08E+07	2.29E+07	2.53E+07	2.80E+07	3.09E+07	3.41E+07	3.77E+07	4.16E+07	4.59E+07	5.07E+07	5.59E+07	6.18E+07	6.82E+07	7.53E+07	8.31E+07	9.18E+07
7	6.62E+06	7.31E+06	8.08E+06	8.92E+06	9.84E+06	1.09E+07	1.20E+07	1.32E+07	1.46E+07	1.61E+07	1.78E+07	1.97E+07	2.17E+07	2.40E+07	2.65E+07	2.93E+07	3.23E+07	3.57E+07	3.94E+07	4.35E+07
8	8.51E+06	9.40E+06	1.04E+07	1.15E+07	1.27E+07	1.40E+07	1.54E+07	1.70E+07	1.88E+07	2.08E+07	2.29E+07	2.53E+07	2.79E+07	3.08E+07	3.40E+07	3.76E+07	4.15E+07	4.58E+07	5.06E+07	5.59E+07
......																				
9995	1.87E+07	2.06E+07	2.28E+07	2.51E+07	2.78E+07	3.06E+07	3.38E+07	3.74E+07	4.12E+07	4.55E+07	5.03E+07	5.55E+07	6.13E+07	6.77E+07	7.47E+07	8.25E+07	9.11E+07	1.01E+08	1.11E+08	1.23E+08
9996	4.15E+07	4.59E+07	5.06E+07	5.59E+07	6.17E+07	6.82E+07	7.53E+07	8.31E+07	9.17E+07	1.01E+08	1.12E+08	1.23E+08	1.36E+08	1.50E+08	1.66E+08	1.83E+08	2.03E+08	2.24E+08	2.47E+08	2.73E+08
9997	1.39E+07	1.54E+07	1.70E+07	1.87E+07	2.07E+07	2.28E+07	2.52E+07	2.78E+07	3.07E+07	3.39E+07	3.75E+07	4.14E+07	4.57E+07	5.04E+07	5.57E+07	6.15E+07	6.78E+07	7.49E+07	8.27E+07	9.13E+07
9998	1.48E+07	1.64E+07	1.81E+07	2.00E+07	2.20E+07	2.43E+07	2.69E+07	2.97E+07	3.27E+07	3.61E+07	3.99E+07	4.41E+07	4.86E+07	5.37E+07	5.93E+07	6.55E+07	7.23E+07	7.98E+07	8.81E+07	9.73E+07
9999	2.89E+06	3.19E+06	3.52E+06	3.89E+06	4.30E+06	4.74E+06	5.24E+06	5.78E+06	6.38E+06	7.05E+06	7.78E+06	8.59E+06	9.48E+06	1.05E+07	1.16E+07	1.28E+07	1.41E+07	1.56E+07	1.72E+07	1.90E+07
10000	2.25E+07	2.49E+07	2.75E+07	3.03E+07	3.35E+07	3.70E+07	4.08E+07	4.51E+07	4.98E+07	5.49E+07	6.06E+07	6.70E+07	7.39E+07	8.16E+07	9.01E+07	9.95E+07	1.10E+08	1.21E+08	1.34E+08	1.48E+08

Table 5.13 shows parts of the analysis outcomes of travel time savings in 20 years. Each row has 20 values standing for the TTS of each year in total 20 years. Each column has ten thousand TTS values representing the ten thousand iterations of traffic data simulation. It is hard to recognize useful information from this table, but PDF plots can help have a closer look at the outcomes. Figure 5.3 shows the results of TTS in the base year.

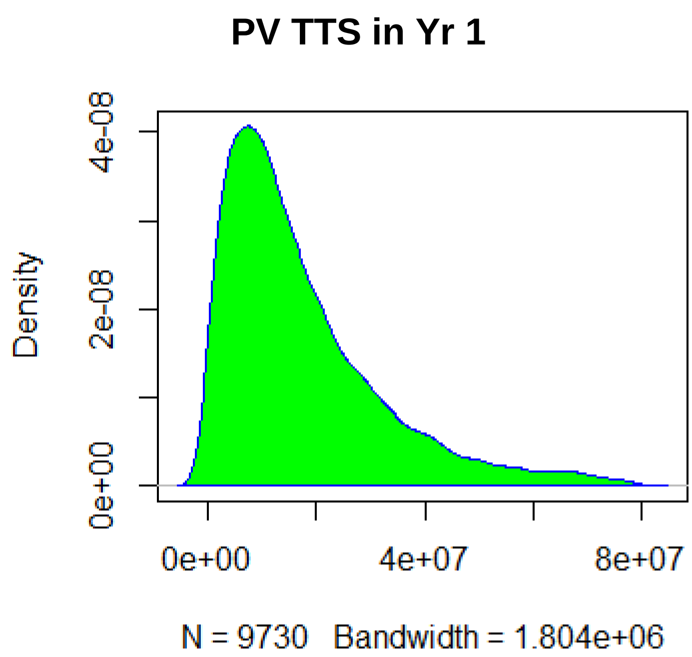

Figure 5.3 Present TTS in Year 1

Figure 5.3 indicates the distribution of travel time savings in the base year has a large right skew. The tail extends slowly on the right side and the majority of data falls on the left area under the curve. This plot indicates, in the base year, it is highly possible to have a TTS in lower percentiles rather than in higher percentiles. The mode is ten million and is located in the thirty-fifth percentile of the data set. Only five percent of the data is larger than 56 million. Sixty percent of data falls between 0 and 22 million.

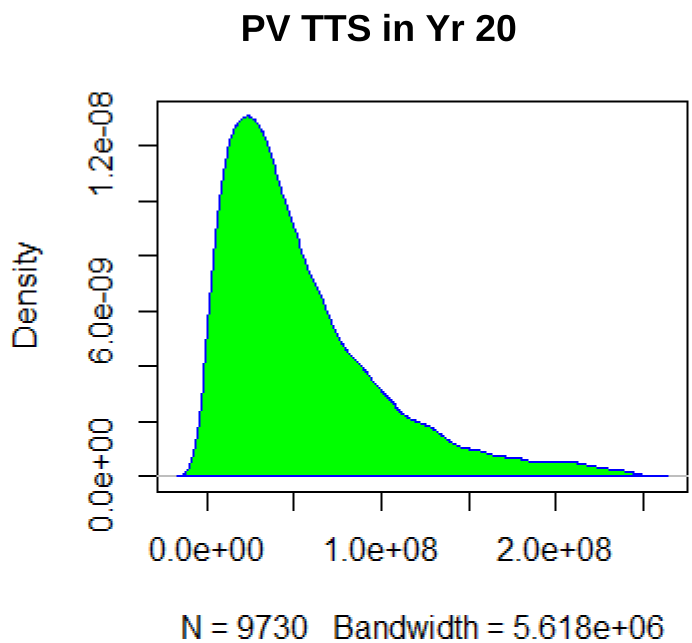

Figure 5.4 Present TTS in Year 20

Figure 5.4 shows TTS results of the twentieth year in terms of present value. The distribution pattern is the same as the base year. It is also right-skewed, increasing very rapidly on its left side and declining much slower after its peak. Only one mode shows in the plot. The mode is 26 million. The area under the curve is bounded by 0 and 1,332 million. About sixty percent of all data falls between 0 million and 67 million. Roughly ninety percent of TTS outcomes fall between 0 million and 121 million. It is unlikely to have a TTS value larger than 181 million in the 20th year because it is in the upper five percentile of the curve with a frequency lower than 0.0024.

Figure 5.5 is the summed present values of TTS in total 20 years.

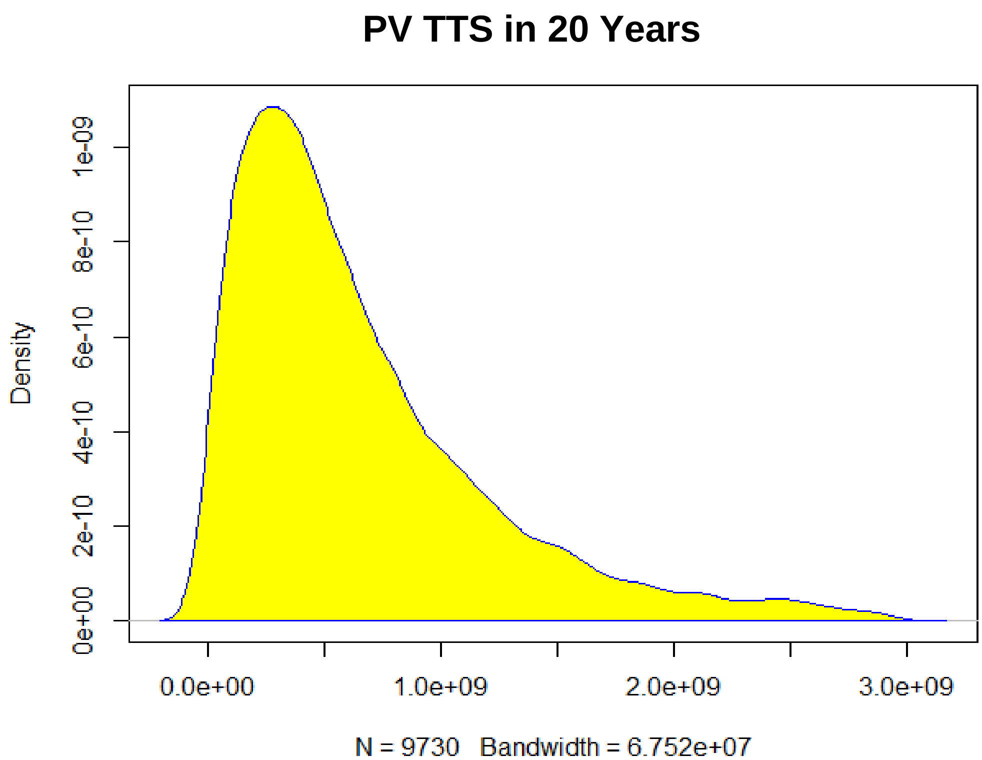

Figure 5.5 Present TTS of 20 Years

The distribution of summed present travel time savings in 20 years is also right-skewed. The values range from 0.99 to 2,210 million (including potential outliers). The mode of TTS in total 20 years is 232 million and is located in the thirty-fifth percentile. Roughly sixty percent of all observations fall between 1 million and 669 million. About ninety percent of all data falls between 1 million and 1560 million. Actual TTS in 20 years barely has a possibility to be larger than 2,000 million, which is within the upper five percentile. Values larger than 2,000 million have frequencies lower than 0.0069.

Cal-B/C using the deterministic method estimated the total TTS in 20 years as 373.2 million. The value is in the forty-first percentile of the distribution and is slightly higher than the mode, though it is likely to happen. This is caused by the ignorance of skewness. The deterministic method used an average daily traffic value to estimate total travel time savings. In a right-skewed distribution, the mean moves to the right side of the mode because of the effect of the long tail. A single value might hide important information. By reading the probabilistic distribution plot, analysts can know the likelihood of a future outcome and its consequences in response to a current action. Analysts could know the outcome is more likely to happen around 223 million rather than 373.2 million, the deterministic result.

5.3.2 Vehicle Operation Costs Savings

Table 5.14 illustrates parts of the analysis outcomes of vehicle operation cost savings in 20 years. Each row has 20 values standing for the VOCS of each year in a total of 20 years. Each column has 10,000 VOCS values representing the 10,000 iterations of traffic data simulation. It is not easy to recognize meaningful information simply by reading this table, but PDF plots can help to understand how these values are distributed. Figure 5.6 shows the results of VOCS in the base year.

Table 5.14 *Vehicle Operation Costs Saving in 20 Years (Part)*

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
1	4491340	4785938	5087158	5393046	5701304	6009278	6313956	6611963	6899584	7172775	7427202	7658284	7861243	8031171	8163106	8252113	8293375	8282283	8214537	8086239
2	7665510	7867450	8036195	8166782	8254279	8293872	8280964	8211265	8080893	7886460	7625163	7294855	6894113	6422288	5879543	5266874	4586114	3839924	3031766	2165861
3	2758825	2979554	3212689	3458034	3715208	3983620	4262444	4550588	4846674	5149011	5455578	5764005	6071565	6375172	6671380	6956406	7226143	7476205	7701963	7898610
4	2468710	2671913	2887536	3115598	3355971	3608354	3872243	4146914	4431386	4724401	5024401	5329500	5637473	5945742	6251366	6551047	6841137	7117659	7376336	7612633
5	4955104	5259217	5566747	5875195	6181703	6483057	6775685	7055685	7318845	7560685	7776505	7961446	8110560	8218890	8281553	8293841	8251306	8149864	7985884	7756278
6	1789387	1946295	2114539	2294511	2486538	2690857	2907604	3136784	3378256	3631701	3896597	4172196	4457494	4751209	5051750	5357205	5665317	5973473	6278702	6577672
7	4526523	4822014	5123908	5430211	5738583	6046327	6350384	6647341	6933439	7204598	7456453	7684394	7883625	8049227	8176236	8259728	8294904	8277188	8202325	8066471
8	2457921	2660445	2875385	3102767	3342471	3594206	3857481	4131583	4415547	4708132	5007794	5312668	5620547	5928872	6234724	6534822	6825539	7102913	7362685	7600331
......																				
9995	4398400	4690512	4989802	5294424	5602193	5910570	6216656	6517195	6808577	7086862	7347805	7586898	7799419	7980496	8125175	8228503	8285614	8291822	8242719	8134266
9996	5846867	6153683	6455654	6749243	7030576	7295468	7539461	7757875	7945861	8098477	8210762	8277824	8294935	8257618	8161751	8003659	7780200	7488850	7127772	6695877
9997	4290603	4579600	4876386	5179239	5486102	5794570	6101882	6404915	6700190	6983888	7251876	7499739	7722830	7916325	8075291	8194765	8269839	8295746	8267964	8182303
9998	6686785	6971106	7239915	7488808	7713147	7908115	8068785	8190191	8267420	8295699	8270490	8187586	8043208	7834092	7557573	7211661	6795101	6307424	5748976	5120940
9999	571336.3	627211.1	688235.7	754825.8	827419.8	906477.4	992478.4	1085920	1187314	1297182	1416053	1544454	1682902	1831898	1991915	2163383	2346678	2542102	2749868	2970077
10000	2867785	3094740	3334024	3585352	3848241	4121985	4405629	4697941	4997388	5302117	5609934	5918290	6224279	6524634	6815737	7093640	7354090	7592575	7804367	7984592

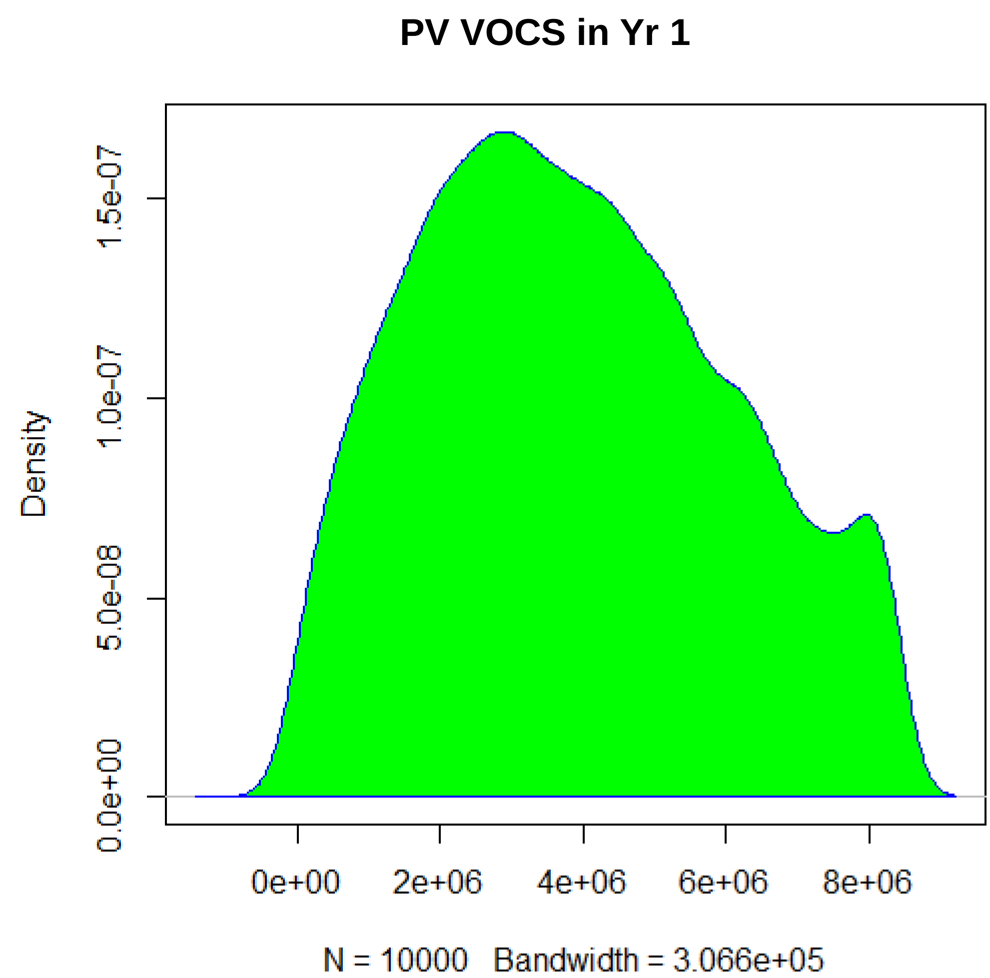

Figure 5.6 Present Values of Vehicle Operation Cost Saving in Year 1

Figure 5.6 demonstrates the vehicle operation cost savings data has at least two peaks in the base year. One is around 2.5 million, while the other is around 8 million. This plot does not show clear skewness. On each side, the data increases and decreases relatively slowly and the tails are not as heavy as those in the TTS plots are. The area under the curve is bounded by -0.4 million (possibly an outlier) and 8.3 million. Ninety percent of data falls between half million and 7.8 million and sixty percent of the data falls between 1.7 million and 6.2 million. The mode of this date set in the base year is 2.6 million. The mean is 3.9 million and the median is 3.7 million. The mean is close to the median, which also proves there is no obvious skewness in this plot.

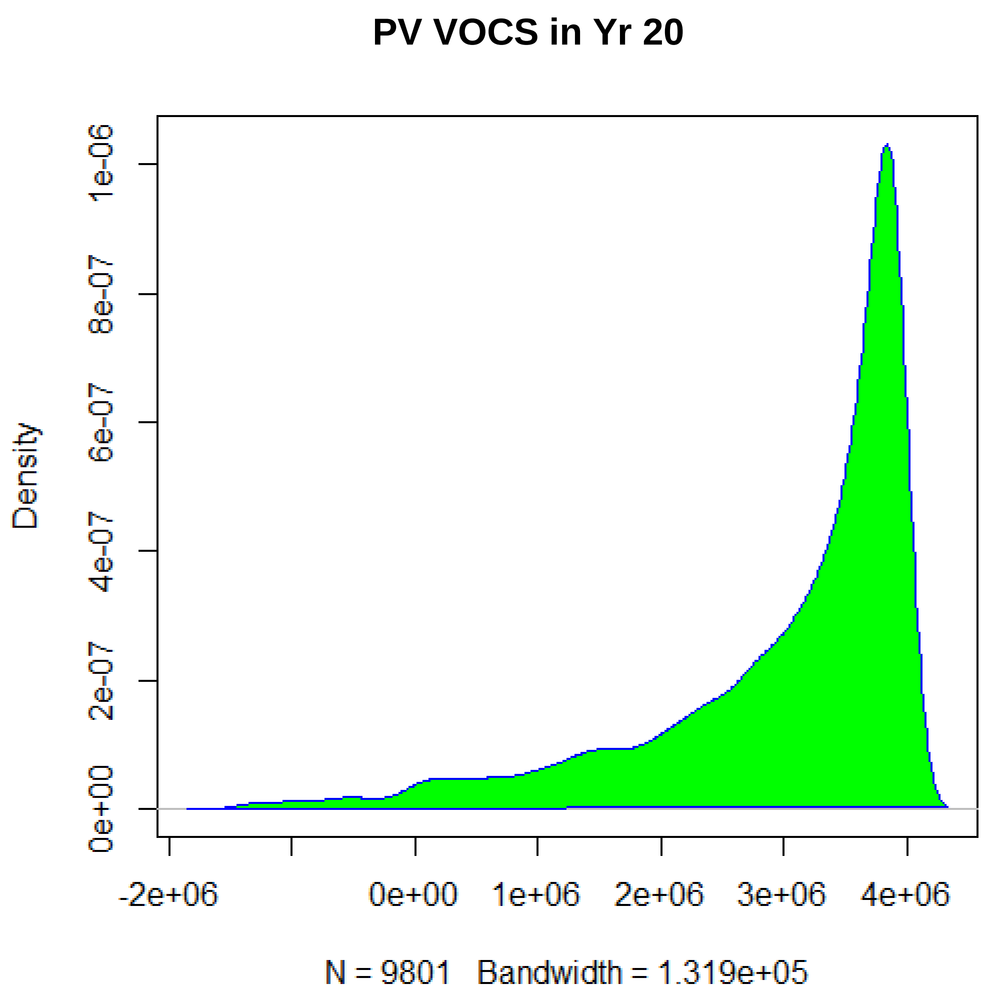

Figure 5.7 Present Values of Vehicle Operation Cost Saving In Year 20

Figure 5.7 shows the results of VOCS (present value) in the 20th year. The distribution pattern is very different from that in the base year. The involvement of Log function may be the cause of this difference. The distribution has a large left skew. The tail extends slowly on the left side and the majority of data falls in the right part under the curve. This plot indicates, in the 20th year, it is highly possible to have a higher VOCS than lower values. The mode is 3.6 million. About ninety percent of the data is larger than 1.1 million. About sixty percent of the data falls between 2.8 million and 3.7 million.

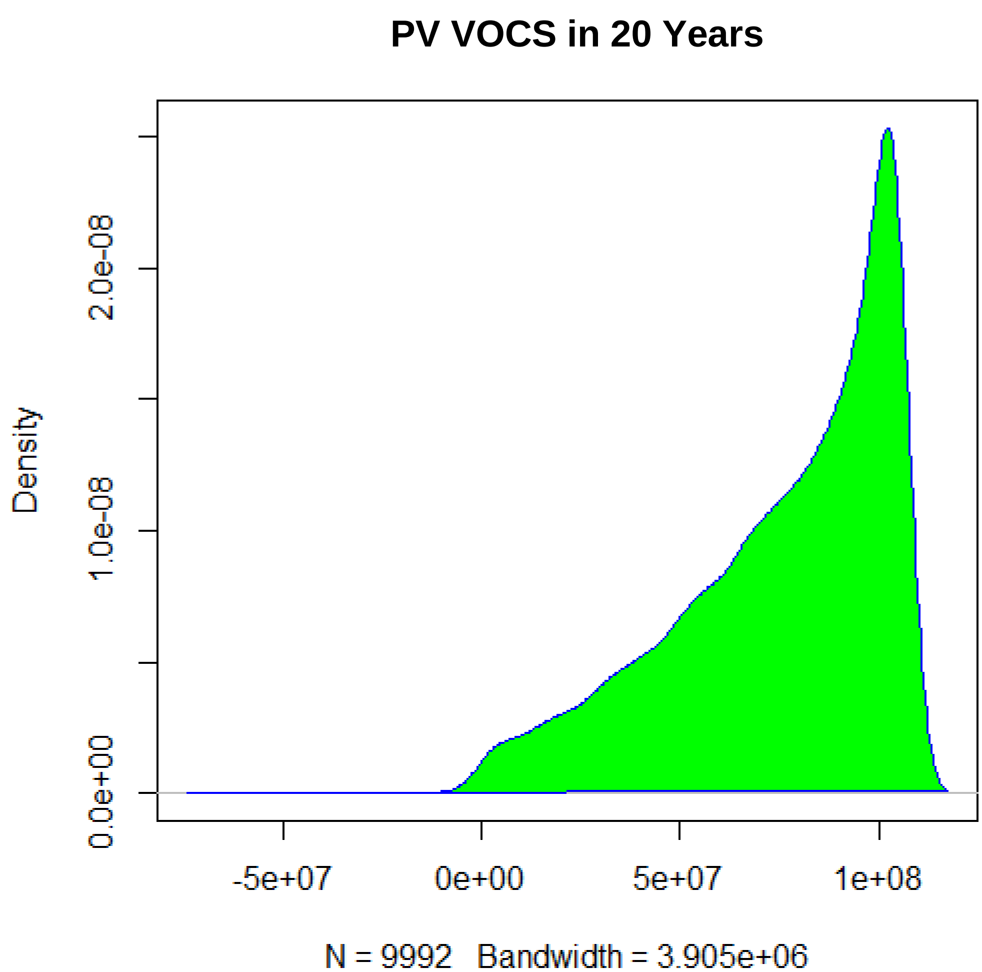

Figure 5.8 Present Values of Vehicle Operation Cost Saving In 20 Years

Figure 5.8 shows the summed present values of VOCS in 20 years. The distribution is also left-skewed. The data in 20 years is bounded by -121 million (possibly an outlier) and 136 million. More than eighty percent of the data is larger than 50 million. The mode of this data set is 102 million and is located in the ninety-nine percentile. About sixty percent of the data falls between 71 million and 121 million. The median of the data is 79 million. The mean is 72 million.

The deterministic value in Cal-B/C is 59 million, which is in about thethirtyth percentile. The deterministic calculation result with a lower percentile is caused by two combined reasons: one is the difference between the deterministic value and the mean, and the other is the ignorance of the skewness of the distribution. After careful examination, the later accounts for the most. A deterministic value is easy to be affected by the long tail on the left side and results in the location of the lower percentiles. In that situation, a decision with the underestimation of the total vehicle operation cost savings would be made. By reading a probabilistic plot, analysts could obtain comprehensive information about project values. In this case, though the result has a relatively high possibility of happening in 59 million, the actual VOCS is most likely to happen around 100 million, which is much higher than 59 million. An underestimation of the analysis outcome may result in the exclusion of this productive alternative.

5.3.3 Safety Savings

Table 5.15 illustrates parts of the analysis outcomes of safety savings in 20 years. Each row has 20 values standing for the SS values of each year in total 20 years. Each column has 10,000 SS values representing the 10,000 iterations of traffic data simulation.

Table 5.15 *Safety Saving in 20 Years (Part)*

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
1	542507	553357	564424	575713	587227	598971	610951	623170	635633	648346	661313	674539	688030	701791	715826	730143	744746	759641	774833	790330
2	674983	688483	702253	716298	730624	745236	760141	775344	790851	806668	822801	839257	856042	873163	890626	908439	926608	945140	964043	983323
3	474189	483673	493347	503214	513278	523543	534014	544694	555588	566700	578034	589595	601387	613414	625683	638196	650960	663980	677259	690804
4	461132	470354	479761	489357	499144	509127	519309	529695	540289	551095	562117	573359	584827	596523	608454	620623	633035	645696	658610	671782
5	559571	570762	582178	593821	605698	617812	630168	642771	655627	668739	682114	695756	709671	723865	738342	753109	768171	783535	799205	815189
6	426776	435312	444018	452898	461956	471196	480619	490232	500036	510037	520238	530643	541256	552081	563122	574385	585872	597590	609542	621733
7	543806	554683	565776	577092	588634	600406	612414	624663	637156	649899	662897	676155	689678	703472	717541	731892	746530	761460	776689	792223
8	460631	469844	479241	488825	498602	508574	518745	529120	539703	550497	561507	572737	584192	595875	607793	619949	632348	644995	657895	671053
......																				
9995	539067	549849	560846	572063	583504	595174	607077	619219	631603	644235	657120	670263	683668	697341	711288	725514	740024	754824	769921	785319
9996	592741	604596	616688	629021	641602	654434	667523	680873	694490	708380	722548	736999	751739	766774	782109	797751	813706	829980	846580	863512
9997	535063	545764	556680	567813	579170	590753	602568	614619	626912	639450	652239	665284	678590	692161	706005	720125	734527	749218	764202	779486
9998	626339	638866	651643	664676	677970	691529	705359	719467	733856	748533	763504	778774	794349	810236	826441	842970	859829	877026	894566	912458
9999	331733	338368	345135	352038	359079	366260	373586	381057	388678	396452	404381	412469	420718	429132	437715	446469	455399	464507	473797	483273
10000	478914	488493	498262	508228	518392	528760	539335	550122	561124	572347	583794	595470	607379	619527	631917	644556	657447	670596	684008	697688

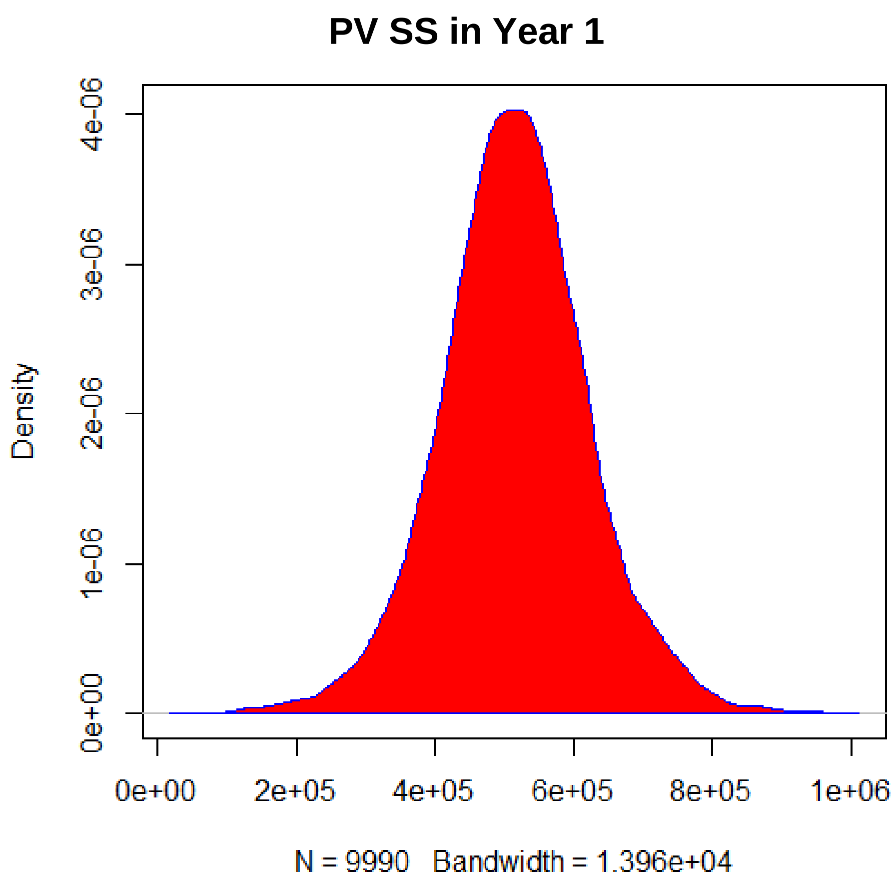

Figure 5.9 Present Values of Safety Saving in Year 1

Figure 5.9 shows the results of SS in the base year. The pattern of safety savings follows a logistic distribution. It is symmetric and has a peaked top. The area under the curve is bounded by -53985 and 1.03 million. The mode is 0.55 million in the base year. About ninety percent of the data falls between 0.35 million and 0.69 million. Sixty percent of the data falls between 0.42 million and 0.6 million. In this plot, the mean is equal to the median and is 0.51 million.

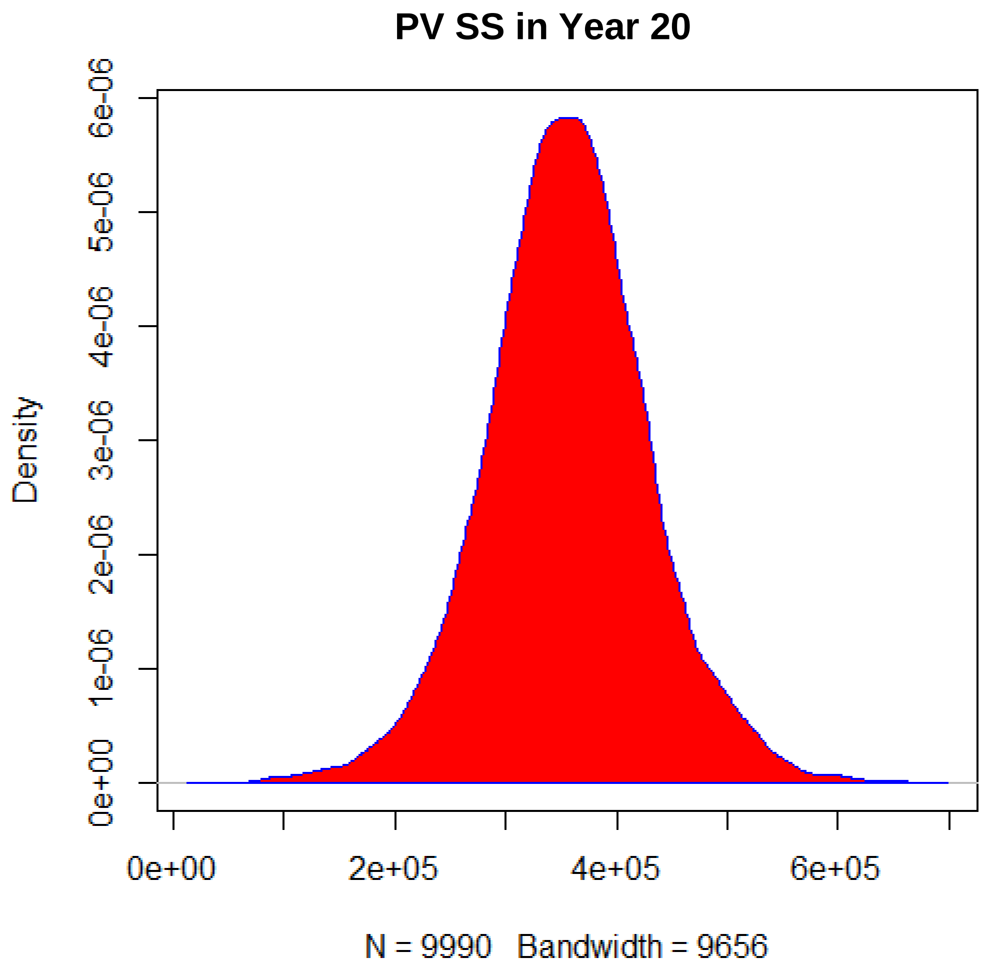

Figure 5.10 Present Values of Safety Saving in Year 20

Figure 5.10 shows the results of safety savings in the 20th year. The distribution pattern in the 20th year is the same as in the base year. The mode is 0.36 million. The distribution is bounded by -35,893 and 0.69 million. The mean value is equal to the median and is 0.34 million. Ninety percent of the data falls between 0.23 million and 0.47 million. Sixty percent of the data falls between 0.285 million and 0.41 million.

Figure 5.11 is the summed present benefits of SS in a total of 20 years.

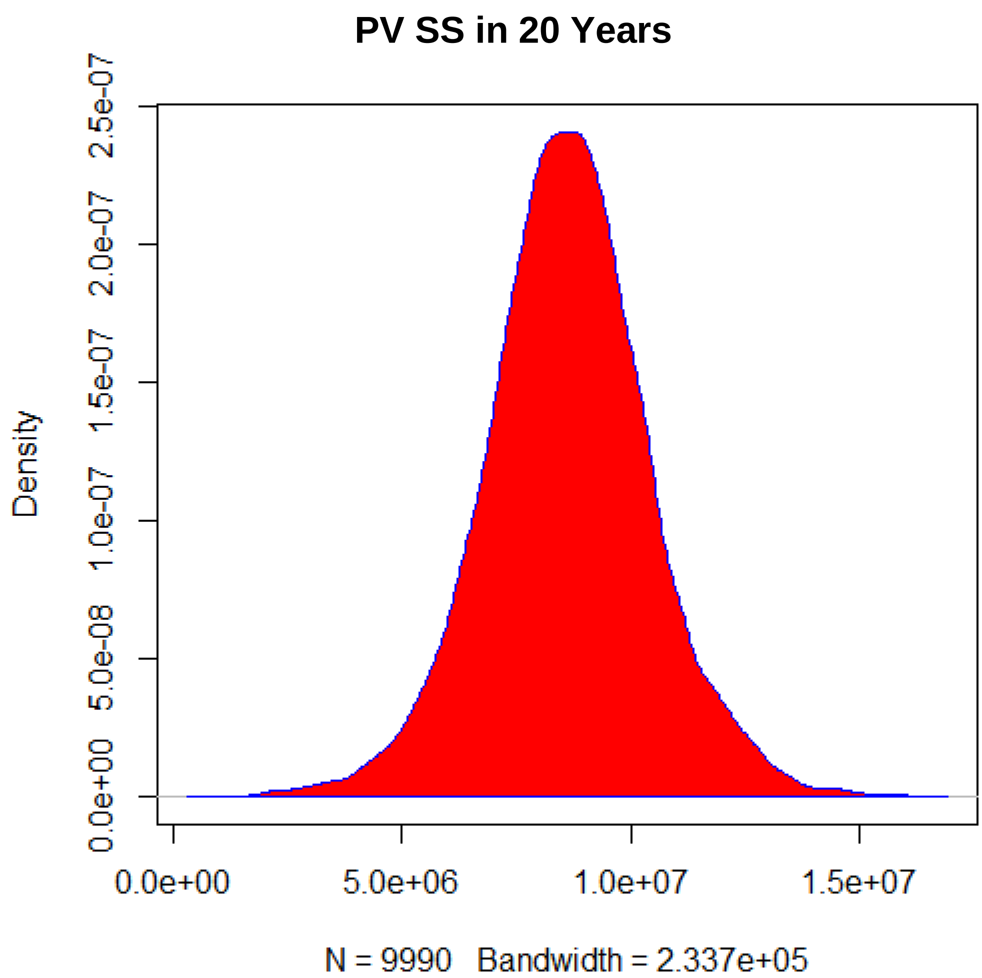

Figure 5.11 Present Values of Safety Saving in 20 Years

The distribution of safety savings in total 20 years roughly follows logistic distribution. The mode is 8.8 million and is located in the sixty fourth percentile. The mean and median is 8.3 million. The difference between mean, median, and mode indicates slight skewness of the data. About ninety percent of the data falls between 5.5 million and 11.2 million. About sixty percent of the data falls between 6.9 million and 9.6 million.

The value calculated by Cal-B/C is 10.8 million, which is within the upper eight percent of the distribution. The results are not comparable because Cal-B/C does not state the cost of accidents. This study used the Crash Costs in Indiana (Jiang et al., 2013; Gkritza, Labi, & Sinha, 2006; Tarko & Kanodia, 2003) to calculate safety savings. The variance in cost selection is the major reason for this difference. Even though the outcomes cannot be compared, the distribution plot can still reveal information concealed by a single result. By reading the plot, analysts can eliminate the uncertainty coming from an estimated input. They can also evaluate the likelihood of the occurrence of an outcome. It could save the time it takes to debate over the uncertainty underneath the traditional deterministic method (FHWA, 1998).

5.3.4 Net Present Value (NPV)

Net present value is defined as the discounting mathematical summed value of future benefits and costs in each year (Wall & Smith, 1998). This study would like to show the impacts of user benefits on project NPV. Agency costs were set as deterministic as a way to control variable. Construction costs are 99 million and occur in the base year. Maintenance costs are ten percent of base year costs occurring in every five years. Rehabilitation occur in every ten years. The costs are $1,500,000 per lane mile. User benefits of this project are computed separately in previous sections. R is using to add these benefits and construction costs together. Figure 5.12 shows the interface of using R to simulate and calculate the project NPV.

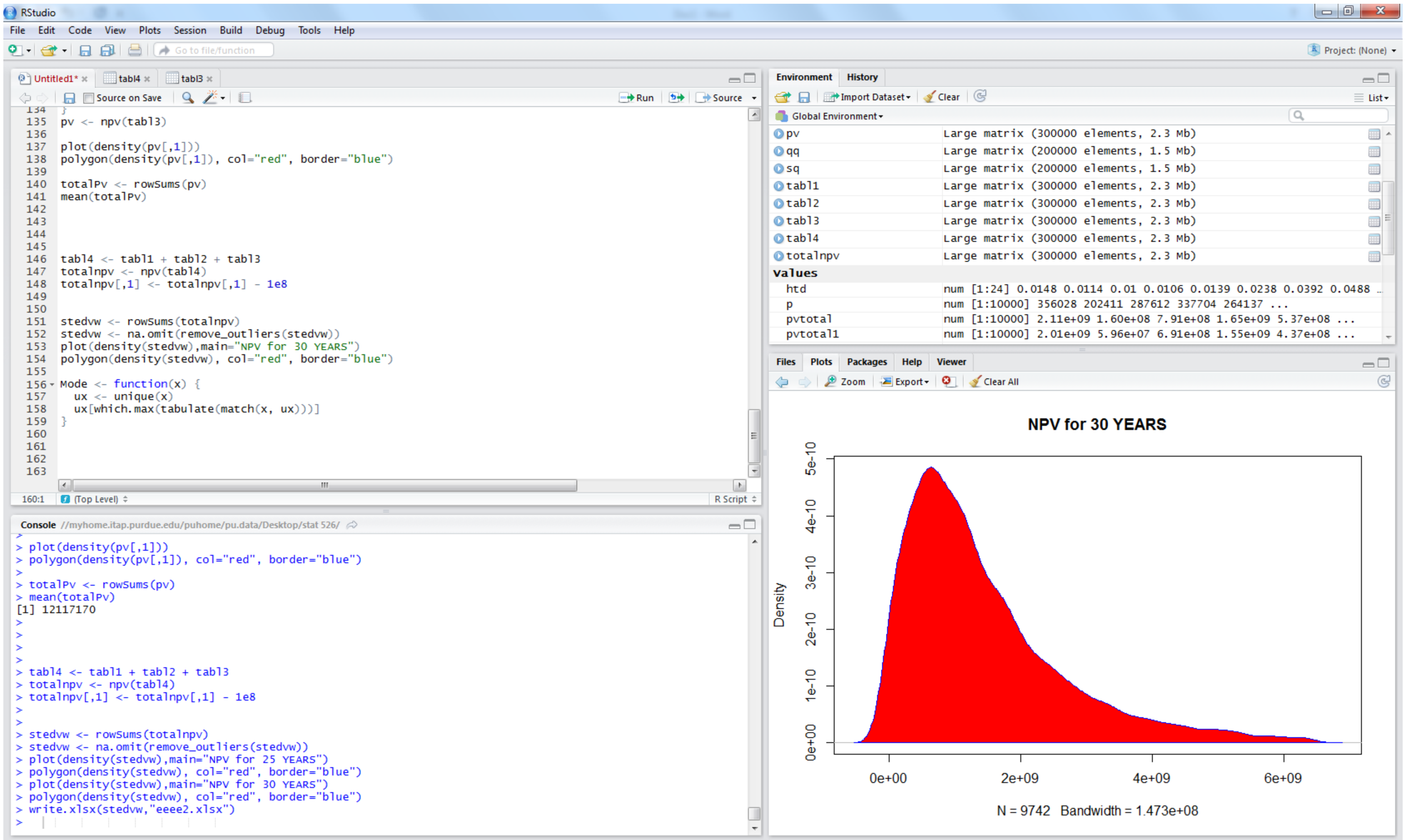

Figure 5.12 Interface of Rstudio

The NPV values in 20 years is partially listed in Table 5.16.

Table 5.16 *NPV Values in 20 Years (Part)*

2.01E+09	9033212.3	337727990.3	155354095.6	869126988.1	278450011.6
59642937	1.851E+09	1338893621	1323339305	151579824.8	1309464665
6.91E+08	401304241	1200274192	532890517.8	242012434	2249543716
1.55E+09	-12866602	1647289584	467293447.1	480463649.9	971544731
4.37E+08	221209051	1294462777	167525208.3	423523485.5	742696499.6
9.46E+08	350210610	787675115.3	553530401.9	695888041.4	248224714.3
2.04E+08	956594232	926954109.7	-4220816.56	867160933.6	522800412
-9.6E+07	464403109	197203635.6	428932711.3	297033055.4	660322748
1.14E+08	840814809	1359349269	1297124399	409856018.1	2266418104
5.96E+08	179014473	574022509.4	173112999.2	275168681.7	110227109.1
6.07E+08	541953792	395425346	147396903.1	279190913.1	69950540.31
2.82E+08	1.19E+09	398612717.7	964186753.6	954424448.3	278938081.8
8.81E+08	344177554	98331194.84	1044063921	219568734.2	-15247785.7
8.83E+08	128374559	-24636701.7	-23764705.3	2329920384	1191456753
4.94E+08	301347804	580791113.4	349449136.4	1071084169	1451363647
3.26E+08	268164028	789436554.7	776471928.2	276408785.7	285964003.3
1.02E+09	430775285	596920951.6	569840673.1	622404060.1	393966877.3
1.89E+09	757137009	-97206771.2	1098084340	2475891469	52040725.36
2.84E+08	517091734	18297344.88	1101764291	197926508	87199261.87
1.04E+09	1.486E+09	408366270.8	663728100.7	242052535.5	2468603.106
1.41E+09	396068534	1213402867	651754771.3	262650146.9	194873998.8
4.04E+08	473776621	276780282.5	360846241.9	473649551.3	808917062.5
1.12E+08	241980599	1327448795	17876394.5	534639056.4	1142298659
5.51E+08	-53222930	555291526.1	592694199.1	477333722.5	1562021695
8.4E+08	878425064	139865086.2	5726708.544	698133798.5	1074197725
4.56E+08	889308031	562250815.2	1116346596	794871121.1	-70248350.5
1.27E+08	145265931	121375322.2	2156094049	236482448.6	498983730.2
2.85E+09	79585080	210093041.1	-83228928.8	580445292.1	233649665.2
1.08E+09	181206399	514445007.1	161714098.1	57415166.58	897012014.7
1.24E+09	24348488	1175794527	364108602.3	1053983379	137461924.9
3.09E+08	740941048	2248773765	565791366.9	260124135.5	383684490.9
1.26E+09	1.662E+09	730563468.7	2076059542	463736803.6	282714864.3
1.08E+09	727408698	-68903786.1	460132000.2	479855259.3	193455012.8
1.53E+09	304591850	886865727.4	3491184.781	217094210.4	315940165.5
6.13E+08	357531469	418745636.2	122061148.4	2188763167	477160664.9
2.18E+08	261500162	188048206.7	237484480	516077416	374425192.1
5.81E+08	831052158	316253901.6	1028609195	1411809451	1495309519
7.64E+08	427680894	1083064315	340199233.6	247180196.6	379427920.5

Compared with reading raw data, analysts may have more interests in how this data is distributed. Figure 5.13 shows the NPV PDF plot of this case project in 20 years.

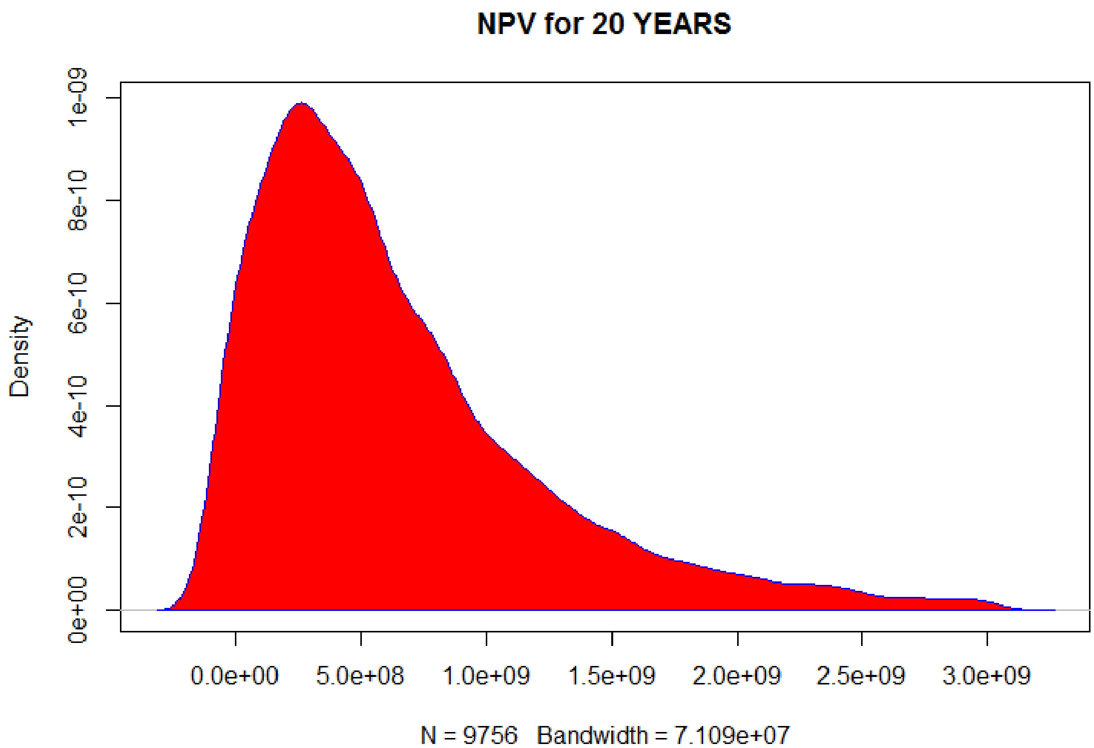

Figure 5.13 NPV for 20 Years

In regular deterministic analysis, analysts calculate an NPV value and assume it distributes normally. However, the probabilistic analysis revealed the risk of making judgments by simply using a single-value result.

The plot in Figure 5.13 is largely right-skewed. Before deleting outliers, the tail extends even longer. The curve increases rapidly on the left side and reaches a peak. The peak in this data set is 253 million and is located in the twenty-eighth percentile. The majority of data distributes on the left area under this curve. About ninety percent of the data falls between 15 million and 1,860 million. Sixty percent of the data falls between -3 million and 704 million. In Cal-B/C, the NPV of this overall project is 355 million, which is located in the thirty-eighth percentile. The result is about a ten percent deviation from the mode. Detailed data examination revealed the NPV actually has a higher probability of falling around the first quantile (25%) rather than other percentiles of the population.

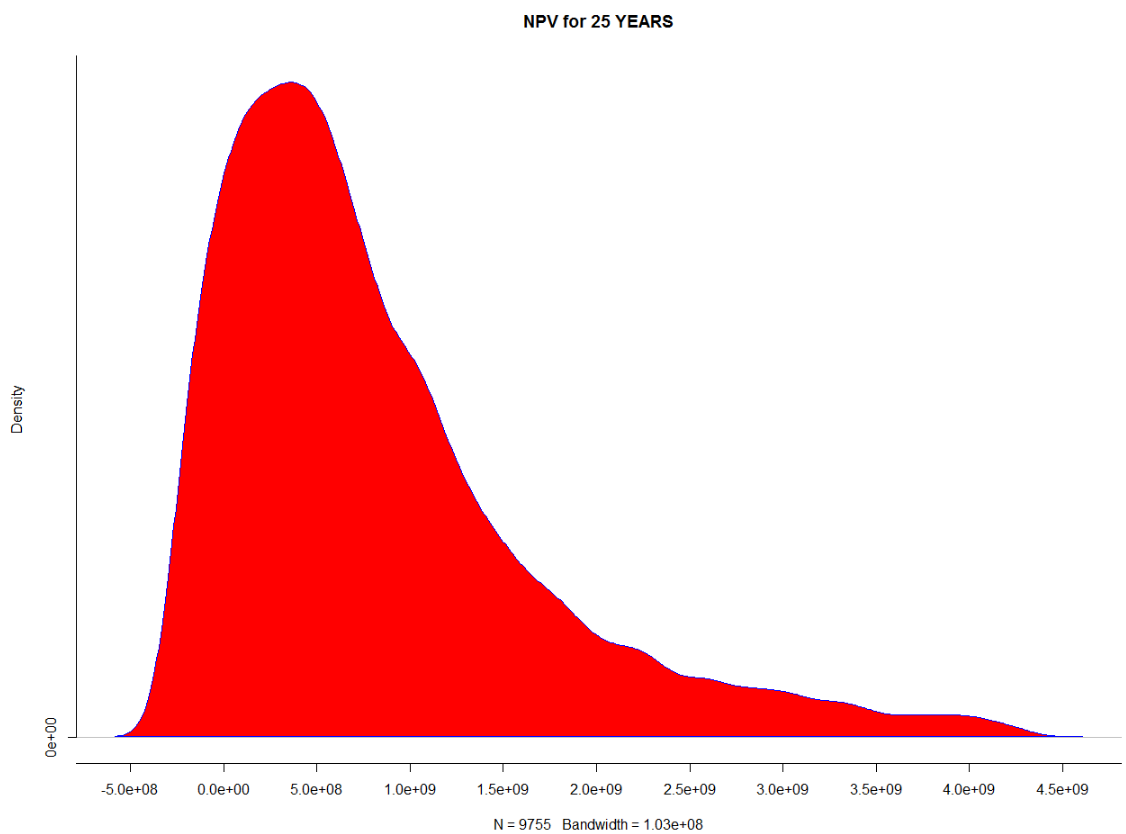

Figure 5.14 NPV for 25 Years

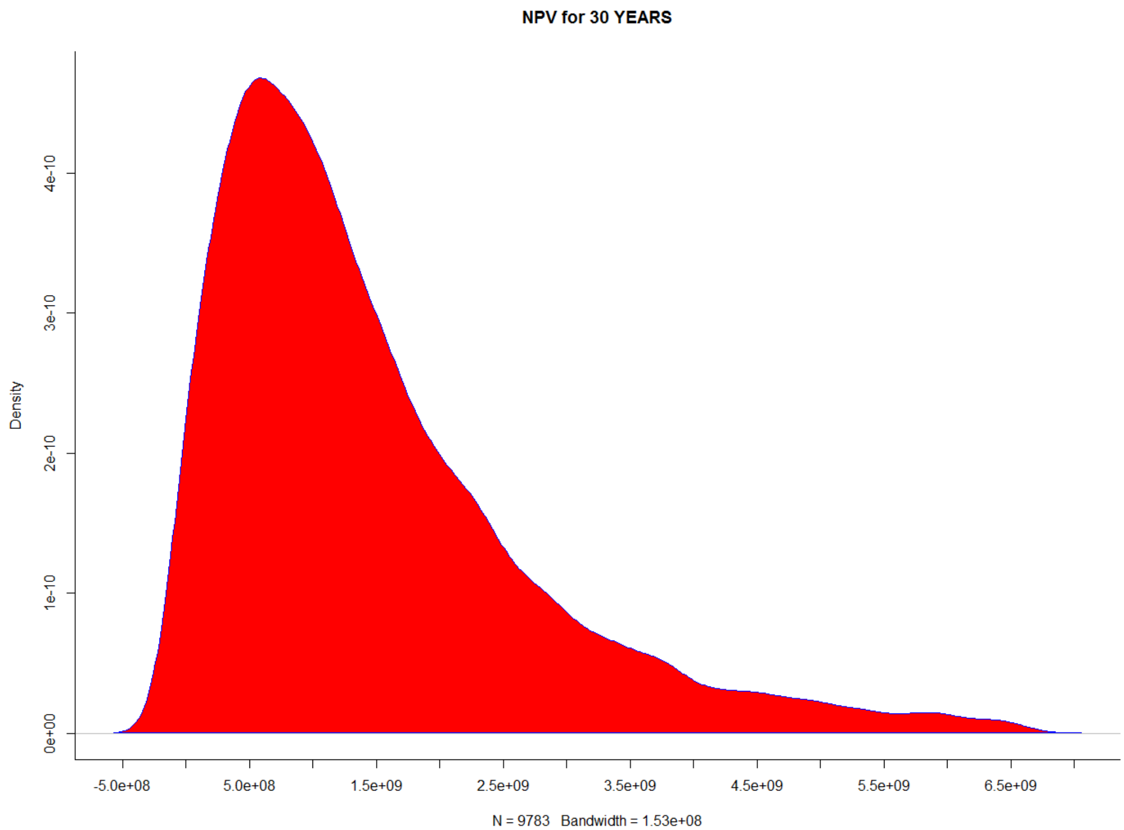

Figure 5.15 NPV for 30 Years

Figure 5.14 and Figure 5.15 are the NPV data distributions of a total of 25 and 30 years. The distributions have similar patterns as that of 20 years. The mode, median, and mean all are increased, as expected. With the extension of the analysis period, this project generates more benefits. In the analysis period of 25 years, the mode is 418 million and is located in the twenty-eighth percentile. About sixty percent of all data falls between 41 million and 1,080 million. Ninety percent of the data falls between -5 million and 2,540 million. In the analysis period of 30 years, the mode is increased to 539 million and is located in the twenty-eighth percentile. About sixty percent of the data falls between 33 million and 1,500 million. Ninety percent of the data falls between 33 million and 3,570 million.

5.4 Chapter summary

This chapter discussed two parts of the probabilistic model. One part is the model building process. The other part is the outcome's analysis by using a case project from Cal-B/C. The analysis in this chapter revealed that the distribution of user benefits followed different patterns according to different calculation algorithms. Another major finding is the distribution of NPV is not generally presumed as normal. The distribution has a large right skew and an obvious peak in all three analysis periods. The majority of the data falls in the left area under the curve rather than being distributed symmetrically. The likelihood of a project NPV higher than 2,000 million in a 20-year analysis period is very low. These findings can help analysts obtain comprehensive information about the project value. Analysts can avoid overestimation of the overall project value and weigh the risks based on their acceptance levels.

CHAPTER 6. CONCLUSIONS AND RECOMMENDATIONS

The primary purpose of this study was to build a probabilistic model for highway construction projects. Presently, most highway construction economic analyses are conducted using deterministic methods, which adopt a single value, such as the mean, to analyze the benefits and costs of projects. However, using an estimated value to analyze a project might cause decision makers to lose sight of the risk. A single value makes too little sense when analysts want to know the risk under the value, the likelihood of its occurrence and the consequences of its occurrence. Lack of information could possibly lead to incorrect judgments or debates over the uncertainty hidden by this value. This study proposed a probabilistic model to solve this problem.

This research first examined the probabilistic nature of daily traffic volume of Interstates in Indiana. Three Interstates (I-64, I-70, and I-80/90) were chosen to present the analysis results. The analysis was using R and its package called "fitdistrplus." The Cullen and Frey Graph and AIC were included in this package to perform quantile fitting, moment fitting, and goodness-of-fit. PDF plots were generated to provide a visual examination of traffic data. The Cullen and Frey Graph results showed the traffic data model had almost zero skewness and a kurtosis of about 4.2. AIC results demonstrate that the logistic distribution model has the least information loss when trying to represent the traffic data model compared with other models such as normal, lognormal, exponential, and gamma models, Logistic distribution has some distinct characteristics. It allows zero skewness, has a peaked top, and has an excess kurtosis of 1.2. PDF plots of daily traffic data showed a clear peak and heavier tails on both sides. Therefore, all three methods indicated that the distribution of daily traffic data followed logistic distribution.

Data simulation was then conducted after the determination of data distribution. The PDF was integrated into a cumulative distribution function (CDF). Uniformly distributed random numbers from 0 to 1 were generated. One iteration of a sample value was applied to one generated number into the CDF. This study conducted 10,000 iterations of data simulation. The simulated data illustrated an excess kurtosis without obvious skewness. The mean and deviation values of simulated data were close to sample values. The data plots showed a clear peak and heavy tails on both sides.

Analysis algorithms were built based on the classification of benefits and costs. In this study, costs were divided into three categories: construction costs, maintenance costs, and rehabilitation costs. Costs were considered to be deterministic values in this study and the information was presented in Table 5.1 and Table 5.2. Benefits in this study were also divided into three categories: travel time savings, vehicle operation cost savings, and safety savings. Traffic volume is the major factor affecting all three benefits. Equations were later established to quantify the relationship between traffic data and all three categories of user benefits. Finalized equations were discussed in Section 5.1.2.

The final step of this probabilistic model was to apply simulated data into benefit and cost equations to see how the outcomes were distributed and how the project NPV was distributed. A case project from the Cal-B/C User Guide was selected to run the analysis to obtain necessary project information and check the accuracy of the proposed probabilistic model. This case project was analyzed using deterministic methods in Cal-B/C so the single-value analysis result should be covered in the results of this probabilistic analysis model.

The plot of travel time savings showed a large right skew in each year and also in a total of 20 years. The majority of the data probably fell in the left area under its curve. A high value of travel time saving, such as 2,000 million, was not reasonable to occur. Cal-B/C calculated a TTS in 20 years as 373.2 million, which was located in the forty-first percentile. The skewness might be the cause of this slightly high value. The distribution patterns of vehicle operation costs were different before and after the 10th year. Distribution patterns in each year from the 1st year to the 10th year was multimodal. The plot had fat tails on both sides. The data was distributed much more evenly compared with the data distribution after the 10th year. After the 10th year, the plot was largely left-skewed. The change may be caused by the involvement of the Log function in equations. The plot of VOCS in a total of 20 years was also left-skewed. The left tail extends shorter than that in the 10th to the 20th year. The curve had heavy tails on both sides. Within its boundary, it was reasonable to expect that a higher VOCS have a higher chance of happening. Cal-B/C calculated VOCS in 20 years as 59 million, which was located in the thirtieth percentile of the data set. The calculation result with a lower percentile was caused by two combined reasons: one was the difference between the deterministic daily traffic value and the mean, while the other was the ignorance the skewness of the distribution. The distribution of safety savings was roughly logistic. The mode is located in the sixty-fourth percentile and each tail decreases symmetrically.

After combining all benefits and costs together, a set of NPV values was computed. In regular deterministic analysis, a single NPV value was calculated and presumed as normal distribution. This presumption means the probabilities of the occurrence of a smaller value or a higher value around this single NPV are supposed to be the same. However, the probabilistic analysis in this study revealed the risk of making judgments by simply using a single-value result. The distribution of project NPV was not normally distributed. It even showed a large right skew after deleting possible outliers. In the NPV distribution of a 20-year analysis period, the mode is located in the twenty-eighth percentile. The interval of this plot is 500 million. Over sixty percent of the data fell within one interval deviated from the mode. The actual NPV had a higher probability to fall around the first quantile (25%), rather than around the mean. The distribution plot of the 25-year and 30-year analysis period showed similar patterns to that of the 20-year analysis period but overall had higher values. The mode moves to 418 million in a 25-year analysis period and 539 million in a 30-year analysis period. Both values are located in the twenty-eighth percentile. Therefore, despite the increase of the analysis period, the probability of the NPV occurrence stayed the same.

The increase of NPV mainly came from user benefits in extended service periods. This study mainly focused on the impacts of user benefits on project NPV. As a way to control variables, costs are set as deterministic. The probability of costs in construction, maintenances, and rehabilitations were not considered in this study. A more comprehensive study could be taken in the future to add the uncertainty of costs into this model.

In summary, this study has proposed a probabilistic model to analyze the risk of highway construction projects from the perspective of traffic volume. This model addressed issues such as the possible occurrences of a short-term action to a long-term decision and the likelihood of these occurrences. This model could be applied to different states and different roads if the hourly traffic distribution data in that state and the daily traffic data of the specific road could be obtained.

REFERENCES

LIST OF REFERENCES

American Association of State Highway and Transportation Officials. (2003). *User benefit analysis for highways*. Washington, DC: American Association of State Highway and Transportation Officials.

Bank of China (BOA). (2014). Updated Loan Prime Rate (LPR) (Y. He, Trans.). Retrieved from: http://www.boc.cn/finadata/lilv/fd32/201310/t20131031_2591219.html

Burns, P. (2006). *R relative to statistical packages: comment on technical report (version1.0) strategically using general purpose statistics packages: A Look at Stata, SAS and SPSS.* [PDF document]. Retrieved from http://www.burns-stat.com/pages/Tutor/R_relative_statpack.pdf

California Department of Transportation. (1999). California Life-Cycle Benefit/Cost Analysis Model (Cal-B/C). Retrieved from http://www.dot.ca.gov/hq/tpp/offices/eab/benefit_files/tech_supp.pdf

California Department of Transportation. (2009). *California life-cycle benefit/cost analysis model (Cal-B/C) technical supplement to user's guide, volume 3: traffic operations consistency, network and corridor analysis, new capabilities, and economic and parameter value updates.* System Metrics Group, Inc. and Cambridge Systematics, Inc.

Cantillo, A. F. (2010). *Shackle: time and uncertainty in economics* [PDF document]. Retrieved from http://www.peri.umass.edu/fileadmin/pdf/conference_papers/newschool/Papaer_GLS_Shackle_Studentship.pdf

Federal Highway Administration (FHWA). (1998). *Life-cycle cost analysis in pavement design -In search of better investment decisions*. Washington, DC: U.S. Department of Transportation.

Federal Highway Administration (FHWA). (2003). *Economic analysis primer.* Washington, DC: U.S. Department of Transportation.

Federal Highway Administration. (2005). *Highway economic requirements system -state version technical report.* Washington, DC: U.S. Department of Transportation.Feldman, R. & Downey, M. L. (1996). In *Transportation expressions.* Washington, DC: U.S. DOT/ Bureau of Transportation Statistics.

Gkritza, K., S. Labi, & K. C. Sinha. *Economic Development Effects of INDOT Transportation Projects* (FHWA/IN/JTRP-2006/37). West Lafayette: Joint Highway Research Program.Gulen, S., Zhu, K., Weaver, J., Shan, J., and Flora, W. F. (2001). *Development of improved pavement performance prediction models for the Indiana pavement management system* (FHWA/IN/JTRP-2001/17). West Lafayette, IN: Joint Transportation Research Program. doi: 10.5703/1288284313192.

Hayter. A. (2012). *Probability and statistics for engineers and scientists*. Boston, MA: Cengage Learning.

Hogg, R. V., McKean, J., & Craig, A. T. (2012). *Introduction to mathematical statistics* (7th ed.). Boston, MA: Pearson Education, Inc.

Jiang, Y., Zhao, G., & Li, S. (2013). *An economic analysis methodology for project evaluation and programming*. (78 msp). Final Report, SPR-3429. Indiana Department of Transportation and Purdue University, Joint Transportation Research Program.

Jiang, Y., Zhao, G., Li, S., Guan, B., & Yang, Y. (2013). A model for life-cycle benefit and cost analysis of highway projects. *International Journal of Pavement Research and Technology*, *6*(5), 633-642.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). *Applied linear statistical models*. New York, NY: McGraw-Hill Irwin.

Lawson, T. (1985). Uncertainty and economic analysis. *Economic Journal, 95*, 909-927

Muller, M. L. D., Dutang, C. Pouillot, R. & Denis, J.B. (2015) *Help to Fit of a Parametric Distribution to Non-Censored or Censored Data.* [PDF document]. Retrieved from http://cran.r-project.org/web/packages/fitdistrplus/fitdistrplus.pdf

Office of Asset Management. (2004). *Life-cycle cost analysis RealCost user manual.* Washington, DC: U.S. Department of Transportation.

Office of the Secretary of Transportation. (2011). *The value of travel time savings: departmental guidance for conducting economic evaluations* (Rev. 2). Washington, DC: U.S. Department of Transportation.

Phillips, E. & Vancleave, R. (2014). Traffic and safety. In The Indiana Design Manual. Retrieved from http://www.in.gov/indot/design_manual/files/Part_5_2013.pdf

Rice, J. A. (1995). *Mathematical statistics and data analysis* (2nd ed.). Pacific Grove, CA: Duxbury Press.

Santos, A., McGuckin, N., Nakamoto, H.Y., Gray, D., & Liss, S. (2011). *Summary of Travel Trends: 2009 National Household Travel Survey* (FHWA-PL-ll-022). Washington, DC: U.S. Department of Transportation.

Shackle, G. L. S. (1949). *Expectation in economics* (2nd ed.). Cambridge, U.K.: Cambridge University Press,

Tarko, A. & M. Kanodia. *Hazard Elimination Program: Volume 1—Research Report, Volume 2—Guidelines for Highway Safety Improvements in Indiana* (FHWA/IN/JTRP-2003/19). West Lafayette, IN: Joint Transportation Research Program.

Tighe, S. (1999). *An integrated model to assess asphalt cement quality on low temperature performance and life cycle cost* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses A&I. (304672172).

Tighe, S. (2001). Guidelines for probabilistic pavement life cycle cost analysis. *Transportation Research Record*, *1769*, 28-38.

Vohra, N. D. (2006). *Quantitative techniques in management*. New Delhi, India: Tata McGraw-Hill Education.

Washington State Department of Transportation. (2004). *Highway Construction Costs.* [PDF document]. Retrieved from http://americandreamcoalition.org/highways/HighwayCosts.pdf

Zhao, G. (2012). An economic analysis methodology for evaluating highway improvement projects (Master Thesis). Retrieved from ProQuest Dissertations & Theses. (UMI 1530493).