# Modelling Risk in Highway Infrastructure Investments: Decision-Theoretic, Bayesian, and Factor Analysis Approaches

By

**Baraa Adil Alfasi**

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements to the degree of

**Doctor of Philosophy**

**in Civil Engineering**

Department of Civil and Environmental Engineering
Carleton University
Ottawa, Ontario

# Abstract

Investments in highway infrastructure are expected to achieve a satisfactory rating for avoiding cost overruns and in fulfilling their role in enhancing the sustainability of cities and regions. Despite challenges, highway infrastructure investment assessments received insufficient research attention in modelling lifecycle cost risk and the associated sustainability effectiveness. The increasing acceptance of the sustainability rating tools for evaluating highway projects implies going beyond former attention to location and design of highway infrastructure guided mainly by functional considerations. Under present high demands from the transportation planning environment, there is emphasis on characterizing risk in the first cost as well as lifecycle costs. Likewise, there is an emphasis on inclusion of sustainability factors, including effective use of resources, in evaluating project alternatives.

The use of infrastructure rating tools such as ENVISION has the potential to improve the effectiveness of investment in highway projects in terms of meeting sustainability criteria, including formal recognition of the importance of lifecycle analysis. To support the application of rating tools, research is needed in modelling risk in lifecycle cost estimates, including the identification and quantification of cost overrun factors. Besides, there is a need for a methodology for joint treatment of multi-attribute criteria that encompass cost and other factors of sustainability.

To go beyond the current state of knowledge, research was carried out on:

- The probability models of cost overruns.
- Treating risk and uncertainty in lifecycle analyses, using decision-theoretic, utility-theoretic, and Bayesian methods for evaluation of investment alternatives.

- Factor Analysis of variables that characterize the causes of cost overruns and logistics regression models based on factor analysis results.

Data were obtained and analyzed on actual projects that may have experienced cost overruns. Also, a questionnaire study was implemented to obtain data from transportation jurisdictions in Canada, the USA, Middle East, and Australia. Following the study of cost overrun probability models, decision-theoretic and utility-theoretic methods were developed and illustrated in the evaluation of investment alternatives while formally treating lifecycle costs and other factors of sustainability. Finally, factor analysis and associated logistic regression models were implemented for characterizing the effect of factors that cause cost overruns.

## Acknowledgements

First and foremost, I must thank Allah, God the Almighty, who enabled me to bring this thesis to its current successful stage.

I would like to take this opportunity to express my deepest gratitude to my supervisor Dr. Ata M. Khan, for his invaluable guidance, patience and encouragement throughout the preparation of this thesis.

This research would not have been possible without the kind support and help of the Ministry of Transportation and Infrastructure in British Columbia, the Saudi Cultural Bureau in Canada and Umm Al-Qura University. I would like to extend my sincere thanks to these institutions.

My most sincere gratitude extends to my parents Adil and Ehssan, my brother Bahaa and my sisters Walaa, Ghaidaa, and Esraa. Special thanks to my wife, Ragdah Mirza, and my wonderful children Hala, Farah, and Adil. Thank you very much for your endless love, encouragement, and support.

Last but certainly not least, many thanks are owed to all of the professors and colleagues in the Department of Civil and Environmental Engineering at Carleton University. Thank you very much for your inspiration and support.

# Table of Contents

# List of Tables

# List of Figures

# List of Appendices

# List of Acronyms

| | |
|---|---|
| **PW** | Present Worth |
| **NPW** | Net Present Worth |
| **EXP (NPW)** | Expected Net Present Worth |
| **EAW** | Equivalent Annual Worth |
| **COR** | Cost Overrun Ratio |
| **LL** | Log-Likelihood |
| **KS** | Kolmogorov-Smirnov |
| **B.C.** | British Columbia |
| **ISI** | Institute for Sustainable Infrastructure |
| **MCS** | Monte Carlo Simulation |
| **LCC** | Life Cycle Cost |
| **FHWA** | Federal Highway Administration |
| **GHG** | Greenhouse Gas |
| **SPSS** | Statistical Package for the Social Sciences |
| **KMO** | Kaiser-Meyer-Olkin |
| **CEEQUAL** | Civil Engineering Environmental Quality Assessment and Award Scheme |
| **LLM** | Log-Logistic Model |
| **CM** | Cauchy Model |

# Chapter 1: Introduction

## 1.1 Problem Definition

The importance of infrastructure cost as a sustainability factor is attracting attention in professional practice (Institute For Sustainable Infrastructure ISI, 2018). If cost estimates of an infrastructure project are subject to overruns and the decision to accept the project for implementation is made without knowledge and analysis of such overruns, the economic and financial sustainability of the project will be adversely affected. Aside from this, the misguided decision to accept the project and allocate scarce resources towards it will adversely affect other projects that are competing for financial resources.

Infrastructure projects of all sizes can potentially experience significant project cost overruns and miss scheduled completion deadlines. When cost overruns and delays occur in mega and large projects, media attention is attracted to these phenomena. Although, this does not mean that medium and small projects do not also encounter these problems. It is a common practice to add a contingency allowance to the project's base cost, but that is not always the case. However, even with a contingency fund built into the estimated costs, cost overruns occur, and these require risk analysis research.

Over the years, avoiding initial construction cost overruns in highway projects has been regarded as an important part of good professional practice and this requirement will continue with the expectation that it will be met. Recently, there is a more demanding requirement to subject life cycle costs in infrastructure investment decisions to risk analysis for sustainability reasons (National Academies of Sciences, Engineering, and Medicine, 2016).

In North America, the generic benefits of life cycle analysis (LCCA) appear to be recognized by transportation agencies; in turn, notable organizations have published documents for training purposes (FHWA, 2015; CALTRANS, 2011; ASCE & ENO Foundation, 2014; US DOT, 2015). Some jurisdictions around the world have also initiated programs to introduce LCCA into the professional field. For example, see Gschösser, and Wallbaum (2013).

Although these programs are beneficial, their scope has been generally limited to pavements and bridges. Research needs were identified by a U.S. National Cooperative Highway Research Program (NCHRP) study in modelling LCCA of assets with full inclusion of all costs, including user and agency costs. The study report notes an improved understanding of the benefits of applying LCCA to all assets (including pavements and bridges) is needed. Also, a notable observation was advanced by this study that *"very little consideration is being given to incorporating resilience goals and uncertainty/risk factors into LCCA applications"* (National Academies of Sciences, Engineering, and Medicine, 2016).

For meeting the sustainability objectives, economic factors are to be studied in association with the social and environmental effects of investment decisions (Institute For Sustainable Infrastructure ISI, 2018). Although the balance of economic benefits and costs is a notable criterion for assessing the sustainability of a project, there is a long list of other criteria that enter in the decision to select a highway project alternative for implementation. Examples are equity and social justice, infrastructure integration/connectivity & accessibility, and greenhouse gas reduction (GHG). Their estimation is subject to risk. Further, for the study of the relative effectiveness of alternatives, there is a need to develop

a multi-attribute utility function within a risk analysis framework. Details of the complexity of this and other parts of the thesis research are provided in later sections of the document.

Given the new wider scope of planning, implementing, and managing highway infrastructure, a number of sustainability rating tools (also called planning and rating tools) were developed that are intended to guide the profession in the sustainability direction. Examples are the CEEQUAL scheme, the Greenroads, and ENVISION (Griffiths et al., 2019). Notable among these is ENVISION, advanced by the Institute for Sustainable Infrastructure (ISI) in association with the Harvard Graduate School of Design's Zofnass Program (Black and Veatch, 2017).

The application of ENVISION and other tools, that were developed for the purpose of improving the sustainability of infrastructure projects, can be enhanced with methods for assisting their informed application (Black and Veatch, 2017; Griffiths et al., 2019). Advancing methods for potentially improving the application of these tools is a research challenge. Specifically, the following factors that relate to ENVISION application require research:

- Life cycle cost risk analysis
- Risk in evaluation criteria attainment levels
- Method for jointly treating diverse sustainability criteria in multi-attribute utility function within a risk analysis framework.

For realism and effective application of a rating tool, risk in life cycle cost (including initial cost) should be formally treated. Likewise, treating risk in evaluation criteria based on multi-attribute utility function is necessary for supporting the application

of a sustainability rating tool. Among other sustainability factors, this thesis research is an attempt to develop methods for quantifying and addressing life cycle cost overrun risks in highway infrastructure.

There is a need to identify potential causes of cost overrun and their effect. To go beyond this step, models are needed to study risk in life cycle costs of infrastructure in support of economic parameters of sustainability. Advanced methods that have the potential to enhance the sustainability of infrastructure include risk analysis models, statistical decision theory including Bayesian models, and risk-based utility-theoretic framework for treating multi-criteria scores achieved by competing alternatives. Also, research is needed in Factor Analysis of causes of cost overruns and associated logistic regression models for predicting the effect of identified factors.

## 1.2 Research Objectives

This research was carried out to meet the following objectives:

- Identify and analyze cost risks in the complete life cycle of infrastructure projects including initial construction and rehabilitation phases.
- Study factors that can explain cost overruns and develop quantitative methods to characterize their importance.
- An advance methodology that can be used to study risk in economic factors in association with other factors of infrastructure sustainability.

## 1.3 Scope

For realism in modelling risk in highway infrastructure investments, this thesis research recognized that:

- A new risk-based life cycle cost analysis approach was necessary. In addition to research in the first cost, other components of the life cycle were included.
- The cost overrun risk model based on completed highway projects was recognized as an appropriate step for characterizing cost overrun risk. To develop such models, real-world data acquisition became necessary.
- To model the lifecycle economic factors in association with other factors of sustainability, decision-theoretic and utility-theoretic formulations were recognized as potential means to model risk on a life cycle basis.
- The need for an advanced version of risk modelling of economic factors called for Bayesian theory models to study the feasibility of acquiring additional information for the purpose of risk reduction.
- Although many studies have discussed causes of cost overruns in highway infrastructure, no attempt has been reported in the literature on a Factor Analysis of numerous variables and the development of predictive models for quantifying the effects of well-researched factors. To advance the state of knowledge, it was decided to approach jurisdictions in Canada, the USA, Middle East, and Australia for data.
- Another notable item in the scope of this thesis research was to show the potential use of its products in preparing inputs to sustainability rating tools such as ENVISION.

## 1.4 Document Organization

This introductory chapter provides the background information for this research study and covers research objectives and scope. Chapter 2 contains a literature review. The research methodology is presented in Chapter 3. Major components of the research framework are described, and their interrelationship is noted. Also, data sources are described.

Chapter 4 covers the analysis of cost overrun risks in highway infrastructure projects based on British Columbia data. Chapter 5 describes how to treat risk in economic and other parameters of sustainability. Specifically, decision-theoretic and utility-theoretic models are described. Chapter 6 illustrates the risk analysis of economic factors on a life cycle basis using the Bayesian model. Chapter 7 covers the development of a questionnaire survey for obtaining data from jurisdictions in Canada, the USA, the Middle East, and Australia for use in Factor Analysis. Chapter 8 and Chapter 9 describe in detail the Factor Analysis and logistics regression models, respectively. Chapter 10 covers the contributions, conclusions, limitations, and recommendations. Among the contributions, the link to ENVISION is included in Chapter 10. Finally, references and appendices are presented.

# Chapter 2: Literature Review

## 2.1 Introduction

This chapter provides an overview of previous research on cost overrun risk in transportation infrastructure, mainly highways, and related sustainability issues. Studies are reviewed on factors influencing costs in infrastructure, causes and explanations of cost overruns, and methods for risk analysis. Also, literature that highlights the need to link risk in infrastructure costs and the application of a rating tool such as ENVISION is reviewed.

## 2.2 Definitions and Classification of Costs

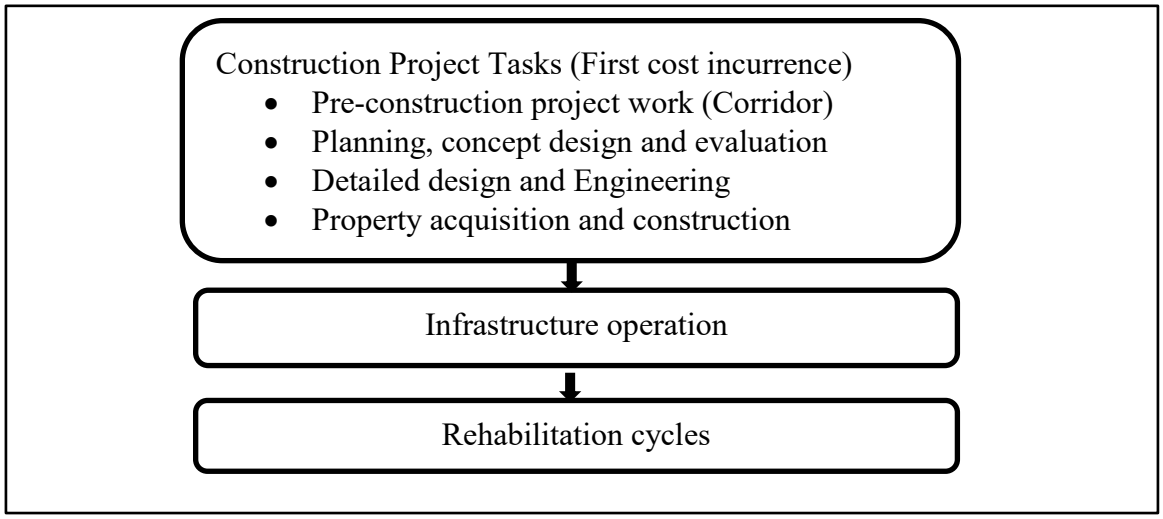
In the subject of "cost", our interest is in risk analysis of life cycle costs of infrastructure, including the most important first cost of constructing the facility. In the literature, frequent references are made to "project cost" which is commonly understood to be the initial construction cost. For an introduction to the principles and practice of project cost estimation (including the "contingency" component), the reader is referred to the Ministry of Transportation and Infrastructure B.C. (2013).

Figure 2.1 shows the context of the first cost within the infrastructure life cycle. The costs that occur following the opening of the facility to service become a part of the overall life cycle costs. Although the professional practice is beginning to appreciate the benefits of life cycle cost analysis, the associated risk and uncertainties do not receive due attention (National Academies of Sciences, Engineering, and Medicine (2016). For the information of the reader, the life cycle costs include the following components:

- Initial cost (commonly referred to as the construction cost)

- Cost of rehabilitation cycles
- Road user costs.

***Figure 2. 1: Infrastructure Life Cycle***




Adapted from Ministry of Transportation and Infrastructure B.C., 2013

The construction project is a mission to build or improve a unique facility, product, or service within the specified time, budget, quality, and scope (Chitkara, 2011). In practice, however, some construction projects experience cost overrun and schedule delay upon completion. Due to large investments, the large and mega projects become notable in the literature. Cost overrun and delays require comprehensive investigation to improve the outputs of the construction industry.

Although cost overrun is common in infrastructure construction projects, the causes of such overruns remain a subject worthy of research. Costs refer to the planned expenditure, which the client initially agrees to pay for the development of the specific construction project (Cantarelli et al., 2013). In the case of public projects, a government agency is responsible to develop the cost estimate. Estimated costs refer to the planned costs that are examined during the construction project approval process. There are two

versions of these planned costs. The first version is used at the planning level and a potentially revised estimate (i.e., the second version) becomes known prior to construction. On the other hand, actual costs refer to the accounted costs which are essentially incurred for constructing infrastructure (Cantarelli et al., 2013). Cost overrun refers to the variation between the actual costs and estimated costs.

The collective costs associated with project construction are usually defined and estimated to certain guidelines that all stakeholders must agree to. Noted earlier, if the reader seeks an introduction to the practice and principles of project cost estimation (including the "contingency" component), they should refer to the Ministry of Transportation and Infrastructure B.C. (2013). Other references on this subject are (Lind and Brunes, 2015; Siemiatycki, 2015; Ramabodu and Verster, 2010).

The construction industry aims to complete projects on time and within the estimated budget. However, this objective may not be achieved in many cases because construction projects are exposed to many uncertainties that impact the first cost of a project.

In infrastructure planning and feasibility study, the cost of rehabilitation cycles is to be estimated and included in life cycle cost analysis (LCCA). Depending upon the design, traffic, and environmental conditions, more than one rehabilitation activity may be planned to extend the service life of the facility. At the end of life, a salvage estimate is required for a feasibility study. Even if such costs and salvage value are estimated in constant dollars, these cannot be estimated with certainty. Therefore, risk analysis is warranted.

## 2.3 Causes and Explanations for Cost Overrun

A review of the literature on this subject presented in this section of the thesis suggests the following observations.

i. Although many causes played a role, it appears that are some recurring factors. These are worth noting from a research perspective.

ii. The need for risk management was identified by several authors.

Early research on cost overruns is reported by Merewitz (1973). This reference notes cost overruns that accrued to the sixty-six American rail and road projects. Relatively more recent studies are reviewed next. Flyvbjerg et al. (2002) compiled data on the size of the cost overruns in several transport infrastructure projects. Additionally, Flyvbjerg (2016) reported a dataset that encompasses 1603 rail and road projects in 2062 infrastructure megaprojects with a value of US$1.5 - 8.5 billion. These projects occurred during the period 1927-2013 within twenty nations. The research revealed that rail and road projects have cost overruns of 24 - 40 percent on average.

It is interesting to note that the variation between high-cost overruns approximated by Flyvbjerg (2016) and the relatively lower levels accruing to some individual-nation studies is quite substantial. According to Lundberg et al. (2011), Swedish road projects experienced cost overruns amounting to 5.5%, and according to Berechman and Wu (2006), Canadian rail and road projects experienced cost overruns of 14% and 5%, respectively.

According to Kaming et al. (2012), cost overrun is too common in infrastructure construction projects. Studies show that in some countries, by the time a project has

completed, the actual costs are much higher than the estimated costs. As per Ramabodu and Verster (2010), the key controllable causes of cost overrun risks in infrastructure projects consist of factors such as deficient formulation of the project, lack of enough information about the project, poor cost estimation, and poor overall lifecycle feasibility evaluation study. The end final results are erroneous investment decisions.

Lind and Brunes (2015) add other causes associated with a cost overrun for infrastructure projects, including poor planning for implementation. This involves insufficient time plan, insufficient resource plan, poor organization skills as well as insufficient equipment supply plan. Lack of appropriate contract planning and management is also a dominant factor associated with cost overruns for projects. This can be caused by incorrect pre-contract actions as well as poor contract awarding. Lack of appropriate project management during implementation is associated with inadequate and inefficient work practices, delays, and changes in work scope (Lind and Brunes, 2015).

Cost estimation can become less reliable due to a lack of well-studied and updated information. This typically leads to significant cost issues that attribute to substantial uncertainties. The need to seek and apply updated cost information becomes critical should there be any substantial delays in starting the construction process (Ramabodu and Verster, 2010).

Cantarelli et al. (2013) argued that cost overrun for infrastructure projects is a significant problem worldwide and Kaming et al. (2012) highlighted the need to predict causes to manage risk. A study conducted by Lind and Brunes (2015) found out that the most common sources of cost escalations in infrastructure projects include inflationary pressures, an increase in the cost of materials, labor payment, intricacies in acquiring

construction materials, construction delays, and deficiencies in cost estimation. Another study conducted by Memon et al. (2011) noted some common causes that include increase in the cost of materials because of inflation, errors in material estimation, and the level of project difficulty. Samarghandi et al. (2016) classified the causes of cost escalation and schedule delay in infrastructure projects in Tehran as a lack of experience of consultant, contractor, and owner as well as policy and legal factors.

Flyvbjerg (2007a) pointed out that uncontrollable cost can be caused by financial issues as well as payment arrangements, problems with contract management, shortages of materials, changes in conditions of the construction sites, changes in the design, administration issues, price fluctuation, and erroneous estimations as well as changes in the contracts. Sinesilassie et al. (2018) stated that the critical factors that impacted the cost of public construction projects in Ethiopia were the lack of experience of the project manager as well as the conflict between project parties (owner, designer, and contractor).

Flyvbjerg et al. (2004) attributed cost overruns to various factors that cannot be controlled or factors that to some extent are uncontrollable. These factors consist of the accuracy of primary estimations, the level of governmental regulation and control, design changes, issues allied to labor, and construction completion delays. This is in line with Memon et al. (2011) who found out that factors associated with the variation between actual and estimated costs include poor project management, changes in the design, unforeseen site conditions, inflation, shortage of construction materials, financing problems, as well as unsuitable contactors. Additionally, lack of appropriate phasing of construction projects can result in overheating of the economy and this causes shortage of construction materials since the demand for materials can be higher than the available

supply. In turn, this results in an increase in the cost of construction materials and obviously results in an increase in the project cost overrun (Siemiatycki, 2015).

Memon et al. (2011) carried out a study on predictors of cost overrun and established that the project size, variation between bidders and the estimated costs by engineers, delivery methods, competition level, as well as administrative aspects are some of the factors that contribute to cost escalation. Research has also illustrated that the size of an infrastructure project impacts the cost overrun (Subramani et al., 2014). Mega/large construction projects are very complex; so, when designing and planning for large projects, it is possible to miss out on some items which in turn, can elevate cost problems. Furthermore, because larger construction projects have much higher stakes, it is important to be cautious from the project's inception until its completion.

Studies have shown that there are conflicting views regarding cost escalation in large projects. According to Cantarelli et al. (2013), the level of cost overrun reduces with the increase in the contract amount. On the other hand, Vu et al. (2016) argue that the level of cost overrun increases as the amount of time constructing the project increases. The cost of infrastructure construction projects is dependent on a cluster of interrelated variables and this means that the project cost can be affected by the construction team, as well as market conditions (Vu et al., 2016).

## 2.4 Issues in Project Budget/Cost Estimation

The most common serious situation is the underestimation of the construction cost. According to Flyvbjerg et al. (2004), it is common to underestimate costs for large infrastructure construction projects in order to acquire approval for funding. They further

stated that nearly all large infrastructure projects have initial cost estimation errors that always lead to the need of more financing to have the project completed successfully. Construction cost overruns normally arise after the construction has begun but basically the cost overrun problem is mostly in the contract cost estimation and tendering phase (Siemiatycki, 2015).

Flyvbjerg et al. (2003) provide explanations of cost under-estimation in three categories that include: political, technical, and economics. Technical explanations have been described in the literature. Numerous studies that compared the final cost and the estimated budget of infrastructure construction projects explain the overruns by foretelling errors that technically include imperfect techniques, insufficient data, honest mistakes, intrinsic problems during future projections, and also lack of experience of those estimating the costs and predicting financial conditions (Vu et al., 2016).

Authors have given numerous explanations of economic nature and how conditions impact the cost of construction projects. Flyvbjerg (2007a) provides two kinds of economic explanations, namely concerning economic self-interest and the other explanation in regard to the public interest. For the self-interest, when a construction project commences, it provides job opportunities for engineers, construction companies, and also various stakeholders to make money from the project. In situations where the stakeholders take part in forecasting process and budget estimation, they have the ability to manipulate the outcomes as well (Siemiatycki, 2015).

When it comes to the public interest, project promoters can intentionally undervalue the project costs for several reasons, including the impression that public money is preserved. Studies have shown that during budget planning for infrastructure construction

projects, it is not right for those planning the budget to keep costs low instead of providing correct estimates, however costly these might be (Flyvbjerg et al., 2004).

Political explanations have also been advanced. According to Flyvbjerg et al. (2004), interests and power can have major impact on the cost of infrastructure construction projects. According to another reference, forecasts and budget estimates can be deliberately biased to serve the interests of the project advocates (Ramabodu and Verster, 2010).

Poor project management is noted in the literature as a potential cause for cost overrun. A project management team is very vital when it comes to the control or management of costs allied to infrastructure construction costs. Poor project management can result in many challenges, including problems allied with costs. Vu et al. (2016) suggest that poor project management affects all phases of the construction process. This in turn, can result in poor coordination and poor planning, ineffective communication between project team members and other stakeholders, inability to identify problems and implement the appropriate changes, and lack of control on both time and cost inputs (Ramabodu and Verster, 2010). It is logical to give credit to good project management for efficient management of costs (i.e., proper cost estimation, cost analysis, and implementation of appropriate measures to correct difficulties allied to cost). On the other hand, poor project management leads to unsatisfactory cost estimation and cost management. Studies have suggested that poor project management is among the leading reasons for cost escalation in construction projects (Love et al., 2010).

Unpredictable site conditions have been noted by studies as an issue in developing cost estimates. In infrastructure construction projects, detailed knowledge of the site is a requirement. But, it may not be possible to assess the entire site and determine the site

conditions until when the excavation is complete (Vu et al., 2016). Unexpected site conditions can sometimes necessitate complete or major redesigning of the construction project, resulting in a significant cost increase. Other additional costs caused by unexpected conditions at the site include the problems of having to move the supplies and machinery away from the site. Labor costs and over-time are also needed in such situations. These result in additional costs and significantly contribute to project overruns (Siemiatycki, 2015).

The cost estimation process is also recognized as an issue. The process involved in the estimation of the costs for construction projects was noted in the literature to be a challenge. This is because various factors influence cost estimates that are difficult to identify or control. Owing to these reasons, some future values of variables are difficult to predict (Siemiatycki, 2015). In addition, preparing estimates for some of these factors is time-consuming and requires much effort in order to estimate the initial costs. But, these are expected to be reasonably exact since decision-making during this phase is likely to significantly affect the successive phases (Kaming et al., 2012).

During the cost estimation process, there are numerous instances in the construction projects where the initial cost estimates are constrained due to failure to include all cost components such as VAT costs (if applicable), professional charges, management costs, statutory costs, among other costs that are not considered as significant. Love et al. (2010) gave some examples where certain problems occur because of not including land acquisition costs within the initial estimations or grossly under-estimating land costs during the cost estimation process. In addition, costs are likely to increase if the current maintenance contract costs within estimations are included while some are excluded. It is

also crucial to include a reasonable “contingency” figure. More importantly, it should be ensured that sufficient costs are included for accommodation works and archaeological tasks for infrastructure construction projects (Kaming et al., 2012). Arif et al. (2015) reported that the contractor's involvement in the estimation process and labor rates were the most significant inaccuracies in cost estimates for infrastructure projects in Pakistan.

A study carried out by Love (2011) showed that large infrastructure construction projects that have significant cost overruns might have contractual problems that even attract media attention. Examples include Denver’s US$5 billion airport project that had a 200 percent cost overrun, the Oresund bridge project that had a 68 percent cost overrun, Scottish Parliament Building project (although not a transportation infrastructure) that had a cost overrun of more than 900 percent (Love et al., 2010). This has also been observed in Australia where numerous large construction projects exhibited significant cost control problems because of poor project management and design errors (Love, 2011). A good example is the Western Australian Perth Arena (not a transportation infrastructure) whose initial cost estimate was $168 million but the actual cost was more than three times the initial cost estimate, indicating a 300% cost overrun for the project. Similarly, in South Africa cost risks have been observed in many large-scale construction projects. For example, during the construction of a soccer city stadium project within Johannesburg (not a transportation infrastructure), the actual cost was 3 billion and the cost overrun for this project was 58%. Similarly, Green-point stadium (not a transportation infrastructure) also had a 50 percent cost overrun of the initial estimated cost (Love, 2011).

According to Flyvbjerg (2007b), the usual cost escalations for infrastructure construction projects amount to 20.4 percent for roads, 34 percent for bridges, and 45

percent for railway projects. On the other hand, Love et al. (2010) found that roads and bridges have relatively low-cost overruns of 14 and 6 percent, respectively.

Flyvbjerg (2009) suggests two basic reasons as to why construction projects have cost issues. One of the reasons is boldly called strategic falsification. It was explained as the deceptive actions that political interests and planners use to make sure that construction projects carry on. The second reason is named the optimism bias which summarizes the methodical propensity for decision-makers to be excessively optimistic regarding the outcome of the planned actions (Siemiatycki, 2015). This materializes in over-estimating the possibility of positive incidents as well as under-estimations of risks and loss. Optimism bias is considered a major problem by many authors during the planning and budgeting of infrastructure construction projects, as well as during detailed cost estimations of these projects. According to Love (2011), some mistakes and inaccuracies in engineering organizations are based on practices that were initially intended to solve a specific problem. But people might repeat unsuitable practices, for example by taking shortcuts and not using the right processes during construction.

The above review of literature is very revealing regarding the causes of cost overruns. A logical question is what can be done as a solution? Vu et al. (2016) suggest that risk assessment within the cost estimation process is important in order to identify aspects that might have been left out as well as identify factors that might contribute to cost overruns. According to Ramabodu and Verster (2010), risk is an intrinsic aspect in construction process but it can be evaluated in advance, controlled, and also integrated during cost estimation of the project. Risk assessment process should be repeated regularly and should involve brainstorming by all pertinent project parties.

## 2.5 Selected Studies of Cost Overrun around the World

A study by Žujo et al. (2010) reported that 81% of 333 medium and large infrastructure projects in Croatia were suffering from a cost overrun problem. In Bosnia and Herzegovina, a research study investigated that 41.23 percent of 178 infrastructure projects did not meet the contract price. Omoregie and Radford (2006) pointed out that 14% is the minimum average percentage of cost overrun in Nigerian infrastructure projects.

Challenges faced by the Malaysian construction industry in completing the construction projects within the estimated budget were reported by (Azis et al., 2013; Ibrahim et al., 2010; Toh et al., 2011). Other authors, namely Endut et al. (2005) found that approximately only 37.20% of 51 private projects and 46.80% of over three hundred public projects were completed within the planned budget, respectively.

A comprehensive study reported by Terrill and Danks (2016) on Australian transport projects completed between 2001 and 2015 contributed three main ideas to improve cost estimates. First, project risk must be carefully evaluated. Second, the process of the cost estimation should be aligned with the previous experience (historical data). Third, cost estimates should reflect predictable patterns in project types that exceed their budgets. Also, they noted that there are three characteristics associated with a high risk of cost overrun. First, large and complex projects; second, road and rail projects; and the third characteristic that affects a project's cost risk is its contract type.

According to a study in the United Kingdom, 30% of the clients complained that the actual cost was higher than the estimated budget in their construction projects (Jackson, 2002). Other selected cost overruns examples are the North-by-pass highway project in Uganda, which faced a cost escalation of more than a hundred percent of the contract price,

and in Slovenia, 51 percent of ninety-two highway and road projects were suffering from cost escalation (Nikic, 1998).

## 2.6 Some Advanced Methods Applied to Study Cost Overruns

Past research in the subject of this thesis is rather scarce. However, a few studies provide a good starting point for advancing the state of knowledge in risk analysis of highway infrastructure life cycle costs.

When investigating the causes of cost overruns and its risks in the construction sector, face-to-face interviews and questionnaire surveys are common techniques that were used in some studies (El-Sayegh and Mansour, 2015; Zhao et al., 2013). Lind and Brunes (2015) explained the causes of cost overruns in infrastructure based on the results of a questionnaire study and proposed a new framework with applications to Sweden. The questionnaire was sent to agents in the field and returns captured their experience and views on this subject. Analyses were carried out with the use of commonly used statistical analyses. Their recommendation for further research is taken into account in this thesis research study.

The cost overrun phenomenon also attracted the modeler's attention. A Working Paper published by the Sauder School of Business, the University of British Columbia, titled "Cost overruns risk analysis in transportation infrastructure investments" investigated the risks and cost overruns in 29 tunnel and bridge projects and 104 highway and road projects in Vancouver (Berechman and Wu, 2006). The authors found that the development of simulation models, performing regression analysis, and the use of

probability distribution fitting approaches help in analyzing risks and identifying the factors that lead to cost overruns.

Abdul Rahman et al. (2013) developed a mathematical model for predicting cost overruns due to some changes that take place during the construction of projects. Berechman and Chen (2011) have reported on the incorporation of cost overrun risks into the decision-making of infrastructure megaprojects. Their contribution is in proposing distribution-fitting models and their use in estimating probabilities of cost overruns. They also illustrate the application of probabilities in a decision-tree model of project selection. The authors recommend that cost overruns should be incorporated into the decision-making and evaluation of transportation infrastructure projects especially megaprojects where a small percentage of cost overrun might lead to a high financial issue.

Heravi and Mohammadian (2019) studied about 72 urban construction projects in Iran. T-test and correlation coefficient analyses were used to analyze the collected data. The authors concluded that, among the seventy-two evaluated projects, the projects completed on time were approximate 8.50%, while those completed within the estimated budget were around 7.0% of evaluated projects. Their findings also revealed that mega construction projects in Iran experienced higher cost overrun than medium and small projects.

Research conducted in the Netherlands by Cantarelli et al. (2012) investigated the impact of three variables on the cost overruns, namely the duration of the projects, the size, and type of the project. In the analysis of the collected data, the authors used different statistical and probabilistic techniques, which included F-test, binomial test, t-test, and

ANOVA. They found cost overruns in 21.7% of fixed link types of projects, 18.6% of roads, and 10.6% of rail projects constructed in the country.

## 2.7 Cost Overrun Risk Analysis Methods

### 2.7.1 Qualitative and Quantitative Risk Analysis

Analysis of risks could be either quantitative or qualitative. Invariably, a quantitative analysis is central to statistical and probabilistic approaches. On the other hand, the qualitative analysis evaluates variable probabilities as well as the repercussions utilizing subjective judgments, experiences, and expert opinions. According to Ezzy (2013), the qualitative analysis approach may be adequate to detect the perils accruing to a system. The quantitative peril analysis approach evaluates the system in greater detail central to risk assessment. The quantitative analysis approach universally provides testable and clear conclusions.

### 2.7.2 Monte Carlo Simulation

Monte Carlo Simulation (MCS) is a mathematical technique used to quantify risk. Random numbers are generated for each uncertain variable based on their respective probability distributions and the model describing the system. This approach addresses the risks involved in input data and output results. The input parameters are represented by their probability distributions (e.g. triangular distribution) (Khan, 2013). Probability distribution and statistics of the model are obtained by repeating the random numbers of the value of the variable several times (Singh et al., 2007).

### 2.7.3 Probability Distributions Used in Risk Analysis

Original research supplemented with literature sources leads to the choice of probability distributions for use in this research. Examples considered are beta distribution and normal distribution. According to Regnier (2005), the beta distribution is the most common probability distribution when analyzing the costs and schedule growth in transportation infrastructure projects. However, as explained in a later chapter, other probability distributions were studied. These are: Cauchy, Normal distribution, Log-logistic, Gamma, and Log-normal. Please refer to Chapter 4 for further information.

According to Frame (2003), the normal distribution is universally deployed in the analogies involving routine processes. However, the rationale of the normal distribution does not apply to many phenomena. The normal distribution lacks applicability logic, notably towards the variables having lower or upper bounds or rather having a probability density function shape that is distinct from the bell shape of normal distribution. Although knowledge gained from literature is useful, in this research, other forms of probability density functions were investigated with the use of actual data and their inherent properties.

### 2.7.4 Risk-based Utility-theoretic Framework for Evaluation of Alternatives

Given that evaluation of investments will treat criteria of sustainability that go beyond economic factors quantified in monetary units (e.g., Net Present Worth expressed in dollars), it becomes necessary to apply utility theory for the development of multi-attribute utility functions. These functions, also called value functions, can transform weighted criteria achievement levels into utility values. In the context of advancing life

cycle risk analysis methods, the probabilities of the occurrence of various criteria achievement levels become a part of the risk-based utility-theoretic framework. The criteria relate to economic, social, and environmental sustainability objectives. An example application of this methodology reported in the literature is a study on policy challenges of automation in driving (Khan et al., 2012).

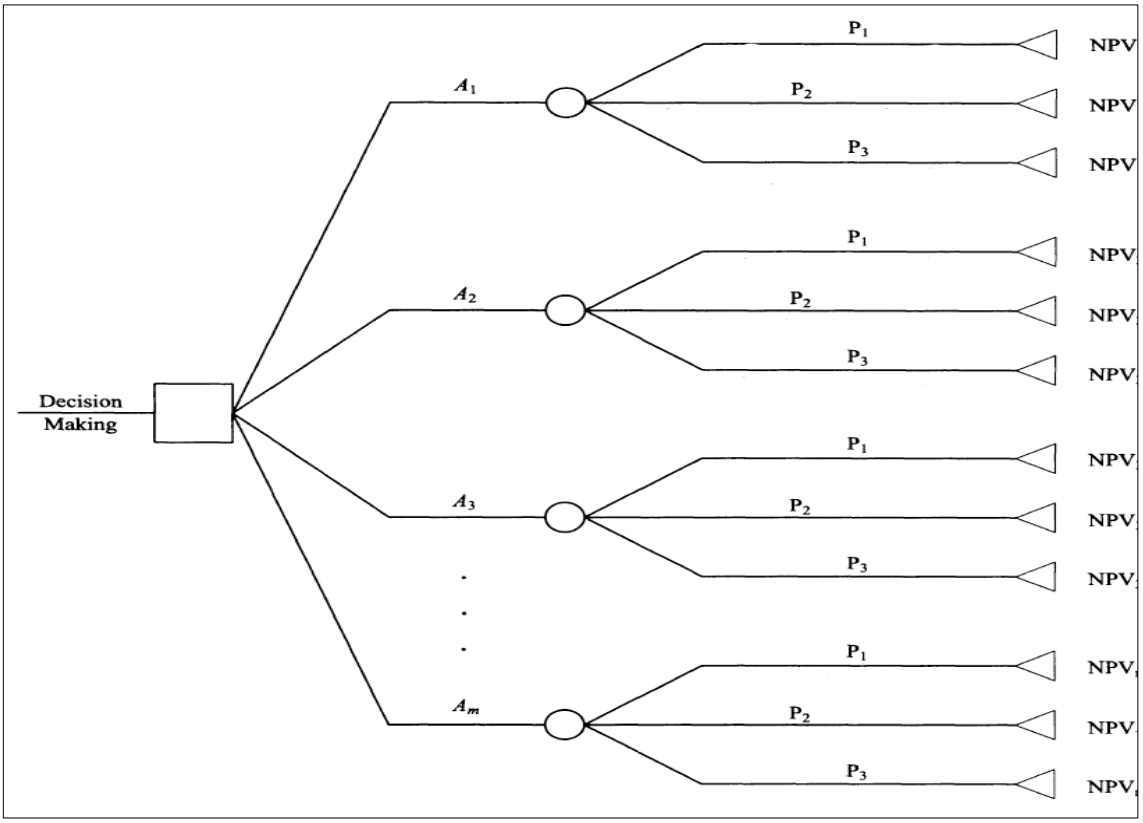
### 2.7.5 Decision-Theoretic Methods

Basic decision-theoretic models usually explained in the form of decision trees provide a tool for risk analysis. These are broadly deployed in practice to represent decision problems in such fields as business and engineering. Well-known applications of these methods are presented by Schlaifer and Raiffa (1968).

Boardman et al. (2001) conducted decision tree analysis notably in two fundamental steps. Firstly, they developed a decision tree in specifying the risk analysis model. Secondly, they described how to solve the decision tree by moving backward, notably from the outcomes towards the initial decision options. The expected net merit values were computed and identified the branch that represented the most optimal net merit value.

A basic example of a decision tree is shown in Figure 2.2, where the NPV is the Net Present Value, (P) is the probability of the occurrence of the random variable, and (A) represents the alternatives. The expected NPV of a given alternative (A) of a project can be calculated as $P_1$ x $NPV_1$ + $P_2$ x $NPV_2$ + $P_3$ x $NPV_3$.

*Figure 2. 2: Scheme of the Decision Tree*




Adapted from Berechman and Chen (2011).

### 2.7.6 Bayesian Statistical Decision Model

The statistical decision theory enables a decision-maker (or a decision system) to identify the optimal course of action in situations involving risk and uncertain future conditions (termed uncertain states of nature in the decision theory terminology). That is, the optimal alternative is identified with the application of probabilities to uncertain outcomes (Schlaifer 1959, Pratt and Schlaifer 1995). A Bayesian method is an advanced form of statistical decision theory. It allows the analyst to use prior information and also offers the ability to update probabilities as a result of new information. These probabilities serve important roles in risk analysis (Khan, 2013). In addition to this, it provides an

estimate of the amount of money that can be spent on additional information to reduce risk (Congdon, 2007; Woodberry et al., 2004). An example application of the Bayesian method reported in the literature is a toll road life cycle analysis study (Khan, 2013).

### 2.7.7 Regression Models for Risk Analysis

Regression models have been attempted for the study of risk determinants. They are effective in explaining the causes of risk and quantification of their impacts. Another advantage of regression analysis is that these can be easily interpreted. Flyvbjerg et al. (2004) deployed simple linear regression frameworks in the analysis of the correlations between the cost overrun of transport project and three potential risk determinants, namely project duration, project magnitude, and project ownership. The deployed model evaluated the presumed impacts accruing to the three potential risk factors, notably on the cost overrun of the project. The regression models were found to assist in the provision of accurate and intuitive interpretations regarding such impacts.

### 2.7.8 Additional Methods with Potential Role in Modelling Risk

Literature sources suggest that a large number of causes of cost overruns can be investigated in risk studies. A method that is well suited for an orderly study of many variables is Factor Analysis. Further, the results of Factor Analysis can be used to develop regression models for quantifying the effect of factors. Although these methods have much potential in explaining and quantifying risk, these have not been investigated by other researchers. This thesis research has successfully applied these methods for contributing new knowledge.

## 2.8 Effect of Cost Overruns on Infrastructure Sustainability

### 2.8.1 Sustainable Construction

Amongst the several investment activities that contribute to sustainable development is sustainable construction. According to Mateus and Bragança (2011), a construction project can be planned according to sustainability principles. The First Global Conference on Sustainable Construction that took place in Florida defined these as "creating a healthy built environment using resource-efficient, ecologically-based principles" (Kibert, 1994). The implication is to focus on economic, environmental, and social factors of sustainability in infrastructure construction.

In practice, all stakeholders must upgrade their methodologies and information that is necessary to meet the requirements of sustainable construction (Rahim et al., 2014). That is, sustainability factors have to be built-in infrastructure projects that require specialized knowledge for this purpose. There is evidence of increasing acceptance of sustainable development objectives in providing engineered transportation infrastructure.

However, experience around the world indicates that the development of highway infrastructure presents many challenges in all phases of planning and implementation, including completion within the planned time and budgeted cost. Unless due attention is paid, the end product may not always be sustainable (Hampson and Brandon, 2004).

Hussin et al. (2013) show that the construction sector is amongst the most substantial industries that yield contribution towards a territory's socio-economic growth, especially for the developing nations. However, prevailing approaches to linking

environmental, social, and economic factors in infrastructural project planning and implementation lack consensus on details (Nkurunziza and Irumva, 2018).

Sustainability of a large/mega infrastructure project calls for the treatment of economic, social, and environmental factors. Given that financial resources are scarce, the best value for money must be obtained when infrastructure investment decisions are made. Studies carried out for the Government of Ontario have emphasized the value for money (VFM) approach to project design and evaluation (WSP, 2015). This emphasis on cost-effectiveness implies that project planners, designers and managers should avoid cost overruns with risk analysis and aim for projects with high value-for-money potential.

### 2.8.2 Life Cycle Costing Towards Sustainable Construction

Cost is one of the primary considerations throughout a project's life cycle. The application of Life Cycle Costing (LCC) in the construction sector is crucial to satisfy all the sustainability principles. In the project development step, the LCC principles and approaches target the provision of the optimal value notably from an entire life view (Rahim et al., 2014).

A useful definition of life cycle cost advanced by White and Ostwald (1976) is as follows. "The life-cycle cost of an item is the sum of all funds expended in support of the item from its conception and fabrication through its operation to the end of its useful life". Also, LCC is defined by Fuller and Petersen (2006) as total costs that result from asset acquisition, operation, and disposal for a specified time-period. It should be noted that the cost of planning and design is a part of the lifecycle cost, but it is not included in the cost overrun study. This cost becomes known before construction begins.

For use in this thesis research, the need for a detailed risk based LCC model was defined to help stakeholders make decisions that balance sustainable development and financial viability. It is recognized that the construction part of the overall life cycle is of much importance due to large capital investments. However, other components are also of importance. An example of a public-private-project (P3) toll road risk assessment is provided by (Khan, 2013).

In practice, according to recent literature, minimal effort is devoted to ensuring the sustainability of transportation system, including highways. One reason is the lack of knowledge of potential adverse effects (Adnan, 2012; Naidu, 2008). Although there is a growing desire to highlight the sustainability objective in transportation infrastructure project planning and implementation, many developers and operators lack well-researched methods that incorporate sustainability factors in guiding decisions that involve large capital investment.

According to current thinking, the application of life cycle cost (LCC) should become a part of the overall sustainability rating of a major project (Institute For Sustainable Infrastructure ISI, 2018). For this reason, the combination of sustainability concept and an appropriate LCC approach to the evaluation of infrastructure projects will be a new challenge that will make a significant difference due to the scale of these projects and the level of funding involved. According to List (2007), LCC models help users to figure out the most sustainable growth path for the common infrastructure as well as to assess competing initiatives.

It is useful to repeat here that the initial construction cost accounts for a very high part of the overall life cycle cost and that cost overruns with adverse effects can occur

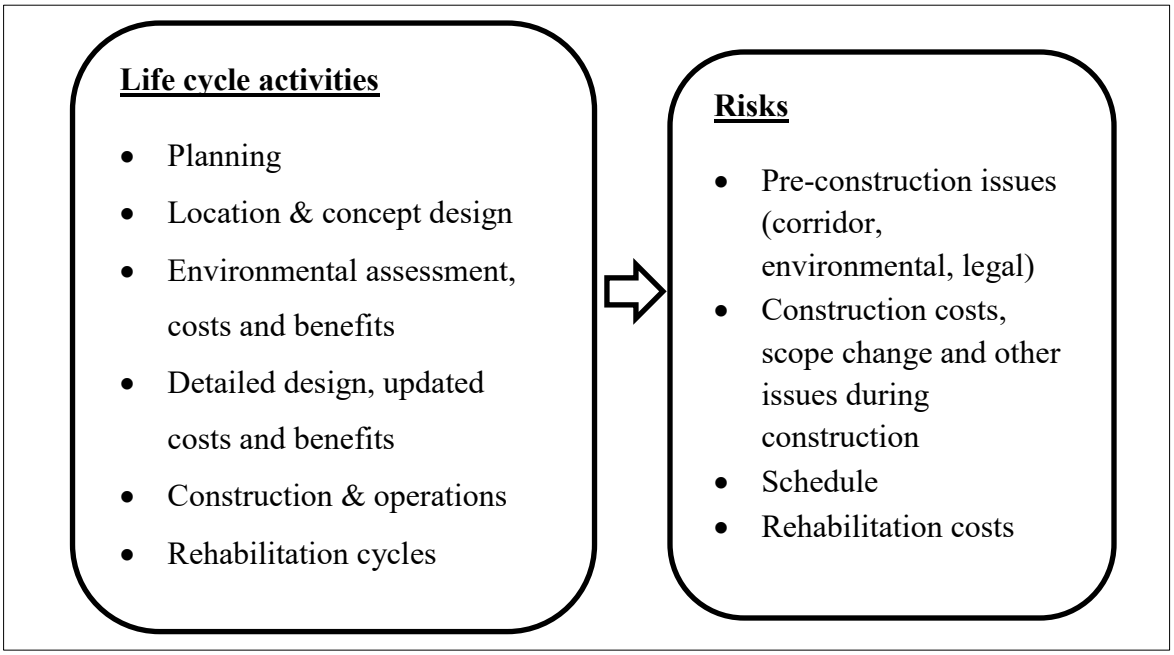
(Abdul Rahman et al., 2013; Ahmed and Hasan, 2014; Alinaitwe et al., 2013; Ramabodu and Verster, 2010). Even a modest cost overrun can impact a project's economic sustainability. Therefore, controlling project costs within a time-span between project commencement and completion of the project construction is a key challenge (Lind and Brunes, 2014).

## 2.9 Study of Life Cycle Investment Tasks and Associated Risks: Guidance from Literature

Stakeholders and decision-makers expect all projects (large, medium, and small) to be on-budget and on-time. As noted earlier, the complete life cycle of a highway facility includes planning, design, construction, and rehabilitation (Figure 2.3). In the case of public-private-projects (P3s) that are intend to be toll facilities, analysis of the full life cycle risks includes the additional risk in obtaining an acceptable rate of return (Federal Government of Canada, 2015; Khan, 2013). However, the risk in the feasibility of a toll facility is beyond the scope of this thesis research.

A highway project can be subjected to risk assessment as the project progresses from planning to design and implementation stages (Figure 2.3). Literature sources covered in this chapter have the potential to provide a starting point for identification and analysis of risk on a life cycle basis for non-toll as well as toll facilities. But, the non-toll type of highway infrastructure is of interest in this research.

*Figure 2. 3: Life Cycle Activities and Risks of a Non-Toll Type Project*




Adapted from Khan (2013).

Although there are many uncertain items in a large project, these commonly affect two key risk areas: cost and schedule. Here, our focus is on cost factors. Previous experience in the risk analysis field implies that highway investment projects should be characterized in terms of factors that cause risk and risk assessment should be carried out on a life cycle basis.

As illustrated in Figure 2.3, there are several stages in the overall life cycle of a project that is likely to benefit from risk assessment. The construction part of the overall life cycle is a key candidate due to large capital investments. However, other components are also of importance. An example of a P3 toll road risk assessment is provided by Khan (2013) and Khan and LaFontaine (1986) provide insights on the management of highway rehabilitation.

It is useful to acquire knowledge that leads to risk identification and quantification. In construction planning, a number of risk factors can potentially increase expected cost and duration (Galway, 2004; Kleine, 2009; Maher et al., 2006; Reilly, 2008). These could belong to the following risk categories:

- Budget risk (deviation from an estimated budget for several reasons that include material price escalation, limited competition, labor shortage).
- Event risk (risk of internal or external events that cause extra time and/or money e.g., environment-induced, extreme weather, unforeseen subsurface conditions, public pressure, and public resistance).
- Scope risk (external pressures to ask for significant changes to project scope, e.g., stakeholder pressure for additions, permitting agency demand).

As previously noted, in this thesis research, issues in identification and quantification of cost risks are of interest.

Table 2.1 presents several risk factors that increase expected cost. However, in rare cases, an opportunity may arise that may decrease expected cost.

*Table 2. 1: Example Risk Factors and Their Expected Impact on Cost*

| Risk Factor | Expected increase in cost | Expected decrease in cost |
|---|---|---|
| Controversy on environmental grounds expected. | Yes | |
| Lack of sufficient number of bidders. | Yes | |
| Other construction projects in region limit supply of labor, bid costs are higher as a result. | Yes | |
| Different site conditions. | Yes | |
| Negative community impacts expected. | Yes | |
| Context sensitive solutions. | Yes | |
| Change in final alignment geometry. | Yes | |
| Formal consultation regarding environmental impact assessment. | Yes | |
| Threat of lawsuits. | Yes | |
| Material availability. | Yes | May be yes |
| Failure to obtain necessary utility agreements or acquisitions. | Yes | |
| Unanticipated cultural or archaeological findings. | Yes | |
| Objections to right of way appraisal take more time and or money. | Yes | |
| Unanticipated hazardous material or contaminated soils. | Yes | |
| Mitigation Bank not available. | Yes | |
| Unexpected geotechnical issues. | Yes | |
| Force Majeure. | Yes | |
| Design incomplete or in error (during design). | Yes | |
| Railroad issues. | Yes | |
| Historic site, endangered species, wetland present. | Yes | |

Adapted from Larocque and Parker (2009).

According to cost estimation practice, a base cost estimate is developed based on known conditions and the assumption of certainty. To account for uncertainties, a well-defined contingency fund is reserved to account for uncertainties. The higher the uncertainty, the higher the contingency fund (Ministry of Transportation and Infrastructure

B.C., 2013). Likewise, schedule risks are handled by assigning a contingency item in the cost estimate.

Owing to the limitation of the traditional contingency approach in addressing risk, the present practice is attempting to move to the identification and analysis of risks. Attempts can be made by transportation departments to subject projects to risk analysis as well as to the broader-scope risk management (Roberds and Mcgrath, 2006; US FHWA, 2006; Washington State Department of Transportation, 2014).

In the risk-based estimation, uncertainties and risk can be identified, probabilistic approaches are then applied to those items that cannot be predicted with certainty. Uncertainties can be addressed at the planning stage, design stage, and construction stage until project completion. The actual cost of construction becomes known at that time, and of course, the adherence to schedule also becomes known at the end of construction. Likewise, the actual cost of rehabilitation becomes known only when these occur.

If a transportation agency has already established the practice of risk assessment of projects, the typical steps followed are noted below.

- Risk identification
- Risk assessment/analysis (qualitative and quantitative analysis)
- Risk response plan
- Risk monitoring and control

Following risk identification, a baseline risk assessment can be carried out that may lead to further risk assessment studies. Next are risk response plan, risk monitoring and control. During implementation of the project, it may become necessary to loop back to risk response plan.

There are issues in the above-noted practice that research can identify. Potentially interested researchers can take the following actions.

- Search archived data and case study reports and identify key risk factors in projects. Note estimates of cost and duration and deviations from actual experience.
- Define the extent to which identified risks were common to projects.
- Record their known impacts on cost and duration (subject to availability of sources).
- Record any other information of value to the identification and impacts of risk.

Since costs and completion time are future milestones, at the planning stage these are forecasts. Even if the focus is only on the construction part of the overall life cycle of a highway facility, better forecasts of costs and schedule must be developed at a detailed level for efficiency and accountability (Holmes, 2016; Larocque and Parker, 2009).

However, as previously noted, experience around the world in completing large projects on time and within budget has been a partial success at best. According to a World Bank study of 258 transportation megaprojects, on average the actual costs are 28 percent higher than the budget/estimated costs and the percentage of overruns increases with the size of the project (Flyvbjerg, 2007b). Of course, in many instances, scope changes and other circumstances beyond the control of planners and engineers may be responsible for cost overruns. But methodological and input data deficiencies are likely to have an effect.

## 2.10 Sustainability Rating Tools and their Informed Application

An introduction to the sustainability rating tools was provided earlier. These were developed for guiding the profession in the sustainability direction while planning and evaluating infrastructure projects. The well-known tools are the CEEQUAL scheme, the

Greenroads, and ENVISION (Griffiths et al., 2019). In North America, ENVISION is widely accepted by all levels of government (Black and Veatch, 2017) and there is a growing list of countries around the world where it is used.

ENVISION was developed as a result of a collaboration between the Zofnass Program for Sustainable Infrastructure at Harvard University Graduate School of Design and the Institute for Sustainable Infrastructure (ISI) located in Washington, D.C. It was developed to serve as a planning and design guide for increasing the sustainability attributes of infrastructure projects. In the latest ENVISION V3, a total of 64 sustainability criteria (termed credits) are defined to address the following five impact areas (Institute For Sustainable Infrastructure ISI, 2018):

- Quality of life: 14 credits
- Leadership: 12 credits
- Resource allocation: 14 credits
- Natural world: 14 credits
- Climate and resilience: 10 credits

ENVISION focuses on the infrastructure's sustainability attributes (i.e., balance of social, economic, and environmental aspects of the project) and actively promotes risk analysis and resilience. Since it focuses on the planning and design phases of a project and also can be useful in the construction phase, it serves as a useful reference for planning and guiding designs so as to improve sustainability attributes.

According to the Institute For Sustainable Infrastructure ISI (2018), the benefits of using ENVISION in infrastructure include the following:

- Lower cost
- Long term viability
- Potential to save owners money over time
- Few negative impacts on the community

Due to ENVISION's promotion of resilience and risk, it encourages consideration of the longer-term implication of infrastructure use and its use of resources on a long-term basis. Therefore, it guides infrastructure owners in longer-term thinking (Griffiths et al., 2019).

In order to benefit from the use of this innovative infrastructure sustainability rating tool, informed use of ENVISION is critical (Griffiths et al., 2019). Of course, the same applies to other rating tools noted above. That is, there is a need to develop and illustrate methods that can help practicing professionals to prepare inputs to ENVISION for informed applications. The methods developed in this research have the potential to address this knowledge gap.

# Chapter 3: Research Methodology and Data

## 3.1 Overview

This chapter presents the research methodology used to collect the data and address the research questions. The chapter begins with an important section, which contains a description of the research method followed by a description of the different stages of the research process (see Figure 3.1). This chapter presents and justifies the method adopted to implement and validate the proposed framework strategy for developing models to quantify and address life cycle cost overrun risks in highway infrastructure by taking into account sustainability factors. Analytical models, including the probability model of cost overruns, decision-theoretic model, utility-theoretic model, and Bayesian analysis model for evaluation of investment alternatives. Besides, factor analysis and associated logistic regression models were implemented for characterizing the effect of factors that cause cost overruns based on an international questionnaire survey.

## 3.2 Research Method

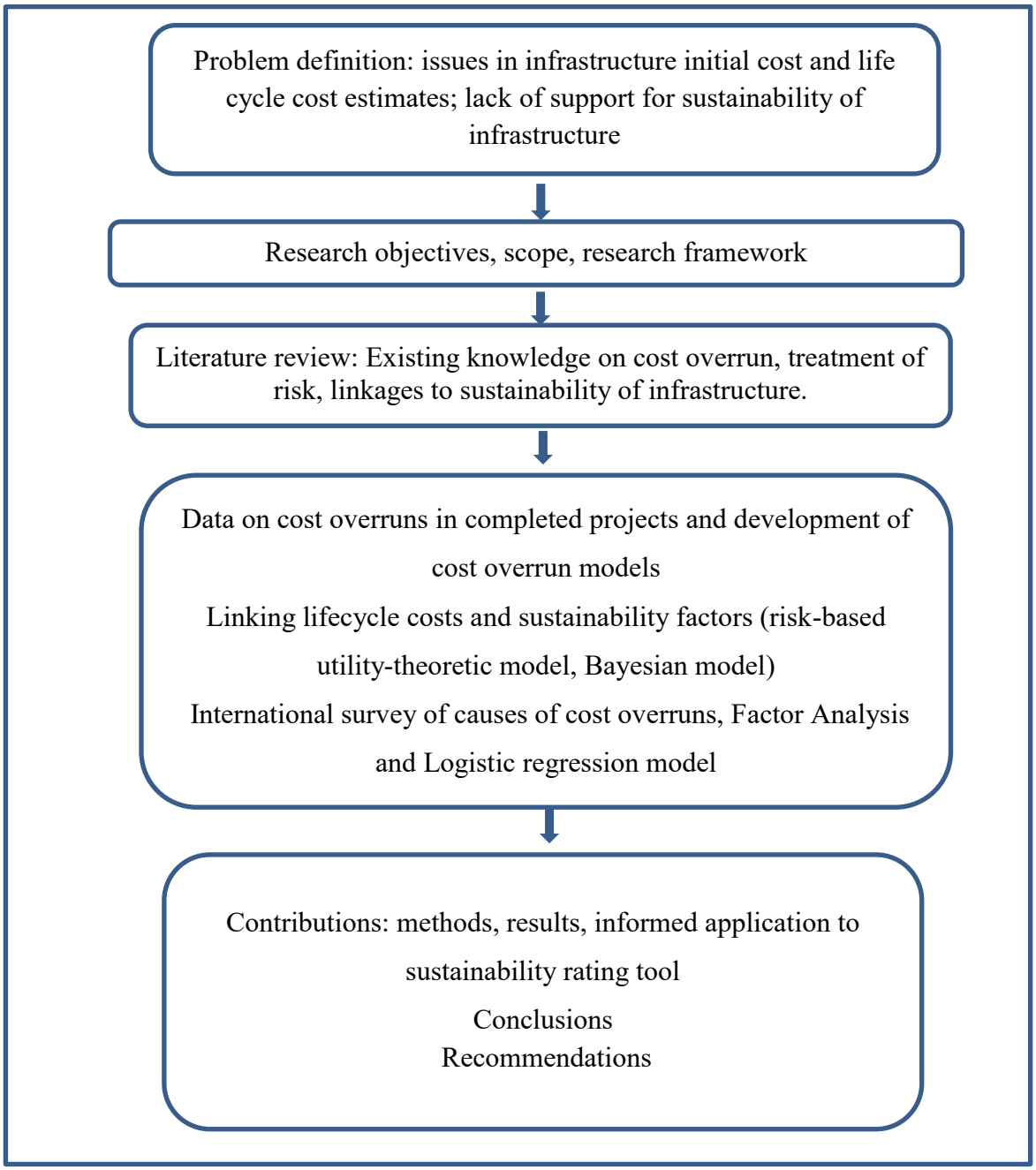
The research framework presented in Figure 3.1 reflects the philosophical foundation of this thesis research. It includes methods for treating risk in life cycle costs of infrastructure and modelling the sustainability factors within a risk framework. The development of the research framework was motivated by the need to advance risk models and to provide means for informed application of sustainability rating tools.

The starting point is the problem definition. As noted in Chapter 1 (in Problem Definition section), the following are the dimensions of the problem to be solved:

- Issues in infrastructure initial cost as well as life cycle cost estimates due to lack of formal treatment of risk.
- Even if an attempt is made to study informally cost overrun in the first cost and to provide a contingency fund (as is commonly done), these actions cannot be regarded as a scientific approach to address the risk problem. Also, there is not much evidence that rehabilitation costs are modelled at the planning stage.
- Necessary support for the sustainability of infrastructure that calls for the treatment of cost and other factors of sustainability.
- Need for a scientific study that can identify the factors that cause cost-overruns and the availability of models that can predict the effect of cost overrun variables.
- Need for a method that can potentially improve the inputs to a sustainability rating tool such as ENVISION.

Following the problem definition step, research objectives and scope were defined (see Chapter 1). As a follow-up, the research framework was put together as illustrated in Figure 3.1 and used as the road map for the entire research study.

***Figure 3. 1: Research Framework***




A detailed search of the literature was carried out to find out available knowledge on all facets of this research, including cost overrun in infrastructure projects. Following the literature review, the following tasks were completed:

- For improving knowledge of cost overruns in completed highway infrastructure projects, data were acquired from the province of British Columbia and models were developed that can be used for predicting cost overruns in highway projects. The purpose of developing the cost overrun model is to select the optimal probability distribution that best fits the British Columbia data and estimate the probability of cost overrun.
- Highway lifecycle cost variables were modelled using a decision-theoretic method. To go beyond the treatment of cost factors, a risk-based utility model was developed that enables the inclusion of environmental, social, and resiliency factors in analyzing risk in highway infrastructure.
- An advanced treatment of life cycle risk was carried out using the Bayesian model which enables the study of the value of additional information in reducing risk.
- The utility-theoretic and Bayesian decision-theoretic methods were used to show how to account for risk in highway costs and incorporate sustainability factors in decision-making. The application of models to the planning and evaluation roles of a sustainability rating tool was illustrated. Included were the first cost overruns as well as life cycle cost models.
- In order to address the knowledge gap of the role of factors that cause cost overruns, a survey of transportation agencies in Canada, the USA, the Middle East, and Australia was carried out on their experience in cost overruns and reasons for such overruns. The survey data were used for Factor Analysis of variables and the development of logistic regression models that can explain cost overruns. Conceptually, when we consider the use of Neural Nets as a predictive tool, but the

logistic regression model is better suited because it has the ability to explain the effect of the underlying factors of the cost overruns.

- In the last part of the research framework, contributions of the thesis research were defined (Figure 3.1). These include methods developed and their results for use by other researchers and persons engaged in professional practice. To highlight a notable contribution to the field, an explanation is provided on how the findings of this research can potentially improve the informed application of a sustainability rating tool such as ENVISION.
- In addition to defining contributions of this research, notable conclusions covering all parts of this research are highlighted and recommendations are advanced for improving the field of risk analysis in highway infrastructure investments.

Given the above-described challenge, this research study was structured in well-defined stages in order to progress from problem definition to results that include findings from data analyses as well as methods that can be applied by other researchers and professional persons with the responsibility to evaluate projects and make recommendations to decision-makers.

## 3.3 Data Requirements

The literature study provided evidence that cost overruns do occur in highway infrastructure projects. However, as expected, most cited references covered examples of very high-cost overruns. Also, a detailed database was not available that could be used for modelling purposes. Given that this research is intended to advance knowledge on the

extent of cost overruns as well as to develop risk models for further use, it was necessary to seek assistance from a Canadian jurisdiction with a well-developed information system.

A report published by Berechman and Wu (2006) provided an insight on locating data for cost overruns. Since the University of British Columbia (UBC) study used a large well-organized database from Vancouver Island projects, it was decided to request the Province of British Columbia (B.C.) to enable us to further advance the state of knowledge. The Government of B.C. (Transportation and Infrastructure Department) accepted our request and provided data for use in this research in developing models of cost overruns (Table 3.1).

The data received from B.C. include the extent of cost and schedule overruns and also reasons for such overruns for a few major projects. Following the scope of this research, these data were used for developing cost overrun risk models, and schedule overrun information was viewed with interest, but not modelled. For modelling life cycle costs including decision-theoretic, utility-theoretic, and Bayesian models, archived data were sourced from the US Federal Highway Administration (FHWA) (1998), analyses on cost overruns reported in this thesis, and from the Ontario Hot Mix Producers Association (OHMPA) (1998) report which was written for the Ministry of Transportation, Ontario.

*Table 3. 1: Highway Infrastructure Data Used for Risk Model Development*

| **Data Obtained from B.C.** | |
|---|---|
| Project ID | ✓ |
| Project location | ✓ |
| Project type | ✓ |
| Road/Highway; Bridge; Tunnel | There are no tunnel projects on the list |
| Ownership | ✓ |
| Public (Federal); provincial; municipal | ✓ |
| Investment make-up | ✓ |
| Private | ✓ |
| Public–Private-Partnership (PPP) | ✓ |
| Public | ✓ |
| Year started | ✓ |
| Year completed | ✓ |
| Delay (years) in completing the project | ✓ |
| Reason(s) for delay | N/A |
| Budget (for implementation) & actual cost | ✓ |
| At the planning stage | ✓ |
| Just prior to implementation | ✓ |
| Actual cost | ✓ |
| Reasons for cost overruns* | Available for a few major projects |

* The reasons for cost overruns analyzed in this thesis research were obtained from an international questionnaire study.

Data required for the Factor Analysis and associated Logistic Regression Model are obtained from a new questionnaire type of survey. The questionnaire and analyses have the approval of the Carleton University Research Ethics Board. Highway transportation agencies in Canada, the USA, and the Middle East contributed sufficient completed returns that enabled analyses and modelling tasks reported in this thesis document. Australia also contributed information of value to this research.

## 3.4 Relationship of This Research to Sustainability Rating Tools

The application of a rating tool such as ENVISION plays a useful role in the perspective of understanding the requirements that are considered necessary. For example, although ENVISION has identified many design and evaluation criteria (64 in total), the user of this tool can benefit from the availability of detailed methods for preparing inputs to the sustainability rating tool. The sustainability enhancement of a project can be successful through the study of inputs before using these in the rating tool. In the case of risks in the project's first cost as well as other costs covering the entire lifecycle, it was found at the early stage of this research that there is a lack of guides and published sources of information on how to characterize these.

As noted earlier, the ENVISION V3 contains 64 sustainability criteria (also called credits). These contribute to five credit categories and 14 sub-categories (i.e., categories of achievements) as shown in Figure 3.2. The achievement categories suggest that this rating system is intended to go beyond environmental sustainability by including community, economic, and financial considerations.

***Figure 3. 2: ENVISION's 64 Sustainability Criteria***


Adapted from Institute For Sustainable Infrastructure ISI (2018).

This thesis research can potentially enhance ENVISION applications in a number of achievement areas (e.g., Quality of Life, Leadership, Climate and Resilience). Table 3.2 provides application ideas in these areas. Knowledge and tools developed in this research can potentially be used by planners, engineers, and analysts so that their projects can be guided during the planning and design stage in order to receive higher ENVISION rating at the evaluation stage.

***Table 3. 2: Linkages Between ENVISION and Thesis Research***

| Action area | Need for advanced risk analysis methods (Selected items shown here) |
|---|---|
| **Quality of life (Community)** | • How to include equity & social justice in investment evaluation? |
| **Leadership (Planning & Economy)** | • How to account for risk in conducting a lifecycle-based economic evaluation?<br>• How to study the economic factors in association with other factors of sustainability? |
| **Climate and Resilience (Emissions)** | • How to include reduction of greenhouse gas emissions in investment evaluation? |
| **Resilience** | • How to include infrastructure integration in the evaluation of a project? |

# Chapter 4: Cost Overrun Risk Analysis Models

## 4.1 Introduction

As noted in the literature review chapter, transportation infrastructure projects experience cost overruns that lead to substantial financial issues. For example, according to Kaming et al. (2012), cost overrun is too common in infrastructure construction projects, including highways. Most of the reviewed literature described some causes of cost overruns, though few of them mentioned how to estimate risks associated with overruns and how to incorporate risk estimates into project assessment.

A review of a research report published by Berechman and Wu (2006), provides an insight into the analytical aspect of risk analysis using a large database from Vancouver Island projects. In order to further advance the state of knowledge, a request was made to the Government of British Columbia (Transportation and Infrastructure Department) to contribute data. The received B.C. data consisting of about ninety-seven completed highway infrastructure projects were analyzed and models of cost overruns were developed as explained below.

In accordance with the objectives of this research, the models are intended for estimating cost overrun risk so that these can contribute information for project evaluation and decision making. These probability distribution fitting models can predict the probability of cost overrun risks.

## 4.2 Assessing the Risk of Cost Overrun

"Cost overrun" can be defined as the difference between the actual cost and the planned/estimated budget. The actual cost is the total funds that the spending agency has paid for the construction. Whereas the estimated budget is the money assigned to the project before its commencement. There is another estimate that is produced before initiating construction.

In this analysis, we began by defining Cost Overrun Ratio (COR) which is the ratio of actual costs to estimated budget. To illustrate further, if the actual cost is (2.5$) billion and the budget for a project is (2.0$) billion, then COR = 1.25; if the actual cost is (0.5$) billion and the planned budget is (1.0$) billion then COR = 0.5 (<1), which means no cost overrun. However, we define the cost overrun risk in three different levels as indicated below.

$S_1$ = Low risk of cost overrun (COR $\leq$ 1) (i.e., no risk)

$S_2$ = Medium risk of cost overrun (1 $<$ COR $\leq$ 1.2) and

$S_3$ = High risk of cost overrun (COR $>$ 1.2).

If the COR exceeds one, the project has a cost overrun issue, otherwise, project costs are controlled. The identified COR ranges were estimated based on data received from B.C. There is no standard practice in setting COR ranges; therefore, in this research decision was made by the researcher.

Most project evaluators have an interest in the expected COR for a new project and the probability of COR exceeding a specific value. The approach that project evaluators use to predict costs and budgets for a new project is by applying their knowledge and experience from previous similar projects. In spite of this, knowledge and experience

cannot be quantified properly since they are subjective. As a result, other methods should be used such as well-found statistical models to have reliable information as well as accurate project costs estimates.

To improve the practice of cost and budget setting, this research identified a probability distribution model that best explains the distribution of a specified COR. This was made possible by developing models that quantify the probability of cost overrun ratio exceeding a specific value (e.g., COR > 1.2), which was defined as the project's risk level.

In this research, the best fit analysis was conducted on two datasets, namely Cost Overrun Ratio (COR) at the planning phase and the second dataset was the Cost Overrun Ratio (COR) prior to construction phase.

It was noted that the budget at the planning phase was the approved budget at the end of the planning phase and the beginning of the design step. The budget prior to construction (estimated cost) was the approved budget to start construction. The budgets and costs are in dollar values.

The data analyzed consists of 97 completed infrastructure projects each costing over one million dollars. The projects consist of roads, highways, and bridges. Table 4.1 shows aggregate cost overrun ratio data ranges for both types of data (i.e., planning phase data & prior to construction phase data). As can be seen in the table, of the 97 infrastructure projects, 37 projects have experienced cost overruns (COR > 1) at planning phase data and 33 projects have experienced cost overruns (COR > 1) prior to construction phase data.

*Table 4. 1: Cost Overrun Ratio Ranges of B.C. Data*

| Data | Number of projects based on COR | | | COR | |
|---|---|---|---|---|---|
| | COR≤1 | 1<COR≤ 1.2 | COR>1.2 | Min | Max |
| **Planning Phase** | 60 | 24 | 13 | 0.53 | 2.16 |
| **Prior to Construction Phase** | 64 | 29 | 4 | 0.64 | 1.80 |

## 4.3 Probability Distributions Fitting Models of Cost Overrun Risk

Distribution fitting is an analytical method that uses empirical data to find the best fit probability distribution model. In many statistical analysis tools, distribution fitting functions are available for various distribution types, including Normal, Gamma, Log-Logistic, and others. For statistical analysis in this research, MATLAB software was used to find the best fit model. The general steps of a distribution fitting method are:

1. Using valid data, plot the histogram (e.g., cost overrun ratio).
2. Based on the histogram, determine the distribution type.
3. Use distribution fitting functions to estimate the relevant parameters.

Since the COR data is continuous, the candidate distributions must all be continuous and thus to be well suited to model the cost overrun. These distribution candidates include:

- Log-Logistic Distribution

  A log Logistic distribution has been widely used in many areas such as reliability, economics, and engineering. The distribution shape of the log-logistic can be defined by the location parameter and scale parameter.

- Cauchy Distribution

Cauchy distribution is a continuous probability distribution that consists of three parameters (location, scale, and shape). The location parameter reflects the mean of the distribution and the scale parameter results in a shorter or taller graph. The smaller the scale parameter, the taller and thinner the curve (Krishnamoorthy, 2016).

- Log-Normal Distribution

  A log-normal distribution is one of the common distributions and consists of two parameters. The shape of the log-normal distribution is defined by the location parameter and the scale parameter, the scale parameter is the standard deviation for the log-normal.

- Gamma Distribution

  The gamma distribution is another widely used distribution that has two parameters (shape and scale). In application, the shape parameter is labeled as ($\gamma$) and the scale parameter is labeled as ($\beta$). The Gamma distribution does not have a location parameter (Krishnamoorthy, 2016).

- Normal Distribution

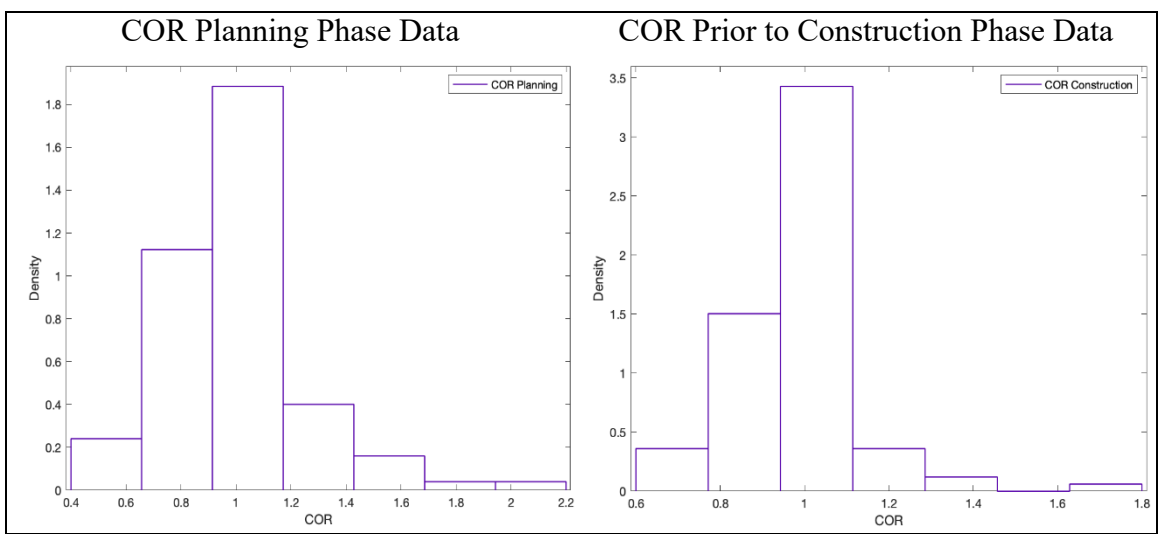
  The normal distribution is the most commonly used distribution to model data from an experiment or a population. The mean ($\mu$) is the location parameter and the standard deviation ($\sigma$) is the scale parameter. These two parameters completely define the normal distribution. The log-logistic, log-normal and normal distributions do not have a shape parameter. The location and scale parameters can be used to define the shape of these three distributions (Krishnamoorthy, 2016).

The purpose of including in the analysis multiple types of distributions is to compare these distributions and to find the distribution that best fits the B.C. data. It is important to have a distribution that accurately reflects the available data.

## 4.4 Cost Overrun Ratio (COR) Histogram

The histograms of the empirical data are given below (in Figure 4.1) for COR at the planning phase and COR prior to construction phase. The histograms are noted to be not symmetrical and appeared to be slightly skewed to the right. Based on this observation, it was believed to be very unlikely that the sample data could come from a normal distribution "bell curve". It was also apparent from these histograms that the cost overrun ratio during the planning phase and prior to construction phase was more likely to be low or medium as compared to a high level. To study further, five theoretical distributions were investigated to find the best obtainable fit.

***Figure 4. 1: Cost Overrun Ratio Histogram for B.C. Data***



## 4.5 Budget at Planning Phase Data

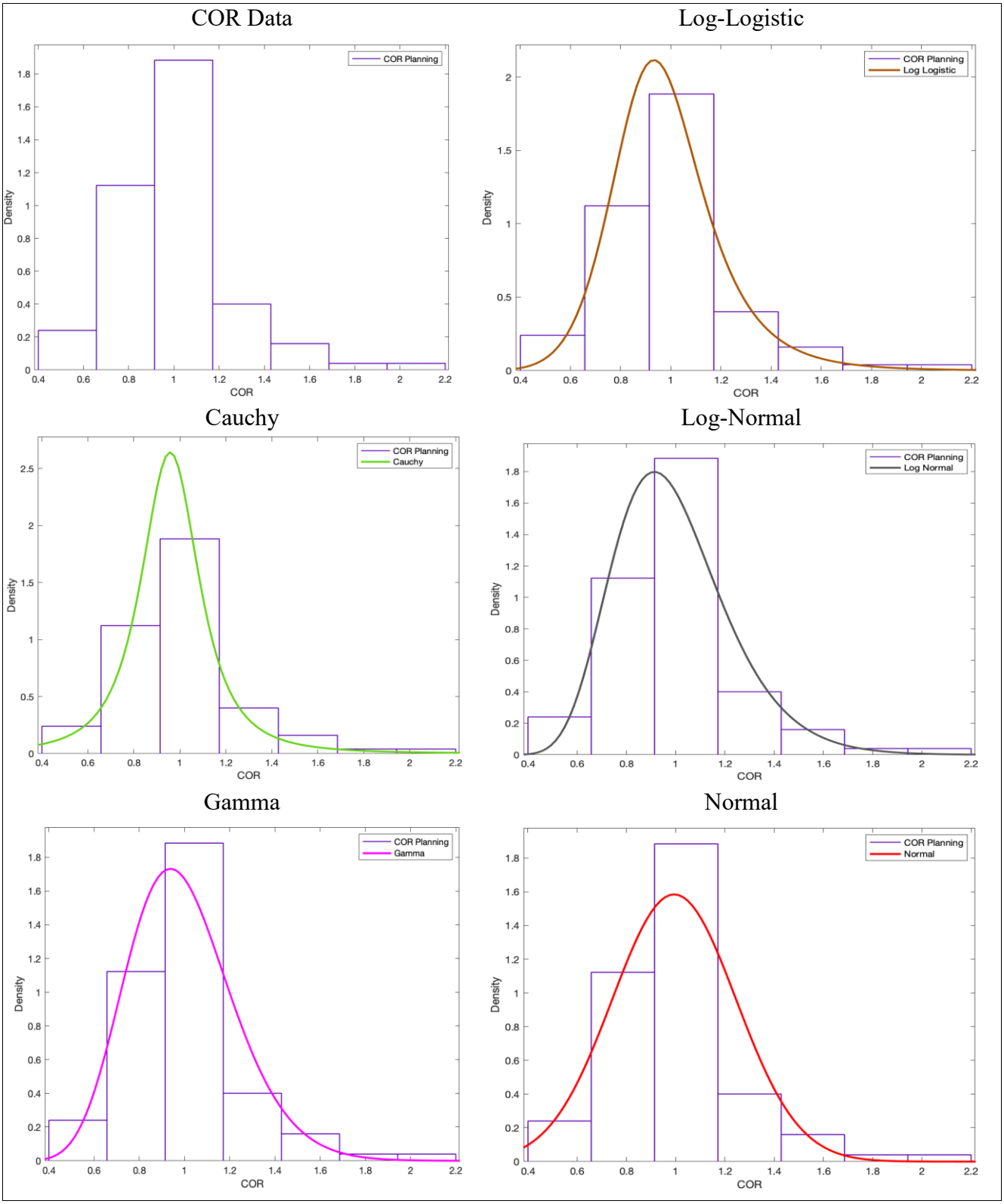
### 4.5.1 Fitting Cost Overrun Ratio Distribution Function

Several statistical techniques can be used to estimate the distribution's parameters from a set of data. The MATLAB software uses the maximum likelihood estimation (MLE) technique and this was used to estimate the parameters for the five candidate distributions as shown in Table 4.2. The candidate distributions are Log-Logistic, Cauchy, Log-Normal, Gamma, and Normal. From the estimated parameters, the distribution plots for the five candidate distributions were performed as shown in Figure 4.2. In the figure, the horizontal axes show the COR, and the vertical axes illustrate the frequency density. Just remember that the density is proportional to the chance that any value in your data is approximately equal to that value. In fact, for a frequency histogram, the density is calculated from the counts, so the only difference between a histogram with frequencies and one with densities, is the scale of the y-axis.

***Table 4. 2: Distribution Parameters (Budget at Planning Phase Data)***

| **Parameters** | **Observed** | **Distribution** | | | | |
|---|---|---|---|---|---|---|
| | | **Log-Logistic** | **Cauchy** | **Log-Normal** | **Gamma** | **Normal** |
| **Mean** | 0.994 | 0.987 | 0.957 | 0.993 | 0.993 | 0.994 |
| **Std Dev** | 0.252 | 0.230 | 0.309 | 0.237 | 0.235 | 0.252 |
| **Parameters** | | Location -0.038 Scale 0.124 | Location 0.957 Scale 0.134 Shape 2.172 | Location -0.034 Scale 0.236 | Shape 17.74 Scale 0.056 | Location 0.994 Scale 0.252 |

***Figure 4. 2: Histogram and Distribution Plots (Budget at Planning Phase Data)***



To determine which candidate distribution best fits empirical data, the log-likelihood (LL) value of each model was estimated, then, the Kolmogorov-Smirnov (KS), and Chi-square goodness of fit tests were conducted, as portrayed in Table 4.3.

When using the Maximum Likelihood Estimation (MLE) technique to estimate the parameters of the distribution model, the likelihood value can be used to measure the fit of the distribution to the dataset. Hence, as the observed distribution (empirical histogram) becomes similar to the model distribution (such as Gamma, Normal, or Log-normal), the log-likelihood increases or becomes higher. However, the distribution with the largest log-likelihood value is the best fit statistically (Montgomery and Runger, 2010).

Chi-square and Kolmogorov–Smirnov (KS) goodness of fit tests also can be used to measure the quality of the distribution. They are used to compare the significant difference between the observed distribution (empirical histogram) and the expected probability distribution (such as Gamma, Normal, or Log-normal). Chi-square and Kolmogorov–Smirnov (KS) goodness of fit tests determine how well theoretical distribution fits the empirical histogram.

Besides, the selection of the optimal probability distribution is also based on the p-value of the Chi-square and Kolmogorov–Smirnov tests. The p-value is the probability that the empirical data deviates from the distribution model by chance alone. As the p-value gets smaller than or equal to the level of significance (0.01), it becomes unlikely that the distribution fits the available data (Montgomery and Runger, 2010). We should then expect the p-value to be relatively high when the histogram is very close or similar in shape to the model. In applications, the most commonly used levels of significance are (5% and 1%). In this thesis research (1%) was used.

Table 4.3 shows the goodness of fit results for the five candidate distributions. The table shows that the log-likelihood value for the log-logistic distribution is the highest compared with the other distributions. In the table it is clearly shown that the LL (Log-logistic) > LL (Cauchy) > LL (Log-normal) > LL (Gamma) > LL (Normal).

Along with this observation, the results further show that the p-values for the Log-logistic distribution are significantly higher than that for the other four distributions. Therefore, we conclude that the Log-logistic distribution best fits the cost overrun ratio as compared to the other four distributions.

***Table 4. 3: Goodness of Fit Statistics (Budget at Planning Phase Data)***

| **Fitting Statistics** | | **Distribution** | | | | |
|---|---|---|---|---|---|---|
| | | **Log-Logistic** | **Cauchy** | **Log-Normal** | **Gamma** | **Normal** |
| **Log-likelihood** | | 9.3263 | 8.4752 | 6.2166 | 4.3234 | 3.3374 |
| **Kolmogorov-Smirnov** | P-Value | 0.034 | <0.01 | <0.01 | <0.01 | 0 |
| **Chi-Square** | P-Value | 0.023 | <0.01 | <0.01 | <0.01 | 0 |

Next, the probabilities of various levels of cost overrun ratio under each of the five distributions were estimated. Table 4.4 and Figure 4.3 show the estimated and observed probabilities for different cost overrun risk levels (S1, S2, and S3) under the five probability distributions (Log-logistic, Cauchy, Log-normal, Gamma, and Normal). The three risk levels are:

$S_1$ = Low risk of cost overrun (COR $\leq$ 1) (i.e., no risk)

$S_2$ = Medium risk of cost overrun (1 $<$ COR $\leq$ 1.2), and

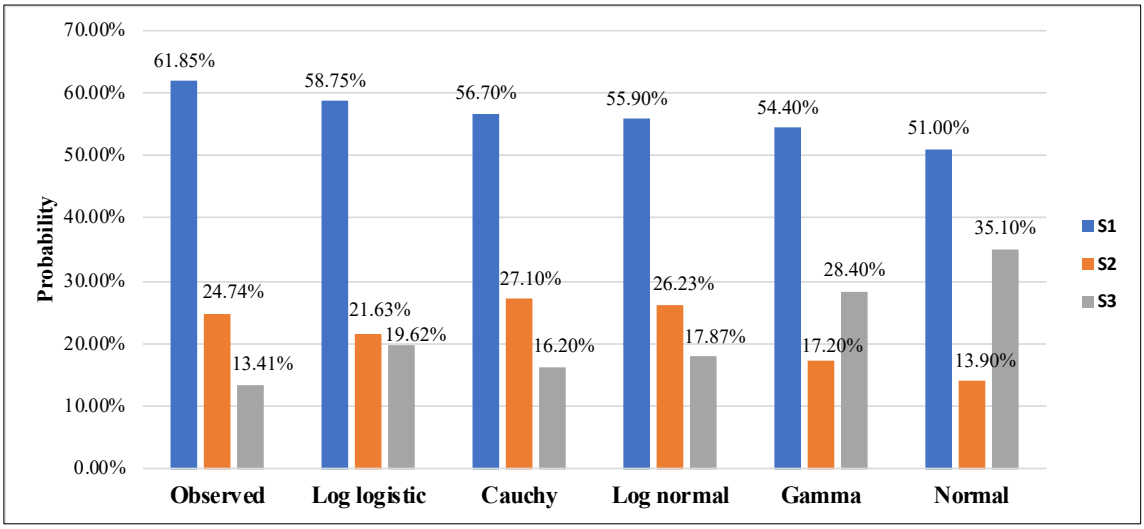
$S_3$ = High risk of cost overrun (COR $>$ 1.2).

Table 4.4 shows that according to the Log Logistic model (LLM) prediction, there is about 58.75% chance that the COR will be small, 21.63% chance that the COR will be medium, and about 19.62% chance that the COR will be high. Thus, according to the (LLM), it is more likely to expect a low cost overrun. Since the (LLM) fits the data better than any other model, we only consider the results of the (LLM) as it reflects the available data better than other models.

***Table 4. 4: Cost Overrun Ratio Probability (Budget at Planning Phase)***

| COR | Observed | Log-logistic | Cauchy | Log-normal | Gamma | Normal |
|---|---|---|---|---|---|---|
| **COR ≤ 1** | 61.85% | 58.75% | 56.70% | 55.9% | 54.40% | 51.0% |
| **1 < COR ≤ 1.2** | 24.74% | 21.63% | 27.1% | 26.23% | 17.20% | 13.90% |
| **COR > 1.2** | 13.41% | 19.62% | 16.2% | 17.87% | 28.40% | 35.1% |
| **Total** | 100 | 100 | 100 | 100 | 100 | 100 |

***Figure 4. 3: Cost Overrun Ratio Probability (Budget at Planning Phase)***



## 4.6 Budget Prior to Construction Phase Data

### 4.6.1 Fitting Cost Overrun Ratio Distribution Function

In this section, the results of the second part of B.C. data are presented and discussed. The Maximum Likelihood Estimation (MLE) technique was used to estimate the parameters for five candidate distributions (Cauchy, Log-logistic, Log-normal, Gamma, and Normal). The results are shown in Table 4.5.

***Table 4. 5: Distribution Parameters (Budget Prior to Construction Phase Data)***

| **Parameters** | **Observed** | **Distribution** | | | | |
|---|---|---|---|---|---|---|
| | | **Cauchy** | **Log-Logistic** | **Log-Normal** | **Gamma** | **Normal** |
| **Mean** | 0.971 | 0.970 | 0.972 | 0.971 | 0.971 | 0.971 |
| **Std Dev** | 0.146 | 0.185 | 0.122 | 0.141 | 0.140 | 0.146 |
| **Parameters** | | Location 0.970 Scale 0.058 Shape 1.638 | Location -0.035 Scale 0.068 | Location -0.039 Scale 0.144 | Shape 47.838 Scale 0.020 | Location 0.971 Scale 0.146 |

From the estimated parameters, the distribution diagrams for the five candidate distributions were developed. These are illustrated in Figure 4.4.

*Figure 4. 4: Histogram and Distribution Plots (Budget Prior to Construction Phase Data)*

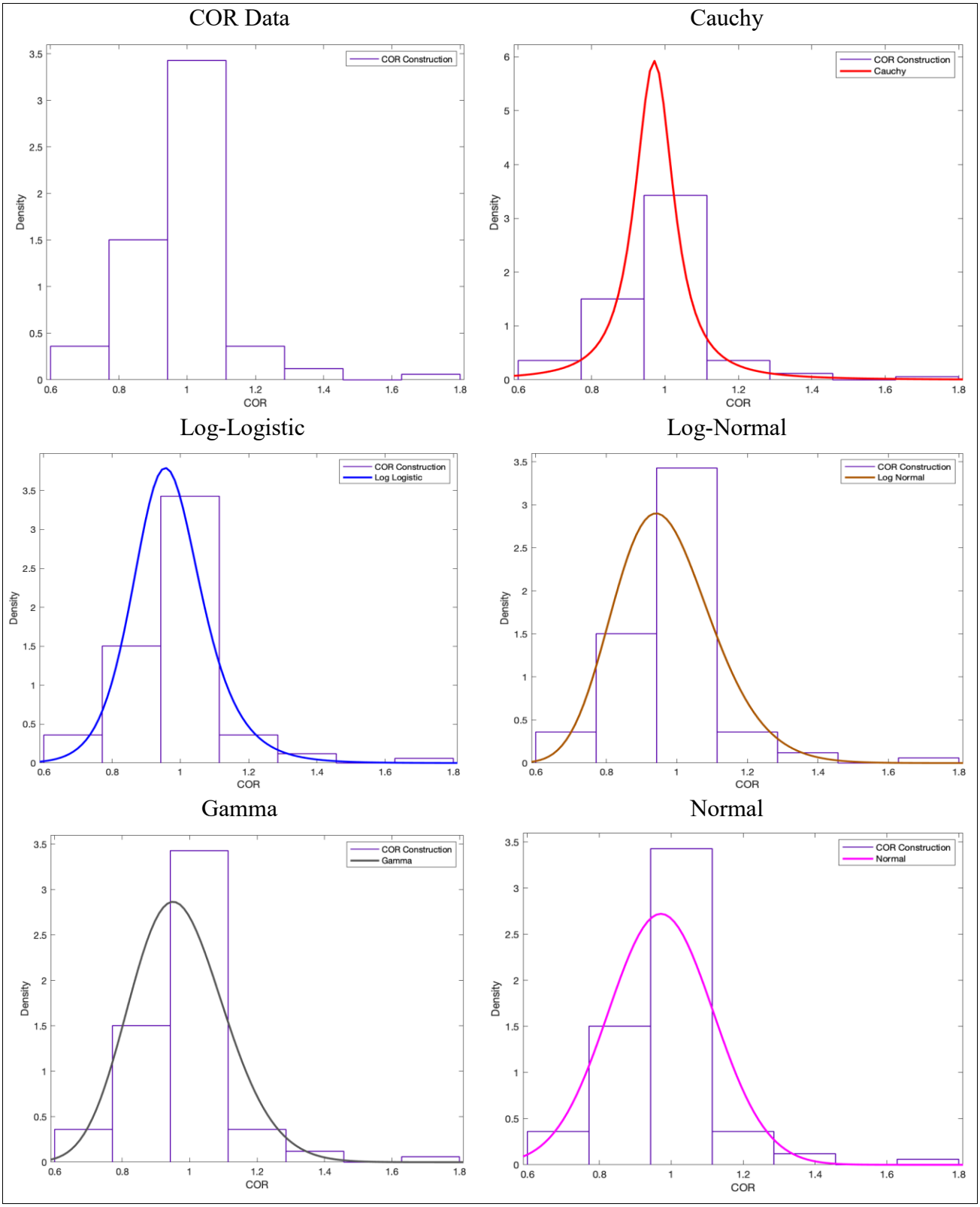


Table 4.6 summarizes the goodness of fit test results. Three best-fit tests were conducted to assess the quality of the fitted distributions, namely log-likelihood (LL), Kolmogorov–Smirnov (KS) test, and Chi-square test. It is noted that the p-value becomes smaller as the distribution shifts from the Cauchy model to the Normal model. It is thus unlikely that the normal distribution fits the available data. However, based on the log-likelihood and p-values, the Cauchy distribution followed by the Log Logistic appears to fit the data better than other distributions. It is obviously evident from these results that LL (Cauchy) > LL (Log-logistic) > LL (Log-normal) > LL (Gamma) > LL (Normal). Therefore, we conclude that the Cauchy distribution best fits the cost overrun ratio and is therefore better than the other four distributions.

***Table 4. 6: Goodness of Fit Statistics (Budget Prior to Construction Phase Data)***

| **Fitting Statistics** | | **Distribution** | | | | |
|---|---|---|---|---|---|---|
| | | **Cauchy** | **Log-Logistic** | **Log-Normal** | **Gamma** | **Normal** |
| **Log-likelihood** | | 73.5438 | 64.5232 | 54.7627 | 53.8181 | 49.3246 |
| **Kolmogorov-Smirnov** | P-Value | 0.076 | <0.01 | <0.01 | <0.01 | <0.005 |
| **Chi-Square** | P-Value | 0.059 | <0.01 | <0.01 | <0.01 | <0.005 |

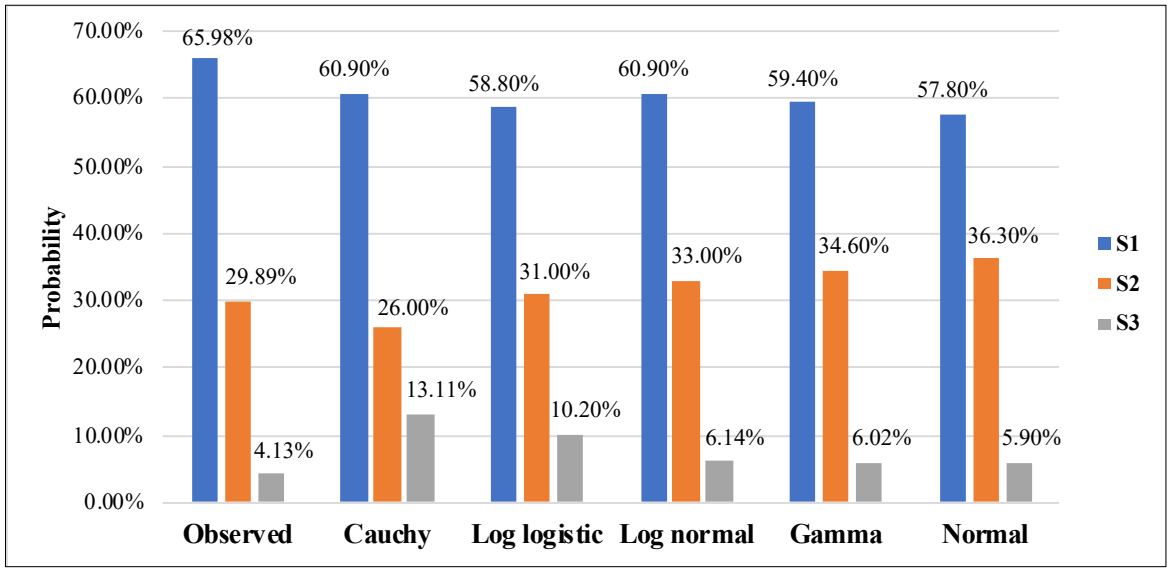
Next, the probabilities of various levels of cost overrun ratio under each of the five distributions were estimated. Table 4.7 and Figure 4.5 show the estimated and observed probabilities for different cost overrun risk levels (S1, S2, and S3) under the five probability distributions (Cauchy, Log-logistic, Log-normal, Gamma, and Normal).

*Table 4. 7: Cost Overrun Ratio Probability (Budget Prior to Construction Phase Data)*

| COR | Observed | Cauchy | Log-logistic | Log-normal | Gamma | Normal |
|---|---|---|---|---|---|---|
| **COR ≤ 1** | 65.98% | 60.9% | 58.8% | 60.9% | 59.4% | 57.8% |
| **1 < COR ≤ 1.2** | 29.89% | 26.0% | 31.0% | 33.0% | 34.6% | 36.3% |
| **COR > 1.2** | 4.13% | 13.11% | 10.2% | 6.14% | 6.02% | 5.9% |
| **Total** | 100 | 100 | 100 | 100 | 100 | 100 |

The Cauchy model (CM) predicts that, for the prior to construction phase data, there is about 60.9% chance that the COR (cost overrun) will be small, 26% chance that the COR will be medium, and about 13.11% chance that the COR will be high. So, it is more likely to expect a small (or low) cost overrun during prior to construction phase of the project. Since the (CM) fits the data better than any other model, we only consider the results of the (CM) as it reflects better the available data than other models.

*Figure 4. 5: Cost Overrun Ratio Probability (Budget Prior to Construction Phase Data)*



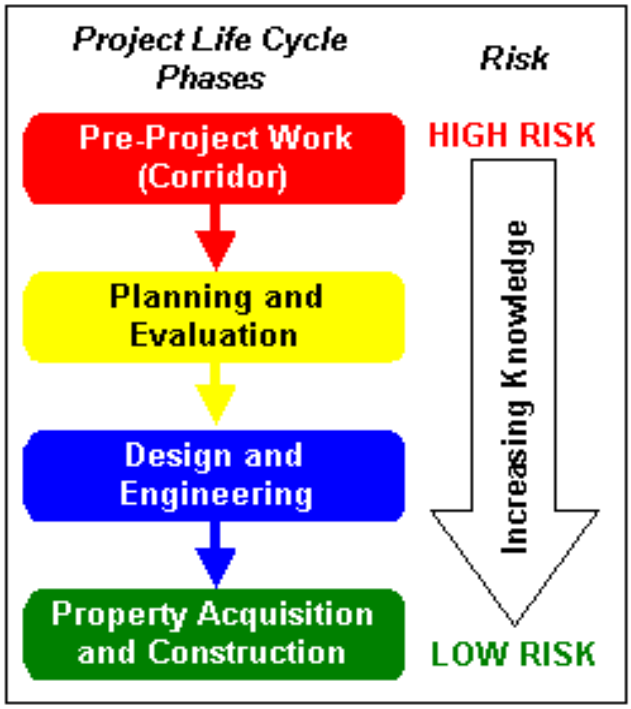
## 4.7 Relative Risk Levels at Planning Phase vs. Pre-Construction Phase

Due to lack of information on sources of uncertainties at the pre-project (corridor) level and even at the planning level, the risk in cost estimates is higher than estimates produced just prior to construction. Results of cost overrun risk models presented in Figures 4.3 and 4.5 confirm this observation. As demonstrated in our analysis, the budget at the construction phase tends to be more accurate than the budget at the planning phase due to having sufficient knowledge and information.

Likewise, Figure 4.6 sourced from the Ministry of Transportation and Infrastructure B.C. (2013) illustrates clearly that risk is higher at the initial phases of the project as compared to later phases. As knowledge increases during project life cycle phases, risk typically decreases. The budgets are always updated and refined during the project life cycle as information becomes available. The more the available knowledge the better the accuracy of the budget (Ministry of Transportation and Infrastructure B.C., 2013).

Another observation is that cost estimates made just prior to construction are not free of risk. Therefore, reliance on risk analysis and risk models for predicting risk is logical. The same observation applies to life cycle costs that include the construction cost. Additionally, project cost estimation must be carried out using the best available knowledge (i.e., guiding principles and the latest information) and by experienced personnel. Further, to cover any risks and uncertainties to the possible extent, the project cost estimate should always contain contingency. The base cost estimate is developed on the basis of known conditions and the assumption of certainty. The higher the uncertainty, the higher the fixed contingency. Likewise, schedule risks are handled by assigning a fixed contingency.

*Figure 4. 6: Project Life Cycle Phases*




Adapted from (Ministry of Transportation and Infrastructure B.C., 2013)

## 4.8 Discussion

Based on data analysis, the best probability models were identified. There could be a difference between the risk of cost overrun estimated prior to the construction phase and the risk that is estimated at the planning phase. The estimation of COR becomes more accurate when additional information becomes available. Moreover, risk analysis should be implemented on an on-going basis during the project life cycle, including planning, design, construction, and operation.

Five probability distributions were studied as candidates in the analysis. These are Log-logistic, Cauchy, Log-normal, Gamma, and Normal distributions. The maximum likelihood estimation (MLE) approach was used to estimate the parameter of the distributions. Techniques of checking how well these distributions match the available data

were also applied. The goodness of fit test includes the log-likelihood, Kolmogorov-Smirnov (KS) test and Chi-square. The higher the value of Log-likelihood the better is the fit of the model to the data. Also, a high p-value means that the assumption is correct, and the data does fit the distribution.

## 4.9 Conclusions

The purpose of developing the cost overrun model is to select the optimal probability distribution that best fits the British Columbia data and estimate the probability of cost overrun. The developed model can be used by any transportation agency around the world. In other words, the developed methodology can be used for highway infrastructure cost overruns experienced by provinces other than British Columbia as predictive tools.

However, the Log-logistic distribution was found to be the best based on "budget at the planning phase" data. The Log-likelihood value and p-value were the highest for this distribution as compared with the other four distributions. On the other hand, the Cauchy distribution was found to be the best for the "budget prior to the construction phase" data. That is, its Log-likelihood value and p-value were the highest compared with the other four distributions.

Since the range of the Cost Overrun Ratio (COR) in the dataset is from (0.53 - 2.16) and from (0.64 – 1.80), distributions other than the ones used are not suitable. For example, we are unable to use Beta distribution because it ranges from zero to one.

The "no-cost overruns" are not labelled as negative cost overruns in order not to confuse the reader. These are simply labelled as projects with a cost overrun ratio (COR) less than 1.

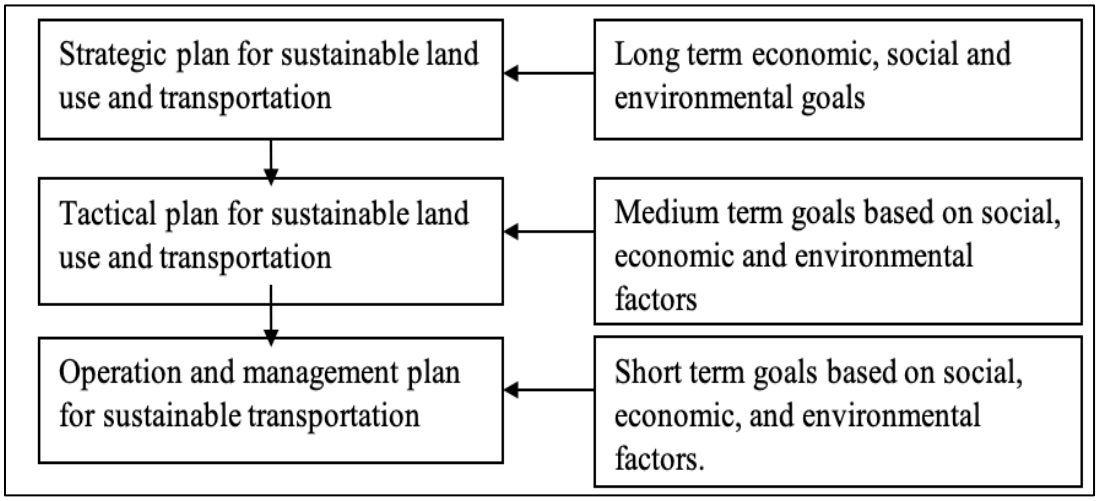
# Chapter 5: Role of Risk Analysis in Enhancing Parameters of Sustainability: Decision-Theoretic and Utility-Theoretic Models

## 5.1 The Link between Life Cycle Cost Risks and Sustainability

As noted in the previous chapters, decisions regarding transportation infrastructure must be guided by sustainability consideration. Since infrastructure changes have to be planned on a long-term basis and these serve as instruments for regional and national economic development, decisions are made as a part of the strategic planning process (Figure 5.1). Next, the planning process guided by the sustainability principles proceeds to the tactical and operational levels.

Within the framework provided by the strategic plan, a medium-term tactical plan is required to refine the transportation and development components and initiate infrastructure implementation. At the operational level, the transportation system has to be operated and managed according to the sustainability goals.

***Figure 5. 1: Levels of Planning***



At the strategic planning level, the systems approach has to be used to ensure that the economic development and transportation plans are shaped according to the values and goals structure. The systems approach to planning consists of problem definition, generation of alternatives, analysis of alternatives, evaluation of alternatives, and choice of the preferred alternative. The follow-up activities of detailed design, project, and program planning take place at a tactical level.

As a part of problem definition, values, goals, objectives, criteria, and standards are to be defined. The sustainability factors of economic, social, and environmental types are to be formally taken into account in every step of the systematic approach to infrastructure planning, design, construction, and rehabilitation.

Sustainability criteria are performance measures or indices of measurement capable of defining the degree to which the sustainability goals/objectives have been attained. These criteria/performance measures (or rates in the ENVISION language) are used for infrastructure design and evaluation. In the systems approach, all relevant criteria are included, and contradictory criteria can be sorted out. Also, feedback may be incorporated to improve the systems approach (e.g., from evaluation back to goals/objectives/criteria to clarify these). Although economic sustainability factors continue to be important, infrastructure investment decisions should take into account social and environmental factors.

These sustainability criteria can be grouped into three categories, namely, economic, social, and environmental factors. An examination of these factors leads to the following observations:

- First, several interest groups are involved.

- Second, many factors cannot be quantified or valued in the common metric of the dollar.
- Third, a number of these can be quantified only in subjective terms.
- Fourth, it is indeed a major challenge to develop methodology and generate data so that a large number of these factors can be quantified.
- Fifth, the conventional evaluation methodologies (e.g., benefit-cost method) are not capable of handling all criteria due to their highly diverse nature. As noted earlier, many cannot be quantified in dollar terms and there are other criteria that can be quantified only in subjective terms.

The state of the art in the economic evaluation of transportation systems has improved considerably over the years. The benefit-cost analysis methodology has already been refined for application to mutually exclusive alternatives of a given system, based on net present worth NPW (or equivalent annual worth EAW) calculations on total investment and their associated incremental approaches. However, the evaluation process needs to be better refined from a sustainability perspective. Specifically, in the context of sustainable transportation, there is a need for:

- Information on how to deal with impact-incidence (the equity – i.e., distributional issues)
- Knowledge on how to quantify or at least to express in qualitative terms, factors that do not enter in benefit-cost analysis,
- Information on the methodology for incorporating strategic economic, social, and environmental impacts in the evaluation of the mutually exclusive alternatives.

## 5.2 Building Blocks for the Quantification of Effectiveness

For a sustainable transportation system, the evaluation process requires the following information:

- An identification of all interest groups (stakeholders)
- A definition of their perceived needs and concerns
- A definition of the goal-objective-criteria hierarchy
- Results of analysis of alternatives to establish:
    - (i) Criteria achievement level, and
    - (ii) Capital and other costs for each alternative
- "Valuation" of the criteria achievement levels.

It is essential to incorporate relevant economic, social and environmental effects of transportation systems in the planning, design and evaluation tasks. Also, the perceived needs and concerns of all relevant interest groups should be considered. The evaluation of alternatives, including do-nothing (if applicable), has to be based on all criteria, in order to establish their relative value to society. Based on the evaluation results, the choice of an alternative is made for implementation.

Economic evaluation of mutually exclusive alternatives can be carried out in the form of a benefit-cost analysis. The best alternative can be identified based on net present worth (NPW) of total investment. Since numerous impacts cannot be valued in dollar terms, it becomes necessary to find a way to review the results of benefit versus cost in conjunction with other indicators.

For sustainable transportation systems, the following types of criteria have to be incorporated in the evaluation framework.

- Criteria that can be quantified in dollar terms (e.g., costs)
- Non-dollar type of criteria that can be quantified in their measurement units, but it is difficult to impute a dollar value (e.g., network effect of strategic importance, noise and emissions)
- Criteria that can be expressed in subjective ordinal terms (e.g., social impacts such as community cohesion)

The economic parameters of sustainability can be enhanced if risk is identified and is formally treated in cost-effectiveness of infrastructure investment and asset management decisions within the overall framework of working with all sustainability-related criteria as defined by the ENVISION scheme. This can be achieved by developing methods that can take into account risk in life cycle costs and integrate results with other criteria (rates) defined by ENVISION. In the following sections of this chapter, methods are discussed for modelling risk.

## 5.3 Decision-Theoretic and Utility-Theoretic Models

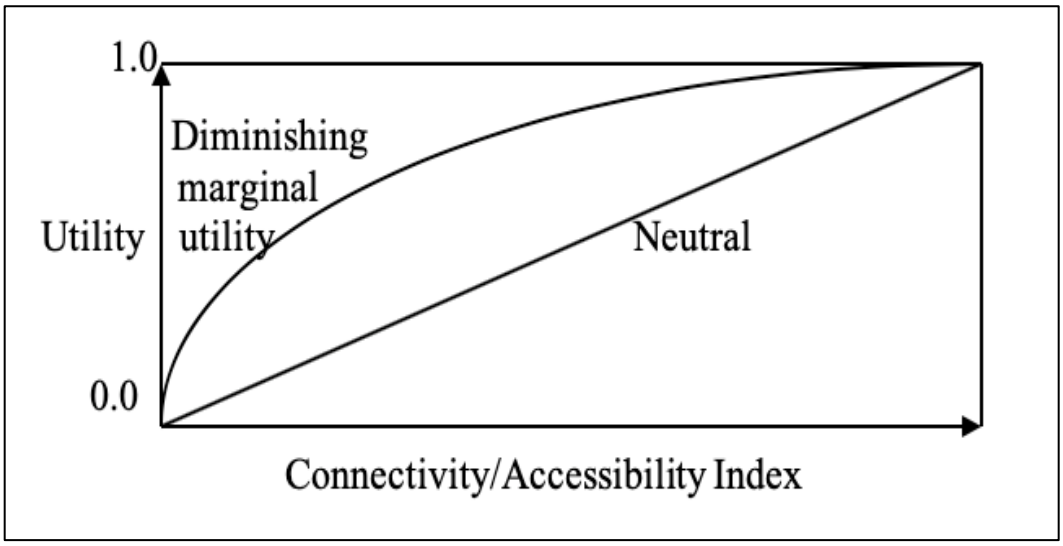
Decision-theoretic models were developed that treat cost levels as probabilistic and initially, the method works with the criteria that can be quantified in dollar terms. As a follow-up, the method is extended in the form of a comprehensive methodology which can treat uncertain factors and uses utility theory to establish an overall expected utility score for each alternative. This methodology is based on the principles of risk-based multi-attribute utility theory. It enables the treatment of uncertain cost overruns, multiple evaluation criteria, differential weighting of criteria, diminishing marginal utility property (if applicable), and risk averse/neutral/risk seeking viewpoint of the decision-maker.

A major challenge in the incorporation of the sustainable transportation criteria is the important task of quantifying their achievement levels (i.e., impacts) or expressing these in qualitative terms. Although there may be much temptation to assign effectiveness scores or to attribute qualitative statements subjectively based on general knowledge of the study area, there is much merit to initiating special studies for obtaining the necessary information that could be used as a basis for predicting criteria achievement levels.

Weights can be used as indicators of the relative importance of corresponding evaluation criteria. A well-known method in Operations Research, namely the Churchman-Ackoff method, can be used to convert criteria ranks to weights. On the other hand, criteria weights can be expressed as rates on a continuous scale of 1 to 5. A number of committee members can participate in assigning criteria weights and these can be combined into a single set of weights.

The level of achievement of a criterion, originally measured in its raw units (e.g., dollars) or, in some cases expressed subjectively (e.g., excellent, very good, good, etc.), can be mapped on a relative value or utility-scale of 0 to 1.0. If a linear value function is used, it would imply the absence of the diminishing utility property (Figure 5.2). For risk analysis, the probability of success of criterion achievement can be expressed based on available historical frequency data or a probability distribution function or can be assigned subjectively.

*Figure 5. 2: Utility of "Importance to Overall Network"*




The criteria achievement levels quantified in their original units are obtained from the analysis of alternatives. In the absence of such information, these can be expressed subjectively on the basis of surveys and judgment of a committee. As for life cycle costs, their present worth or equivalent annual worth can be found by using standard cost estimation and discounting methodologies. The treatment of operation and salvage can be in accordance with the standard practice in the profession. Criteria weights, noted above, can be used to modify achievement levels.

In the utility-theoretic methodology, utility functions relate raw values of criteria achievement with their worth to the stakeholder(s). See Figure 5.2. To the extent possible, the utility functions should be based on known relationships (equations). Experience with utility functions indicates that some of these may exhibit diminishing marginal utility. In addition to the use of empirical equations, threshold and/or recommended values can be investigated.

The effectiveness of an alternative in achieving an objective or satisfying a certain evaluation criterion such as maximizing net present worth or maximizing equity and social justice or greenhouse gas reduction can be found as a utility number. Alternatively, it can be expressed in terms of units of any criterion (e.g., dollars used to quantify economic effectiveness).

Theoretical foundation and practical applications of this approach can be found in (Dodgson et al., 2009; Khan et al., 2012; Li, Pollard et al., 2009; Lifson, 1968; Savelsberg, 2008; Seiler, 1969).

Principles of utility theory have enabled advancements in transportation planning and engineering. For example, the fast-growing modelling field of modal choice relies on utility-theoretic and probability models. Advancements in utility theoretic models have a detailed history and the future looks bright. In the context of this research, without the utility-theoretic method in association with discrete (or continuous) probability distribution functions, the sustainability criteria could not be studied together for the assessment of relative merits of investment alternatives. Although utility-theoretic methods in association with probability models have been applied elsewhere, the model reported in this research is original and meets the requirements of utility theory axioms. References cited describe applications to different research areas.

In the risk-based utility-theoretic framework, the effectiveness of an investment alternative can be expressed in terms of multi-criteria and probability of criteria achievement, as shown below (Khan et al., 2012).

$$e(A) = \sum_{i=1}^{n} p_i(Cr_i).e(Cr_i) \qquad (5.1)$$

*Where*

$e(A)$ = the effectiveness of the alternative

$Cr_i$ = achievement level of a criterion $i$ by the alternative, expressed in the original units,

$i = 1,2, \ldots, q$.

$e(Cr_i)$ = the utility of achieving $Cr_i$ (e.g., net present worth, equity and social justice)

$p_i(Cr_i)$ = probability that $Cr_i$ will be achieved by the alternative that is being evaluated.

The method allows to take into account differential effects or achievement of an evaluation criterion for various interest or impact groups (e.g., accessibility of an impact group or a geographic area). In such a case, the $h_{th}$ level of criterion $g$, $Cr_{gh}$, which can occur due to the implementation of an alternative, can be expressed as:

$Cr_{gh}$ = the $h_{th}$ level of criterion $g$ (e.g., accessibility and network connectivity), weighted for all impact or interest groups (e.g., displaced businesses or families).

$$= k_1 Cr_g^1 + k_2 Cr_g^2 + k_3 Cr_g^3 + \ldots + k_s Cr_g^s \qquad (5.2)$$

*Where*

$Cr_g^s$ = the level of achievement of criterion $g$ for group $s$ (e.g., accessibility for group $s$ – rural residents, etc.)

$k_s$ = a weight, reflecting the importance of the interest group $s$ with respect to criterion $Cr_g$, and can be determined from the societal (community's) preference.

As noted above, for the application of the methodology, values for various levels of criteria achievement are to be found. This can be done by the use of value or utility theory. In the context of this research, the utility is a measure of the degree that each criterion is achieved by each alternative that is to be evaluated. An alternative can score a

high utility concerning a particular criterion if it satisfies the need expressed by that criterion better than other alternatives (e.g., NPW for different alternatives).

As noted above, in the context of this research, linear transformations of criterion achievement levels into the relative value or utility-scale can be used. However, the methodology allows the use of utility curves that can be derived from stated preference or revealed preference surveys of interest groups. The non-linear curves may exhibit the property of diminishing marginal utility (e.g., increasing accessibility and network connectivity offered by an alternative). However, due to the difficulty of identifying representative segments of the overall market, these are difficult to obtain, and their linear approximations should be acceptable under conditions when alternatives under study do not show very wide differences in achieving a criterion (e.g., NPW).

For the non-linear case, the process of mapping criterion achievements into relative values is illustrated in Figure 5.3.

$e(Cr_i) = L[Cr_i\ (Cr_i \text{ measured on the original value scale})]^z$ for $z<1$ and $i= 1,2, \ldots,q$ *(5.3)*

*Where*

$e(Cr_i)$ = the utility measure, in transformed units, for criterion $Cr_i$.

$L$ and $z$ are constants found from a calibration study.

For the general case of linear transformation:

$e(Cr_i) = m_i Cr_i + b_i$ *for* $i = 1,2, \ldots, q$ *(5.4)*

*Where*

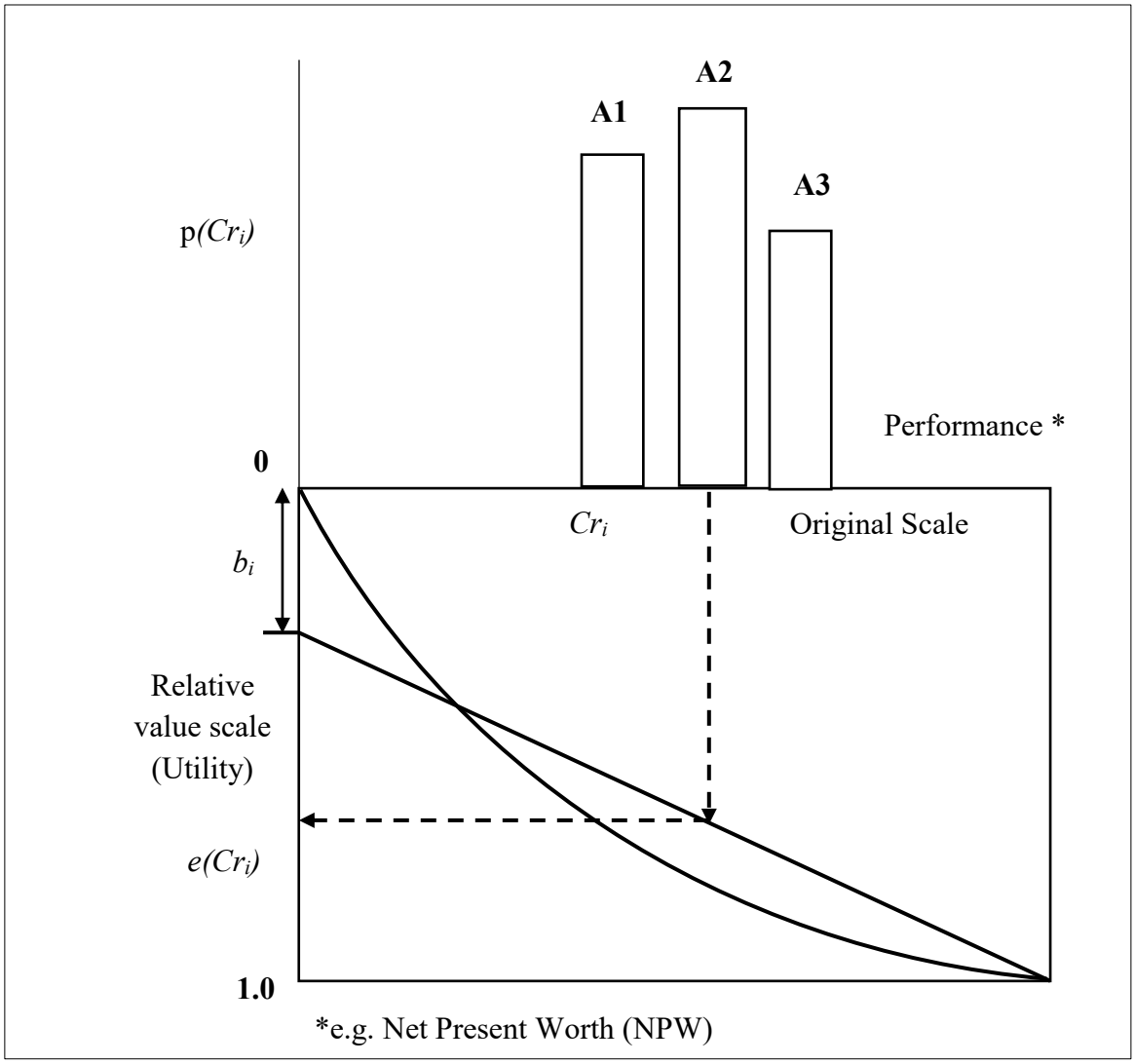
$e(Cr_i)$ = the utility measure, in transformed units, for criterion $Cr_i$.

$Cr_i$ = performance/achievement level, corresponding to criterion i, in original units (e.g., dollars, reduction in gms of average greenhouse gas emissions (GHG)/passenger-km)

$m_i$ = slope of the transformation curve for $Cr_i$

$b_i$ = vertical axis intercept (if applicable).

***Figure 5. 3: Transformation of Performance Measures from Original Measures to Relative Values (Non-Linear Case)***




Adapted from Khan et al (2012).

In the applications reported in this chapter, since no utility step function is applicable, $b$ = 0.

The relative value scale can be set in any desirable way. For the example evaluations presented in this thesis, $0 \leq e(Cr_i) \leq 1.0$.

The effectiveness of an alternative is obtained as:

$$e(A) = \sum_{i=1}^{q} wi(Cr_i) \qquad (5.5)$$

*Where*

$e(A)$ = expected effectiveness of the alternative, weighted for all interest groups (if applicable) and also weighted for criteria used

$W_i$ = weight assigned to criterion $i$

For general application, the risk in achieving actual effectiveness could be expressed by showing the utility of the expected value $\pm 1\sigma_{ei}$. Here $\sigma$ represents standard deviation. The values for the measures $e(Cr_i)$ and $\sigma_{ei}$ are obtained by using the applicable probability distribution (e.g., normal or triangular or other distribution of the probability of the criterion achievement level). The application of the Monte Carlo simulation method for estimating expected value and σ was suggested by the US DOT for the IntelliDrive benefit-cost analysis (US DOT, 2010). Alternatively, the equations for the applicable probability distributions function can be used to find the expected value and corresponding σ. It should be noted that the above description of the model applies to situations when a large number of alternatives are to be evaluated.

However, in applications that involve a limited number of alternatives, such as the one used in this thesis research, discrete probabilities can be used as shown in Figure 5.3.

As noted above, the analyst can express the effectiveness values in relative value units (utils) or units of a particular criterion. The equation that can be used for this purpose is noted next. In this equation, utils can be transformed into units of any criterion $q$.

$$\textit{Effectiveness (e)} \text{ in units of } Cr_q = \frac{ej \text{ in utils}}{wq.mq} - \frac{bq}{mq} \qquad (5.6)$$

*Where*

$w_q$ = weight of criterion q

$m_q$ = slope of the utility function

$b_q$ = vertical axis intercept (set equal to zero in analyses presented in this research)

In the case of cost factors, their present worth or equal annual worth is to be found. The utility of cost could be based on risk-averse, neutral, or risk taker (seeking) viewpoints. The utility functions for project costs applied in this research are based on a "neutral" viewpoint.

In this research, the starting point for the application of the decision-theoretic and utility-theoretic models was to work with life cycle economic factors. These were expressed as the Expected Net Present Worth (NPW) and equal time periods were used for analysis. Since alternatives have equal lives, it was not necessary to use Equivalent Annual Worth (EAW) in the decision models.

Initially, the decision model was formulated on the basis of dollars using impacts that can be quantified in dollar terms. In a later step, utils were used in order to include non-dollar type of impacts in the analysis.

The variables of the decision model as applied in this research are:

- Alternatives (A1, A2, A3)
- The states of cost overruns (S1, S2, S3)
- The NPW for each A & S combination.

The states of cost overruns were defined as follows:

- S1 = Low risk of cost overrun (COR $\leq$ 1) (i.e., no cost overrun).

  (1.0) was used with probability P1

- S2 = Medium risk of cost overrun (1 $<$ COR $\leq$ 1.2).

  (1.2) was used with probability P2

- S3 = High risk of cost overrun (COR $>$ 1.2).

  (1.5) was used with probability P3

For decision based on Expected Net Present Worth, the following probability-weighted expected values were found:

$$\mathrm{EXP}(\mathrm{NPW_i}) = \sum_{j=1}^{n} [\mathrm{NPW_{ij}} \times \mathrm{P_j}] \qquad (5.7)$$

*Where*

EXP ($\mathrm{NPW_i}$) = the expected Net Present Worth of alternative $A_i$; i = 1, 2,…, m

$\mathrm{NPW_{ij}}$ = the NPW of alternative $A_i$, under state of nature $S_j$; j = 1, 2,…, n

$P_j$ = probability of state of nature (COR risk level) $S_j$

An example application of the decision model is presented next. In this life cycle risk analysis problem, three mutually exclusive alternatives (i.e., A1, A2, and A3) for a highway transportation project "A" were evaluated. The data for this risk analysis example were sourced from the US Federal Highway Administration (FHWA) (1998), analyses on cost overruns reported in this thesis, and from the Ontario Hot Mix Producers Association (OHMPA) (1998) report which was written for the Ministry of Transportation, Ontario.

Tables 5.1 to 5.3 present lifecycle costs. Table 5.4 presents the present worth (PW) of costs ($M). Table 5.5 shows the lifecycle Expected Net Present Worth EXP(NPW). The

probabilities for the occurrence of S1, S2 and S3 shown in Tables 5.4 and 5.5 are based on Cauchey's probability distribution function found in this research and Berechman and Chen (2011). The alternatives are of equal length of 50 kms.

***Table 5. 1: Alternative A1 Lifecycle Costs***

| **Activity** | **Year** | **Cost ($M) (Current dollars)** | **PW of Cost @4% $M** | **PW of Costs @4% x 1.0 agency cost overrun S1** | **PW of Costs @4% x 1.2 agency cost overrun S2** | **PW of Costs @4% x 1.5 agency cost overrun S3** |
|---|---|---|---|---|---|---|
| Construction (Agency) | 0 | 0.975x50 = 48.75 | 48.750 | 48.750 | 58.5 | 73.125 |
| User cost | 0 | 0.200x50 =10.00 | 10.000 | 10.000 | 10.00 | 10.00 |
| Rehab#1 (Agency) | 10 | 0.200x50 =10 | 6.750 | 6.750 | 8.1 | 10.125 |
| User cost#1 | 10 | 0.269x50 =13.45 | 9.100 | 9.100 | 9.1 | 9.10 |
| Rehab#2 (Agency) | 20 | 0.200x50 =10 | 4.550 | 4.550 | 5.46 | 6.825 |
| User cost #2 | 20 | 0.361x50 =18.05 | 8.250 | 8.250 | 8.25 | 8.25 |
| Rehab#3 (Agency) | 30 | 0.200x50 =10 | 3.100 | 3.100 | 3.72 | 4.65 |
| User cost #3 | 30 | 0.485x50 =24.250 | 7.500 | 7.500 | 7.5 | 7.5 |
| Salvage | 35 | -0.100x50 = -5.00 | -1.250 | -1.250 | -1.25 | -1.25 |
| Total | | | 96.75 | 96.75 | 109.38 | 128.325 |

*@ 4%, 35 years of analysis.
*Cost overruns are applied to transportation agency costs only. Cost overruns multipliers are obtained from this research and Berechman and Chen (2011).

*Table 5. 2: Alternative A2 Lifecycle Costs*

| Activity | Year | Cost ($M) (Current dollars) | PW of Cost @4% $M | PW of Costs @4% x 1.0 agency cost overrun S1 | PW of Costs @4% x 1.2 agency cost overrun S2 | PW of Costs @4% x 1.5 agency cost overrun S3 |
|---|---|---|---|---|---|---|
| Construction (Agency) | 0 | 1.100x50 = 55.00 | 55.000 | 55.000 | 66.00 | 82.5 |
| User cost | 0 | 0.300x50 =15 | 15.000 | 15.000 | 15.00 | 15.00 |
| Rehab#1 (Agency) | 15 | 0.325x50 =16.25 | 9.000 | 9.000 | 10.8 | 13.5 |
| User cost#1 | 15 | 0.269x50 =13.45 | 7.45 | 7.45 | 7.45 | 7.45 |
| Rehab#2 (Agency) | 30 | 0.325x50 =18.05 | 5.00 | 5.00 | 6.0 | 7.5 |
| User cost #2 | 30 | 0.361x50 =18.05 | 5.55 | 5.55 | 5.55 | 5.55 |
| Salvage | 35 | -0.217x50 = -10.85 | -2.75 | -2.75 | -2.75 | -2.75 |
| Total | | | 94.25 | 94.25 | 108.05 | 128.75 |

*@ 4%, 35 years of analysis.
* Cost overruns are applied to transportation agency costs only. Cost overruns multipliers are obtained from this research and Berechman and Chen (2011).

*Table 5. 3: Alternative A3 Lifecycle Costs*

| Activity | Year | Cost ($M) (Current dollars) | PW of Cost @4% $M | PW of Costs @4% x 1.0 agency cost overrun S1 | PW of Costs @4% x 1.2 agency cost overrun S2 | PW of Costs @4% x 1.5 agency cost overrun S3 |
|---|---|---|---|---|---|---|
| Construction (Agency) | 0 | 1.11x50 = 55.5 | 55.5 | 55.5 | 66.6 | 83.25 |
| User cost | 0 | 0.267x50 =13.35 | 13.35 | 13.35 | 13.35 | 13.35 |
| Rehab#1 (Agency) | 15 | 0.366x50 =18.31 | 10.17 | 10.17 | 12.204 | 15.255 |
| User cost#1 | 15 | 0.3044x50 =15.22 | 8.455 | 8.455 | 8.455 | 8.455 |
| Rehab#2 (Agency) | 30 | 0.3664x50 =18.32 | 5.65 | 5.65 | 6.78 | 8.475 |
| User cost #2 | 30 | 0.3204x50 =16.02 | 4.94 | 4.94 | 4.94 | 4.94 |
| Salvage | 35 | - 0.217x50 =10.85 | -2.75 | -2.75 | -2.75 | -2.75 |
| Total | | | 95.315 | 93.315 | 109.579 | 130.975 |

*@ 4%, 35 years of analysis.

* Cost overruns are applied to transportation agency costs only. Cost overruns multipliers are obtained from this research and Berechman and Chen (2011).

*Table 5. 4: Lifecycle Analysis Present Worth (PW) of Costs ($M)* *

| Alternative | S1<br>Cost overrun multiplier = 1.0 | S2<br>Cost overrun multiplier = 1.2 | S3<br>Cost overrun multiplier = 1.5 |
|---|---|---|---|
| $A_1$ | 96.750 | 109.38 | 128.325 |
| $A_2$ | 94.25 | 108.05 | 128.750 |
| $A_3$ | 95.315 | 109.579 | 130.975 |

*@ 4%, 35 years of analysis. Alternatives A1 has 3 rehabilitations and Alternatives A2 and A3 have two rehabilitations.
∑ (Initial cost + cost of rehabilitation cycles + routine maintenance cost + user costs) in present worth @ 4% interest/discount rate.
Cost overruns are applied to transportation agency costs only. Cost overruns multipliers are obtained from this research as well as from Berechman and Chen (2011).
**Life cycle cost for "Do-nothing" option is $131.00M (in present worth). It consists of Agency and user costs of existing route (with inadequate capacity and low quality of ride). Net Present Worth (NPW) for each alternative shown in Table 5.5 is found by the [PW of cost of do-nothing – PW of cost of the alternative].

*Table 5. 5: Lifecycle Economic Evaluation NPW & Expected (NPW) ($M) **

| Alternative | S1<br>P(S1) = 0.42 | S2<br>P(S2) = 0.55 | S3<br>P(S3) = 0.03 | EXP(NPW) |
|---|---|---|---|---|
| $A_1$ | 34.25 | 21.62 | 2.67 | 26.356 |
| $A_2$ | 36.75 | 22.95 | 2.25 | 28.125+ |
| $A_3$ | 35.685 | 21.421 | 0.03 | 26.770 |

*@ 4%, 35 years of analysis.
+ Choice based on life cycle economic evaluation

The criteria that could be measured in dollar terms were used to carry out the cost-benefit analysis. The results of cost-benefit analysis as well as information on three additional criteria/impacts are shown in Table 5.6.

***Table 5. 6: Highway Life Cycle Analysis: Mutually Exclusive Alternatives and Performance Scores (Selected ENVISION Criteria)***

| | | | **ENVISION Sustainability Criteria** | | | |
|---|---|---|---|---|---|---|
| | | | **LD 3.3**[+] | **QL 3.1**[++] | **CR 2.6**[#] | **CR 1.2**[##] |
| Alternative | Life cycle cost* ($M in PW) | Life cycle cost of do-nothing** ($M in PW) | Life Cycle (NPW) | Equity and Social Justice | Infrastructure Integration Connectivity/ Accessibility index | Greenhouse Gas Emissions (gm/pass-km) |
| $A_1$ | 96.750 | 131.000 | 34.250 | 70 | 30 | 197.7 |
| $A_2$ | 94.250 | 131.000 | 36.750 | 90 | 60 | 87.0 |
| $A_3$ | 95.315 | 131.000 | 35.685 | 60 | 90 | 111.5 |

* ∑(Initial cost + cost of rehabilitation cycles + routine maintenance cost + user costs) in present worth @ 4% interest/discount rate.

** Agency and user cost of existing route (with inadequate capacity and low quality of ride).

+ NPW = [PW Life cycle cost of "do-nothing" – PW Life cycle cost of alternative]. Relates to Envision's Economy criterion. Leadership (LD 3.3 Conduct a life cycle economic evaluation)

++ Relates to Envision's Community Criterion. Quality of Life (QL 3.1 Advance equity and social justice). (% of Committee members voted the achievement of this criterion; for illustration purposes)

# Connectivity and accessibility index (for illustration purposes); related to Envision's Resilience criterion. Climate and Resilience (CR 2.6 Improve infrastructure integration)

## Relates to Envision's Emissions Criterion. Climate and Resilience (CR 1.2 Reduce Greenhouse Gas Emissions; for illustration purposes)

The equity and social justice were quantified on the assumption that a percentage of members of a committee of experts assessed the achievement of this criterion by each alternative (an assumption). Here, 70% of committee members thought that A1 meets the equity criterion whereas 90% of committee members indicated A2 and 60% voted for A3. The infrastructure integration (or connectivity/accessibility) index values were assumed to be estimated by network analysis. The greenhouse gas impact indicator was based on

assumed gms of emissions per passenger-km. The purpose of these rates assigned to the non-monetary criteria is to illustrate the application of the method.

The transformed values are shown in Table 5.7. The mapping of impacts into utilities was carried out using the concept and approach presented in Figure 5.3. For connectivity/accessibility criterion, a diminishing marginal utility function was assumed.

For linear transformations, Equation 5.4 was applied. Since NPW, connectivity/accessibility, and equity and social justice criteria are to be maximized, the slope $m_i$ is used as noted in the equation. For example, in the case of NPW, $m_i$ is 1/28.125 and a $Cr_i$ value of 26.77 million dollars was transformed into [(1/28.125) x (26.77)] = 0.95 utils. Since the intercept $b_i$ is zero, it does not enter into calculations. The NPW of 28.125 million is the highest magnitude of the NPW. See Table 5.7.

***Table 5. 7: Evaluation Based on Utility Theory***

| Alternative | Expected (NPW) ($M)* | Expected NPW (in relative value units) | Equity and Social Justice | Infrastructure integration/ Connectivity & accessibility | GHG reduction | Total |
|---|---|---|---|---|---|---|
| $A_1$ | 26.356 | 0.94 | 0.78 | 0.39 | 0.44 | 2.55 |
| $A_2$ | 28.125+ | 1.00 | 1.00 | 0.71 | 1.00 | 3.71++ |
| $A_3$ | 26.770 | 0.95 | 0.67 | 1.00 | 0.78 | 3.4 |

* Decision model results (Relative Value Scores Assigned to Criteria).
+ Choice based on life cycle economic evaluation
++ Choice based economic, social and environmental criteria (utility values).

The following figure illustrates the utility theoretic evaluation results.

On the other hand, given that the GHG emissions are to be minimized, the slope is used in the reverse manner. For example, the $Cr_i$ value of 197.7 gms/passenger-km was converted to utils as shown next: [(87.0) x (1/197.7)] = 0.44 utils. The lowest value of GHG emissions is 87.0 gm/passenger-km. See Table 5.7.

Next, criteria weights were defined on a scale 1 to 5: NPW 2, equity and social justice 2, connectivity/accessibility 3, and GHG emissions 3. The final utility scores are presented in Table 5.8. The results show that the use of infrastructure integration criterion (i.e., resilience criterion) alone would lead to alternative A3 as the choice. However, alternative A2 is the choice under all other criteria. Also, A2 is choice on the basis of sum of weighted utility scores.

***Table 5. 8: Evaluation Based on Weighted Utilities***

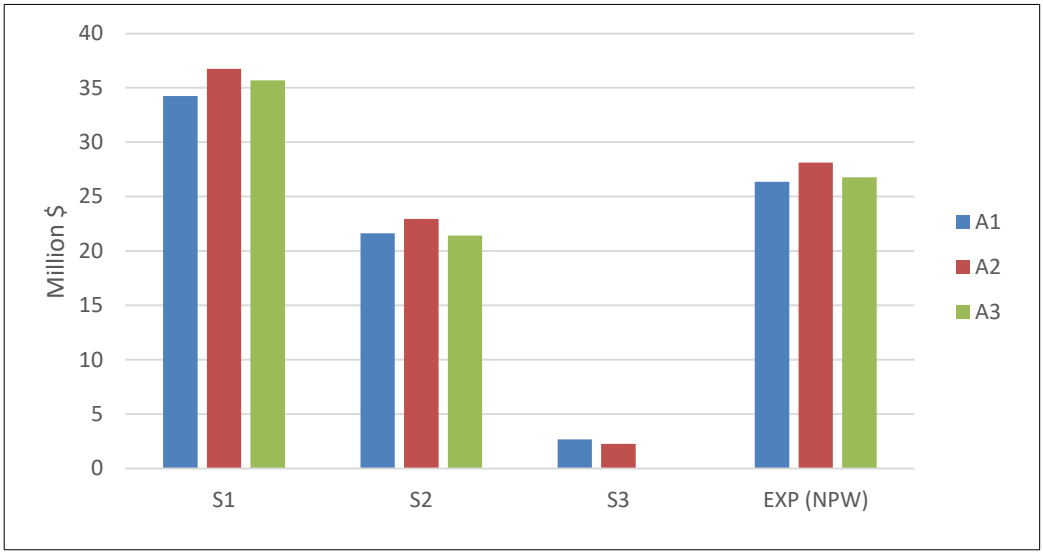
| Alternative | Expected (NPW) @4%* | Expected NPW (in relative value units) | Equity and Social Justice | Infrastructure integration/ Connectivity & accessibility | GHG reduction | Total |
|---|---|---|---|---|---|---|
| Weight | | 2 | 2 | 3 | 3 | |
| $A_1$ | 26.356 | 1.88 | 1.56 | 1.17 | 1.32 | 5.93 |
| $A_2$ | 28.125+ | 2.00 | 2.00 | 2.13 | 3.00 | 9.13++ |
| $A_3$ | 26.770 | 1.90 | 1.34 | 3.00 | 2.34 | 8.56 |

* Decision model results.
+ Choice based on life cycle economic evaluation.
++ Choice based on weighted economic, social and environmental criteria (weighted utility values)

Figure 5.4 shows the results of the lifecycle economic evaluation in terms of NPW & EXP (NPW). For each alternative (A1, A2, and A3), three risk levels (S1, S2, and S3) are taken into account. Alternative (A1) has 3 rehabilitations and Alternatives (A2) and (A3) have two rehabilitations see Tables 5.1 to 5.4. However, based on the NPW values,

alternative (A2) is the best at (S1 and S2). At (S3, high cost overrun) the results show that alternative (A1) is the best. Using probabilities for the cost overruns states, alternative A2 is the best on the basis of EXP (NPW).

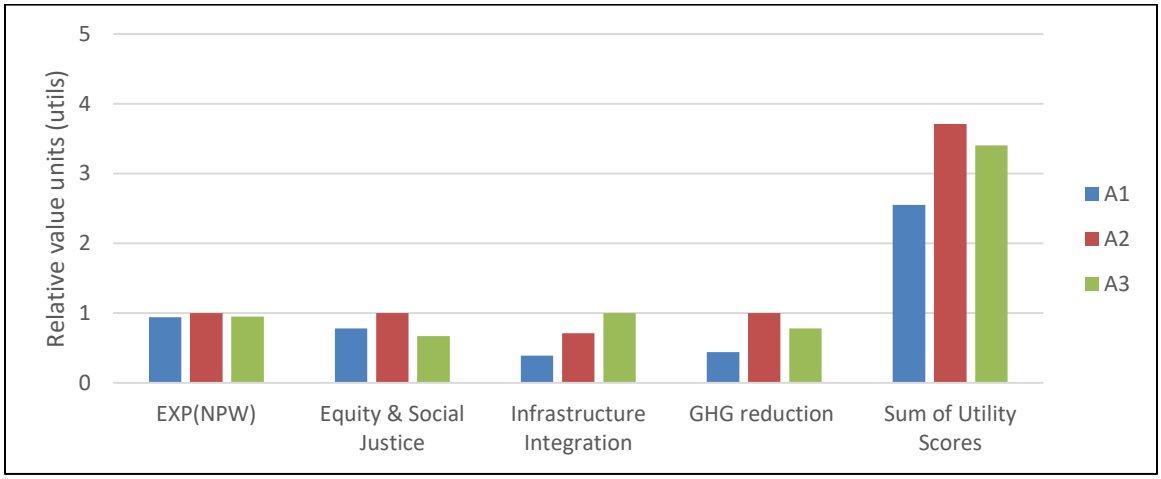
***Figure 5. 4: Lifecycle Economic Evaluation NPW and Expected (NPW)***




Evaluation of alternatives based on the Expected NPW addresses the economic sustainability objective. The methods used have met the requirement of risk-based life cycle analysis. However, the additional requirement of formally including other sustainability objectives in the analysis should also be addressed as shown in Tables 5.7 and 5.8.

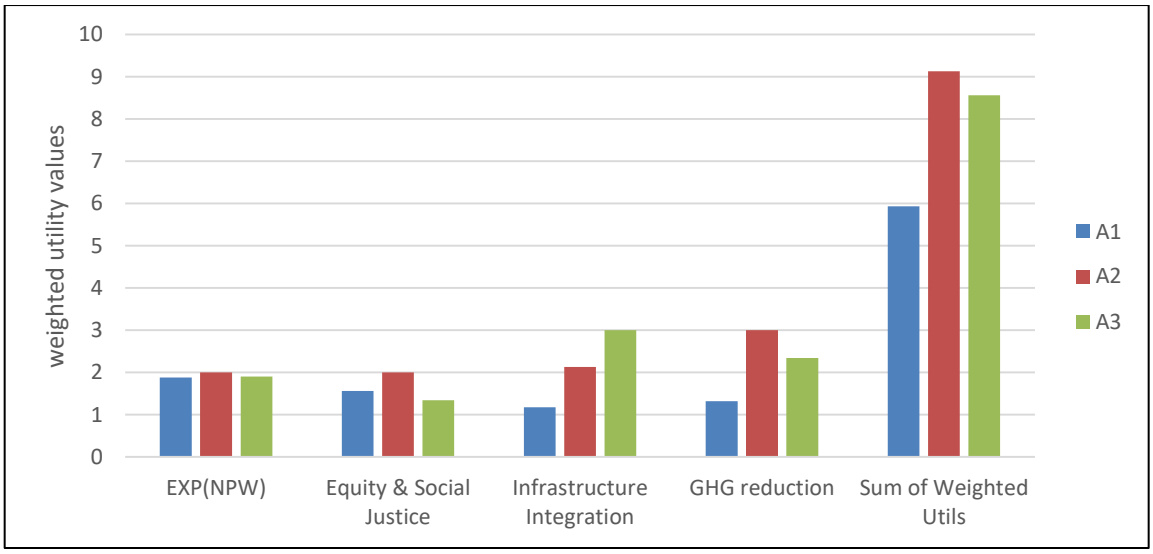
The results of the utility-theoretic approach presented in Table 5.7 are illustrated in Figure 5.5. An examination of this figure shows that based on EXP(NPW), equity and social justice, and GHG reduction, (A2) is the best. On the other hand, alternative (A3) has the highest effectiveness for the infrastructure integration criterion. Taking into account all criteria, the utility theoretic method indicates that alternative (A2) should be chosen.

***Figure 5. 5: Evaluation Based on Utility Theory***




As shown in Table 5.8, the risk-based utility-theoretic method allows weighting of utilities in the infrastructure evaluation process. Figure 5.6 illustrates results of evaluation based on weighted utilities. It can be seen that (A2) is the best at EXP (NPW), equity and social justice, and GHG reduction. On the other hand, (A3) has higher effectiveness for infrastructure integration. Taking into account all criteria, the best alternative/choice based on weighted utilities is (A2).

***Figure 5. 6: Evaluation Based on Weighted Utilities***



## 5.4 Discussion

The decision-theoretic method analyzes economic factors. The addition of utility-theory enables the treatment of economic as well as other factors of sustainability. In this research, an additional requirement is to account for risk in all components of the utility function, including economic criteria. The development and application of the combined risk-based decision-theoretic and utility-theoretic framework is a contribution to knowledge in the field.

In the context of an infrastructure investment project (e.g., a highway), a limited number of alternatives are generated using sustainability criteria as aids. These go through the analysis and evaluation steps and the preferred alternative moves to the detailed design phase. As explained in a later section on the link to ENVISION (added to the conclusions, contributions, and recommendations chapter), the methods advanced in this chapter can potentially assist in identifying the preferred alternative that can be used as the input to the sustainability rating framework.

Utility (U), estimated in sustainability terms, is a function of the combination of an Alternative (A) to be assessed and the Stochastic State-of-Nature (S) under which the Alternative will provide intended service. That is, U = f (A, S). There are always a limited number of well-studied alternatives and also the relevant States-of-Nature (e.g., traffic demand to be served, cost-overrun) are discrete (i.e., not continuous).

The states-of-nature are usually characterized in terms of high, medium, and low future predictions (but never continuous or infinite states). At the planning phase, a limited number of alternatives are analyzed and assessed using sustainability criteria. The methods

described in this chapter have the potential to identify the "preferred alternative" based on economic and other factors of sustainability, while treating risks.

In this research, the states-of-nature (in decision-theory terminology) are defined in terms of cost overrun (COR) bands (i.e., S1, S2, and S3). This is a realistic approach in research, especially when these bands are based on the results of cost overruns. The COR bands are not chosen arbitrarily, these are based on real-life data analysis. In this chapter as well as Chapter 4, cost overruns states are treated as stochastic, and methods are advanced to analyze these.

## 5.5 Conclusions

- For incorporating the sustainability objective in highway infrastructure planning and implementation, the life cycle approach is necessary.
- For use in life cycle analysis, the formal treatment of risk and uncertainty in cost models will add realism in highway infrastructure investment planning and evaluation.
- Given that a substantial part of life cycle costs is accounted for by the initial investment cost, it is prudent that the risk of cost overrun should be investigated. In this research, models for this purpose are described.
- Decision-theoretic and utility-theoretic methods that are advanced and illustrated in the evaluation of investment alternatives can formally treat life cycle costs and other factors of sustainability. That is, these methods are able to analyze lifecycle-based economic factors (i.e., net present worth) of alternatives as well as other factors of sustainability while formally treating risk. An additional methodological feature is

the treatment of differential effects on incidence groups and applicable probabilities of impacts.

# Chapter 6: Bayesian Statistical Decision Model: Life Cycle Analysis

## 6.1 Introduction

A more advanced version of the decision-theoretic model covered in Chapter 5 is the Bayesian statistical decision model for probabilistic life cycle analysis of highway infrastructure investments. It can serve the same function as the decision-theoretic model and additionally offers the opportunity to the decision-maker to assess the value of acquiring additional information to enhance knowledge about uncertain states of nature (the uncertain cost overruns in the context of this thesis research).

The principles of statistical decision theory enable a decision-maker (or a decision system) to identify the optimal course of action in situations when the outcomes are not known with certainty. Decision making under uncertainty is a specialized subject, invoking probabilistic states of nature and gains/payoffs that have to be defined for the applicable actions and states of nature combinations. Bayesian analysis is a statistical approach that allows one to use prior information and offers the ability to update probabilities as a result of new information. These probabilities serve important roles in risk analysis.

The Bayesian approach is a logical and common-sense approach. It quantifies the trade-offs between various decisions using probabilities and costs that accompany such decisions. As noted above, it enables the decision-maker to take into account new data/information in modifying the initial probabilities.

The field of Bayesian decision theory as a part of the broader statistical decision theory has enabled numerous disciplines to model their problems and obtain logical answers. Numerous applications have been reported in the business and engineering fields.

If a decision is to be made under uncertainty and there is the opportunity to learn from new observations in order to modify probabilities of the uncertain phenomenon, the Bayesian theory can assist modelling the problem. However, certain axioms of utility theory have to be met if complex problems are to be modelled. Examples include Bayesian Artificial Intelligence (Korb and Nicholson 2011), design of traveler information system (Khan 2010), toll road life cycle analysis (Khan 2013).

## 6.2 Variables and Basic Formulation

The basic data required to solve a decision problem as defined by Raiffa and Schlaifer (1961) and Schlaifer and Raiffa (1968) are noted below:

1. A set of alternative terminal actions: $A1, A2, \ldots \varepsilon\, a$, wherein the decision is to be made by selecting a single act $A$, from some domain $a$ of potential acts.
2. The possible states of nature: $S_1, S_2, \ldots \varepsilon$ s. The states of nature are probabilistic. That is, probabilities have to be assigned to each state in the set of states of nature. The decision maker or the decision-making system follows the philosophy that the selection of an alternative $A$ from the set $a$ is dependent upon the actual state of nature which is not known with certainty.
3. The set of alternative information acquisition means (termed experiments in the decision theory terminology): $e_0, e_1, \ldots \varepsilon\, E$. The decision-maker (or system) may elect to use one experiment from the set $E$ for the purpose of obtaining more information about the actual state of nature prior to the selection of an action.
4. The sample space or the set of outcomes or results of information acquisition experiment: $r_0, r_1 \ldots\ldots \varepsilon\, R$. For each experiment $e$, there is a space of possible outcomes

for that experiment. The space of possible outcomes *R,* is defined to encompass any outcome of any *e* in *E*.

5. The values, utilities, or gains *G (e, r, A, S)* represent the decision maker's preferences for all *e, r, A, S* combinations. The sequence of a course of action is as follows. The decision-maker selects an information acquisition means *e*, observes a result *r*, selects a particular *A*, and then a particular state of nature, S, occurs. The space of all possible combinations is (*e, r, A, S*).

   It is assumed that a single-valued gain function (utility function) is defined over this space, expressed as *G(e, r, A, S)*, and that in accordance with the principles of utility/value theory, the gains *G(e, r)* and *G(A, S)* are additive. These gains could be in dollars or in utils (value scores).

6. The probability distributions of *P'(s)* and *P(r|s, e)*, portray a joint probability measure *P(S,r|e)* over the space for each information acquisition tactic. The *P'(S)* (a prior probability) represents the decision maker's judgment about the relative likelihood of values of S, and *P(r|S, e)* (a conditional probability) characterizes each information acquisition tactic. It is the probability that the outcome *r*, will be observed if the experiment *e*, is performed and S is the value of the state of nature variable. Further information on probability measures is provided next.

### 6.3 Probability Measures

For solving the Bayesian decision problem, it is required to assign, either directly or indirectly, a joint probability *P(S,r|e)* to the joint distribution of S, *r*, over the space *s* x *R* for each information acquisition tactic *e.* This implies that the decision-maker (or the

decision system) should define the reliability of each possible information outcome *r* in predicting the true state of nature S, for each information acquisition tactic *e*.

From the joint probability measure, four other probability measures are defined, as described below.

1. The prior measure $P'(S)$ on the states of nature that the decision system would assign to S before observing the outcome r of tactic (experiment) *e*.
2. The conditional measure $P(r|S, e)$ on the space *R* the probability that the outcome *r* will be observed if the tactic (experiment) *e* is performed and S is the true value of the state of nature.
3. The marginal measure $P(r|e)$ on the space *R* for all *S* or the probability of observing outcome *r* from tactic *e*. It is computed using the following:

$$P(r|e) = \Sigma P'(S)\, P(r|S, e) \qquad (6.1)$$

4. The posterior measure $P''(S|r, e)$ on the space *S*, wherein this is the likelihood of different states S, given *r* and *e*. The decision-maker assigns this probability measure to the space *s* after knowing the outcome *r* of experiment *e*. It is computed using the following equation:

$$P''(S|r, e) = \frac{P(r|S, e)\ \ P'(S)}{P(r|e)} \qquad (6.2)$$

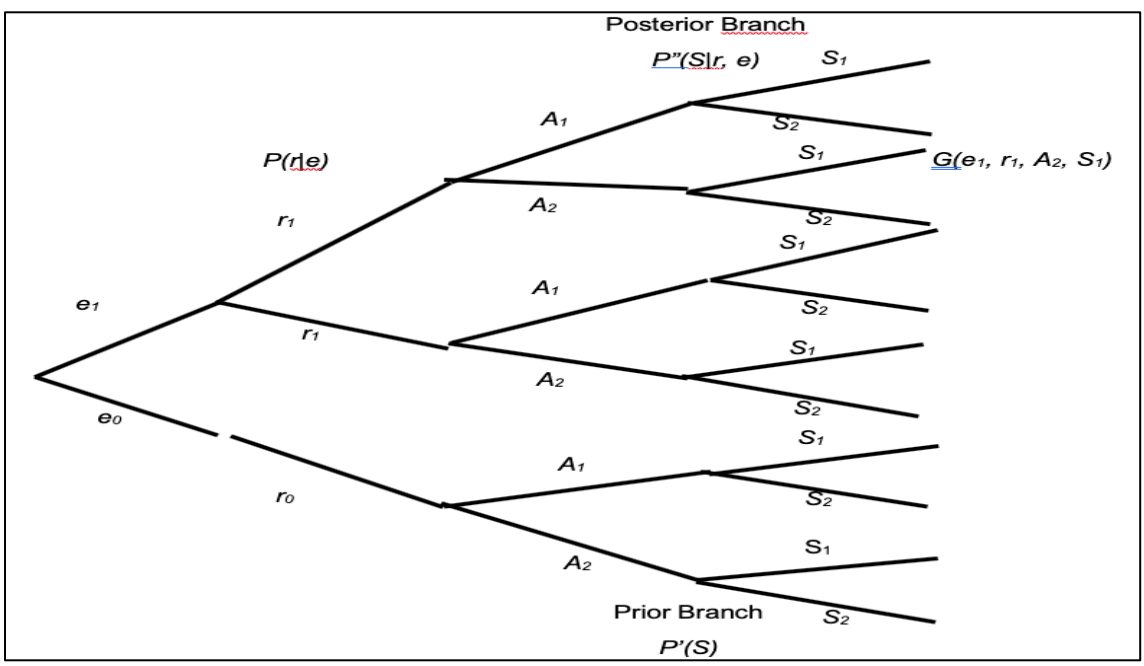
The Bayesian philosophy is that each *e* can be characterized by a conditional probability distribution $P(r|S, e)$, such that the relationship between the prior and posterior distribution is given as Bayesian Theorem.

## 6.4 Solving the Decision Problem

A decision tree is usually constructed to represent a decision problem (Figure 6.1). Any sequence of (*e, r, A, S*) can be represented through this tree. The desirability of a course of action (i.e., the sequence) is represented by the gain *G (e, r, A, S).*

The decision-maker wishes to choose an information acquisition tactic e (including the "null" tactic $e_0$) and therefore has to evaluate all the possible sequences or courses of action. The method of analysis that involves the evaluation of alternative courses of action to determine the most desirable information acquisition tactic *e* is known as pre-posterior analysis. In this thesis, to determine if additional information should be obtained, the pre-posterior analysis mode is used. However, if it is intended to find the optimal *e* and the purpose of the analysis is to compare results of analysis under prior and posterior probability distribution, then the mode of analysis can be termed posterior analysis.

***Figure 6. 1: Decision Tree***




Adopted from Schlaifer and Raiffa (1968).

## 6.5 Pre-posterior Analysis

The decision-maker is interested in choosing an information acquisition tactic/experiment that includes the 'null' experiment ($e_o$) and therefore has to evaluate all the possible courses of action. The mode of analysis concerned with the evaluation of alternative courses of action to determine the most appropriate information acquisition alternative *e* is known as pre-posterior analysis.

In this thesis, the pre-posterior analysis is used to establish if it is desirable to obtain additional information on the uncertain states (i.e., cost overrun states) and the amount of money that can be spent on information acquisition. Also, the optimal design/investment alternative can be identified by using the expected gain (i.e., Expected NPW).

The pre-posterior analysis is a sequence of operations required in making a decision (Schlaifer and Raiffa 1968). These are as follows:

1. The likelihood of different states of nature, *S*, is expressed in the form of prior probability distribution *P'(S).*
2. The conditional probability characteristics *P(r|S,e)*, is determined for each experiment.
3. The marginal measures *P(r|e),* for each experiment is computed as noted above, wherein for the null alternative ($e_o$), the probability is equal to one $P(r_o|e_o) = 1.0$.
4. The posterior probability distribution *P''(S|r, e)* is computed for each combination as defined above.
5. For each combination of *e, r, A, S*, its relative gain is found: *G (e, r, A, S)*.
6. The expected gain for each action *A*, for each *(e, r)* combination is as follows:

The expected gain for the posterior branch:

$$G^*(A, r, e) = \sum_S P''(S|r, e)\, G(e, r, A, S) \tag{6.3}$$

However, for the prior branch, where no new information is acquired,

$$G^*(A, r_0, e_0) = \sum_s P'(S)\, G(e_0, r_0, A, S) \tag{6.4}$$

For each ($e$, $r$) combination, the optimal action is determined, and its associated gain is noted:

$$G^*(r, e) = Max_A\, G^*(A, r, e) \tag{6.5}$$

For each information acquisition tactic $e$, the expected gain can be computed:

$$G^*(e) = \sum_r P(r|e)\, G^*(r, e) \tag{6.6}$$

The optimal experiment $e^*$ is that $e$ for which $G^*(e)$ is a maximum. That is:

$$G^*(e^*) = Max_e\, G^*(e) \tag{6.7}$$

## 6.6 Value of Information

In the pre-posterior analysis, if the optimal experiment is determined not to be the 'null' case, the value of information acquired from sampling should be calculated. In this analysis, an assumption is made that the sampling and terminal gains/utilities are additive. Then, instead of calculating the gain $G(e, r)$ for each possible experimental outcome, it is possible to compute the increase in gain, which would result for each $r$ if the prior choice of a terminal action after learning about it was changed. A weighted average of these gain increases can be calculated as follows:

$$G(e, r, A, S) = G_s(e, r) + G_t(A, S) \tag{6.8}$$

Where subscript S is used for "additional data" gain and $t$ is used for "terminal" gain.

The expected gain of an information acquisition tactic *e* as defined by Equation 6.6 is:

$$G^*(e) = \sum_r P(r|e)\, G^*(r, e) = \sum_r P(r|e)\, [Max_A \sum_s P''(S|r,e)\, G(e, r, A, S)] \qquad (6.9)$$

From Equations 6.8 and 6.9:

$$G^*(e) = \sum_r P(r|e)\, [Max_A \sum_s P''(S|r,e)\, [G_s(e, r) + G_t(A, S)]]$$

$$= \sum_r P(r|e)\, [G_s(e, r) + Max_A \sum_s P''(S|r, e)\, G_t(A, S)] \qquad (6.10)$$

Or

$$G_s^*(e) = \sum_r P(r|e)\, G_s(e, r) \qquad (6.11)$$

and

$$G_t^*(e) = \sum_r P(r|e)\, [Max_A \sum_s P''(S|r,e)\, G_t(A, S)] \qquad (6.12)$$

If *A'* is an optimal action under prior conditions of *S:*

$$\sum_S P'(S)\, G_t(A', S) \geq \sum_S P'(S)\, G_t(A, S) \qquad (6.13)$$

Letting *Ar* be defined as the optimal action under the posterior conditions of *S:*

$$\sum_S P''(S|r,e)\, G_t(Ar, S) = Max_A \sum_S P''(S|r,e)\, G_t(A, S) \qquad (6.14)$$

If instead of choosing the action *A* directly (i.e., based on prior probabilities), the decision maker performs tactic *e*, observes *r*, and then chooses *Ar*, the terminal gain increases by:

$$V_t(e, r) = \sum_S P''(S|r,e)\, G_t(A_r, S) - \sum_S P''(S|r,e)\, G_t(A', S) \qquad (6.15)$$

The term $V_t(e, r)$ is known as the conditional value of additional information *r*. But before *r* is observed, the expected value of additional information is given by:

$$V_t^*(e) = \sum_r P(r|e)\, V_t(e, r) \qquad (6.16)$$

It can also be interpreted as the amount of money that can be spent on acquiring additional information for the purpose of reducing risk.

Now, the expected net gain of information acquisition is therefore determined to be the expected value of additional information less the expected cost of obtaining it. In equation form:

$$v^*(e) = V_t^*(e) - c_s^*(e) \tag{6.17}$$

where,

$$c_s^*(e) = G_s^*(e) = \sum_r P(r|e)\, G_s(e, r) \tag{6.18}$$

**Notes:**

$V_t^*(e)$ can be considered as the expected reduction in risk. Also, for decision cases when all variables are expressed in monetary terms, it is considered as the maximum amount of money that can be spent on additional information acquisition.

If $V_t^*(e)$ is equal to zero or it is found to be negative, then $e_0$ is the best course of action (i.e., it is not desirable to acquire additional information in support of decision-making).

Please see Chapter 5 for data, the investment alternatives, the net present worth of alternatives, and the states-of-nature. These are used in the Bayesian method application to investigate risk in highway infrastructure investment.

## 6.7 Example Applications

Example 1

The inputs are:

- The Gain (i.e., NPW) matrix (See Table 5.5)
- Prior probabilities (See Table 5.5).
- The conditional probabilities are set on the basis that if additional information is obtained, it can be reasonably reliable (i.e., $P(r|s = 0.7))$.

The software computes:

- Marginal probabilities and Posterior probabilities
- Expected Gain based on prior probabilities & maximum gain obtainable and corresponding alternative
- Expected gain based on posterior probabilities, maximum gain obtainable based on the result of additional information acquisition *e* (i.e., *r1, r2, r3*) and corresponding alternative
- Value of information
- Expected NPW (taking into account marginal probabilities)

An examination of the Gain (NPW) matrix shows that alternative A2 is the choice under *S1* (no cost overrun) and *S2* (moderate cost overrun), but A1 is the choice under *S3* (high cost overrun state). The Bayesian model results show that Alternative A2 should be the choice and risk cannot be reduced by acquiring additional information. This is the same answer as obtained with the Decision Model that does not include the pre-posterior analysis capability.

*Figure 6. 2: Pre-Posterior Analysis Example 1*

**Prior Probabilities: P'(S)**

| | |
|---|---|
| P'(S1) | 0.42 |
| P'(S2) | 0.55 |
| P'(S3) | 0.03 |
| Sum | 1 |

**Conditional Probabilities P(r\|S, e)**

| | S1 | $S2_c$ | S3 |
|---|---|---|---|
| r1 | 0.7 | 0.2 | 0.1 |
| r2 | 0.2 | 0.7 | 0.2 |
| r3 | 0.1 | 0.1 | 0.7 |
| Sum | 1 | 1 | 1 |

**Marginal Probabilities P(r|e)**

| | |
|---|---|
| r1 | 0.407 |
| r2 | 0.475 |
| r3 | 0.118 |
| Sum | 1 |

**Posterior Probabilities P"(S|r,e)**

| | S1 | S2 | S3 | Sum |
|---|---|---|---|---|
| r1 | 0.722 | 0.270 | 0.007 | 1 |
| r2 | 0.177 | 0.811 | 0.013 | 1 |
| r3 | 0.356 | 0.466 | 0.178 | 1 |

**Gain Matrix**

| | S1 | S2 | S3 |
|---|---|---|---|
| A1 | 34.25 | 21.62 | 2.67 |
| A2 | 36.75 | 22.95 | 2.25 |
| A3 | 35.685 | 21.421 | 0.03 |

**Expected Gain**

| Priors | | Max Gain |
|---|---|---|
| G*(A1) | 26.3561 | 28.125 |
| G*(A2) | 28.125 | |
| G*(A3) | 26.77015 | |

| Posteriors | | Max Gain |
|---|---|---|
| G*(A1, r1,e1) | 30.604 | 32.766 |
| G*(A2, r1,e1) | 32.766 | |
| G*(A3, r1,e1) | 31.567 | |
| G*(A1, r2,e1) | 23.614 | 25.129 |
| G*(A2, r2,e1) | 25.129 | |
| G*(A3, r2,e1) | 23.673 | |
| G*(A1, r3,e1) | 22.743 | 24.178 |
| G*(A2, r3,e1) | 24.178 | |
| G*(A3, r3,e1) | 22.691 | |

**Value of Information Table**

| | r1 | r2 | r3 |
|---|---|---|---|
| $A_r$ | 32.766 | 25.129 | 24.178 |
| A' | 28.125 | 28.125 | 28.125 |
| $A_r - A'$ | 4.641 | -2.996 | -3.947 |
| P(r\|e) | 0.41 | 0.48 | 0.12 |

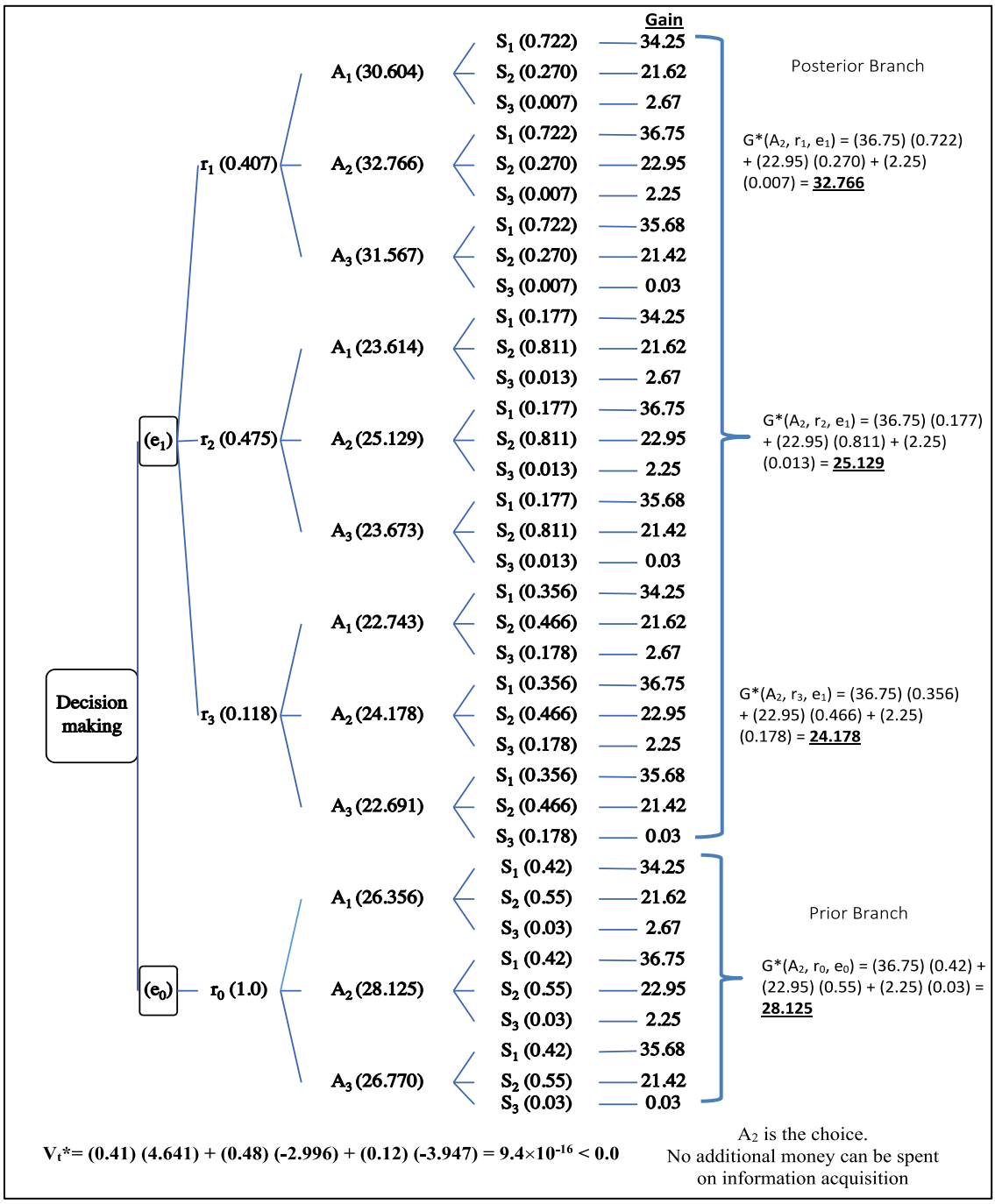
| | | |
|---|---|---|
| $V_t^*$ = | 9.4E-16 | **< 0.0** |

E(NPW) 28.125 A2 is the choice
No additional money can be spent on information acquisition

Since $V_t^*(e)$ is < 0, the optimal strategy is $e_0$, prior branch (i.e., no additional money will be spent on information acquisition). The Bayesian model results show that Alternative "A2" should be the choice and risk cannot be reduced by acquiring additional information, and therefore the expected NPW (taking into account marginal probabilities) is EXP(NPW)

= 28.125. Figure 6.3 below illustrates the above process (i.e., pre posterior analysis) in a decision tree.

***Figure 6. 3: Decision Tree (Pre-Posterior Analysis)***



Example 2

Same as Example 1, except the prior probabilities are set to be equal (i.e., $P(S1) = P(S2) = P(S3) = 0.333$). See Table 5.5 for some input. The results show that:

- The expected gains are lower than in Example 1 due to relatively high probabilities assigned to states with cost overruns.
- Alternative A2 is the choice.
- Risk can be reduced to the extent of \$0.021M with additional information acquisition, and therefore \$0.021M can be spent on additional studies.

*Figure 6. 4: Pre-Posterior Analysis Example 2*

**Prior Probabilities: P'(S)**

| | |
|---|---|
| P'(S1) | 0.333 |
| P'(S2) | 0.333 |
| P'(S3) | 0.333 |
| Sum | 0.999 |

**Conditional Probabilities P(r\|S, e)**

| | S1 | $S2_c$ | S3 |
|---|---|---|---|
| r1 | 0.7 | 0.2 | 0.1 |
| r2 | 0.2 | 0.7 | 0.2 |
| r3 | 0.1 | 0.1 | 0.7 |
| Sum | 1 | 1 | 1 |

**Marginal Probabilities P(r|e)**

| | |
|---|---|
| r1 | 0.333 |
| r2 | 0.3663 |
| r3 | 0.2997 |
| Sum | 0.999 |

**Posterior Probabilities P"(S|r,e)**

| | S1 | S2 | S3 | Sum |
|---|---|---|---|---|
| r1 | 0.700 | 0.200 | 0.100 | 1 |
| r2 | 0.182 | 0.636 | 0.182 | 1 |
| r3 | 0.111 | 0.111 | 0.778 | 1 |

**Gain Matrix**

| | S1 | S2 | S3 |
|---|---|---|---|
| A1 | 34.25 | 21.62 | 2.67 |
| A2 | 36.75 | 22.95 | 2.25 |
| A3 | 35.685 | 21.421 | 0.03 |

**Expected Gain**

| Priors | | Max Gain |
|---|---|---|
| G*(A1) | 19.49382 | 20.629 |
| G*(A2) | 20.62935 | |
| G*(A3) | 19.026288 | |

| Posteriors | | Max Gain |
|---|---|---|
| G*(A1, r1,e1) | 28.566 | 30.540 |
| G*(A2, r1,e1) | 30.540 | |
| G*(A3, r1,e1) | 29.267 | |
| G*(A1, r2,e1) | 20.471 | 21.695 |
| G*(A2, r2,e1) | 21.695 | |
| G*(A3, r2,e1) | 20.125 | |
| G*(A1, r3,e1) | 8.284 | 8.383 |
| G*(A2, r3,e1) | 8.383 | |
| G*(A3, r3,e1) | 6.368 | |

**Value of Information Table**

| | r1 | r2 | r3 |
|---|---|---|---|
| $A_r$ | 30.540 | 21.695 | 8.383 |
| A' | 20.629 | 20.629 | 20.629 |
| $A_r - A'$ | 9.911 | 1.066 | -12.246 |
| P(r\|e) | 0.33 | 0.37 | 0.30 |

$V_t^* =$ 0.0206

E(NPW) 20.629

A2 is the choice
$0.021M can be spent on
additional information

Example 3

The Gain (NPW) matrix and the conditional probabilities are the same as in Examples 1 & 2, but the prior probabilities are highly skewed. The state with the highest cost overrun is assigned 0.9 probability. See Table 5.5 for some input.

The results show that:

- Expected gains (NPW) are lower than in Examples 1 and 2

- Alternative A1 is the choice
- Risk can be reduced to the extent of $0.063M with additional information. Therefore, this is the highest amount of money that can be spent on additional information acquisition.

***Figure 6. 5: Pre-Posterior Analysis Example 3***

**Prior Probabilities: P'(S)**

| | |
|---|---|
| P'(S1) | 0.05 |
| P'(S2) | 0.05 |
| P'(S3) | 0.9 |
| Sum | 1 |

**Conditional Probabilities P(r\|S, e)**

| | S1 | $S2_c$ | S3 |
|---|---|---|---|
| r1 | 0.7 | 0.2 | 0.1 |
| r2 | 0.2 | 0.7 | 0.2 |
| r3 | 0.1 | 0.1 | 0.7 |
| Sum | 1 | 1 | 1 |

**Marginal Probabilities P(r|e)**

| | |
|---|---|
| r1 | 0.135 |
| r2 | 0.225 |
| r3 | 0.64 |
| Sum | 1 |

**Posterior Probabilities P"(S|r,e)**

| | S1 | S2 | S3 | Sum |
|---|---|---|---|---|
| r1 | 0.259 | 0.074 | 0.667 | 1 |
| r2 | 0.044 | 0.156 | 0.800 | 1 |
| r3 | 0.008 | 0.008 | 0.984 | 1 |

**Gain Matrix**

| | S1 | S2 | S3 |
|---|---|---|---|
| A1 | 34.25 | 21.62 | 2.67 |
| A2 | 36.75 | 22.95 | 2.25 |
| A3 | 35.685 | 21.421 | 0.03 |

**Expected Gain**

| Priors | | Max Gain |
|---|---|---|
| G*(A1) | 5.1965 | 5.197 |
| G*(A2) | 5.01 | |
| G*(A3) | 2.8823 | |
| **Posteriors** | | **Max Gain** |
| G*(A1, r1,e1) | 12.261 | 12.728 |
| G*(A2, r1,e1) | 12.728 | |
| G*(A3, r1,e1) | 10.858 | |
| G*(A1, r2,e1) | 7.021 | 7.021 |
| G*(A2, r2,e1) | 7.003 | |
| G*(A3, r2,e1) | 4.942 | |
| G*(A1, r3,e1) | 3.065 | 3.065 |
| G*(A2, r3,e1) | 2.681 | |
| G*(A3, r3,e1) | 0.476 | |

**Value of Information Table**

| | r1 | r2 | r3 |
|---|---|---|---|
| $A_r$ | 12.728 | 7.021 | 3.065 |
| A' | 5.197 | 5.197 | 5.197 |
| $A_r - A'$ | 7.531 | 1.825 | -2.132 |
| P(r\|e) | 0.14 | 0.23 | 0.64 |

$V_t^* =$ 0.063

E(NPW) 5.2595

A1 is the choice
$0.063M can be spent on
additional information

Example 4

Here, the Gain (NPW) matrix and conditional probabilities are the same as in other examples. The prior probabilities are symmetrical with a very high probability of 0.9 is assigned to *S2*. See Table 5.5 for some input. The results show that:

- Expected gains (NPW) are higher than in Example 3 since the state of moderate cost run (i.e., *S2*) receives a prior probability of 0.9 and the state with a high cost overrun is assigned a very low probability.
- Alternative A2 is the choice.
- Risk cannot be reduced with additional studies and therefore no money should be spent on additional information.

## *Figure 6. 6: Pre-Posterior Analysis Example 4*

**Prior Probabilities: P'(S)**

| | |
|---|---|
| P'(S1) | 0.05 |
| P'(S2) | 0.9 |
| P'(S3) | 0.05 |
| Sum | 1 |

**Conditional Probabilities P(r\|S, e)**

| | S1 | $S2_c$ | S3 |
|---|---|---|---|
| r1 | 0.7 | 0.2 | 0.1 |
| r2 | 0.2 | 0.7 | 0.2 |
| r3 | 0.1 | 0.1 | 0.7 |
| Sum | 1 | 1 | 1 |

**Marginal Probabilities P(r|e)**

| | |
|---|---|
| r1 | 0.22 |
| r2 | 0.65 |
| r3 | 0.13 |
| Sum | 1 |

**Posterior Probabilities P"(S|r,e)**

| | S1 | S2 | S3 | Sum |
|---|---|---|---|---|
| r1 | 0.159 | 0.818 | 0.023 | 1 |
| r2 | 0.015 | 0.969 | 0.015 | 1 |
| r3 | 0.038 | 0.692 | 0.269 | 1 |

**Gain Matrix**

| | S1 | S2 | S3 |
|---|---|---|---|
| A1 | 34.25 | 21.62 | 2.67 |
| A2 | 36.75 | 22.95 | 2.25 |
| A3 | 35.685 | 21.421 | 0.03 |

**Expected Gain**

| Priors | | Max Gain |
|---|---|---|
| G*(A1) | 21.304 | 22.605 |
| G*(A2) | 22.605 | |
| G*(A3) | 21.06465 | |

| Posteriors | | Max Gain |
|---|---|---|
| G*(A1, r1,e1) | 23.199 | 24.675 |
| G*(A2, r1,e1) | 24.675 | |
| G*(A3, r1,e1) | 23.204 | |
| G*(A1, r2,e1) | 21.523 | 22.844 |
| G*(A2, r2,e1) | 22.844 | |
| G*(A3, r2,e1) | 21.311 | |
| G*(A1, r3,e1) | 17.004 | 17.908 |
| G*(A2, r3,e1) | 17.908 | |
| G*(A3, r3,e1) | 16.211 | |

**Value of Information Table**

| | r1 | r2 | r3 |
|---|---|---|---|
| $A_r$ | 24.675 | 22.844 | 17.908 |
| A' | 22.605 | 22.605 | 22.605 |
| $A_r - A'$ | 2.070 | 0.239 | -4.697 |
| P(r\|e) | 0.22 | 0.65 | 0.13 |

$V_t^* =$ -1E-15

E(NPW) 22.605

A2 is the choice
No money can be spent on additional information

Example 5

In this risk analysis case, the gain (NPW) matrix is the same as in Examples 1 to 4. But equal prior probabilities are assigned to the state of cost overrun and the conditional probabilities are lowered to reflect the concept that although additional information may not be as reliable as in other examples, it is still worthwhile to spend money in order to reduce risk. See Table 5.5 for some input.

The result show that:

- As compared to Example 2 which assigned higher reliability to additional information than in this example, the expected gains (NPW) are somewhat lower.
- Alternative A2 is the choice
- Risk can be reduced to the extent of $0.204M with additional information and therefore cost of additional information should not exceed this amount.

*Figure 6. 7: Pre-Posterior Analysis Example 5*

**Prior Probabilities: P'(S)**

| | |
|---|---|
| P'(S1) | 0.33 |
| P'(S2) | 0.33 |
| P'(S3) | 0.33 |
| Sum | 0.99 |

**Conditional Probabilities P(r\|S, e)**

| | S1 | $S2_c$ | S3 |
|---|---|---|---|
| r1 | 0.5 | 0.25 | 0.2 |
| r2 | 0.3 | 0.5 | 0.3 |
| r3 | 0.2 | 0.25 | 0.5 |
| Sum | 1 | 1 | 1 |

**Marginal Probabilities P(r|e)**

| | |
|---|---|
| r1 | 0.3135 |
| r2 | 0.363 |
| r3 | 0.3135 |
| Sum | 0.99 |

**Posterior Probabilities P"(S|r,e)**

| | S1 | S2 | S3 | Sum |
|---|---|---|---|---|
| r1 | 0.526 | 0.263 | 0.211 | 1 |
| r2 | 0.273 | 0.455 | 0.273 | 1 |
| r3 | 0.211 | 0.263 | 0.526 | 1 |

**Gain Matrix**

| | S1 | S2 | S3 |
|---|---|---|---|
| A1 | 34.25 | 21.62 | 2.67 |
| A2 | 36.75 | 22.95 | 2.25 |
| A3 | 35.685 | 21.421 | 0.03 |

**Expected Gain**

| Priors | | Max Gain |
|---|---|---|
| G*(A1) | 19.3182 | 20.444 |
| G*(A2) | 20.4435 | |
| G*(A3) | 18.85488 | |

| Posteriors | | Max Gain |
|---|---|---|
| G*(A1, r1,e1) | 24.278 | 25.855 |
| G*(A2, r1,e1) | 25.855 | |
| G*(A3, r1,e1) | 24.425 | |
| G*(A1, r2,e1) | 19.896 | 21.068 |
| G*(A2, r2,e1) | 21.068 | |
| G*(A3, r2,e1) | 19.477 | |
| G*(A1, r3,e1) | 14.305 | 14.961 |
| G*(A2, r3,e1) | 14.961 | |
| G*(A3, r3,e1) | 13.166 | |

**Value of Information Table**

| | r1 | r2 | r3 |
|---|---|---|---|
| $A_r$ | 25.855 | 21.068 | 14.961 |
| A' | 20.444 | 20.444 | 20.444 |
| $A_r - A'$ | 5.412 | 0.625 | -5.483 |
| P(r\|e) | 0.31 | 0.36 | 0.31 |

$V_t^* =$ 0.2044

E(NPW) 20.444

A2 is the choice
$0.204M can be spent on
additional information

## 6.8 Discussion

The Bayesian method is a part of decision-theoretic methods. It is an advanced tool since it enables the use of posterior probabilities as compared to the use of prior probabilities in the "non-Bayesian" decision methods. Explanation of the variables,

formulas, and example applications are presented in this chapter. The Bayesian model for decision making in highway infrastructure investments presented in this thesis enables the study of the role of additional information in reducing the risk of cost overrun.

The conditional probability part of the Bayesian model (i.e., $P(r|S,e)$) is considered as the reliability of the additional information. This is a strength of the Bayesian method. This probability measure enables the conversion of Prior Probabilities into Posterior Probabilities.

## 6.9 Conclusion

In adopting the Bayesian approach, the decision-maker has the opportunity to update the probabilities of uncertain states. The developed method for decision analysis based on the Bayesian method evaluates the anticipated reduction in risk using the potential role of additional data. It is an advanced form of the decision-theoretic method due to its ability to treat stochastic states of nature and also offers the decision-maker the opportunity to find out how much can be spent on additional information in support of decision-making. The illustrative examples have produced logical and useful results.

# Chapter 7: Questionnaire Survey

## 7.1 Development of the Questionnaire

Surveys are a useful tool for collecting data about personal experience, services, attitudes, and values. Surveys may take the form of postal questionnaire, face-to-face interview, telephone interview, email, and most recently online questionnaire (Gilbert, 2008). Sekaran and Bougie (2003) reported that the most effective data collection instruments that would enhance the better measurement of variables are survey questionnaires.

For descriptive and analytical surveys, online questionnaires are commonly used to find facts, opinions, and views to investigate and analyze research topics. Researchers also face the challenge of finding the most useful and relevant information in formulating and designing the survey questions. These questions need to be based on the literature review and the ways in which they are phrased and the order in which they are placed needs to be addressed (Gilbert, 2008; Naoum, 1998).

The questionnaire as a tool for data collection has the following advantages as reported by Bryman (2012) and Gilbert (2008):

- Lots of information can be obtained from a large number of participants in a short period of time and relatively cheaply.
- It can be analyzed more scientifically and objectively than other research methods.
- Allows respondents to fill out the questionnaires at their own convenience.
- They are relatively easy to analyze if constructed correctly.

However, this method of data collection has its disadvantages as reported by Bryman (2012) and Gilbert (2008):

- Meaning of questions may not be clear to some respondents.
- Lower response rates.
- There is a possibility that the questionnaires might be answered by someone else other than the right person.
- Participants may read every question differently and, thus respond based on their own interpretation of the question.
- It is difficult to ask a lot of questions as they could cause 'respondent fatigue'.

## 7.2 Choice of Variables

The research was initiated with a comprehensive and detailed literature review in order to develop a preliminary list of variables related to cost overrun factors in transportation infrastructure projects. The principal literature sources (as noted in Chapter 2) were conference papers, journal papers, and internet sources (i.e., papers, reports). The review of available references resulted in important information on the transportation infrastructure industry and helped to identify relevant material for developing the survey questionnaire and implementation approach.

## 7.3 Design of Questionnaire and Implementation Approach

Appendix A describes the process followed in developing the questionnaire and obtaining the University approval for implementation. The self-questionnaire survey was prepared with 53 generally recognized factors that are potentially responsible for cost overruns. Necessary steps were followed regarding checks on the wording of questions and

the structure of the questionnaire. The Carleton University Ethics Review personnel were consulted before it was forwarded to about 150 professional infrastructure agencies/managers in Canada, the USA, Australia, and the Middle East. The survey population targeted were professional persons and organizations involved in infrastructure-related tasks.

Fifty-three completed questionnaires (data sets) were received. These were used to prepare the database for analysis using the Statistical Package for Social Sciences (SPSS). The survey questions were phrased to ask the respondents to rate factors that contribute to cost overrun in highway infrastructure projects.

The final version of the questionnaire, organized into three parts, consisting of 11 pages is shown in Appendix A. These parts are as follows:

- The first part of the questionnaire: General Information: This part requested background information of the respondents and their projects, such as the respondent's position, years of industrial/professional experience, type of sector (federal/provincial or state/private), region, and the project type.
- The second part of the questionnaire: Causes of Cost Overruns: This is the principal part of the questionnaire in which the respondent was asked to rate the factors that contribute to cost overrun. The impact level was measured on a 5-point Likert scale, (very high, high, medium, low, and very low). The respondents were requested to answer the questions based on actual experiences. In addition to rating fifty-three factors, the respondents were requested to add any comments they considered important.

- The last part of the questionnaire required participants to comment if they had any additional information about the cost overrun factors that had not been mentioned in the questionnaire.

It should be noted that the questionnaire design, plan for analysis, and measures to safeguard the confidentiality of data were approved by Carleton University's Research Ethics Board.

## 7.4 Sampling Approach

Recognizing that the cost overrun is a rather sensitive subject and therefore a high level of return could not be relied upon, the questionnaire was sent to all provinces in Canada and all states in the U.S.A. In the Middle East, construction managers with experience were requested to complete the questionnaire. Following the advice of professional contacts in Australia, associations with intimate knowledge of the highway construction industry were contacted.

The sampling approach defined was based on the following considerations:

(1) Coverage of major jurisdictions responsible for making infrastructure decisions (i.e., provinces/states/regions),

(2) Respondents knowledgeable about a large number of factors that can potentially cause cost overruns. The intent was to focus on top professionals/managers/experts with experience and knowledge in the subject.

Data analysis in this study was carried out using a number of statistical tools, namely, Factors Analysis and logistic regression analysis. These tools are discussed in detail in the upcoming chapters. The analyses were conducted with the help of the

Statistical Package for the Social Sciences (SPSS) version 25. Table 7.1 presents the list of variables and their labels.

## 7.5 Discussion

The rationale and approach to the questionnaire survey research method used in this study are discussed in this chapter and Appendix A. For developing a list of variables, the comprehensive and detailed literature review reported in Chapter 2 was useful.

All aspects of the questionnaire study, including the wording of the questions regarding 53 variables, the agencies and their professional personnel to approach for data, analysis methods, etc. were checked by experts in the Carleton University Ethics Review Board. Following their approval, the questionnaire was forwarded to about 150 professional infrastructure agencies/managers in Canada, the USA, Australia, and the Middle East. The survey population targeted were professional persons and organizations involved in infrastructure-related tasks.

According to the study plan, the contents of the completed responses were used to form a database using the SPSS package. Checks were carried out so as to ensure that no coding errors occurred. The survey data were used to carry out three versions of Factor Analysis and the corresponding Logistic Regression analysis. These are reported in Chapter 8 and Chapter 9, respectively.

Table 7.1 presents variable names and descriptions. Table 7.2 shows a summary of raw data.

*Table 7. 1: Variable Name and Description*

| Variable Number | Variable Description |
|---|---|
| **POLICY** | |
| **V1** | Changes in government funding policies |
| **V2** | Deal termination due to changes in law, government policy or protocols |
| **V3** | Change in regulations |
| **DESIGN, CONSTRUCTION AND SCHEDULING** | |
| **V4** | Complexity of the project (e.g., Project size, Project type, scope of work) |
| **V5** | Design changes during construction work |
| **V6** | Re-work due to the construction errors |
| **V7** | Unexpected technical problem |
| **V8** | Design errors that represent insufficient deliverables |
| **V9** | Changes by owner on the completion date of the project |
| **V10** | Scope changes by Owner during construction |
| **V11** | Delays related to owner or owner representative (e.g., stop work) |
| **V12** | Unrealistic project scheduling |
| **V13** | Acceleration to maintain schedule |
| **V14** | Delays in sending important documents to construction site (e.g., drawings, design changes) |
| **V15** | Type of construction contract (e.g., unit price contract) |
| **V16** | Unnecessary practices, specifications, procedures and documentation requirements forced onto the construction site workers |
| **V17** | Replacing unsatisfactory subcontractors from site by hiring new subcontractors |
| **V18** | Delay by subcontractor |
| **SITE CONDITIONS/ENVIRONMENT** | |
| **V19** | Poor site management |
| **V20** | Unexpected weather conditions |
| **V21** | Accidents due to poor site safety |
| **MATERIALS AND EQUIPMENT** | |
| **V22** | Shortage of materials & equipment on site |
| **V23** | Damages in materials and equipment in transit to the construction site |
| **V24** | Late delivery of materials & equipment at the construction site |
| **V25** | Equipment Selection Changes |
| **V26** | Construction variations due to equipment selection |
| **V27** | Defective materials |
| **LABOUR, STAFF AND VENDOR** | |
| **V28** | Shortage of skilled labor |
| **V29** | Lack of staff at the time of construction |

| | |
|---|---|
| **V30** | labor strikes & vendor strikes |
| **PERMITS AND APPROVALS** | |
| **V31** | Delays and approval of shop drawings and installation procedures |
| **V32** | Building Permit to the construction contractor |
| **V33** | Government/ Municipal Approvals |
| **ESTIMATION/BUDGET AND FINANCIAL** | |
| **V34** | Lack of expertise in setting the budget |
| **V35** | The approved budget was too low |
| **V36** | Absence of a detailed Estimate Plan |
| **V37** | Changes in prices of items that have already been approved |
| **V38** | Economic and financial factors |
| **V39** | Inappropriate and inadequate procurement (e.g., payment terms, pricing) |
| **V40** | Shortage of contingency and management reserve funds |
| **V41** | Unaddressed overtime work or multiple shifts that was not included in the base estimate |
| **V42** | Bankruptcy of subcontractors and vendors during construction work |
| **V43** | Currency fluctuations |
| **OTHER FACTORS** | |
| **V44** | Bad luck |
| **V45** | Lack of technical qualifications of the client |
| **V46** | Overly high expectations |
| **V47** | Poor communication and coordination between all parties |
| **V48** | Disputes between parties (designer, contractor, owner) |
| **V49** | Political Factors |
| **V50** | Technological risk |
| **V51** | Land acquisition issues within right-of-way |
| **V52** | Quality assurance and quality control |
| **V53** | Inexperienced Project Managers, Estimators and Planners |

***Table 7. 2: Percentage of Responses Based on the Likert Scale****

| Very high | High | Medium | Low | Very low | Total |
|---|---|---|---|---|---|
| 4.34 | 7.08 | 15.59 | 16.63 | 56.35 | 100.00 |

***** NOTE: 53 completed responses were received.

# Chapter 8: Factor Analysis for the Study of Determinants of Cost Overrun Factors

## 8.1 Factor Analysis: Why and How?

Factor Analysis is an advanced statistical tool available to researchers for reducing a large number of variables to a small number of factors from the database (Field, 2009). In Factor Analysis, variables that are inter-correlated can be grouped together so that they can be interpreted as one combined variable (called a factor) rather than a series of separate variables (Syed and Khan, 2001). The resulting smaller numbers of factors are used to describe the data with minimum loss of information rather than the original variables (Hair et al., 1998). For research studies, Factor Analysis provides answers on the factors (clusters of variables) as well as inherent variance characteristics (Blaikie, 2003). These findings can then be used to develop predictive models. There are two types of factor analysis that are used for different kinds of research and analysis, namely exploratory and confirmatory factor analysis. In this thesis research, exploratory factor analysis was used.

The research objectives and associated research framework required detailed Factor Analysis and to apply results in the predictive modelling part of the research. Following the study of the philosophical and theoretical rationale for these analyses, the question of software selection was addressed. Statistical software such as SPSS has the capability to perform Factor Analysis and detailed user guides are available. In this thesis research, the SPSS was used.

## 8.2 Factor Analysis Versions

Three versions of Factor Analysis noted below were carried out. Detailed explanations of this tool are provided in this chapter and in Chapter 9, logistic regression models based on outputs of Factor Analysis are presented.

- **Version 1**: 43 variables out of a total 53 variables were analyzed in Version 1-1 and the remaining 10 variables were studied as Version 1-2.
- **Version 2**: 31 variables were analyzed that resulted from filtering of correlation coefficient.
- **Version 3**: 27 variables were used that were rated by respondents as having very high, high, and medium importance as a potential source of cost overrun.

The rationale for subjecting data to three versions of Factor Analysis is to learn to the highest degree possible the nature of factors that can explain the cost overrun phenomenon as viewed by survey respondents.

The Version 1 analysis started with all 53 variables without violating any guidelines provided by the methodology. However, an advisory comment from the SPSS suggested that all 53 variables could not be analyzed for technical reasons noted later in this chapter. Therefore, it was decided to divide the data into 43 variables that belonged to specific categories with explanatory labels (e.g., policy) and the 10 remaining variables that were categorized as "other variables". Please see Table 7.1 for a list of variables. These two analyses are termed as Version 1-1 and Version 1-2 of the overall Factor Analysis study.

Although the Version 1 analyses produced logical results, it was decided to filter the data by using advice from literature sources regarding correlation coefficients. According to literature, if there is no correlation between variables, a variable can be removed from analysis. In Version 2 of Factor Analysis, variables with a weak correlation of less than (0.3) were removed. This step resulted in removing 22 variables out of a total of 53 variables and the remaining 31 variables were used in the second version of the Factor Analysis. Once again, the results were useful in providing further insights on factors that can potentially cause cost overrun.

Following the completion of Version 2 analysis, it was decided to use a different approach to filter data and benefit from the results of a Version 3 analysis. Here, 27 variables were selected on the basis of rates (i.e., very high, high, and medium) assigned by survey respondents. The results of Version 3 analysis further strengthened the identification of cost overrun factors.

## 8.3 Factor Analysis Processes

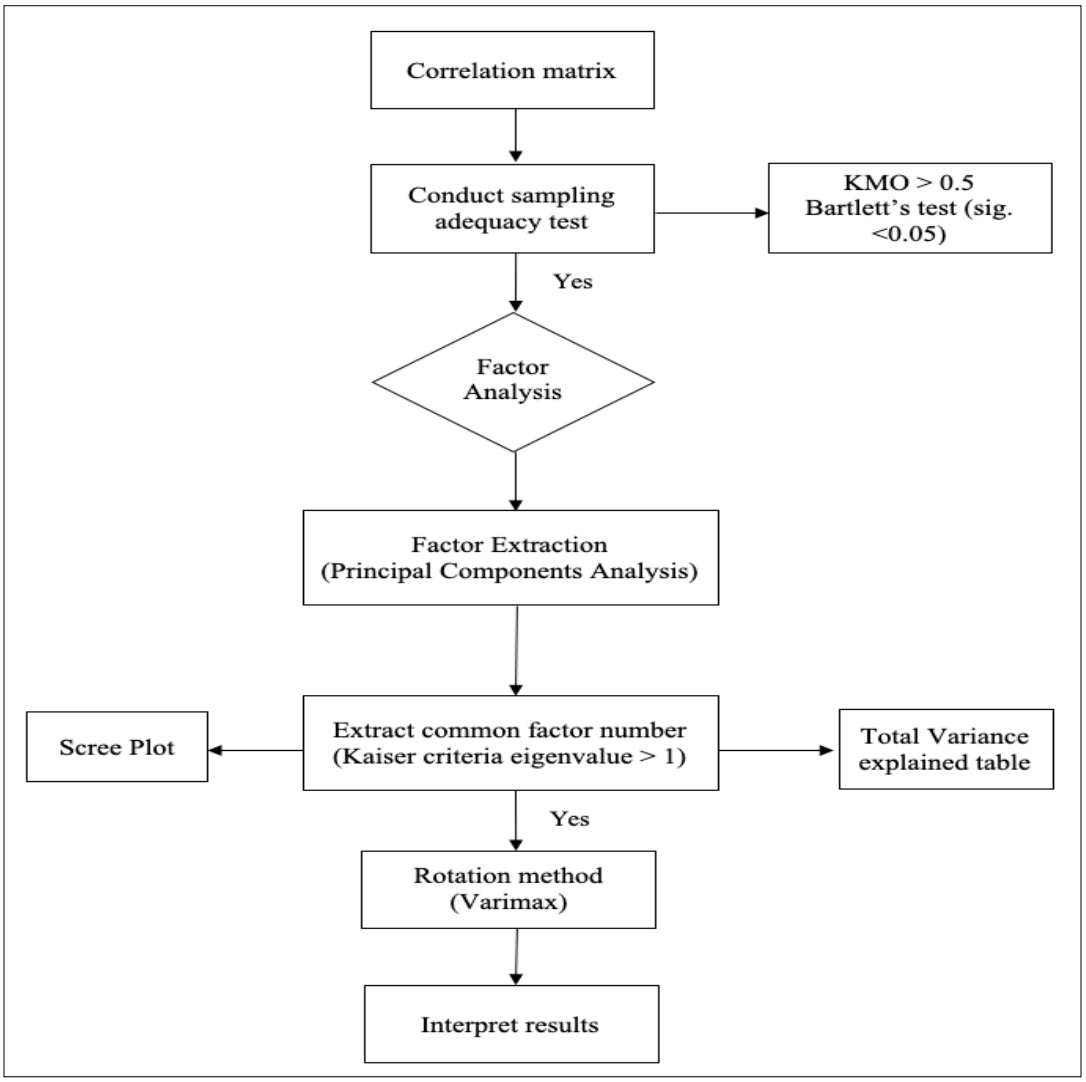
Figure 8.1 shows the flow chart of Factor Analysis procedures. The steps to implement Factor Analysis are described next.

Assessing the correlation between variables is the first step. The correlation matrix is examined so as to identify variables that correlate well for forming common factors.

Before conducting Factor Analysis, it is crucial to measure the suitability of the data for this purpose. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy test is used for this purpose. Large values of the KMO indicate that Factor Analysis is appropriate. Also, the strength of the relationship between variables is a consideration in

Factor Analysis. This is examined by performing Bartlett's test. When Bartlett's test is statistically significant, Factor Analysis is appropriate. Further information on testing data adequacy is provided in a later section of this chapter.

***Figure 8. 1: Flow of Factor Analysis Procedures***




Next, factors are extracted and an efficient number of factors are identified for further analysis using rotation method. In this research, the Principal component analysis was used for factor identification and Varimax rotation method provides support in the study of factor extraction and quantification of factor loadings.

Factor Analysis, as a multivariate technique, requires the examination of correlation matrices of the original variables. It uses the correlations between the variables to create a new set of a more closely related set of variables. The sharing of common factors is due to the strength of the correlation between variables.

The varimax rotation method was used to determine which variables are loaded on which factor (sometime called cluster). The number of factors that are considered for inclusion into the model depends on the magnitude of the proportion of the total variance explained by each factor or eigenvalues. In practice, those factors having an eigenvalue of greater than or equal to one are included. These eigenvalues are also plotted with the variables in a two-dimensional plot called the 'Scree Plot'. The cut-off point indicated by the scree plot determines the number of factors that should be included in the factor extraction process.

Following the determination of factors, factor loadings of variables can be studied. These factor loadings give us an idea about how much a variable has contributed to the factor; the larger the factor loading, the more the variable has contributed to that factor (Field, 2009).

Factor loadings define the strength of the relationships when interpreting the factors. In a model structure, the indicator variables that have lightly loaded are discarded and the highest loadings are retained.

Table 8.1 is an example of rotated factor loadings. If we sort the loadings in order of magnitude and blank out the lower values of loadings, the distinguishable factors (clusters) should emerge.

*Table 8. 1: Example Rotated Factor Loadings with Lower Values Blocked*

| Variables | Factor 1 loading | Factor 2 loading |
|---|---|---|
| **X1** | 0.9 | |
| **X3** | 0.6 | |
| **X5** | 0.5 | |
| **X2** | | 0.8 |
| **X4** | | 0.6 |

It can be seen from the above table that X1, X3, and X5 form a cluster under factor 1, and X2 and X4 create a second cluster with factor 2. That is, the variables X1, X3, and X5 are explained by factor 1, and the variables X2 and X4 are explained by factor 2. Thus, five observed variables have been explained by two factors.

## 8.4 Reliability of Scales

A reliability measure used to ensure that the Likert scale applied as the survey instrument is consistently reflecting the construct it is measuring. Cronbach's Alpha (α) is the most commonly used scale of reliability (Field, 2009). Therefore, it was adopted to test whether the respondents responded to all of the questions (53) variables consistently.

Cronbach's Alpha values range from zero to one. The higher the value (close to one), the more reliable the scale is. The Cronbach's Alpha for the 53 independent variables in this research is 0.978 (Table 8.2). This value suggests that the research instrument used is reliable.

*Table 8. 2: Reliability of Scale*

| Cronbach's Alpha | Number of Items |
|---|---|
| 0.978 | 53 |

## 8.5 Analysis Version 1-1: Examining the Underlying Dimensions of Cost Overrun

### 8.5.1 Sample Adequacy

We started our Factor Analysis by deriving the correlation matrix of 43 variables noted earlier. The correlation matrix is presented in Appendix B. The matrix was examined for the development of a factor model. To share common factors, there should be a high correlation between variables. But, using a relatively large correlation matrix such as the one we have, it is difficult to assess by visual examination the adequacy of the sample for Factor Analysis.

The Kaiser-Meyer-Olkin (KMO) measure of sample adequacy is an index that is sometimes used by researchers to observe if the data are suitable for Factor Analysis or not. The KMO statistic varies between 0 and 1. Small values of the KMO indicate that Factor Analysis may not be a good approach. Kaiser (1974) suggests accepting values greater than 0.5. He characterizes measures between 0.5 and 0.7 as mediocre but acceptable, values between 0.7 and 0.8 as good, values between 0.8 and 0.9 as great, and value above 0.9 as superb. Table 8.3 shows that the KMO was 0.662, which is in a mediocre but acceptable range.

Another test used for assessing the suitability of data for Factor Analysis is the Bartlett Test of Sphericity. It focuses on testing the hypothesis that the correlation matrix is an identity matrix. If it is an identity matrix, that means there are no correlations between the observed variables, which indicate that the data are unsuitable for Factor Analysis (Field, 2009). The value of Bartlett's test should be large and the associated significance level should be less than 0.05 for Factor Analysis to be appropriate (Field, 2009). In this

research, as shown in Table 8.3, the value of Bartlett's test of sphericity is (2933.042) and the associated significance level is small (0.000), suggesting that the correlation matrix is not an identity matrix. Therefore, the data are suitable for Factor Analysis.

***Table 8. 3: KMO and Bartlett's Test (Analysis Version 1-1)***

| Kaiser-Meyer-Olkin Measure of Sampling Adequacy | | 0.622 |
|---|---|---|
| Bartlett's Test of Sphericity | Approx. Chi-Square | 2933.042 |
| | df | 903 |
| | Sig. | 0.000 |

The data collected meet the requirements of both the KMO and Bartlett's tests; hence Factor Analysis has been considered a suitable statistical tool for this research.

### 8.5.2 Factor Extraction

The objective of factor extraction is to determine how many factors (components) should be retained in the analysis. Principal Components Analysis (PCA) was used to identify the underlying factors. The eigenvalue of each factor is an indication of the substantive importance of that factor. Therefore, in an analysis only factors with large eigenvalues are retained. Kaiser (1974) recommended that all factors with eigenvalues greater than one should be retained and SPSS by default uses this criterion. Table 8.4 presents the factor extraction results.

*Table 8. 4: Factor Extraction Results (Analysis Version 1-1)*

| Factor | Initial Eigen values | | |
|---|---|---|---|
| | Total | % of Variance | Cumulative % |
| **1** | **22.885** | **53.22** | **53.22** |
| **2** | **2.366** | **5.503** | **58.724** |
| **3** | **1.979** | **4.603** | **63.327** |
| **4** | **1.838** | **4.274** | **67.601** |
| **5** | **1.473** | **3.427** | **71.028** |
| **6** | **1.419** | **3.299** | **74.327** |
| **7** | **1.18** | **2.745** | **77.072** |
| **8** | **1.037** | **2.412** | **79.483** |
| 9 | 0.886 | 2.06 | 81.543 |
| 10 | 0.805 | 1.872 | 83.415 |
| 11 | 0.744 | 1.73 | 85.145 |
| 12 | 0.735 | 1.71 | 86.855 |
| 13 | 0.596 | 1.385 | 88.24 |
| 14 | 0.578 | 1.345 | 89.586 |
| 15 | 0.523 | 1.217 | 90.803 |
| 16 | 0.481 | 1.119 | 91.922 |
| 17 | 0.458 | 1.065 | 92.987 |
| 18 | 0.409 | 0.952 | 93.939 |
| 19 | 0.359 | 0.836 | 94.774 |
| 20 | 0.303 | 0.704 | 95.479 |
| 21 | 0.281 | 0.653 | 96.131 |
| 22 | 0.238 | 0.553 | 96.685 |
| 23 | 0.222 | 0.516 | 97.201 |
| 24 | 0.199 | 0.463 | 97.664 |
| 25 | 0.164 | 0.381 | 98.045 |
| 26 | 0.157 | 0.364 | 98.41 |
| 27 | 0.113 | 0.264 | 98.673 |
| 28 | 0.1 | 0.232 | 98.905 |
| 29 | 0.088 | 0.204 | 99.109 |
| 30 | 0.075 | 0.175 | 99.284 |
| 31 | 0.059 | 0.137 | 99.422 |
| 32 | 0.054 | 0.125 | 99.547 |
| 33 | 0.048 | 0.111 | 99.658 |
| 34 | 0.041 | 0.094 | 99.753 |
| 35 | 0.033 | 0.076 | 99.829 |
| 36 | 0.026 | 0.061 | 99.89 |
| 37 | 0.019 | 0.044 | 99.934 |
| 38 | 0.014 | 0.033 | 99.968 |
| 39 | 0.007 | 0.017 | 99.985 |
| 40 | 0.003 | 0.007 | 99.992 |

| | | | |
|---|---|---|---|
| 41 | 0.002 | 0.005 | 99.998 |
| 42 | 0.001 | 0.002 | 100 |
| 43 | 0 | 0 | 100 |

The total variance explained by each factor is listed in the first column of the initial Eigenvalues. In the second column, '% of variance' shows the percentage of the total variance attributable to each factor. For instance, factor 1 has a variance of 22.885, which accounts for 53.22% of the total variance. The results also show that 79.483% of the total variance is attributable to the first 8 factors. The remaining 35 factors together account for only 20.512% of the total variance. Thus, the first 8 factors have an eigenvalue greater than one which are the underlying (extracted) factors. Therefore, the table confirms that an 8 factors model should be sufficient for the research model.

### 8.5.3 Communality

Communality is the proportion of common variance within a variable. In SPSS software, before the extraction process, the initial communality is assumed to be one for all variables (all variance is common). After extraction, some of the factors are retained, others are dismissed. However, the true communality of a variable has been calculated by squaring the factor loadings of the concerned variable and adding them together. Now, the amount of variation in each variable explained by the factors is the communality, as shown in Table 8.5. For instance, the calculated communality for variable 1 shown in the third column means 82% of variable 1 can be predicted based on the knowledge of the underlying factors.

*Table 8. 5: Communalities (Analysis Version 1-1)*

| Variable | Initial | Extraction |
|---|---|---|
| **V1** | 1.000 | 0.821 |
| **V2** | 1.000 | 0.664 |
| **V3** | 1.000 | 0.787 |
| **V4** | 1.000 | 0.876 |
| **V5** | 1.000 | 0.759 |
| **V6** | 1.000 | 0.731 |
| **V7** | 1.000 | 0.639 |
| **V8** | 1.000 | 0.787 |
| **V9** | 1.000 | 0.878 |
| **V10** | 1.000 | 0.786 |
| **V11** | 1.000 | 0.702 |
| **V12** | 1.000 | 0.683 |
| **V13** | 1.000 | 0.827 |
| **V14** | 1.000 | 0.867 |
| **V15** | 1.000 | 0.770 |
| **V16** | 1.000 | 0.852 |
| **V17** | 1.000 | 0.871 |
| **V18** | 1.000 | 0.796 |
| **V19** | 1.000 | 0.744 |
| **V20** | 1.000 | 0.600 |
| **V21** | 1.000 | 0.856 |
| **V22** | 1.000 | 0.796 |
| **V23** | 1.000 | 0.809 |
| **V24** | 1.000 | 0.886 |
| **V25** | 1.000 | 0.743 |
| **V26** | 1.000 | 0.865 |
| **V27** | 1.000 | 0.783 |
| **V28** | 1.000 | 0.865 |
| **V29** | 1.000 | 0.833 |
| **V30** | 1.000 | 0.762 |
| **V31** | 1.000 | 0.806 |
| **V32** | 1.000 | 0.812 |
| **V33** | 1.000 | 0.834 |
| **V34** | 1.000 | 0.834 |
| **V35** | 1.000 | 0.709 |
| **V36** | 1.000 | 0.860 |
| **V37** | 1.000 | 0.603 |
| **V38** | 1.000 | 0.810 |
| **V39** | 1.000 | 0.857 |
| **V40** | 1.000 | 0.832 |
| **V41** | 1.000 | 0.884 |

| | | |
|---|---|---|
| **V42** | 1.000 | 0.820 |
| **V43** | 1.000 | 0.878 |

### 8.5.4 Eigenvalue

The eigenvalue individually was calculated from the factor matrix by squaring and adding the loadings under each factor, as shown in Table 8.6. These operations were carried out using SPSS, which can handle a database as large and as complicated as this one. As a rule of thumb, the number of factors was determined by the number corresponding to eigenvalue equal to or greater than one (Table 8.6).

Table 8.6 presents the eigenvalue for each factor before extraction, after extraction, and after rotation. The SPSS extracted all factors with eigenvalues greater than 1, providing eight factors (components) using the principal component analysis (PCA) technique (recommended for this purpose).

In the initial solution, the first factor accounted for 23.6 % of the total variance after rotation and the subsequent factors explained variance with gradually diminishing percentages. The eighth factor explained 3.4% of the variance.

Table 8.6 also shows that the first eight factors, which have an eigenvalue greater than one, explain 79.48 % of the total variance. The remaining factors, representing the remaining percent of the variance, have eigenvalue less than one and were not included. Therefore, the first eight factors were included in the factor model and their factor loadings were calculated.

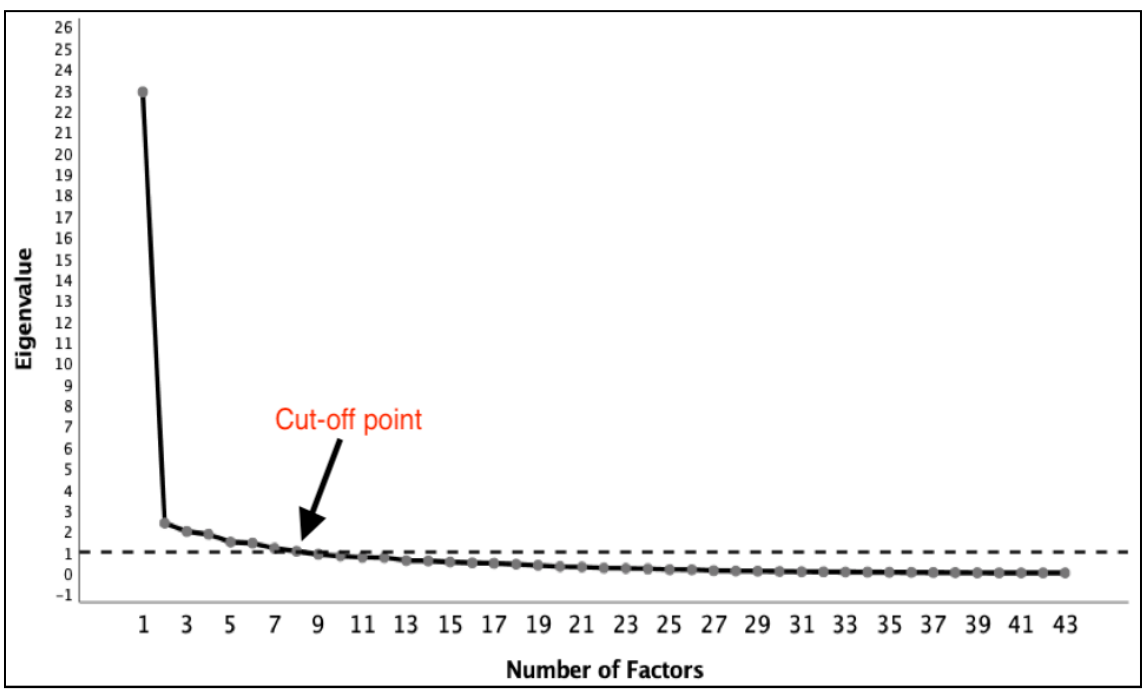
*Table 8. 6: Total variance Explained (Analysis Version 1-1)*

| Component | Initial Eigenvalues | | | Extraction Sums of Squared | | | Rotation Sums of Squared | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| **1** | 22.885 | 53.220 | 53.220 | 22.885 | 53.220 | 53.220 | 10.153 | 23.612 | 23.612 |
| **2** | 2.366 | 5.503 | 58.724 | 2.366 | 5.503 | 58.724 | 6.102 | 14.191 | 37.803 |
| **3** | 1.979 | 4.603 | 63.327 | 1.979 | 4.603 | 63.327 | 5.146 | 11.968 | 49.770 |
| **4** | 1.838 | 4.274 | 67.601 | 1.838 | 4.274 | 67.601 | 3.427 | 7.970 | 57.740 |
| **5** | 1.473 | 3.427 | 71.028 | 1.473 | 3.427 | 71.028 | 3.224 | 7.497 | 65.237 |
| **6** | 1.419 | 3.299 | 74.327 | 1.419 | 3.299 | 74.327 | 2.722 | 6.331 | 71.568 |
| **7** | 1.180 | 2.745 | 77.072 | 1.180 | 2.745 | 77.072 | 1.936 | 4.502 | 76.070 |
| **8** | 1.037 | 2.412 | 79.483 | 1.037 | 2.412 | 79.483 | 1.468 | 3.413 | 79.483 |
| **9** | 0.886 | 2.060 | 81.543 | | | | | | |
| **10** | 0.805 | 1.872 | 83.415 | | | | | | |
| **11** | 0.744 | 1.730 | 85.145 | | | | | | |
| **12** | 0.735 | 1.710 | 86.855 | | | | | | |
| **13** | 0.596 | 1.385 | 88.240 | | | | | | |
| **14** | 0.578 | 1.345 | 89.586 | | | | | | |
| **15** | 0.523 | 1.217 | 90.803 | | | | | | |
| **16** | 0.481 | 1.119 | 91.922 | | | | | | |
| **17** | 0.458 | 1.065 | 92.987 | | | | | | |
| **18** | 0.409 | 0.952 | 93.939 | | | | | | |
| **19** | 0.359 | 0.836 | 94.774 | | | | | | |
| **20** | 0.303 | 0.704 | 95.479 | | | | | | |
| **21** | 0.281 | 0.653 | 96.131 | | | | | | |
| **22** | 0.238 | 0.553 | 96.685 | | | | | | |
| **23** | 0.222 | 0.516 | 97.201 | | | | | | |
| **24** | 0.199 | 0.463 | 97.664 | | | | | | |
| **25** | 0.164 | 0.381 | 98.045 | | | | | | |
| **26** | 0.157 | 0.364 | 98.410 | | | | | | |
| **27** | 0.113 | 0.264 | 98.673 | | | | | | |
| **28** | 0.100 | 0.232 | 98.905 | | | | | | |
| **29** | 0.088 | 0.204 | 99.109 | | | | | | |
| **30** | 0.075 | 0.175 | 99.284 | | | | | | |
| **31** | 0.059 | 0.137 | 99.422 | | | | | | |
| **32** | 0.054 | 0.125 | 99.547 | | | | | | |
| **33** | 0.048 | 0.111 | 99.658 | | | | | | |
| **34** | 0.041 | 0.094 | 99.753 | | | | | | |
| **35** | 0.033 | 0.076 | 99.829 | | | | | | |
| **36** | 0.026 | 0.061 | 99.890 | | | | | | |
| **37** | 0.019 | 0.044 | 99.934 | | | | | | |
| **38** | 0.014 | 0.033 | 99.968 | | | | | | |
| **39** | 0.007 | 0.017 | 99.985 | | | | | | |
| **40** | 0.003 | 0.007 | 99.992 | | | | | | |
| **41** | 0.002 | 0.005 | 99.998 | | | | | | |
| **42** | 0.001 | 0.002 | 100.000 | | | | | | |
| **43** | 0.000 | 0.000 | 100.000 | | | | | | |

### 8.5.5 Scree Plot

A scree plot can be used to examine if the eigenvalue is large enough to represent a meaningful factor or not (Field, 2009). This plot shows a graph of the number of factors on the (X-axis) against the eigenvalue (Y-axis). The point of inflection of this curve represents the cut-off point for selecting factors to be retained in the analysis (Field, 2009). It can be noted that in the scree plot shown in Figure 8.2, there is a negative slope decreasing from factor number 1 through to factor number 3, and then it seemed to have leveled off and sloped further down and leveled off again. The point that corresponds to the scree plot could identify the number of factors that should emerge (Syed and Khan, 2001). However, there may be the necessity sometimes to employ subjective judgment to identify the level-off point.

***Figure 8. 2: Factor Scree Plot (Analysis Version 1-1)***



According to Kim and Mueller (2011), *"At this stage of the analysis, one should not be too concerned with whether the factors extracted are interpretable or meaningful. The chief concern is whether a smaller number of factors can account for a much larger number of variables."*

Here, in this version 1-1 of Factor Analysis, the number of factors extracted is eight from a total of forty-three (43) observed variables. These are based on principal components analysis and varimax rotation with an assumed rule of thumb to determine the number of factors from the eigenvalues to be equal to or greater than one.

### 8.5.6 Factor Rotation

The fact that after the extraction process, most of the variables that have high loading will be clustered around the first factor, and the other factors will have a small loading, which makes interpretation difficult. Therefore, the rotation process technique is used to discriminate between factors. The main goal of the rotation is to attain an optimal simple structure that attempts to have each variable load on as few factors as possible but to maximize the number of high loadings on each variable (Field, 2009). After the rotation process, the software outputs show which variables relate to which factors, and therefore that makes the output more reliable to understand.

As with extraction, there are many rotation methods. The varimax method is the most commonly used method for a rotation where the objective is to minimize the number of variables that have high loadings on a factor (Costello and Osborne, 2011; Syed and Khan, 2001).

After rotation and to assist interpretation, the rotated factor loadings are to be sorted with the intention that these variables with high loadings appear together on the same factor. The factors extracted are those with the loadings greater than (0.5), representing the significant associating variables, whereas less than (0.5) are considered unsubstantial. Field (2009) further suggests that some components need to be dropped from the analysis if the factor loading is not much above (0.5).

Table 8.7 shows the rotated factor matrix where the eight factors (groups of clusters) are portrayed. For full details of theoretical aspects of Factor Analysis and factor rotation, the reader is referred to Field (2009).

*Table 8. 7: Extracted Factors Using Varimax Method (Analysis Version 1-1)*

| % of variance | Variable | Factor | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 23.61 | V41 | 0.829 | | | | | | | |
| | V26 | 0.784 | | | | | | | |
| | V24 | 0.776 | | | | | | | |
| | V43 | 0.763 | | | | | | | |
| | V25 | 0.706 | | | | | | | |
| | V22 | 0.702 | | | | | | | |
| | V14 | 0.694 | | | | | | | |
| | V39 | 0.666 | | | | | | | |
| | V42 | 0.642 | | | | | | | |
| | V29 | 0.639 | | | | | | | |
| | V36 | 0.620 | | | | | | | |
| | V40 | 0.611 | | | | | | | |
| | V28 | 0.610 | | | | | | | |
| | V19 | 0.564 | | | | | | | |
| 14.19 | V4 | | 0.867 | | | | | | |
| | V15 | | 0.730 | | | | | | |
| | V9 | | 0.689 | | | | | | |
| | V5 | | 0.631 | | | | | | |
| | V16 | | 0.589 | | | | | | |
| | V8 | | 0.574 | | | | | | |
| | V12 | | 0.533 | | | | | | |
| | V31 | | 0.509 | | | | | | |
| 11.96 | V17 | | | 0.773 | | | | | |
| | V38 | | | 0.694 | | | | | |
| | V30 | | | 0.665 | | | | | |
| | V3 | | | 0.651 | | | | | |
| | V1 | | | 0.549 | | | | | |
| | V23 | | | 0.526 | | | | | |
| 7.97 | V33 | | | | 0.778 | | | | |
| | V35 | | | | 0.648 | | | | |
| | V11 | | | | 0.627 | | | | |
| 7.49 | V37 | | | | | 0.686 | | | |
| | V18 | | | | | 0.504 | | | |
| 6.33 | V7 | | | | | | 0.728 | | |
| | V20 | | | | | | 0.579 | | |
| | V13 | | | | | | 0.531 | | |
| 4.5 | V27 | | | | | | | 0.660 | |
| | V6 | | | | | | | 0.621 | |
| 3.41 | V10 | | | | | | | | 0.825 |
| Cumulative variance explained 79.48% | | | | | | | | | |

## 8.6 Descriptive Results (Factor Analysis Version 1-1)

### 8.6.1 Correlation Matrix

During the analysis of the questionnaire data, the correlation matrix of the variables relevant to the cost overrun was computed, as shown in Appendix B. The variables are correlated to each other and their acceptability was verified through the Kaiser-Meyer-Olkin (KMO) values. These were found to be in the acceptable range for factor extraction. Therefore, we proceeded with the analysis on the basis of the KMO values indicative of a high degree of correlation amongst the independent variables. Also, as noted earlier, another test, namely the Bartlett's test showed acceptable results. Therefore, the data were assessed to be suitable for Factor Analysis.

### 8.6.2 Rotated Factor Matrix

The aim of pursuing procedures like rotation was to attain a simple structure so that the interpretation of such factors became easier. This was done first using Varimax rotation. This rotation method, however, did not change the degree of fit between data and factor structure. This structure was obtained by following the definition of the simple structure, as discussed earlier. The lower magnitude ($\leq 0.5$) loadings were blocked out in the rotated matrix table for showing the clusters more distinctly.

### 8.6.3 Varimax Rotated Solution

The varimax rotated factor matrix yielded the following factors, which were first interpreted and later used for calibrating the logistic regression models described in the next chapter.

**Factor 1**

The rotated factor matrix gives the following 14 observed variables under factor 1:

- Unaddressed overtime work or multiple shifts that were not included in the base estimate
- Construction variations due to equipment selection
- Late delivery of materials & equipment at the construction site
- Currency fluctuations
- Equipment Selection Changes
- Shortage of materials & equipment on site
- Delays in sending important documents to the construction site (e.g. drawings, design changes)
- Inappropriate and inadequate procurement (e.g. payment terms, pricing)
- Bankruptcy of subcontractors and vendors during construction work
- Lack of staff at the time of construction
- Absence of a detailed Estimate Plan
- Shortage of contingency and management reserve funds
- Shortage of skilled labor
- Poor site management

It is useful to note that the above observed variables are the outcome of an international questionnaire survey designed to extract feedback from experts to identify the factors that cause cost overrun in highway transportation infrastructure projects.

In Table 8.8, factor 1 cluster consisting of the original observed variables and factor loading for each variable are shown.

*Table 8. 8: Factor 1 Cluster (Analysis Version 1-1)*

| Factor 1 (Cluster) | | |
|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** |
| **V41** | Unaddressed overtime work or multiple shifts that were not included in the base estimate | 0.829 |
| **V26** | Construction variations due to equipment selection | 0.784 |
| **V24** | Late delivery of materials & equipment at the construction site | 0.776 |
| **V43** | Currency fluctuations | 0.763 |
| **V25** | Equipment Selection Changes | 0.706 |
| **V22** | Shortage of materials & equipment on site | 0.702 |
| **V14** | Delays in sending important documents to construction site (e.g. drawings, design changes) | 0.694 |
| **V39** | Inappropriate and inadequate procurement (e.g. payment terms, pricing) | 0.666 |
| **V42** | Bankruptcy of subcontractors and vendors during construction work | 0.642 |
| **V29** | Lack of staff at the time of construction | 0.639 |
| **V36** | Absence of a detailed Estimate Plan | 0.620 |
| **V40** | Shortage of contingency and management reserve funds | 0.611 |
| **V28** | Shortage of skilled labor | 0.610 |
| **V19** | Poor site management | 0.564 |
| ***Percentage of Variance 23.61%*** | | |

The first factor accounts for 23.6% of the variance, according to the analysis. The factor loadings are all positive within the cluster. A close look at the responses of each of the observed variables reveals the attitude of the respondents about the cost overrun factors. Here, the impact of these factors on infrastructure projects is appreciated. There are 14 variables that have clustered together to produce a single underlying entity. These 14 observed variables may not be important individually, but they acted like proxy variables to produce a factor explaining the causes of cost overruns in infrastructure projects. Therefore, in accordance with the clustered variables, we interpret and label the factor as [**Financial, Materials, Labor, and Equipment related factors**]. The impact of this factor

is examined later in the logistic regression models intended to find its impact and significance regarding the phenomenon of cost overrun.

**Factor 2**

From the rotated factor matrix, factor 2 is found to consist of the following eight observed variables:

- Complexity of the project (e.g., project size, project type, scope of work)
- Type of construction contract (e.g., unit price contract)
- Changes by owner on the completion date of the project
- Design changes during construction work
- Unnecessary practices, specifications, and procedures
- Design errors that represent insufficient deliverables
- Unrealistic project scheduling
- Delays and approval of shop drawings and installation procedures.

Table 8.9 shows the factor 2 cluster with a description of the observed variables.

***Table 8. 9: Factor 2 Cluster (Analysis Version 1-1)***

| **Cluster 2** | | |
|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** |
| **V4** | Complexity of the project (e.g., Project size, Project type, scope of work) | 0.867 |
| **V15** | Type of construction contract (e.g., unit price contract) | 0.730 |
| **V9** | Changes by owner on the completion date of the project | 0.689 |
| **V5** | Design changes during construction work | 0.631 |
| **V16** | Unnecessary practices, specifications, and procedures | 0.589 |
| **V8** | Design errors that represent insufficient deliverables | 0.574 |
| **V12** | Unrealistic project scheduling | 0.533 |
| **V31** | Delays and approval of shop drawings and installation procedures | 0.509 |
| **Percentage of Variance 14.19%** | | |

The percentage of variance explained by the second factor is 14.19%. This factor was represented by eight variables and most of the variables are related to design, construction, and scheduling. Therefore, it is reasonable to label this factor as **[Design, Construction, and Scheduling related factors]**.

**Factor 3**

The third factor explains 11.96% of the total variance. It includes six observed variables noted next. Please see Table 8.10 for factor loadings. The variables are: replacing unsatisfactory subcontractors, economic and financial factors, labor strikes & vendor strikes, change in regulations, changes in government funding policies, and damages in materials and equipment in transit to the construction site. Table 8.10 presents factor 3 variables that give us the cluster of **[Policy, budget, and operation-related factors]**.

***Table 8. 10: Factor 3 Cluster (Analysis Version 1-1)***

| **Cluster 3** | | |
|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** |
| **V17** | Replacing unsatisfactory subcontractors from site by hiring new subcontractors | 0.773 |
| **V38** | Economic and financial factors | 0.694 |
| **V30** | labor strikes & vendor strikes | 0.665 |
| **V3** | Change in regulations | 0.651 |
| **V1** | Changes in government funding policies | 0.549 |
| **V23** | Damages in materials and equipment in transit to the construction site | 0.526 |
| **Percentage of Variance 11.96%** | | |

**Factor 4**

The contribution of factor 4 explained 7.97% of the total variance (Table 8.11). The variables are government/municipal approvals, the approved budget was too low, and delays related to the owner. Therefore, the factor is labeled as **[Decision Making related factors]**. Table 8.11 shows Factor Analysis results, including factor loadings.

***Table 8. 11: Factor 4 Cluster (Analysis Version 1-1)***

| Cluster 4 | | |
|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** |
| **V33** | Government/ Municipal Approvals | 0.778 |
| **V35** | The approved budget was too low | 0.648 |
| **V11** | Delays related to owner or owner representative (e.g. stop work) | 0.627 |
| **Percentage of Variance 7.97%** | | |

**Factor 5**

There are two observed variables clustered around factor 5 and accounted for 7.49% of the total variance (Table 8.12). The variables are noted next and their factor loading are presented in the table: changes in prices of items that have already been approved, and the second variable is delayed by a subcontractor. This cluster can be labeled as **[Material Price Fluctuation and Work Performance related factors]**.

*Table 8. 12: Factor 5 Cluster (Analysis Version 1-1)*

| Cluster 5 | | |
|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** |
| **V37** | Changes in prices of items that have already been approved | 0.686 |
| **V18** | Delay by subcontractor | 0.504 |
| **Percentage of Variance 7.49%** | | |

**Factor 6**

Table 8.13 presents variables, factor loadings, and percent of the variance. All three variables represent unforeseen or unexpected circumstances. The contribution of factor 6 explains 6.33% of the total variance. The nature of variables suggests that this factor can be labeled as **[Unexpected related factors]**.

*Table 8. 13: Factor 6 Cluster (Analysis Version 1-1)*

| Cluster 6 | | |
|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** |
| **V7** | Unexpected technical problem | 0.728 |
| **V20** | Unexpected weather conditions | 0.579 |
| **V13** | Acceleration to maintain schedule | 0.531 |
| **Percentage of Variance 6.33%** | | |

**Factor 7**

Table 8.14 presents variables and statistical information. This factor includes two observed variables: defective materials and re-work due to construction errors. This cluster explained 4.5% of the total variance. On the basis of the nature of variables, we interpret and label the factor as [**Quality Control and Quality Assurance**].

***Table 8. 14: Factor 7 Cluster (Analysis Version 1-1)***

| Cluster 7 | | |
|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** |
| **V27** | Defective materials | 0.660 |
| **V6** | Re-work due to the construction errors | 0.621 |
| **Percentage of Variance 4.5%** | | |

**Factor 8**

Under this factor, only one original variable has been clustered namely scope changes by the owner during construction (Table 8.15). This cluster represents 3.41% of the total variance and the factor loading has been calculated as 0.825, as shown in the table. This cluster can be labeled as **[Owner/client related factor]**.

***Table 8. 15: Factor 8 Cluster (Analysis Version 1-1)***

| Cluster 8 | | |
|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** |
| **V10** | Scope changes by owner during construction | 0.825 |
| **Percentage of Variance 3.41%** | | |

### 8.6.4 Interpretation of Factors

The following table gives a summary of the retained factors (Common factors of cost overrun) after interpretation and were used in a logistic regression analysis to find the odds of cost overrun occurring or not for each of the underlying factors.

***Table 8. 16: Interpreted Factors (Analysis Version 1-1)***

| Factor | % of variance | Factor Label |
|---|---|---|
| **Factor 1** | 23.61 | Financial, Materials, labor, and Equipment related factors |
| **Factor 2** | 14.19 | Design, construction, and scheduling related factors |
| **Factor 3** | 11.96 | Policy, budget, and operation related factors |
| **Factor 4** | 7.97 | Decision making related factors |
| **Factor 5** | 7.49 | Material prices fluctuation and work performance factors |
| **Factor 6** | 6.33 | Unexpected related factors |
| **Factor 7** | 4.5 | Quality control & quality assurance related factors |
| **Factor 8** | 3.41 | Owner/client related factor |

## 8.7 Analysis Version 1-2: Examining the Underlying Dimensions of Cost Overrun

As noted earlier, out of 53 variables, 10 variables were analyzed separately in order not to affect the results of 43 variables. In this second Version of analysis 1 (i.e., Version 1-2), the sample adequacy was studied to measure the suitability of the data for Factor Analysis. The results obtained are shown in Table 8.17.

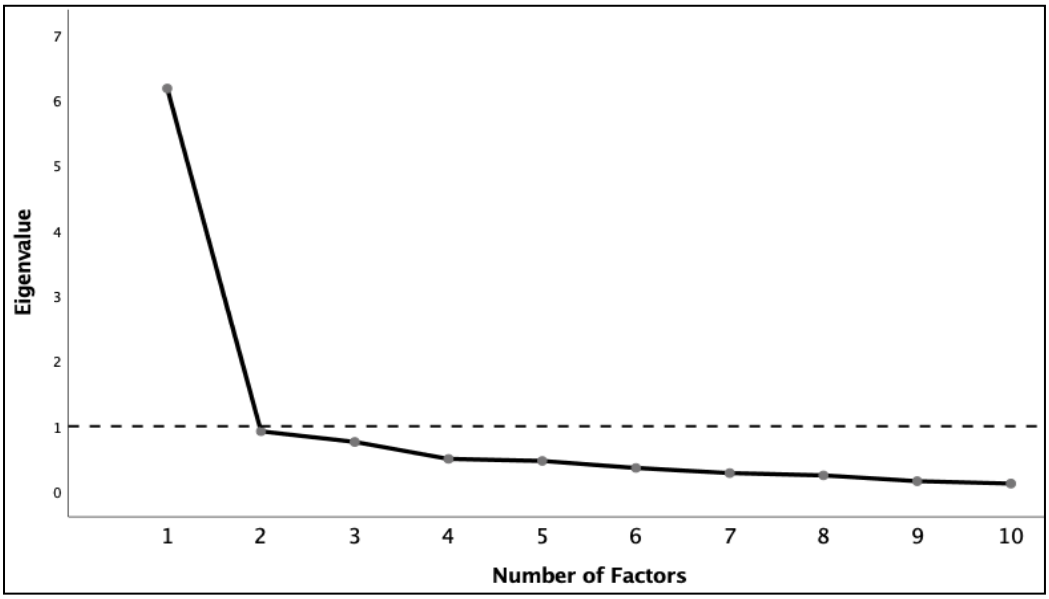
Since the KMO value is greater than 0.5 and the Bartlett test value is significant; therefore, the data are suitable for Factor Analysis. Kaiser (1974) suggests that the KMO value between 0.8 and 0.9 is great.

Based on the eigenvalue and scree plot shown in Figure 8.3, only one factor has an eigenvalue greater than one. This indicates that out of 10 variables, only one factor was extracted. However, the percentage of the variance of this factor was found to be 61.79%. Since only one factor was extracted, the solution cannot be rotated.

***Table 8. 17: KMO and Bartlett's Test (Analysis Version 1-2)***

| | | |
|---|---|---|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy | | 0.886 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 364.486 |
| | df | 45 |
| | Sig. | 0.000 |

***Figure 8. 3: Factor Scree Plot (Analysis Version 1-2)***



## 8.8 Analysis Version 2: Examining the Underlying Dimensions of Cost Overrun

Although Version 1 analyses are statistically sound and no guidelines are violated, for the sake of curiosity type of research, this second version of the Factor Analysis was carried out. For this purpose, based on guidance from literature, it was decided to remove the variables that showed a weak correlation of less than (0.3) (Field, 2009; Tabachnick and Fidell 2007). This process resulted in removing 22 variables from the analysis.

### 8.8.1 Sample Adequacy

Before conducting Factor Analysis, the Kaiser-Meyer-Olkin (KMO) measure of sample adequacy is used to measure the suitability of the data for Factor Analysis. Table 8.18 shows the KMO was 0.816 (KMO>0.5), which is in the acceptable range based on Kaiser's (1974) criteria. Also, the Bartlett Test of Sphericity was performed to measure the strength of the relationship between variables. The value of the Bartlett Test of Sphericity is 2189.718 and the associated significance level is 0.000 (P<0.05). These suggest that the correlating matrix is not an identity matrix. Therefore, the data of 31 variables are suitable for Factor Analysis.

***Table 8. 18: KMO and Bartlett's Test (Analysis Version 2)***

| Kaiser-Meyer-Olkin Measure of Sampling Adequacy | | 0.816 |
|---|---|---|
| Bartlett's Test of Sphericity | Approx. Chi-Square | 2189.718 |
| | df | 465 |
| | Sig. | 0.000 |

### 8.8.2 Factor Extraction

The Total Variance Explained results shown in Table 8.19 enabled the identification of significant factors. These factors are arranged in descending order, based on the explained variance. Additional explanation of the process of factor identification is provided next.

***Table 8. 19: Total Variance Explained (Analysis Version 2)***

| Component | Initial Eigenvalues | | | Extraction Sums of Squared | | | Rotation Sums of Squared | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 19.699 | 63.545 | 63.545 | 19.699 | 63.545 | 63.545 | 5.777 | 18.635 | 18.635 |
| 2 | 1.879 | 6.061 | 69.605 | 1.879 | 6.061 | 69.605 | 5.586 | 18.02 | 36.655 |
| 3 | 1.358 | 4.38 | 73.986 | 1.358 | 4.38 | 73.986 | 4.682 | 15.103 | 51.758 |
| 4 | 1.133 | 3.654 | 77.64 | 1.133 | 3.654 | 77.64 | 4.578 | 14.769 | 66.527 |
| 5 | 1.055 | 3.403 | 81.043 | 1.055 | 3.403 | 81.043 | 4.5 | 14.516 | 81.043 |
| 6 | 0.813 | 2.624 | 83.667 | | | | | | |
| 7 | 0.701 | 2.262 | 85.928 | | | | | | |
| 8 | 0.59 | 1.903 | 87.831 | | | | | | |
| 9 | 0.532 | 1.717 | 89.548 | | | | | | |
| 10 | 0.457 | 1.473 | 91.021 | | | | | | |
| 11 | 0.412 | 1.328 | 92.349 | | | | | | |
| 12 | 0.377 | 1.216 | 93.565 | | | | | | |
| 13 | 0.329 | 1.062 | 94.628 | | | | | | |
| 14 | 0.289 | 0.931 | 95.559 | | | | | | |
| 15 | 0.233 | 0.751 | 96.31 | | | | | | |
| 16 | 0.199 | 0.642 | 96.952 | | | | | | |
| 17 | 0.171 | 0.55 | 97.502 | | | | | | |
| 18 | 0.14 | 0.452 | 97.954 | | | | | | |
| 19 | 0.123 | 0.396 | 98.35 | | | | | | |
| 20 | 0.095 | 0.306 | 98.656 | | | | | | |
| 21 | 0.09 | 0.29 | 98.946 | | | | | | |
| 22 | 0.075 | 0.243 | 99.189 | | | | | | |
| 23 | 0.059 | 0.189 | 99.378 | | | | | | |
| 24 | 0.051 | 0.164 | 99.542 | | | | | | |
| 25 | 0.035 | 0.113 | 99.655 | | | | | | |
| 26 | 0.032 | 0.102 | 99.757 | | | | | | |
| 27 | 0.026 | 0.084 | 99.841 | | | | | | |
| 28 | 0.022 | 0.072 | 99.913 | | | | | | |
| 29 | 0.016 | 0.051 | 99.964 | | | | | | |
| 30 | 0.008 | 0.025 | 99.989 | | | | | | |
| 31 | 0.004 | 0.011 | 100 | | | | | | |

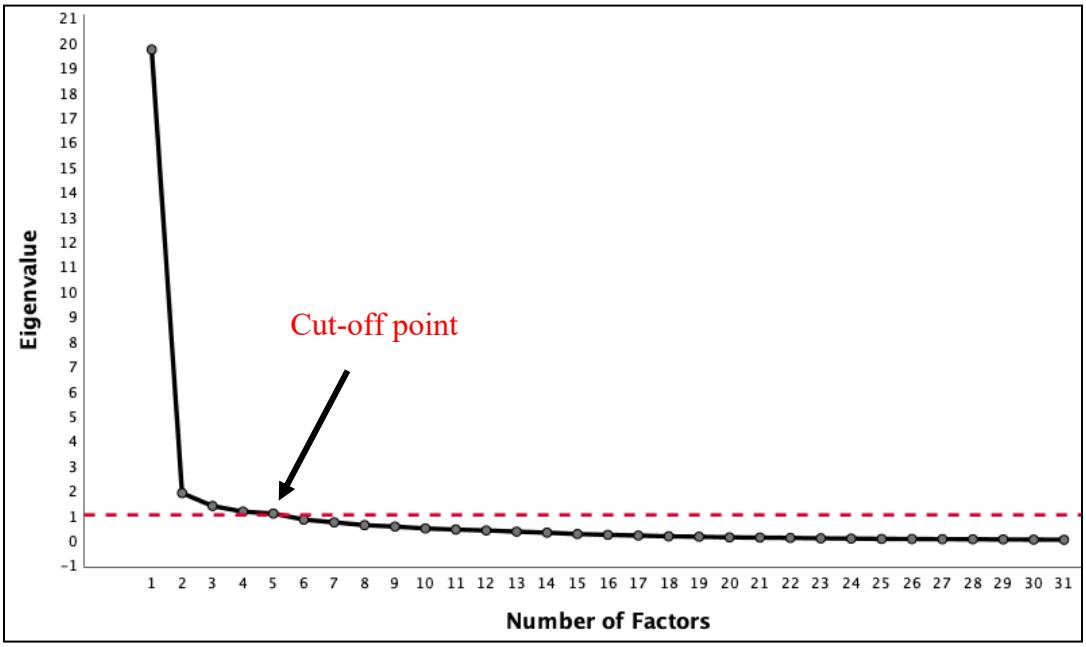
The table consists of three columns that show the eigenvalues and variance before extraction, after extraction, and after rotation. This table and the scree plot (Figure 8.4) were used to determine the number of extracted factors. Based on Kaiser's (1974) criteria, factors that have eigenvalues greater than one were extracted with the use of SPSS software. Therefore, five factors were identified with the use of the principal component analysis (PCA) technique.

In the initial eigenvalue column (before extraction), the first factor has a variance of 19.69, which accounts for 63.54% of the total variance. The results also show that 81.04% of the total variance is attributable to the first five factors. The remaining 26 factors together account for only 18.96% of the total variance. However, the table confirms that a five factors model should be sufficient for the research model.

After rotation, the first factor accounted for 18.63% of the variance and the last factor explained 14.51% of the variance. The results also show that 81.04% of the total variance is attributable to the first five factors. Therefore, the first five factors were included in the factor model and their factor loadings were calculated.

The eigenvalues and scree test (i.e., scree plot) are used to determine how many factors to retain. Figure 8.4 shows the number of extracted factors based on Kaiser's criterion. The point of inflection of this curve represents the cut-off point for selecting factors to be retained in the analysis (Field, 2009). As can be seen from the figure, there is a negative slope decreasing from factor 1 through to factor 4, and then it seemed to have leveled off and sloped further down and leveled off again. However, the number of factors extracted is five from a total of thirty-one observed variables. These are based on principal components analysis and varimax rotation.

*Figure 8. 4: Factor Scree Plot (Analysis Version 2)*




### 8.8.3 Factor Rotation

Given that factors are rotated for better interpretation, the Factor Analysis results are viewed on this basis. By rotating the factors, each factor will have a set of variables that are loaded highly. Table 8.20 represents the rotation results.

*Table 8. 20: Extracted Factors Using Varimax Method (Analysis Version 2)*

| % of variance | Variable | Factor | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| 18.63 | V3 | 0.776 | | | | |
| | V17 | 0.740 | | | | |
| | V38 | 0.735 | | | | |
| | V1 | 0.651 | | | | |
| | V16 | 0.634 | | | | |
| | V42 | 0.579 | | | | |
| | V14 | 0.564 | | | | |
| | V40 | 0.546 | | | | |
| 18.01 | V4 | | 0.837 | | | |
| | V15 | | 0.785 | | | |
| | V53 | | 0.748 | | | |
| | V5 | | 0.683 | | | |
| | V8 | | 0.656 | | | |
| | V9 | | 0.547 | | | |
| | V13 | | 0.521 | | | |
| | V31 | | 0.514 | | | |
| 15.10 | V21 | | | 0.799 | | |
| | V52 | | | 0.738 | | |
| | V6 | | | 0.515 | | |
| | V19 | | | 0.507 | | |
| 14.76 | V36 | | | | 0.688 | |
| | V46 | | | | 0.659 | |
| | V32 | | | | 0.603 | |
| | V18 | | | | 0.577 | |
| | V39 | | | | 0.572 | |
| | V2 | | | | 0.570 | |
| | V34 | | | | 0.555 | |
| 14.51 | V41 | | | | | 0.700 |
| | V26 | | | | | 0.666 |
| | V28 | | | | | 0.623 |
| | V24 | | | | | 0.599 |
| **Cumulative variance explained 81.04%** | | | | | | |

### 8.8.4 Varimax Rotated Solution

The varimax rotated factor matrix yielded the following factors, which were first interpreted and later used for calibrating the logistic regression models described in the next chapter.

**Factor 1**

From the rotated factor matrix, factor 1 is found to consist of eight observed variables, as shown in the following table.

***Table 8. 21: Factor 1 Cluster (Analysis Version 2)***

| Cluster 1 | | |
|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** |
| **V3** | Change in regulations | 0.776 |
| **V17** | Replacing unsatisfactory subcontractors from site by hiring new subcontractors | 0.740 |
| **V38** | Economic and financial factors | 0.735 |
| **V1** | Changes in government funding policies | 0.651 |
| **V16** | Unnecessary practices, specifications and procedures | 0.634 |
| **V42** | Bankruptcy of subcontractors and vendors during construction work | 0.579 |
| **V14** | Delays in sending important documents to construction site (e.g., drawings, design changes) | 0.564 |
| **V40** | Shortage of contingency and management reserve funds | 0.546 |
| ***Percentage of Variance 18.63%*** | | |

According to the analysis, the first cluster accounts for 18.63% of the variance. Therefore, in accordance with the clustered variables, we interpret and label the factor as **[Design, Construction, Financial, and Funding policy-related factors]**. The impact of this factor is examined later in the logistic regression models intended to find its impact and significance upon the concept of cost overrun.

**Factor 2, Factor 3, and Factor 4**

We have combined factor 2, factor 3, and factor 4 together because they represent inexperience or lack of qualifications of parties involved in the projects, as shown in Table

8.22. In fact, the lack of qualifications of project parties is one of the most causes that contribute to the cost overrun in transportation infrastructure projects.

However, factor 2 accounts for 18.01% of the variance, factor 3 accounts for 15.10% of the variance, and factor 4 accounts for 14.76% of the variance. The total variance is 47.87% for all factors combined together.

Table 8.22 presents factor 2, factor 3, and factor 4 variables and gives us the cluster of **[Inexperienced or lack of qualification of (designer, contractor, client, and project manager)]**.

***Table 8. 22: Factor 2, Factor 3, and Factor 4 Cluster (Analysis Version 2)***

| **Cluster 2** | | | |
|---|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** | **Variance Explained** |
| V4 | Complexity of the project (e.g., Project size, Project type, scope of work) | 0.837 | **18.01%** |
| V15 | Type of construction contract (e.g., unit price contract) | 0.785 | |
| V53 | Inexperienced Project Managers, Estimators and Planners | 0.748 | |
| V5 | Design changes during construction work | 0.683 | |
| V8 | Design errors that represent insufficient deliverables | 0.656 | |
| V9 | Changes by owner on the completion date of the project | 0.547 | |
| V13 | Acceleration to maintain schedule | 0.521 | |
| V31 | Delays and approval of shop drawings and installation procedures | 0.514 | |
| **Cluster 3** | | | |
| **Variable** | **Original Variable Descriptions** | **Factor Loading** | **Variance Explained** |
| V21 | Accidents due to poor site safety | 0.799 | **15.10%** |
| V52 | Quality assurance and quality control | 0.738 | |
| V6 | Re-work due to the construction errors | 0.515 | |
| V19 | Poor site management | 0.507 | |

| Cluster 4 | | | |
|---|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** | **Variance Explained** |
| V36 | Absence of a detailed Estimate Plan | 0.688 | **14.76%** |
| V46 | Overly high expectations | 0.659 | |
| V32 | Building Permit to the construction contractor | 0.603 | |
| V18 | Delay by subcontractor | 0.577 | |
| V39 | Inappropriate and inadequate procurement (e.g. payment terms, pricing) | 0.572 | |
| V2 | Deal termination due to changes in law, government policy or protocols | 0.570 | |
| V34 | Lack of expertise in setting the budget | 0.555 | |
| ***Cumulative variance explained*** | | | **47.87%** |

**Factor 5**

As shown in Table 8.23, factor 5 explains 14.51% of the total variance. It includes the following variables: unaddressed overtime work or multiple shifts that were not included in the base estimate, construction variations due to equipment selection, shortage of skilled labor, and late delivery of materials & equipment at the construction site. This factor is labeled as **[Site Management related factors]**. The following table gives the Factor Analysis results.

***Table 8. 23: Factor 5 Cluster (Analysis Version 2)***

| Cluster 5 | | |
|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** |
| **V41** | Unaddressed overtime work or multiple shifts that was not included in the base estimate | 0.700 |
| **V26** | Construction variations due to equipment selection | 0.666 |
| **V28** | Shortage of skilled labor | 0.623 |
| **V24** | Late delivery of materials & equipment at the construction site | 0.599 |
| ***Percentage of Variance 14.51%*** | | |

### 8.8.5 Interpreted Factors

The following table gives a summary of the retained factors in this version of the analysis. The results were used in a logistic regression analysis to find the odds of cost overrun occurring or not for each of the underlying factors.

***Table 8. 24: Interpreted Factors (Analysis Version 2)***

| Factor | % of variance | Factor Label |
|---|---|---|
| **Factor 1** | 18.63 | Design, Construction, Financial, and funding policy related factors |
| **Combined Factor (based on Factor 2 + Factor 3 + Factor 4)** | 47.87 | Inexperienced or lack of qualification of designer, contractor, Client, and project manager related factors |
| **Factor 3** | 14.51 | Site Management related factors |

## 8.9 Version 3 Analysis

In Version 3 analysis, we filtered the data of the questionnaire based on the response rates of each question (variable). The variables that received a low response rate were removed. The following steps illustrate the filtration process:

- The percent of the response rates to each question were calculated.
- The Likert scale was divided into two groups. The first group consisted of three scales (very high, high, and medium) and labelled as YES, which implies the factor causes cost overrun. For example, when the participant selected (very high, high, and medium), she/he agreed that the factor causes cost overrun. The

second group consisted of two scales (low and very low) and labelled as NO, which implies the factor does not cause cost overrun.

- The cumulative percent for (very high, high, and medium) and (low and very low) were calculated.
- The cumulative percent of the first group was ranked to show only the top-ranked (highly rated) variables. The sorting of the percent was in descending order.
- After sorting the variables, we found that the top-ranked variables amount to 27 variables. Table 8.25 shows the variable number and description.
- The Factor Analysis outputs of top 27 variables are presented in this chapter. These were used to develop Logistic models reported in Chapter 9.

***Table 8. 25: Variable Number and Description (Analysis Version 3)***

| Variable Number | Variable Description |
|---|---|
| **V4** | Complexity of the project (e.g., Project size, Project type, scope of work) |
| **V5** | Design changes during construction work |
| **V6** | Re-work due to the construction errors |
| **V7** | Unexpected technical problem |
| **V8** | Design errors that represent insufficient deliverables |
| **V9** | Changes by owner on the completion date of the project |
| **V10** | Scope changes by Owner during construction |
| **V11** | Delays related to owner or owner representative (e.g., stop work) |
| **V12** | Unrealistic project scheduling |
| **V13** | Acceleration to maintain schedule |
| **V15** | Type of construction contract (e.g., unit price contract) |
| **V18** | Delay by subcontractor |
| **V19** | Poor site management |
| **V20** | Unexpected weather conditions |
| **V28** | Shortage of skilled labor |
| **V31** | Delays and approval of shop drawings and installation procedures |
| **V33** | Government/ Municipal Approvals |
| **V34** | Lack of expertise in setting the budget |

| | |
|---|---|
| **V35** | The approved budget was too low |
| **V36** | Absence of a detailed Estimate Plan |
| **V37** | Changes in prices of items that have already been approved |
| **V45** | Lack of technical qualifications of the client |
| **V47** | Poor communication and coordination between all parties |
| **V48** | Disputes between parties (designer, contractor, owner) |
| **V51** | Land acquisition issues within right-of-way |
| **V52** | Quality assurance and quality control |
| **V53** | Inexperienced Project Managers, Estimators and Planners |

## 8.10 Analysis Version 3: Examining the Underlying Dimensions of Cost Overrun

A correlation test was conducted on 27 variables; also, a sample adequacy test was used to measure the suitability of the data. Table 8.26 shows that the KMO was 0.825 (KMO>0.5), which is in the acceptable range and the Bartlett Test was at a significance level at 0.000 (P<0.05). Therefore, the data of 27 variables seemed suitable for Factor Analysis.

***Table 8. 26: KMO and Bartlett's Test (Analysis Version 3)***

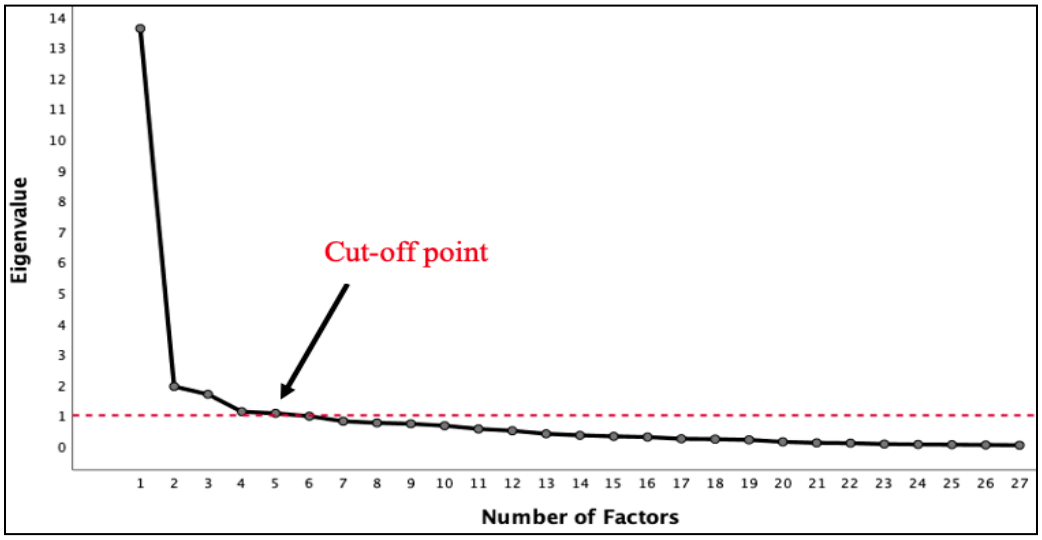
| | | |
|---|---|---|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.825 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 1251.089 |
| | df | 351 |
| | Sig. | 0.000 |

The principal component analysis was used to identify the number of significant factors. Table 8.27 presents the eigenvalue for each factor before extraction, after extraction, and after rotation. Also, Figure 8.5 shows the number of extracted factors and the cut-off point could identify the number of significant factors.

*Table 8. 27: Total Variance Explained (Analysis Version 3)*

| Component | Initial Eigenvalues | | | Extraction Sums of Squared | | | Rotation Sums of Squared | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| **1** | 13.613 | 50.417 | 50.417 | 13.613 | 50.417 | 50.417 | 6.703 | 24.824 | 24.824 |
| **2** | 1.943 | 7.198 | 57.614 | 1.943 | 7.198 | 57.614 | 5.235 | 19.387 | 44.211 |
| **3** | 1.687 | 6.247 | 63.862 | 1.687 | 6.247 | 63.862 | 3.576 | 13.244 | 57.455 |
| **4** | 1.120 | 4.149 | 68.010 | 1.120 | 4.149 | 68.010 | 2.018 | 7.472 | 64.927 |
| **5** | 1.069 | 3.959 | 71.969 | 1.069 | 3.959 | 71.969 | 1.901 | 7.042 | 71.969 |
| **6** | 0.973 | 3.604 | 75.573 | | | | | | |
| **7** | 0.810 | 3.001 | 78.574 | | | | | | |
| **8** | 0.756 | 2.800 | 81.375 | | | | | | |
| **9** | 0.726 | 2.688 | 84.062 | | | | | | |
| **10** | 0.663 | 2.454 | 86.516 | | | | | | |
| **11** | 0.557 | 2.064 | 88.580 | | | | | | |
| **12** | 0.498 | 1.846 | 90.426 | | | | | | |
| **13** | 0.398 | 1.476 | 91.901 | | | | | | |
| **14** | 0.349 | 1.294 | 93.196 | | | | | | |
| **15** | 0.315 | 1.168 | 94.364 | | | | | | |
| **16** | 0.293 | 1.084 | 95.448 | | | | | | |
| **17** | 0.237 | 0.879 | 96.327 | | | | | | |
| **18** | 0.224 | 0.830 | 97.157 | | | | | | |
| **19** | 0.201 | 0.746 | 97.904 | | | | | | |
| **20** | 0.136 | 0.505 | 98.408 | | | | | | |
| **21** | 0.105 | 0.389 | 98.797 | | | | | | |
| **22** | 0.095 | 0.354 | 99.150 | | | | | | |
| **23** | 0.066 | 0.245 | 99.395 | | | | | | |
| **24** | 0.054 | 0.201 | 99.596 | | | | | | |
| **25** | 0.045 | 0.167 | 99.763 | | | | | | |
| **26** | 0.036 | 0.133 | 99.897 | | | | | | |
| **27** | 0.028 | 0.103 | 100.000 | | | | | | |

*Figure 8. 5: Factor Scree Plot (Analysis Version 3)*




The number of factors extracted is five from a total of 27 variables. The extraction of the factors was done using principal components analysis and varimax rotation method. The following table presents the results of factor rotation.

*Table 8. 28: Varimax Factor Rotation Results (Analysis Version 3)*

| Cluster 1 | | | |
|---|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** | **Variance Explained** |
| **V4** | Complexity of the project (e.g., Project size, Project type, scope of work) | 0.789 | 24.82% |
| **V9** | Changes by owner on the completion date of the project | 0.786 | |
| **V15** | Type of construction contract (e.g., unit price contract) | 0.721 | |
| **V53** | Inexperienced Project Managers, Estimators and Planners | 0.690 | |
| **V47** | Poor communication and coordination between all parties | 0.680 | |
| **V8** | Design errors that represent insufficient deliverables | 0.660 | |
| **V12** | Unrealistic project scheduling | 0.588 | |
| **V36** | Absence of a detailed Estimate Plan | 0.581 | |
| **V28** | Shortage of skilled labor | 0.565 | |
| **V45** | Lack of technical qualifications of the client | 0.563 | |

| | | | |
|---|---|---|---|
| **V31** | Delays and approval of shop drawings and installation procedures | 0.562 | |
| **V48** | Disputes between parties (designer, contractor, owner) | 0.558 | |
| **V13** | Acceleration to maintain schedule | 0.542 | |
| **V10** | Scope changes by Owner during construction | 0.513 | |
| **Cluster 2** | | | |
| **Variable** | **Original Variable Descriptions** | **Factor Loading** | **Variance Explained** |
| **V6** | Re-work due to the construction errors | 0.733 | 19.38% |
| **V18** | Delay by subcontractor | 0.719 | |
| **V52** | Quality assurance and quality control | 0.716 | |
| **V19** | Poor site management | 0.662 | |
| **V34** | Lack of expertise in setting the budget | 0.603 | |
| **Cluster 3** | | | |
| **Variable** | **Original Variable Descriptions** | **Factor Loading** | **Variance Explained** |
| **V35** | The approved budget was too low | 0.802 | 13.24% |
| **V33** | Government/ Municipal Approvals | 0.745 | |
| **V11** | Delays related to owner or owner representative (e.g., stop work) | 0.649 | |
| **V51** | Land acquisition issues within right-of-way | 0.613 | |
| **Cluster 4** | | | |
| **Variable** | **Original Variable Descriptions** | **Factor Loading** | **Variance Explained** |
| **V7** | Unexpected technical problem | 0.841 | 7.47% |
| **Cluster 5** | | | |
| **Variable** | **Original Variable Descriptions** | **Factor Loading** | **Variance Explained** |
| **V20** | Unexpected weather conditions | 0.745 | 7.04% |
| **V5** | Design changes during construction work | 0.533 | |
| **Cumulative variance explained 71.96%** | | | |

The variables under cluster 1 are related to two main categories of variables. These are design, construction, and scheduling-related factors and inexperienced or lack of qualification of designer, contractor, client, and project manager related factor. Since these two factors appeared in Version 1 and Version 2 analyses, no additional investigation is required. Cluster 2 variables are related to work performance-related factors. It is useful to note that this factor was highlighted in the Version 1 analysis. Likewise, there is also no

need for further investigation of cluster 3 since it is related to decision-making tasks that have already been identified in Version 1 analysis. Furthermore, variables in cluster 4 and cluster 5 are related to unexpected problem and design issue factors. These factors were presented and highlighted in Version 1 and Version 2 analyses; hence no additional investigation is required.

## 8.11 Discussion

The inputs to the Factor Analysis method are the statistical data about the variables (V1 to V53) and no information can be provided to the methodology on what these variables represent (e.g., currency fluctuation). The variables are given numbers, but no labels can be provided. Therefore, the variables are placed in clusters by the Factor Analysis method on the basis of statistical properties only.

Since all variables can potentially have a cost overrun effect as viewed by survey respondents, any variable due to its associated statistical property can potentially end up in one out of a number of factors (clusters) in each version of the analysis. In other word, the Factor Analysis method cannot and should not define clusters on the basis of labels.

For the above-noted reasons, the reader is advised to accept any and all collections of reasons for cost overruns grouped together as various factors (cluster of variables). For example, the reason that “currency fluctuations” and “poor site management” have ended up together in Factor 1 of Analysis Version 1 is that both have a cost overrun implication. Likewise, for the same reason, “change in regulations” and “damages in materials and equipment in transit to the construction site” appear together in Factor 3 of Analysis Version 1.

Another word of caution is offered to the reader while reading outputs. The international responders read and interpreted the questions within the "context of their experience". For example, the reason that "the approved budget was too low" did not emerge as an extremely strong variable (problem) in the analysis is that in some jurisdictions, they have not approved very low budgets, and therefore they have not experienced cost overruns for this reason. *Therefore, the reader is requested that the outputs of Factor Analysis should not be read "out-of-context".*

An attempt is made to name the identified factors (as defined by clusters of variables). Given that in this research, a large number of variables with diverse labels are used and the fact that the Factor Analysis method does not recognize labels, the names assigned to factors (clusters) are rather lengthy.

The only way that a group of variables can be studied according to their label is to run a separate analysis for each sub-set of the data (e.g., design, construction, and scheduling). However, in a number of cases, a subset contains no more than 5 variables and using Factor Analysis for a small number of variables is not intended. Further, dividing data according to labels will result in biased and low-quality outputs.

A deeper understanding of cost-overruns was obtained from Factor Analysis. Although a large number of variables on experienced cost-overrun are analyzed, a section is added to the Logistic Regression Models chapter on ways to mitigate the risk of cost overruns.

It should be noted that due to agreement with contributors of data, this thesis document cannot identify specific geographic and jurisdictional locations. Since the

products of this research are addressed to a global technical audience, no specific agencies are noted.

## 8.12 Conclusions

- The Factor Analysis method for data reduction was successful in reducing a large number of variables into a small number of factors (clusters). Analysis Versions 1-1 (based on 43 variables), Version 2 (based on 31 variables), and Version 3 (based on 27 variables) were successful in all aspects. The sampling data adequacy tests, namely KMO and Bartlett tests, were satisfactory for the use of data in Factor Analysis. The Scree Plot test provided a reasonable number of underlying factors for further analysis. The identified underlying factors are those common factors that can potentially cause cost overrun. Given the diverse nature of survey data, the contribution of each factor to explaining variance appears logical.
- The level of variance explained by a factor mainly reflects the number of variables that belong to the factor. A high number of variables represented by a factor is due to the principles of Factors Analysis method (i.e., one factor can explain the effect of a large number of variables in explaining cost overrun.
- The most obvious findings to emerge from this study are as follow:

  Factor Analysis identified eight underlying factors based on Version 1 analysis: (1) Financial, materials, labor, and equipment related factors. (2) Design, construction, and scheduling related factors. (3) Policy, budget, and operation related factors. (4) Decision making related factors. (5) Material prices fluctuation and work performance

factors. (6) Unexpected condition related factors. (7) Quality control & quality assurance related factors, and (8) Owner/client related factors.

Based on Version 2 analysis, three underlying factors of cost overruns were identified: (1) Design, construction, financial, and funding policy-related factors. (2) Inexperienced or lack of qualification of designer, contractor, client, and project manager related factors, and (3) Site management related factors.

Based on Version 3 analysis, five underlying factors were identified. These are: (1) Design, construction, and scheduling-related & inexperienced or lack of qualification of designer, contractor, client, and project manager related factor. (2) Work performance related factors. (3) Decision-making tasks. (4) and (5) Unexpected problem and design issue factors.

- The Version 1-2 analysis based on the combination of 10 variables noted in the questionnaire did not provide much insight in explaining cost overruns. However, some variables in this list became a part of other factors (clusters). This finding suggests that a pre-conceived grouping of variables need not produce meaningful results and it is useful to let the Factors Analysis method identify factors (clusters of variables) for further analysis, as we have done in this research.
- It is a highly significant finding of this research that important causes of cost overruns identified by international respondents (in the context of their experience) do not cluster around one or two issue areas. Therefore, attention should be paid to a large number of causes (variables) in an attempt to avoid cost overruns.
- The relative effect of factors on cost overruns can be inferred from the Logistic regression model results (reported in Chapter 9) and not just on the level of variance

explained by factors. The results of Factor Analysis become input to the regression model development process.

# Chapter 9: Logistic Regression Modelling

## 9.1 Introduction

At the outset when building a model using the factors as independent variables, the multiple linear regression analysis, and the discriminant analysis are two related techniques that come to mind. However, these techniques pose difficulties when the dependent variable is binary. Therefore, logistic regression analysis was chosen as the appropriate technique in our analysis. This chapter explains the development of logistic regression models to study the impact of common factors of cost overrun in transportation infrastructure projects. Several logistic regression models were run using the underlying factors (clusters) of independent variables explained in Chapter 8.

Factor Analysis identified eight underlying factors based on Version 1 analysis:

1. Financial, materials, labor, and equipment related factors
2. Design, construction, and scheduling related factors
3. Policy, budget, and operation related factors
4. Decision making related factors
5. Material prices fluctuation and work performance factors
6. Unexpected condition related factors
7. Quality control & quality assurance related factors, and
8. Owner/client related factors.

In Version 2 analysis, three underlying factors were identified. These are,

1. Design, construction, financial, and funding policy-related factors

2. Inexperienced or lack of qualification of designer, contractor, client, and project manager related factors, and
3. Site management related factors.

In Version 3 analysis, five underlying factors were identified noted next.

1. Design, construction, and scheduling-related & inexperienced or lack of qualification of designer, contractor, client, and project manager related factor.
2. Work performance-related
3. Decision-making tasks
4. & 5. Unexpected problem and design issue factors.

Although factors are now known, the strength of effect on cost overrun phenomenon cannot be inferred. That is, no direct relationships between cost overrun occurring (dependent) and those underlying factors (independent) can be shown simply by Factor Analysis results. Therefore, the logistic regression analysis technique was used to examine the relationships between the dependent variable (i.e., occurrence of cost overrun) and independent factors (cluster of variables). The objective of using logistic regression is to predict the probability that an event will occur. Figure 9.1 shows the flow of logistic regression analysis procedures.

As explained earlier, the underlying factors represented by the Factor Analysis have been treated as if they are independent variables in logistic regression models with a cost overrun dependent variable of binary nature. That is, the cost overrun has only two probable values, one is cost overrun occurring coded as (1) and the other is the cost overrun not occurring coded as (0) in this analysis.

The logistic regression analysis attempts to model the probability of an event occurring based on the values of the independent factors (cluster of variables). The logistic model can be expressed in the form of:

$$\text{Log (P)} = B_0 + B_1X_1 + B_2X_2 + \cdots + B_iX_i \qquad (9.1)$$

*Where:*

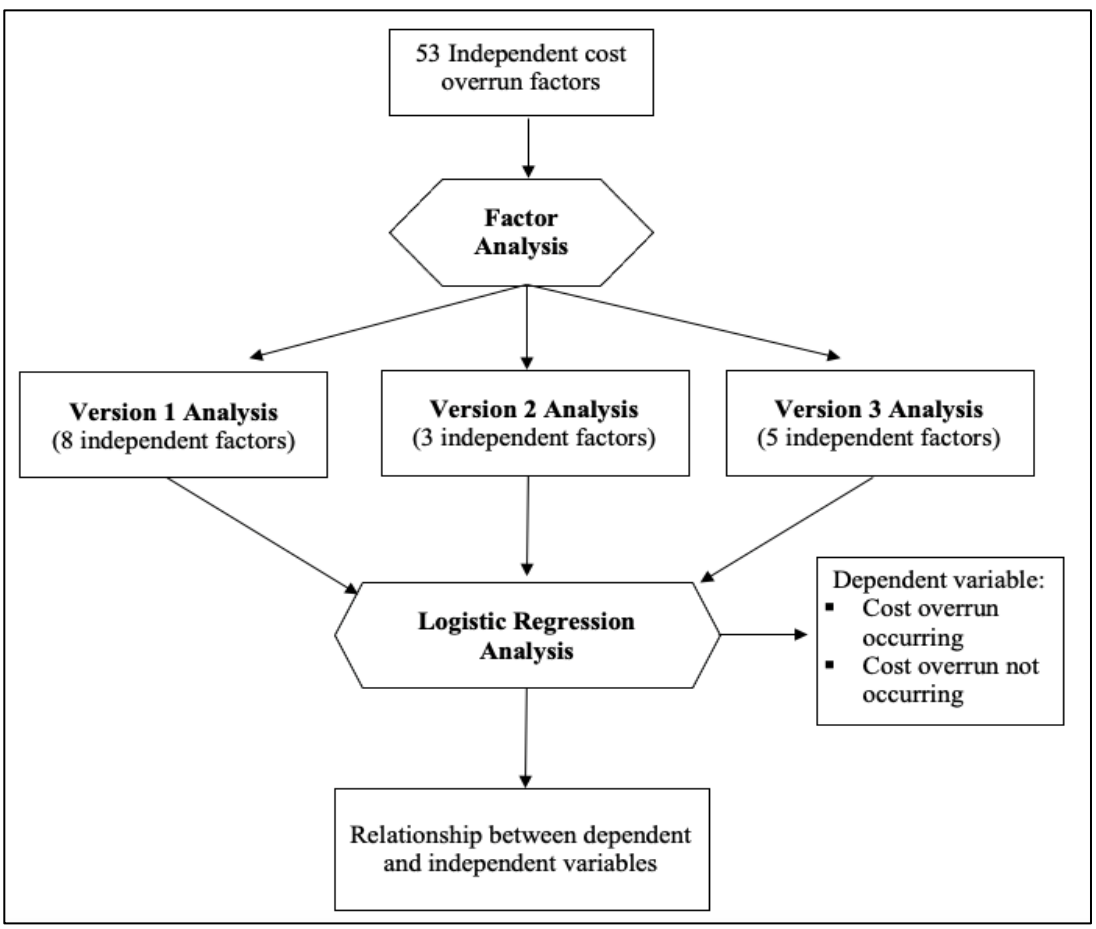
P = is the probability of cost overrun occurring (dependent variable).

$B_0$ = exposure variable or constant.

$B_1, B_2$ *and* $B_i$ = are coefficients.

$X_1, X_2$ *and* $X_i$ = are explanatory factors (also called independent variables).

***Figure 9. 1: Flow of Logistic Regression Analysis Procedures***



The general method of estimating the model parameters is called maximum likelihood (Field, 2009). Details of logistic regression analysis results can be found in Appendix C. These include various aspects of models, statistics, and probabilities that are used to interpret and report logistic regression outputs. The key statistics are included in this chapter.

### 9.1.1 Odds Ratio

In logistic regression, the odds ratio represents the effect of the independent factors (cluster of variables) on the dependent variable in terms of the likelihood of the event coded as 1. In this case, it is a cost overrun occurring. Thus, the logistic regression model dependent variable is coded in terms of the odds of an event occurring. The methodology uses ratio of the probability that it will occur to the probability that it will not. In relative terms, if the odds ratio is greater than one (odds >1), it indicates a strong effect of an independent factor (cluster of variables) on the dependent variable. A value of less than one (odds <1) indicates a weak effect. When the odds ratio is less than one, that does not mean there is no effect, but the factor does not meet the criteria of odds ratio.

When modeling a binary value, it is necessary to interpret the model's outputs in terms of the adjusted odds ratios. It is also necessary to measure the fit of the model to the observed data (Bangdiwala, 2018). It can be noticed that the odds ratio is what SPSS reports as Exp(B).

### 9.1.2 Pseudo R-Square

The Pseudo R-Square seeks to measure the proportion of explained variance in the logistic regression model. There are two types of Pseudo R-Square in logistic regression

analysis, which are Nagelkerke and Cox and Snell. In general, there is no strong guidance in the literature on how these should be used or interpreted (Costello and Osborne, 2011; Hahs-Vaughn and Lomax, 2013; Smith and McKenna, 2013; Stevens, 2012).

Smith and McKenna (2013) noted other commonly used Pseudo-R-Square indices (McFadden's index, Cox-Snell index with or without Nagelkerke). According to Field (2009), the Nagelkerke R-Square is the most commonly used statistic when interpreting the logistic regression model and the Cox and Snell R-Square only provides an approximate value of variance. According to Campbell (2008), both the Nagelkerke and Cox and Snell R-Square values can be interpreted similarly; hence, the higher the R-Square value, the better the model fits.

The Cox and Snell's R-Square is calculated from the following equation:

$$R^2_{CS} = 1 - \exp\left(\frac{-2LL_{new} - (-2LL_{Initial})}{N}\right) \qquad (9.2)$$

*Where:*

N: the sample size

LL: Log-likelihood

The Nagelkerke R-Square is calculated from:

$$R^2_N = \frac{R^2_{CS}}{1 - \exp\left(-\frac{-2LL_{Initial}}{N}\right)} \qquad (9.3)$$

Caution must be applied to the reader when reading the Pseudo R-Square values. The Pseudo R-Square values should not read as R-Square in linear regression.

### 9.1.3 Log-Likelihood

The Log-likelihood (LL) can be used to assess the significance of logistic regression model. However, the smaller the value of the log-likelihood, the better the fit of the model. The large value of the log-likelihood indicates poorly fitting models (Field, 2009). Sometimes the log-likelihood is referred to as (−2LL).

### 9.1.4 Chi-Square

Chi-square, another goodness of fit test measures the significance of the model. In SPSS software, the Chi-square model is labeled as the "Omnibus Test of model Coefficients." The chi-square model is sometimes called the traditional fit measure. A significant value of less than (0.05) indicates that the model fits well.

### 9.1.5 Confidence Interval

Another quantity that can be used to interpret a logistic regression model is the confidence interval (i.e., the boundaries within which we believe the true value of the mean will fall). Field (2009) notes that a 95% confidence interval means that for 95% of the time the true value of the population mean will fall within these limits. The confidence interval values for all logistic regression models can be found in Appendix C.

## 9.2 Logistic Regression Models Using Underlying Factors

### 9.2.1 Models Based on Factor Analysis Version 1

In this analysis, the main objective of the regression model is to find the best fitting model for describing the relationship between the independent variables and dependent variables. As a first step towards calibrating the logistic regression models, the

interpretable factors have been added one by one to the logistic regression model as independent variables to derive their respective unique contribution to cost overrun. The following tables give the regression models with their odds ratio, log-likelihood, Pseudo R-Square, and chi-square values. Models 1 to 9 represent the regression analysis results for Factor Analysis Version 1 and models 10 to 12 represent the regression analysis results for Factor Analysis Version 2. Model 13 represents the regression analysis results for the Factor Analysis Version 3.

**Logistic Regression Statistics for Model 1**

The first underlying factor, namely, financial, materials, labor, and equipment related factors, were used in logistic regression analysis. Table 9.1 shows how the significant predictor variables contributed to the model.

The outcome of the analysis showed a very small odds ratio of less than one, which indicates a weak effect of an independent factor (cluster of variables) on the dependent variable. The -2LL test result was 71.828 with chi-squared of 0.719, and non-significant at 0.397 ($P>0.05$), indicating that the model does not fit well.

The Cox and Snell R-squared is 0.013; this means that about 1.3% of the variation in the dependent variable is explained by the model. The Nagelkerke R-squared was also estimated for the model and found to be 0.018, which indicates that about 1.8% of the variation in the dependent variable is explained by the model.

*Table 9. 1: Logistic Regression Statistics for Model 1*

| Interpreted Factor | Odds Ratio |
|---|---|
| Cluster 1 | 0.779 |
| **Model Fit Information** | |
| -2 Log likelihood (-2LL) | 71.828 |
| Model Chi-square | 0.719 |
| Sig. | 0.397 |
| **Pseudo R-square** | |
| Cox & Snell R-Square | 0.013 |
| Nagelkerke R-Square | 0.018 |

As we proceed with the development of the models, the same tests will be repeated to examine the model fit as well as to see the improvement, if any, to the models due to the inclusion of other factors.

**Logistic Regression Statistics for Model 2**

Two underlying factors, namely, [Financial, Materials, labor, and Equipment related factors] and [Design, construction, and scheduling related factors] were entered in the logistic regression analysis.

The model had a satisfactory fit at ($P<0.05$), as shown by the tabulated statistics. The results of the analysis showed that the odds ratio of cluster 2 is greater than one, which indicates a strong effect of an independent factor (cluster of variables) on the dependent variable. The -2LL decreased from 71.828 to 51.692, and R-Square values have been increased compared to model 1, which implies good improvement. The model accounts for between 32.5% and 43.6% of the variance, as shown by the Cox & Snell R Square and Nagelkerke R Square.

*Table 9. 2: Logistic Regression Statistics for Model 2*

| Interpreted Factor | Odds Ratio |
|---|---|
| Cluster 1 | 0.654 |
| Cluster 2 | 4.938 |
| **Model Fit Information** | |
| -2 Log likelihood (-2LL) | 51.692 |
| Model Chi-square | 20.854 |
| Sig. | 0.000 |
| **Pseudo R-square** | |
| Cox & Snell R-Square | 0.325 |
| Nagelkerke R-Square | 0.436 |

In this model, we can say that the odds ratio of the cost overrun occurring is 4.93 times higher than not occurring due to the inclusion of factor (cluster) 2. In the next model, cluster 3 has been added. The following table illustrates the model.

**Logistic Regression Statistics for Model 3**

Three underlying factors were entered in the logistic regression model as independent variables, namely, [Financial, Materials, labor, and Equipment related factors], [Design, construction, and schedule-related factors], and [policy, budget, and operation-related factors].

The overall test of the model turned out to be significant at ($P< 0.05$), indicating a good fit of the model. The odds of cluster 1 and cluster 3 show weak impacts, whereas the odds of cluster 2 show a strong impact. The -2LL ratio decreased from 51.692 to 51.307, and the chi-square increased from 20.854 to 21.240. The Nagelkerke R-squared of this model was 44.3%, while the Cox and Snell R-squared was 33%. The model fit values and the R-square values are quite similar to model 2; thus, this model shows weak

improvement. Table 9.3 presents the results. Since cluster 3 does not seem to show any significant effect on cost overrun, we proceed with the next model by adding another factor.

***Table 9. 3: Logistic Regression Statistics for Model 3***

| Interpreted Factor | Odds Ratio |
|---|---|
| Cluster 1 | 0.627 |
| Cluster 2 | 5.118 |
| Cluster 3 | 0.814 |
| **Model Fit Information** | |
| -2 Log likelihood (-2LL) | 51.307 |
| Model Chi-square | 21.240 |
| Sig. | 0.000 |
| **Pseudo R-square** | |
| Cox & Snell R-Square | 0.330 |
| Nagelkerke R-Square | 0.443 |

**Logistic Regression Statistics for Model 4**

As shown in Table 9.4, four underlying factors entered in the logistic regression model as independent factors (cluster of variables), namely, [Financial, Materials, labor, and Equipment related factors], [Design, construction, and schedule-related factors], [policy, budget, and operation-related factors], and [decision making related factors].

Four variables were significant to the model fit at 0.000 ($P<0.05$), and the Log-likelihood was 45.627 with a chi-square 26.919, indicating that the model fits well and was a significant improvement over the previous model (model 3) due to the inclusion of factor cluster 4 to the model. Moreover, the model accounts for between 39.8% and 53.4% of the variance (Cox and Snell R-Square and the Nagelkerke R-square), indicating a moderate to a good association between the independent factors (variables) and the dependent variable. The results also show that cluster 4 has an odds ratio greater than one, indicating that its inclusion into the model increases the probability of cost overrun occurring.

*Table 9. 4: Logistic Regression Statistics for Model 4*

| Interpreted Factor | Odds Ratio |
|---|---|
| Cluster 1 | 0.535 |
| Cluster 2 | 7.126 |
| Cluster 3 | 0.780 |
| Cluster 4 | 2.502 |
| **Model Fit Information** | |
| -2 Log likelihood (-2LL) | 45.627 |
| Model Chi-square | 26.919 |
| Sig. | 0.000 |
| **Pseudo R-square** | |
| Cox & Snell R-Square | 0.398 |
| Nagelkerke R-Square | 0.534 |

**Logistic Regression Statistics for Model 5**

The fifth underlying factor, namely, [Material price fluctuation and work performance related factors], was entered in logistic regression analysis. Table 9.5 shows the results.

The odds of cluster 5 show a strong impact, and the model had a satisfactory fit at ($P<0.05$), as shown by the tabulated statistics. The likelihood ratio, R-Square, and chi-square are found to be quite similar to the previous model, respectively, which again shows a slight improvement.

*Table 9. 5: Logistic Regression Statistics for Model 5*

| Interpreted Factor | Odds Ratio |
|---|---|
| Cluster 1 | 0.538 |
| Cluster 2 | 7.087 |
| Cluster 3 | 0.789 |
| Cluster 4 | 2.553 |
| Cluster 5 | 1.125 |
| **Model Fit Information** | |
| -2 Log likelihood (-2LL) | 45.534 |
| Model Chi-square | 27.012 |
| Sig. | 0.000 |

| Pseudo R-square | |
|---|---|
| Cox & Snell R-Square | 0.399 |
| Nagelkerke R-Square | 0.536 |

In the next model, we add cluster 6 and the results are shown in Table 9.6.

**Logistic Regression Statistics for Model 6**

Five clusters were added to the logistic regression model one by one to examine the relationships between the independent factors and dependent variable. Next, one more underlying factor was added, namely, [unexpected related factors]. Table 9.6 presents the results.

The statistics of this model show that cluster 2, cluster 4, and cluster 6 have odds ratio of greater than 1. However, it can be interpreted that the inclusion of cluster 6 in the model increases the probability that cost overrun will occur.

Compared with the previous model, the likelihood value of this model was smaller than the previous model. The smaller the likelihood, the better the fit of the model. On the other hand, the higher the chi-square, the better the fit of the model. The likelihood decreased from 45.534 to 38.361, whereas the chi-square increased from 27.012 to 34.185, respectively.

Additionally, pseudo-R-Square values have been increased compared to model 5, which implies very good improvement. The model accounts for between 47.5% and 63.8% of the variance, as shown by the Cox & Snell R Square and Nagelkerke R Square.

*Table 9. 6: Logistic Regression Statistics for Model 6*

| Interpreted Factor | Odds Ratio |
|---|---|
| Cluster 1 | 0.660 |
| Cluster 2 | 7.158 |
| Cluster 3 | 0.886 |
| Cluster 4 | 2.554 |
| Cluster 5 | 1.186 |
| Cluster 6 | 3.214 |
| **Model Fit Information** | |
| -2 Log likelihood (-2LL) | 38.361 |
| Model Chi-square | 34.185 |
| Sig. | 0.000 |
| **Pseudo R-square** | |
| Cox & Snell R-Square | 0.475 |
| Nagelkerke R-Square | 0.638 |

**Logistic Regression Statistics for Model 7**

Seven underlying factors were entered in the logistic regression model as independent factors (cluster of variables), namely, [Financial, Materials, labor, and Equipment related factor], [Design, construction, and scheduling related factor], [policy, budget, and operation-related factor], [decision making related factor], [Material prices fluctuation and work performance related factor], [unexpected circumstances-related factor], and [Quality control & quality assurance factor]. See Table 9.7.

The results show that the inclusion of cluster 7 makes the odds ratio greater than one. Therefore, it is concluded that the inclusion of factor (cluster) 7 will increase the odds that cost overrun will occur. Moreover, the Cox and Snell R-squared is 0.494; this means that the model explains about 49.4% of the variance. The Nagelkerke R-squared was also estimated for the inclusion of cluster 7 and found to be 0.662, which indicates that the model explains about 66.2% of the variance in the dependent factors (cluster of variables).

These values indicate a strong association between the independent factors and the dependent variable.

***Table 9. 7: Logistic Regression Statistics for Model 7***

| Interpreted Factor | Odds Ratio |
|---|---|
| Cluster 1 | 0.619 |
| Cluster 2 | 8.642 |
| Cluster 3 | 0.868 |
| Cluster 4 | 2.784 |
| Cluster 5 | 1.101 |
| Cluster 6 | 3.100 |
| Cluster 7 | 1.773 |
| **Model Fit Information** | |
| -2 Log likelihood (-2LL) | 36.494 |
| Model Chi-square | 36.052 |
| Sig. | 0.000 |
| **Pseudo R-square** | |
| Cox & Snell R-Square | 0.494 |
| Nagelkerke R-Square | 0.662 |

**Logistic Regression Statistics for Model 8**

We finally add the factor (cluster) 8) to the model, and the results are shown in Table 9.8. Eight interpretable factors have been added one by one to the logistic regression model as independent factors to derive their respective unique contribution to cost overrun. The interpretable factors are as follow:

- Factor (Cluster) 1: Financial, materials, labor, and equipment related factors
- Factor (Cluster) 2: Design, construction, and scheduling related factors
- Factor (Cluster) 3: Policy, budget, and operation-related factors
- Factor (Cluster) 4: Decision making related factors
- Factor (Cluster) 5: Material prices fluctuation and work performance related factors
- Factor (Cluster) 6: Unexpected related factors

- Factor (Cluster) 7: Quality control & quality assurance related factors and
- Factor (Cluster) 8: Owner/client-related factors

In Table 9.8, the Factor (cluster) 8 variable shows an odds ratio of less than one, indicating that its inclusion into the model does not increase the odds of cost overrun occurring. Also, factor (cluster) 8 does not seem to show any significant effect on cost overrun. However, its presence in the model is not detrimental.

Compared with the first model, in this model, the -2LL decreased from 71.828 to 36.472, respectively. The lower value of -2LL indicates that the model predicts the outcome variable more accurately and fits well. The large value of the log-likelihood indicates poorly fitting models (Field, 2009). The chi-square increased from 0.719 for Model 1 to 36.075 for Model 8. As noted earlier, the higher the chi-square, the better the model fit. Also, the Cox & Snell R-Square increased from 1.3% to 49.4%, and the Nagelkerke R Square increased from 1.8% to 66.2%, indicating high improvement.

***Table 9. 8: Logistic Regression Statistics for Model 8***

| Interpreted Factor | Odds Ratio |
|---|---|
| Cluster 1 | 0.628 |
| Cluster 2 | 8.801 |
| Cluster 3 | 0.878 |
| Cluster 4 | 2.803 |
| Cluster 5 | 1.088 |
| Cluster 6 | 3.094 |
| Cluster 7 | 1.796 |
| Cluster 8 | 0.932 |
| **Model Fit Information** | |
| -2 Log likelihood (-2LL) | 36.472 |
| Model Chi-square | 36.075 |
| Sig. | 0.000 |
| **Pseudo R-square** | |
| Cox & Snell R-Square | 0.494 |
| Nagelkerke R-Square | 0.662 |

In conclusion, the inclusion of factor 1, namely, [Financial, Materials, labor, and Equipment related factors] in the logistic regression analysis does not increase the probability of cost overruns occurring. Whereas the inclusion of factor 2 [Design, construction, and scheduling related factors] gives an odds ratio of greater than one 8.801, which indicates the probability of cost occurrence is higher. It should be noted here that the variables under cluster 2 are important in transportation infrastructure investments. The inclusion of factor 3, namely, [policy, budget, and operation-related factors], shows a weak impact with an odds ratio of less than 1 (i.e., 0.878).

The odds ratio of factor 4 [decision making related factors] is estimated to be more than one 2.803, which indicates a strong effect. On the other hand, the inclusion of factor 5 [Material price fluctuation and work performance related factors] in the model shows no effect with an odds ratio equal to one. That is, if the odds ratio is equal to one, there is no relationship between independent and dependent variables.

There are two likely strong effects by including factors 6 and 7 in the model (i.e., unexpected factors and quality control and quality assurance related factors). The results of the analysis show that the odds ratio of factor 6 is 3.094, and the odds ratio of factor 7 is found to be 1.796. This indicates a strong effect of an independent factor on the dependent variable, and therefore there is a possibility of cost overrun to occur. Lastly, the factor 8 variable shows an odds ratio of less than 1, indicating that its inclusion into the model does not increase the odds of cost overrun occurring.

**Logistic Regression Statistics for Model 9**

In the Factor Analysis chapter, it was noted that as a part of the first version of the Factor Analysis, we divided the data obtained from the questionnaire study into two parts.

Table 9.9 shows the logistic regression analysis model for the second part of the datasets, which consists of 10 variables. The results do not appear to be impressive.

***Table 9. 9: Logistic Regression Statistics for Model 9***

| Interpreted Factor | Odds Ratio |
|---|---|
| Other Factors | 2.932 |
| **Model Fit Information** | |
| -2 Log likelihood (-2LL) | 60.466 |
| Model Chi-square | 12.080 |
| Sig. | 0.001 |
| **Pseudo R-square** | |
| Cox & Snell R-Square | 0.204 |
| Nagelkerke R-Square | 0.273 |

### 9.2.2 Models Based on Factor Analysis Version 2

As reported in the Factor Analysis chapter, Version 2 of Factor Analysis was carried out after removing weak correlation variables. This decision resulted in removing 22 variables out of 53 from the data. The Second Version analysis resulted in the identification of three underlying factors. These are (1) design, construction, financial, and funding policy-related factor, (2) inexperienced or lack of qualification of designer, contractor, client, and project manager related factor, and (3) site management related factor.

Subsequently, a logistic regression analysis was performed using these factors, and the results are presented as models 10 to 12.

**Logistic Regression Statistics for Model 10**

Results based on factor 1, shown in Table 9.10, indicate an odds ratio of less than one. This result implies that the cost overrun is not occurring due to the inclusion of this factor in the regression model. Other results (i.e., the log-likelihood of 72.391, a very small

chi-square of 0.155, non-significant at 0.693 (P>0.05) indicate that the model does not fit well. Therefore, it can be inferred that factor 1 does not seem to show any significant effect on cost overrun. We proceed with the next model by adding another factor.

***Table 9. 10: Logistic Regression Statistics for Model 10***

| Interpreted Factor | Odds Ratio |
|---|---|
| Cluster 1 | 0.895 |
| **Model Fit Information** | |
| -2 Log likelihood (-2LL) | 72.391 |
| Model Chi-square | 0.155 |
| Sig. | 0.693 |
| **Pseudo R-square** | |
| Cox & Snell R-Square | 0.003 |
| Nagelkerke R-Square | 0.004 |

**Logistic Regression Statistics for Model 11**

In the Factor Analysis chapter, it was noted that original factor 2, original factor 3, and original factor 4 were combined and labeled as [inexperienced or lack of qualification of designer, contractor, client, and project manager related factors]. This combined factor was named as factor 2 in the Version 2 of Factor Analysis. The results of adding this factor to the regression model are reported in Table 9.11 as Model 11.

This model is significant at 0.000 (P<0.05). The Log-likelihood decreased from 72.391 to 53.664, and the chi-square increased from 0.155 to 18.882, indicating that the model fitted well and was a significant improvement over the previous model (model 10) due to the inclusion of three factors together to the model.

Moreover, the model accounts for between 30% and 40.2% of the variance (Cox and Snell R-Square and the Nagelkerke R-square), indicating a moderate to a good association between the independent factors (cluster of variables) and the dependent

variable. The results also show that the combined factor has an odds ratio greater than one, indicating that its inclusion into the model increases the probability of the cost overrun occurring.

***Table 9. 11: Logistic Regression Statistics for Model 11***

| Interpreted Factor | Odds Ratio |
|---|---|
| Factor 1 | 0.848 |
| Factor 2 (combination of original factors 2, 3, and 4) | 13.626 |
| **Model Fit Information** | |
| -2 Log likelihood (-2LL) | 53.664 |
| Model Chi-square | 18.882 |
| Sig. | 0.000 |
| **Pseudo R-square** | |
| Cox & Snell R-Square | 0.300 |
| Nagelkerke R-Square | 0.402 |

**Logistic Regression Statistics for Model 12**

Model 12 was developed by adding the last underlying factor (i.e., site management variables) to the regression analysis. The results are shown in Table 9.12.

The overall test of the model turned out to be significant at ($P< 0.05$), indicating a good fit of the model. The odds of factor 1 and factor 5 show a weak impact, whereas the odds of the combined cluster show a strong impact.

The -2LL ratio decreased from 53.664 to 52.674, and the chi-square increased from 18.882 to 19.872. The Nagelkerke R-squared of this model was 41.9%, while the Cox and Snell R-squared was 31.3%. The model fit values and the R-square values are quite similar to the previous model (i.e., model 11); thus, this model shows weak improvement.

*Table 9. 12: Logistic Regression Statistics for Model 12*

| Interpreted Factor | Odds Ratio |
|---|---|
| Factor 1 | 0.854 |
| Factor 2 (combination of original factors 2, 3, and 4) | 16.305 |
| Factor 5 | 0.715 |
| **Model Fit Information** | |
| -2 Log likelihood (-2LL) | 52.674 |
| Model Chi-square | 19.872 |
| Sig. | 0.000 |
| **Pseudo R-square** | |
| Cox & Snell R-Square | 0.313 |
| Nagelkerke R-Square | 0.419 |

### 9.2.3 Model Based on Factor Analysis Version 3

Table 9.13 shows the results of the final model of Logistic regression analysis for all factors (clusters of variables) described in the Version 3 analysis part of Chapter 8.

*Table 9. 13: Logistic Regression Statistics for Model 13*

| Interpreted Factor | Odds Ratio |
|---|---|
| Factor 1 (Cluster) | 2.611 |
| Factor 2 (Cluster) | 1.332 |
| Factor 3 (Cluster) | 2.181 |
| Factor 4 (Cluster) | 1.442 |
| Factor 5 (Cluster) | 2.162 |
| **Model Fit Information** | |
| -2 Log likelihood (-2LL) | 52.683 |
| Model Chi-square | 19.863 |
| Sig. | 0.001 |
| **Pseudo R-square** | |
| Cox & Snell R-Square | 0.312 |
| Nagelkerke R-Square | 0.419 |

The model has a satisfactory fit ($P<0.05$), as shown by the tabulated statistics. The analysis results show that the odds ratio of each cluster is greater than one, which indicates a strong effect of an independent factor (cluster of variables) on the dependent variable.

The model accounts for between 31.2% and 41.9% of the variance, as shown by the Cox & Snell R-Square and Nagelkerke R-Square. These statistics are satisfactory.

## 9.3 Logistic Regression Model Results: Potential Causes of Cost Overruns

Results of Models 8, 12, and 13 are summarized in Tables 9.14 and 9.15. Also presented in these tables is the classification information for variables and their resulting factors (clusters). As expected, well-defined factors are formed based on variables that mainly characterize design, construction, scheduling, and related variables. Other factors are classified as permits and approvals, estimation/budget and financial issues, site conditions/environment, materials and equipment, and expertise and qualifications of personnel. Further factors have also been identified, although infrequently, that relate to policy factors, and quality assurance and quality control.

***Table 9. 14: Logistic Regression Model Results: Potential Causes of Cost Overruns (Based on Analysis Versions 1 and 2)***

| *Factor 2 Cluster - (Analysis Version 1-1): Design, Construction, Scheduling and Related Tasks Odds Ratio 8.801 in Regression Model 8* | | | |
|---|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** | **Classification** |
| **V4** | Complexity of the project (e.g., Project size, Project type, scope of work) | 0.867 | Design, construction, scheduling |
| **V15** | Type of construction contract (e.g., unit price contract) | 0.730 | Design, construction, scheduling |
| **V9** | Changes by owner on the completion date of the project | 0.689 | Design, construction, scheduling |
| **V5** | Design changes during construction work | 0.631 | Design, construction, scheduling |
| **V16** | Unnecessary practices, specifications, and procedures | 0.589 | Design, construction, scheduling |
| **V8** | Design errors that represent insufficient deliverables | 0.574 | Design, construction, scheduling |

| | | | |
|---|---|---|---|
| **V12** | Unrealistic project scheduling | 0.533 | Design, construction, scheduling |
| **V31** | Delays and approval of shop drawings and installation procedures | 0.509 | Permits and approvals |
| **Percentage of Variance 14.19%** | | | |

| *Factor 4 Cluster - (Analysis Version 1-1): Decision Making Tasks Odds Ratio 2.803 in Regression Model 8* | | | |
|---|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** | **Classification** |
| **V33** | Government/ Municipal Approvals | 0.778 | Permits and approvals |
| **V35** | The approved budget was too low | 0.648 | Estimation/budget and financial |
| **V11** | Delays related to owner or owner representative (e.g., stop work) | 0.627 | Design, construction and scheduling |
| **Percentage of Variance 7.97%** | | | |

| *Factor 6 Cluster - (Analysis Version 1-1): Unexpected Problems and Related Issues Odds Ratio 3.094 in Regression Model 8* | | | |
|---|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** | **Classification** |
| **V7** | Unexpected technical problem | 0.728 | Design, construction and scheduling |
| **V20** | Unexpected weather conditions | 0.579 | Site conditions/ environment |
| **V13** | Acceleration to maintain schedule | 0.531 | Design, construction and scheduling |
| **Percentage of Variance 6.33%** | | | |

| *Factor 7 Cluster - (Analysis Version 1-1): Quality Control, Quality Assurance and Related Issues Odds Ratio 1.796 in Regression Model 8* | | | |
|---|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** | **Classification** |
| **V27** | Defective materials | 0.660 | Materials and equipment |
| **V6** | Re-work due to the construction errors | 0.621 | Design, construction and scheduling |
| **Percentage of Variance 4.5%** | | | |

| *Combined Factors 2, 3, and 4 Cluster from Analysis Version 2: Inexperienced or lack of qualification of designer, contractor, client, and project manager related factors Odds Ratio 16.305 in Regression Model 12* | | | |
|---|---|---|---|
| **Variable** | **Original Variable Descriptions** | **Factor Loading** | **Classification** |
| **V4*** | Complexity of the project (e.g., Project size, Project type, scope of work) | 0.837 | Design, construction and scheduling |
| **V15*** | Type of construction contract (e.g., unit price contract) | 0.785 | Design, construction and scheduling |
| **V53** | Inexperienced Project Managers, Estimators and Planners | 0.748 | Other factors |
| **V5*** | Design changes during construction work | 0.683 | Design, construction and scheduling |
| **V8** | Design errors that represent insufficient deliverables | 0.656 | Design, construction and scheduling |
| **V9*** | Changes by owner on the completion date of the project | 0.547 | Design, construction and scheduling |
| **V13*** | Acceleration to maintain schedule | 0.521 | Design, construction and scheduling |
| **V31*** | Delays and approval of shop drawings and installation procedures | 0.514 | Permits and approvals |
| **Variable** | **Original Variable Descriptions** | **Factor Loading** | **Classification** |
| **V21** | Accidents due to poor site safety | 0.799 | Site conditions/Environment |
| **V52** | Quality assurance and quality control | 0.738 | Other factors |
| **V6*** | Re-work due to the construction errors | 0.515 | Design, construction and scheduling |
| **V19** | Poor site management | 0.507 | Site conditions/ Environment |
| **Variable** | **Original Variable Descriptions** | **Factor Loading** | **Classification** |
| **V36** | Absence of a detailed Estimate Plan | 0.688 | Estimation/Budget and financial |
| **V46** | Overly high expectations | 0.659 | Other factors |
| **V32** | Building Permit to the construction contractor | 0.603 | Permits and approvals |

| | | | |
|---|---|---|---|
| **V18** | Delay by subcontractor | 0.577 | Design, construction and scheduling |
| **V39** | Inappropriate and inadequate procurement (e.g., payment terms, pricing) | 0.572 | Estimation/Budget and financial |
| **V2** | Deal termination due to changes in law, government policy or protocols | 0.570 | Policy |
| **V34** | Lack of expertise in setting the budget | 0.555 | Estimation/Budget and financial |
| ***Cumulative variance explained 47.87%*** | | | |

* Also appears in a factor identified in Analysis Version 1

***Table 9. 15: Logistic Regression Model Results: Potential Causes of Cost Overruns (Based on Analysis Version 3)***

| *Factors (Clusters) 1 to 5* | | | |
|---|---|---|---|
| **Factor (Cluster)** | **Variables** | **Odds Ratio** | **Classification** |
| **1** | See Table 8.28 | 2.611 | Factor 1 variables relate to two main categories of variables: design, construction and scheduling-related factors and inexperienced or lack of qualification of designer, contractor, client, and project manager related factor. These appeared in Version 1 and Version 2 analyses. |
| **2** | See Table 8.28 | 1.332 | Factor 2 variables are related to work performance-related variables. It is useful to note that this factor has common variables with Version 1 and Version 2 results. |
| **3** | See Table 8.28 | 2.181 | Factor 3 variables relate to decision-making tasks that have already been identified in Version 1 analysis. |
| **4** | See Table 8.28 | 1.442 | Factor 4 variable relates to unexpected technical problem. It has been highlighted in Version 1 analysis. |
| **5** | See Table 8.28 | 2.162 | Factor 5 variables relate to unexpected weather condition and design change issues. These were highlighted in Version 1 analysis. |

## 9.4 Discussion

Three versions of Factor Analysis described in Chapter 8 identified factors (i.e., clusters of variables) for further study so that their effects are quantified. The reason for further study is that the use of Factor Analysis alone does not go far enough in identifying a direct relationship between those factors (i.e., formed out of independent variables) and the occurrence of cost overrun (dependent variable). Hence, a logistic regression analysis was conducted to estimate the probability of the occurrence of cost overrun. Several logistic models were developed in order to find any significant associations between independent factors and the dependent variable using SPSS software.

The goodness of fit of the models was studied by using the log-likelihood, R-squared, and Chi-square. The change in Pseudo R-square, due to the inclusion of factors into the regression model, is an indicator of the unique information provided by variables. The interpretation of regression results is based on the odds ratio. In relative terms, if the odds ratio is greater than one (odds >1), it indicates a strong effect of an independent factor (cluster of variables) on the dependent variable. A value of less than one (odds <1) indicates a weak effect. When the odds ratio is less than one, that does not mean there is no effect, but the factor does not meet the criteria of odds ratio.

## 9.5 Conclusions and Mitigation Measures

The summaries of model results are shown in Tables 9.8, 9.12, 9.13, 9.14, and 9.15. Tables 9.14 and 9.15 can be viewed for information on the classification of significant factors identified in this research study.

- As expected, well-defined factors are formed based on variables that mainly characterize design, construction, scheduling, and related factors.
- Other factors are classified as permits and approvals, estimation/budget and financial issues (i.e., decision-making factors), site conditions/environment, materials and equipment, unexpected conditions/events, and expertise and qualifications of personnel.
- Additional other factors have also been identified, although infrequently, that relate to quality assurance, quality control, and related variables.

The subject of what should be done to mitigate cost overruns in highway infrastructure projects is multifaceted for the following reason. As noted in the conclusions part of Chapter 8, it is a highly significant finding of this research that several important causes of cost overruns identified by international respondents (in the context of their experience) do not cluster around just one or two issue areas. Therefore, attention should be paid to a large number of causes (variables) in an attempt to avoid cost overruns. These can be noted as follows:

- As a minimum, the importance of risk analysis should be recognized in planning, design, cost estimation and budgeting, and construction and project management.
- Enhanced education, training, and professional practice are essential.

# Chapter 10: Conclusions, Contributions, Limitations, and Recommendations

## 10.1 Summary and Conclusions

The availability and application of planning and sustainability rating tools such as ENVISION has resulted in the awareness in the profession that major investments in highway infrastructure should achieve satisfactory rating for avoiding cost overruns and in fulfilling their role in enhancing the sustainability of cities and regions. Although the knowledge of these challenges has been conveyed by developers and promoters of sustainability rating tools, assessment of investments in transportation infrastructure including highways has received insufficient research attention in modelling life cycle cost risk and the associated sustainability effectiveness.

At the outset of this research, it was recognized that the increasing acceptance of the sustainability rating tools for evaluating highway projects implies going beyond the past attention on just location and design of highway infrastructure guided mainly by functional considerations. That is, in defining the research problem, it was recognized that the use of infrastructure rating tools such as ENVISION has the potential to improve the effectiveness of investment in highway projects in terms of meeting sustainability criteria, including formal recognition of the importance of probability-based life cycle analysis. In defining the research objectives and plan, it was intended to support the application of rating tools by modelling risk in life cycle cost estimates, including the identification and quantification of cost overrun factors. In addition, the need for risk analysis methodology

for the joint treatment of multi-attribute criteria that encompass economic and other factors of sustainability was recognized.

Consequently, the objectives of this research were shaped by the present highly demanding transportation planning environment that calls for modelling risk in the first cost as well as life-cycle costs. Also, the integration of life cycle cost risk factors with other sustainability criteria for assessing project alternatives was pursued as a research objective.

In order to go beyond the current state of knowledge, the research plan included the following tasks:

- Developing the probability models of cost overruns,
- Treating risk and uncertainty in life cycle analyses, using decision-theoretic, utility-theoretic, and Bayesian methods required for evaluation of investment alternatives,
- Application of Factor Analysis method that characterizes causes of cost overruns and development of logistics models based on Factor Analysis results for quantifying the impact of factors on cost overruns.

Data were obtained on actual projects from the Province of British Columbia and analyzed regarding cost overruns. Besides, a questionnaire study was implemented to obtain data on the causes of cost overruns. The questionnaire was sent to transportation agencies/jurisdictions in Canada, the USA, the Middle East, and Australia.

The developed cost-overrun probability models confirmed that cost overruns do occur. It should be noted that cost estimates include contingency funds and that the cost overruns go beyond these estimates. However, the British Columbia highway project data

did not show the incidence of major cost overruns as reported in the literature. Based on actual data analysis, best probability models were identified. These were calibrated by analyzing data based on the difference between the actual cost and cost estimate developed prior to construction phase. In addition, models were developed based on estimated cost at the planning phase. The computational challenges were well served by MATLAB software.

Five probability distributions were studied as candidates in the analysis. These are Log-logistic, Cauchy, Log-normal, Gamma, and Normal distributions. The maximum likelihood estimation (MLE) approach was used to estimate the parameter of the distributions. However, the Log-logistic distribution was found to be the best based on "budget at the planning phase" data. On the other hand, the Cauchy distribution was found to be the best for the "budget prior to the construction phase" data.

The purpose of developing the cost overrun model is to select the optimal probability distribution that best fits the British Columbia data and estimate the probability of cost overrun. The developed model can be used by any transportation agency around the world as predictive tool.

The decision-theoretic and utility-theoretic methods that were developed and illustrated in the evaluation of investment alternatives were found to be useful for formally treating life cycle costs and other factors of sustainability. The developed methods are able to analyze life cycle-based economic factors (i.e., net present worth) of alternatives as well as other factors of sustainability. Additionally, a method is suggested for taking into account the differential effects on incidence groups and applicable probabilities of impacts. However, the developed methods have the potential to identify the "preferred alternative" based on economic, social, and environmental criteria while treating risks.

Adopting the Bayesian approach, the decision-maker has the opportunity to update the probabilities of unknown states. A method for decision analysis based on the Bayesian approach that evaluates the anticipated reduction in risk using the potential role of additional data is presented and illustrated.

The Factor Analysis and associated logistic regression models based on survey data characterize the effect of factors that cause cost overruns. A total of 53 independent variables were identified and analyzed using three versions of the Factor Analysis method. The computational challenges were well served by SPSS. The methodology enabled the extraction of underlying factors that were subsequently tested in Logistic models for their impact on cost overruns. These models enable the estimation of the probability of cost overrun occurring. The goodness of fit of the models (i.e., their statistical significance) was assessed by the Log likelihood, Pseudo-R-squared, and Chi-square. The odds ratio statistics was used as a guide to identify notable factors that cause cost overruns.

The Factor Analysis and associated logistic regression models based on survey data characterize the effect of factors that cause cost overruns. A total of 53 independent variables were identified and analyzed using three versions of the Factor Analysis method. The computational challenges were well served by SPSS. The methodology enabled the extraction of underlying factors that were subsequently tested in Logistic models for their impact on cost overruns. These models enable the estimation of the probability of cost overrun occurring. The goodness of fit of the models (i.e., their statistical significance) was assessed by the Log likelihood, Pseudo-R-squared, and Chi-square. The odds ratio statistics was used as a guide to identify notable factors that cause cost overruns.

However, based on Version 1 analysis, factor analysis identified eight underlying factors of cost overruns. These are: (1) Financial, materials, labor, and equipment related factors. (2) Design, construction, and scheduling related factors. (3) Policy, budget, and operation related factors. (4) Decision making related factors. (5) Material prices fluctuation and work performance factors. (6) Unexpected condition related factors. (7) Quality control & quality assurance related factors, and (8) Owner/client related factors. Based on Version 2 analysis, three underlying factors of cost overruns were identified: (1) Design, construction, financial, and funding policy-related factors. (2) Inexperienced or lack of qualification of designer, contractor, client, and project manager related factors, and (3) Site management related factors. Based on Version 3 analysis, five underlying factors were identified. These are: (1) Design, construction, and scheduling-related & inexperienced or lack of qualification of designer, contractor, client, and project manager related factor. (2) Work performance related factors. (3) Decision-making tasks. (4) and (5) Unexpected problem and design issue factors.

Hence, a logistic regression analysis was conducted to estimate the probability of the occurrence of cost overrun. Several logistic models were developed in order to find any significant associations between independent factors and the dependent variable. The results of the models are summarized in Tables 9.14 and 9.15.

Detailed conclusions based on results of analyses are presented at the end of Chapters 4, 5, 6, 8, and 9.

The following text shows how the products of this research can be used in professional practice by decision-makers and other researchers.

1- The decision-maker has the opportunity to use the methodology for the cost overrun model. It is a predictive tool to estimate the probabilities of cost overrun and to identify the ranges of cost overrun ratio (COR) based on data from actual completed projects.
2- Also, the decision-maker has the opportunity to evaluate and analyze the project alternatives based on economic factors, including other factors of sustainability by using the decision-theoretic model and utility theory model and then the preferred alternative can be selected.
3- The use of the Bayesian analysis model enables the decision-maker to update or refine the prior probabilities by acquiring additional information for the purpose of risk reduction. Additional information can potentially reduce the risk.
4- The purpose of factor analysis and a logistic regression model is to identify the common factors that cause cost overrun based on an international survey. The decision-maker has the opportunity to take into account these factors during the evaluation of project alternatives in order to reduce the risk of cost overrun.

## 10.2 Contributions of This Research

### 10.2.1 Methodological Advances

This research has made a number of original contributions in methods of potential use to other researchers as well as to persons engaged in the professional practice of developing and evaluating highway investment alternatives. In addition, answers are

provided to outstanding questions in life cycle cost risk analysis of investments in highways.

1. The developed probability models of overruns in first cost are based on actual data that enable the decision-makers to appreciate risk levels.
2. The decision-theoretic and associated utility-theoretic methods serve a number of purposes. These enable life cycle cost risk assessment and the inclusion of other sustainability factors (e.g., network accessibility, greenhouse gas emissions, equity, and social justice) in evaluating investment alternatives. The utility-theoretic method shows how to integrate multi-criteria (based on monetary as well as other factors) for quantifying the effectiveness of alternatives while treating the probabilistic nature of criteria achievement levels.
3. The Bayesian method treats life cycle cost risk and goes beyond this requirement by quantifying the value of obtaining additional information in support of decision-making under risk.
4. The Factor Analysis, based on international survey data, has resulted in the identification of factors of cost overrun and produced results for use in calibrating logistic regression models of cost overruns. These logistic regression models are a contribution as a tool to quantify the cost overrun effect of factors.

Taken together, the methods advanced in this research enable a formal treatment of risk on a life cycle basis in order to enhance highway infrastructure planning and management. The availability of developed models has the potential to enhance the informed application of a rating tool such as ENVISION in support of planning and

evaluation of highway infrastructure projects. Please see Section 10.2.2 for further information on the link to ENVISION. The results of Factor Analysis and associated logistic regression models are also a contribution to knowledge in the field.

### 10.2.2 Link to ENVISION

As noted in this thesis document, ENVISION is a framework for the sustainability rating of infrastructure projects. The purpose of this section is to describe how the products of this thesis research can be used for informed applications of this tool. For the information of the reader, this framework provides a structure that enables users to study the achievement of a project in meeting objective, subjective, quantifiable, and qualitative criteria at various stages of project and system development. It can be applied at the planning, design, and post-construction phases. The verification that the applicant (the project owner) has implemented the intended sustainability measures in the final design can take place either after the design phase (at or after 95% design completion) or after the construction phase (at or after 95% construction completion).

Over the years, as a result of feedback from researchers as well as persons engaged in professional practice, developers of ENVISION made it flexible in application and also added new requirements. The following are a few notable changes that appeared in the 2018 version (titled Envision: Sustainable Infrastructure Framework Guidance Manual, 2018) (Institute For Sustainable Infrastructure ISI, 2018).

1. In addition to its application at post-design and post-construction phases, the applicants are encouraged to apply sustainability criteria at the planning phase. For

application at this phase, approaches and methods for preparing required inputs to ENVISION that characterize a project are expected to treat items 2 to 4 noted below.

2. The inclusion of a "base case" or the "do-nothing" case is a requirement so that the merits of alternatives can be found by comparing their effectiveness with the base case.
3. There is an emphasis on life-cycle analysis, risk analysis, consideration of related sustainability criteria (which implies the use of multi-attribute criteria),
4. Resilience criteria are added that supplement previously defined sustainability criteria.

To suggest how the products of this thesis research can enhance the application of ENVISION, it is necessary to briefly describe the logistics of applying the rating tool to a project. The reader is advised to refer to the 2018 version of ENVISION Manual for details. The applicant (i.e., the owner of the infrastructure project) is encouraged by ENVISION developers to use the latest available knowledge, including methods to characterize the project in terms of the extent to which relevant sustainability criteria are met. Good practices call for the generation of alternative concept level plans/designs that meet sustainability criteria, and identification and enhancement of the preferred alternative. The analysis and evaluation results become inputs to ENVISION. Higher capability methods and new information that become available due to new research can enhance the quality of inputs to ENVISION. This is referred to in the thesis document as "informed application" of the rating tool.

The products of this research (i.e., developed methods and new knowledge) have the potential to enhance ENVISION application at the planning phase of a highway infrastructure project. At this phase, following the generation of alternatives (usually up to three), there is a need to analyze and evaluate these so that the preferred alternative is identified for progressing to the design phase. The lifecycle approach and methods advanced in this thesis are well suited to analyze and evaluate alternatives and identify the preferred alternative for the next phase in the project life cycle. If the project owner is interested in ENVISION application at this stage in the project life cycle, the information generated by the methods advanced in this thesis research can become inputs to the ENVISION rating scheme. These information items, including the preferred alternative, are much more refined for the rating application as compared to the existing practice.

Table 10.1 presents elements of the link to ENVISION. The enhancement occurs primarily in shaping alternatives as well as the identification and improvements to the preferred alternative (i.e., additional features if applicable). These steps and their results become a part of the inputs to ENVISION. Logically, high-quality detailed inputs to the rating tool will result in higher ratings for a project with higher sustainability potential following construction.

The link to ENVISION addresses the requirements of (1) life cycle economic evaluation (LD3.3), and (2) identification of the preferred alternative using related multi-criteria (e.g., LD3.3 in association with QL3.1 equity and social justice, CR2.6 infrastructure integration, CR1.2 greenhouse gas emissions).

The suggested use of research products of this thesis research as aids in enhancing inputs to ENVISION is noted in Table 10.1. Here, explanations are provided on the

elements of the link to ENVISION. Step 1 consists of using products of this research for shaping project alternatives, identification of the preferred alternative, and improvement to the sustainability potential of the preferred alternative. Step 2 requires documentation of methods used and results obtained in the format required by ENVISION. These become inputs to the rating tool. Step 3 calls for the application of ENVISION to obtain credits. The end result is a higher performance potential of the project as compared to the scenario in which the new knowledge and methods are not used in preparing inputs to the rating tool.

It is to be noted that the selection of criteria related to LD3.3 is not limited to the examples used in the thesis for illustration purposes. The utility-theoretic method can handle all relevant criteria (i.e., there is no limit to the number of criteria included in the analysis).

Specifically, the contribution of this research to LD3.3 is as follows. (1) Structuring life cycle costs: initial and other life costs (Chapters 5 and 6). The life cycle economic evaluation approach (based on the costs and benefits) can enhance decisions that lead to sustainable projects. (2) Identification and incorporation of risks into the cost-benefit analysis (according to ENVISION, this is a best practice approach to cost-benefit analysis). Please see Chapters 5 and 6).

*Table 10. 1: Link to ENVISION*

| ENVISION Requirement | Contribution of this Research | Suggested Use of Research Products in ENVISION Application |
|---|---|---|
| Life cycle costs and risk<br>▪ Conduct a life-cycle economic evaluation LD3.3 | Initial cost overrun (probability) models (Chapter 4)<br>Decision-theoretic (including Bayesian) models for life-cycle economic evaluation (Chapters 5 & 6)<br>Cost-overrun factors and logistic predictive models, mitigation measures (Chapters 8 & 9) | Step 1: Use of products of this research to define project alternatives, identification of the preferred alternative, and improvement to the sustainability potential of the preferred alternative.<br>Step 2: Prepare inputs for ENVISION application.<br>Step 3: Obtain ENVISION credits.<br>End Result: Higher sustainability performance potential. |
| Identification of preferred alternative using related criteria<br>Example: Joint use of<br>▪ LD3.3 (life cycle economic evaluation),<br>▪ QL3.1 (equity and social justice),<br>▪ CR2.6 (improve infrastructure integration),<br>▪ CR1.2 (reduce greenhouse gas emissions) | Utility-theoretic method for identifying preferred alternative using economic and other sustainability factors (Chapter 5). | Step 1: Use of this method to define project alternatives, identification of the preferred alternative, and improvement to the sustainability potential of the preferred alternative.<br>Step 2: Prepare inputs for ENVISION application.<br>Step 3: Obtain ENVISION Credits.<br>End Result: Higher sustainability performance potential. |

The initial cost overrun probability models (Chapter 4) provide useful information to the analyst for incorporating risk in the study of alternatives (as demonstrated in Chapters 5 and 6). Likewise, the Factor Analysis and logistic predictive models (Chapters

8 and 9) can serve as guides for the analyst to mitigate the risk of cost-overruns in defining alternatives. The decision-theoretic (including Bayesian) models for life-cycle economic evaluation (Chapters 5 and 6) are intended to include risk factors in the study of relative effectiveness of alternatives.

The utility-theoretic method works with economic and other sustainability factors in a risk-based framework (Chapter 5). It is a contribution to the best practice approach in the use of multi-criteria for evaluating alternatives and identification of the preferred alternative. For an illustration of the use of the method, four sustainability criteria are used. The life cycle economic evaluation criterion (LD3.3) has already been described above. The equity and social justice criterion (QL3.1) addresses issues of differential impacts of the highway project. The infrastructure integration criterion (CR2.6) quantifies the achievement of an alternative in enhancing the operational relationships and strengthening the functional integration of the project into connected, efficient, and strategically important infrastructure systems. The greenhouse gas emission criterion (CR1.2) assesses greenhouse gas emission reduction during the operation of the project.

Column three of Table 10.1 shows that in Step 1 of the use of this method, project alternatives are characterized in relative value terms, the preferred alternative is identified, and attempts can be made to further improve the sustainability potential of the preferred alternative. Step 2 is devoted to the preparation of inputs to ENVISION. Here, details of the method, analyses, and results are to be documented. Next, Step 3 is devoted to the application of ENVISION, and credits are obtained. The end result is expected to be the high-performance potential of the preferred alternative following construction.

The thesis research does not intend to alter the rating process of ENVISION. But it is contended that with the application of contributed methods and knowledge, a project to be rated by ENVISION has much potential to obtain higher credits in sustainability and resilience factors. For example, life cycle risk analysis of economic factors, preferably in association with other factors of sustainability in quantifying the utility values for alternatives, will identify alternatives with the potential to achieve higher rates within the ENVISION framework.

## 10.3 Limitations

This thesis research, like any other, is not without limitations. As such, the models presented in this study would have limitations, some of which are presented as follows:

1. The research focused on all types of highway infrastructure projects together and no analysis was made on the basis of individual types.
2. It would be of interest to obtain data from actual projects which include information about life cycle costs (e.g., rehabilitation costs).
3. In the case of B.C data, no very high-cost overruns were found. It would be of interest to obtain data from other actual projects in other jurisdictions that encountered very high-cost overruns.
4. It would be of interest to increase the sample size of the questionnaire returns.
5. The purpose of factor analysis is to reduce a large number of variables to a small number of factors. Naming the factors can be difficult, especially when there are many clustered variables.

6. If additional data could be obtained, it would be of interest to run a separate analysis of factor analysis and logistic regression model based on location of jurisdictions (i.e., Canada, the USA, Middle East, and Australia).

The aim of this section was to note some limitations while recognizing data constraints. However, the author believes that the above discussed limitations are not considered as being severe and do not show a fundamental weakness of this study. In fact, some of the stated issues are beyond the scope of this research and might be used to identify possible areas for further research.

## 10.4 Recommendations for Future Research Works

1. This research is focused on lifecycle cost overruns. It is recommended to develop similar methods for schedule delay risks.
2. Given that in the future, large investments in public transit (i.e., rail rapid transit, light rail transit, bus rapid transit) will be made, it will be useful to carry out similar research on cost overruns in public transit projects. The results will help in taking steps to avoid cost overruns.
3. The rating tools such as ENVISION apply a large number of sustainability rating factors. This research has contributed knowledge on how to integrate life cycle costs overrun with other sustainability factors. There is a need for research in developing utility functions that exhibit diminishing marginal utility.
4. The development of new specialized software for predicting the probability of cost overruns will enable an agency to predict the probability of cost overrun efficiently.

The use of general-purpose software such as MATLAB requires complex coding steps.

5. The cost overrun probability models developed in this research are based on data contributed by the Province of British Colombia, Canada. Further research will be useful based on data acquired from different provinces or countries.

# References


Abdul Rahman, I., Memon, A. H., Karim, A., & Tarmizi, A. (2013). Significant Factors Causing Cost Overruns in Large Construction Projects in Malaysia. Journal of Applied Sciences, 13(2), 286–293. https://doi.org/10.3923/jas.2013.286.293

Adnan, A. (2012). Sustainable Development of The Rail System: The Case Of Malaysia.

Ahmed, S., & Hasan, B. (2014). Deviation in the Cost of Projects. In Construction Macroeconomics Conference.

Alinaitwe, H., Apolot, R., & Tindiwensi, D. (2013). Investigation into the causes of delays and cost overruns in Uganda's public sector construction projects. Journal of Construction in Developing Countries, 18(2), 33–47.

American Society of Civil Engineering (ASCE) and Eno Center for Transportation (2014). Maximizing the Value of Investment Using Life Cycle Cost Analysis. ASCE, Reston, Va.,

Arif, F., Lodi, S. H., & Azhar, N. (2015). Factors influencing accuracy of construction project cost estimates in Pakistan: Perception and reality. International Journal of Construction Management, 15(1), 59–70. https://doi.org/10.1080/15623599.2015.1012141

Azis, A. A. A., Memon, A. H., Rahman, I. A., & Karim, A. T. A. (2013). Controlling cost overrun factors in construction projects in malaysia. Research Journal of Applied Sciences, Engineering and Technology, 5(8), 2621–2629. https://doi.org/10.19026/rjaset.5.4706

Bangdiwala, S. I. (2018). Regression: binary logistic. International Journal of Injury Control and Safety Promotion, 25(3), 336–338. https://doi.org/10.1080/17457300.2018.1486503

Berechman, J., & Chen, L. (2011). Incorporating Risk of Cost Overruns into Transportation Capital Projects Decision-Making. Journal of Transport Economics and Policy, 45(1), 83–104.

Berechman, J., & Wu, Q. (2006). Cost Overruns Risk Anaysis in Transportation Infrastructure Investments. https://doi.org/10.1139/L06-077

Black, & Veatch. (2017). Planning Sustainable Infrastructure Projects with ISI Envision. Retrieved January 5, 2020, from https://www.bv.com/perspectives/planning-sustainable-infrastructure-projects-isi-envision

Blaikie, N. (2003). Analyzing Quantitative Data From Description To Explanation. https://doi.org/10.4135/9781849208604

Boardman, A. E., Greenberg, D. H., Vining, A. R., & Weimer, D. L. (2001). Cost-benefit analysis : concepts and practice. Upper Saddle River, New Jersey: Prentice Hall.

Bryman, A. (2012). Social research methods (4th ed.). Oxford, New York: Oxford University Press.

Caltrans (2011). Using Life Cycle Cost Analysis in Highway Project Development. CTC & Associates LLC, for Caltrans Division of Research and Innovation, Sacramento, 2011,

Campbell, M. J. (2008). Statistics at Square Two: Understanding Modern Statistical Applications in Medicine. https://doi.org/10.1002/9780470755839

Cantarelli, C. C., Van Wee, B., Molin, E. J., & Flyvbjerg, B. (2012). Different cost performance: Different determinants?. The case of cost overruns in Dutch transport infrastructure projects. Transport Policy, 22, 88–95. https://doi.org/10.1016/j.tranpol.2012.04.002

Cantarelli, Chantal C, Chorus, C. G., & Cunningham, S. W. (2013). Explaining cost overruns of large-scale transportation infrastructure projects using a signalling game. Transportmetrica A: Transport Science, 9(3), 239–258. https://doi.org/10.1080/18128602.2011.565817

Chitkara, K.K. (2011), Construction Project Management: Planning, Scheduling, and Controlling, 2nd Edition, McGraw Hills.

Congdon, P. (2007). Bayesian statistical modelling. John Wiley & Sons.

Costello, A. B., & Osborne, J. W. (2011). Best Practices in Exploratory Factor Analysis: four recommendations for getting the most from your analysis. Practical Assessment, Research, and Evaluation, 10(1), 1–7. https://doi.org/10.4135/9781412995627.d8

Dodgson, J. S., Spackman, M., Pearman, A., & Phillips, L. D. (2009). Multi-criteria analysis : a manual. In Appraisal. https://doi.org/10.1002/mcda.399

El-Sayegh, S. M., & Mansour, M. H. (2015). Risk assessment and allocation in highway construction projects in the UAE. Journal of Management in Engineering, 31(6). https://doi.org/10.1061/(ASCE)ME.1943-5479.0000365

Endut, I. R., Akintoye, A., & Kelly, J. (2005). Cost and Time Overruns of Projects in Malaysia. ICONDA Prooceedings of the 2nd Scottish Conference for Postgraduate Researchers of the Built and Natural Environment (PRoBE), (2001), 243–252.

Ezzy, D. (2013). Qualitative analysis. Journal of Chromatography Library, 5(C), 229–245. https://doi.org/10.1016/S0301-4770(08)61253-1

Federal Government of Canada. (2015). Public-Private-Projects. Retrieved December 30, 2019. fromhttps://lop.parl.ca/sites/PublicWebsite/default/en_CA/ ResearchPublications/201550E

Federal Highway Administration (FHWA) (2015). Improving Transportation Investment Decisions through Life-Cycle Cost Analysis. Washington, D.C., updated Aug. 2015 [Online]. Available: http:// www.fhwa.dot.gov/infrastructure/asstmgmt/lccafact. cfm.

Field, A. (2009). Discovering statistics using SPSS (3rd edn). Sage Publications Ltd., London. https://doi.org/10.5860/choice.43-3433

Flyvbjerg, B. (2007a). Cost overruns and demand shortfalls in urban rail and other infrastructure. Transportation Planning and Technology, 30(1), 9–30.

Flyvbjerg, B. (2007b). Policy and planning for large-infrastructure projects: Problems, causes, cures. Environment and Planning B: Planning and Design, 34(4), 578–597. https://doi.org/10.1068/b32111

Flyvbjerg, B. (2009). Survival of the unfittest: Why the worst infrastructure gets built-and what we can do about it. Oxford Review of Economic Policy, 25(3), 344–367. https://doi.org/10.1093/oxrep/grp024

Flyvbjerg, B. (2016). The Fallacy of Beneficial Ignorance: A Test of Hirschman's Hiding Hand. World Development, 84, 176–189. https://doi.org/10.1016/j.worlddev.2016.03.012

Flyvbjerg, B., Holm, M. S., & Buhl, S. (2002). Underestimating Costs in Public Works Projects: Error or Lie? Journal of the American Planning Association, 68(3), 279–295. https://doi.org/10.1080/01944360208976273

Flyvbjerg, B., Skamris Holm, M. K., & Buhl, S. L. (2003). How common and how large are cost overruns in transport infrastructure projects. Transport Reviews, 23(1), 71–88. https://doi.org/10.1080/0144164022000016667

Flyvbjerg, B., Skamris Holm, M. K., & Buhl, S. L. (2004). What causes cost overrun in transport infrastructure projects? Transport Reviews, 24(1), 3–18.

Frame, J. (2003). Managing risk in organizations: A guide for managers. John Wiley & Sons.

Fuller, S. K., & Petersen, S. R. (2006). Life Cycle Costing Manual for the Federal Energy Management Program (NIST Handbook 135).

Galway, L. (2004). Quantitative Risk Analysis for Project Management - A Critical Review. In RAND Corporation.

Gilbert, N. (2008). Researching Social Life (3rd ed.). London: Sage Publications Ltd.

Griffiths, K., Boyle, C., & Henning, T. F. P. (2019). Infrastructure sustainability rating tools–exploring opportunities for the engineering profession based on international research. Australian Journal of Civil Engineering, 1–11. https://doi.org/10.1080/14488353.2019.1688590

Gschösser, F. and H. Wallbaum, H. (2013). Life Cycle Assessment of Representative Swiss Road pavements for National Roads with an Accompanying Life Cycle Cost Analysis. Vol. 47, No. 15, 2013, pp. 8453–8461.

Hahs-Vaughn, D. L., & Lomax, R. G. (2013). An Introduction to Statistical Concepts. https://doi.org/10.4324/9780203137819

Hair, J. F., Black, W. C., Babin, B.J., Anderson, R.E., & Tatham, R. L. (1998). Multivariate Data Analysis (5th ed.). Vol. 5, No. 3, pp. 207-219. Upper Saddle River, NJ: Prentice hall.

Hampson, K., & Brandon, P. (2004). Construction 2020 : A Vision for Australia's Property and Construction Industry. CRC Construction Innovation. Retrieved from https://eprints.qut.edu.au/40762/

Heravi, G., & Mohammadian, M. (2019). Investigating cost overruns and delay in urban construction projects in Iran. International Journal of Construction Management, 1–11. https://doi.org/10.1080/15623599.2019.1601394

Holmes, S. (2016). The Economic Impact of Canadian P3 Projects: Why building infrastructure "on time" matters. Retrieved from https://www.pppcouncil.ca/web/pdf/cancea_report_economic_impact_of_p3s.pdf

Hussin, J. M., Abdul Rahman, I., & Memon, A. H. (2013). The Way Forward in Sustainable Construction: Issues and Challenges. International Journal of Advances in Applied Sciences, 2(1). https://doi.org/10.11591/ijaas.v2i1.1321

Ibrahim, A. R. Bin, Roy, M. H., Ahmed, Z., & Imtiaz, G. (2010). An investigation of the status of the Malaysian construction industry. Benchmarking, 17(2), 294–308. https://doi.org/10.1108/14635771011036357

Institute For Sustainable Infrastructure (I.S.I) (2018). ENVISION: Sustainable Infrastructure Framework Guidance Manual V3. Retrieved January 5, 2020, from Institute For Sustainable Infrastructure website: https://sustainableinfrastructure.org/

Jackson, S. (2002). Project Cost Overruns and Risk Management. Proceedings of Association of Researchers in Construction Management 18th Annual ARCOM Conference, Newcastle, Northumber University, UK (pp. 1-10).

Kaiser, H. F. (1974). An index of factorial simplicity. Psychometrika, 39(1), 31–36. https://doi.org/10.1007/BF02291575

Kaming, P. F., Olomolaiye, P. O., Holt, G. D., & Harris, F. C. (2012). Factors influencing construction time and cost overruns on high-rise projects in Indonesia. Construction Management and Economics, 15(1), 83–94. https://doi.org/10.1080/014461997373132

Khan, A.M. (2013). Risk factors in toll road life cycle analysis, Transportmetrica A: Transport Science, 9:5, 408-428, DOI: 10.1080/18128602.2011.587134

Khan, A. M., Bacchus, A., & Erwin, S. (2012). Policy challenges of increasing automation in driving. IATSS Research, 35(2), 79–89. https://doi.org/10.1016/j.iatssr.2012.02.002

Khan, A. M., & LaFontaine, P. (1986). The Management of Highway Infrastructure Rehabilitation: A Forward Look. 459–479. Vancouver: World Conference on Transport Research.

Khan, A.M. (2010). Bayesian predictive travel time methodology for advanced traveller information systemTraveller Information System. Journal of Advanced Transportation. Published online in Wiley Online Library. DOI: 10.1002/atr.147

Kibert, C. J. (1994). Final session of first international conference of CIB TG 16 on sustainable construction. Book Final Session of First International Conference of CIB TG, 16.

Kim, J.-O., & Mueller, C. (2011). Introduction to Factor Analysis. In Introduction to Factor Analysis. https://doi.org/10.4135/9781412984652

Kleine, R. J. (2009). Better Projects Through Risk Analysis, Case Study – Highway 69 Four-Laning. Ottawa: CSVA Conference.

Korb, K. B., and Nicholson, A. E., (2011). Bayesian Artificial Intelligence. Chapman & Hall/CRC, Boca Rston, Florida: CRC Press

Krishnamoorthy, K. (2016). Handbook of statistical distributions with applications. In Handbook of Statistical Distributions with Applications. https://doi.org/10.1198/tech.2007.s494

Larocque, S., & Parker, J. (2009). An integrated process for cost & risk analysis and value engineering. CSVA Conference.

Li, J., Pollard, S., Kendall, G., Soane, E., & Davies, G. (2009). Optimising risk reduction: An expected utility approach for marginal risk reduction during regulatory decision making. Reliability Engineering and System Safety, 94(11), 1729–1734. https://doi.org/10.1016/j.ress.2009.05.005

Lifson, M. W. (1968). Cost-effectiveness: the economic evaluation of engineered systems. New York: John Wiley & Sons.

Lind, H., & Brunes, F. (2014). Policies to avoid cost overruns in infrastructure projects: Critical evaluation and recommendations. Australasian Journal of Construction Economics and Building, 14(3), 74–85. https://doi.org/10.5130/ajceb.v14i3.4151

Lind, H., & Brunes, F. (2015). Explaining cost overruns in infrastructure projects: A new framework with applications to Sweden. Construction Management and Economics, 33(7), 554–568. https://doi.org/10.1080/01446193.2015.1064983

List, G. (2007). A model for life cycle evaluation of highway investments. Structure and Infrastructure Engineering, 3(2), 95–101. https://doi.org/10.1080/15732470600590903

Love, P. (2011). Plugging the gaps' between optimum bias and strategic misrepresentation and infrastructure cost overruns. Procedia Engineering, 14, 1197–1204. https://doi.org/10.1016/j.proeng.2011.07.150

Love, Peter, Edwards, D. J., Watson, H., & Davis, P. (2010). Rework in civil infrastructure projects: Determination of cost predictors. Journal of Construction Engineering and Management, 136(3), 275–282. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000136

Lundberg, M., Jenpanitsub, A., & Pyddoke, R. (2011). Cost overruns in Swedish transport projects. 1–17.

Maher, M. L. ., & McGoey-Smith, A. . (2006). Risk-Based Cost and Schedule Estimation for Large Transportation Projects. 18–20. In Annual European Transport Conference.

Mateus, R., & Bragança, L. (2011). Sustainability assessment and rating of buildings: Developing the methodology SBToolPT-H. Building and Environment, 46(10), 1962–1971. https://doi.org/10.1016/j.buildenv.2011.04.023

Memon, A. H., Rahman, I. A., & Azis, A. A. A. (2011). Preliminary Study on Causative Factors Leading to Construction Cost Overrun. International Journal of Sustainable Construction Engineering and Technology, 2(1), 57–71. Retrieved from https://www.researchgate.net/publication/256662503

Merewitz, L. (1973). Cost overruns in public works. In Benefit-Cost and Policy Analysis (pp. 277–295).

Montgomery, D. C., & Runger, G. C. (2010). Applied Statistics and Probability for Engineers. In John Wiley & Sons (Vol. 19). https://doi.org/10.1080/03043799408928333

Ministry of Transportation and Infrastructure B.C. (2013). Project Cost Estimating Guidelines. Retrieved from http://gww.th.gov.bc.ca/gwwpmss/Content/costestimating.asp

Naidu, G. (2008). Infrastructure development in Malaysia. International Infrastructure Development in East Asia–Towards Balanced Regional Development and Integration (Vol. 204). IDE-JETRO.

Naoum, S. G. (1998). Dissertation Research and Writing for Construction Students (1st ed.). https://doi.org/10.4324/9780203720561

National Academies of Sciences, Engineering, and Medicine (2016). Life-Cycle Cost Analysis for Management of Highway Assets. Washington, DC: The National Academies Press. https://doi.org/10.17226/23515.

Nikic, R. (1998). Construction Project Risk Management in a Transition Country. master's thesis, Faculty of Civil Engineering, University of Zagreb, Zagreb.

Nkurunziza, D., & Irumva, A. (2018). Sustainability Analysis of Road Construction Projects in Kigali City-Rwanda. Lusaka–Zambia: Commonwealth Association of Surveying and Land Economy (CASLE).

Omoregie, A., & Radford, D. (2006). Infrastructure delays and cost escalation: causes and effects in Nigeria. Proceeding of Sixth International Postgraduate Research Conference Netherlands, Delft University of Technology, 79–93. Retrieved from http://hdl.handle.net/10545/621459

Pratt, J.W., and Schlaifer, R. (1995). Introduction to Statistical Decision Theory. MIT Press.

Raiffa, H. and Schlaifer, R. (1961). Applied Statistical Decision Theory. The MIT Press.

Rahim, F. A., Muzaffar, S. A., Yusoff, N. S. M., Zainon, N., & Wang, C. (2014). Sustainable Construction Through Life Cycle Costing. Journal of Building Performance, 5(1). Retrieved from http://spaj.ukm.my/jsb/index.php/jbp/article/view/132

Ramabodu, M. S., & Verster, J. (2010). Factors Contributing to Cost Overruns of Construction Projects. 5th Built Environment Coference, 131–143.

Regnier, E. (2005). Activity Completion Times in PERT and Scheduling Network Simulation , Part II. DRMI Newsletter, (12). Retrieved from www.nps.navy.mil/drmi/

Reilly, J. (2008). Probable cost estimating and risk management. North American Tunneling 2008 Proceedings, 576–584.

Roberds, B., & Mcgrath, T. (2006). Quantitative Cost and Schedule Risk Assessment and Management for Large Infrastructure Projects. PMICOS Conference, (May).

Samarghandi, H., Tabatabaei, S. M. M., Taabayan, P., Hashemi, A. M., & Willoughby, K. (2016). Studying the reasons for delay and cost overrun in construction projects: The case of Iran. Journal of Construction in Developing Countries, 21(1), 51–84. https://doi.org/10.21315/jcdc2016.21.1.4

Savelsberg, E. (2008). Innovation in European Freight Transportation: Basics, Methodology and Case Studies for the European Markets. Springer Science & Business Media.

Schlaifer, R. (1959). Probability and Statistics for Business Decisions, An Introduction to Managerial Economics Under Uncertainty. McGraw-Hill Book Company.

Schlaifer, R., & Raiffa, H. (1968). Applied statistical decision theory. The MIT Press, Cambridge, Mass., USA.

Seiler, K. (1969). Introduction to systems cost-effectiveness (Vol. 17). https://doi.org/10.1016/B978-0-12-397025-1.09994-2

Sekaran, U., & Bougie, R. (2003). Research methods for business: A skill building approach. Long Range Planning, 26(2), 136. https://doi.org/10.1016/0024-6301(93)90168-f

Siemiatycki, M. (2015). Cost Overruns on Infrastructure Projects: Patterns, Causes, and Cures. IMFG Perspectives, (11), 14.

Sinesilassie, E. G., Tabish, S. Z. S., & Jha, K. N. (2018). Critical factors affecting cost performance: a case of Ethiopian public construction projects. International

Journal of Construction Management, 18(2), 108–119. https://doi.org/10.1080/15623599.2016.1277058

Singh, V. P., Jain, S. K., & Tyagi, A. (2007). Risk and reliability analysis: A handbook for civil and environmental engineers. In American Society of Civil Engineers (ASCE). https://doi.org/10.1061/9780784408919

Smith, T. J., & McKenna, C. M. (2013). A comparison of logistic regression pseudo R2 indices. Multiple Linear Regression Viewpoints, 39(2), 17–26. Retrieved from file:///C:/Users/Nga/Downloads/Smith_v39n2.pdf

Stevens, J. P. (2012). Applied Multivariate Statistics for The Social Sciences. https://doi.org/10.1201/9780429429422-7

Subramani, T. (2014). Causes of Cost Overrun In Construction. IOSR Journal of Engineering, 4(6), 01–07. https://doi.org/10.9790/3021-04640107

Syed, S. I., & Khan, A. M. (2001). Factor Analysis for the Study of Determinants of Public Transit Ridership. Journal of Public Transportation, 3(3), 1–17. https://doi.org/10.5038/2375-0901.3.3.1

Tabachnick, B. G., & Fidell, L. S. (2007). Using Multivariate Statistics. In Boston: Pearson Allyn and Bacon. https://doi.org/10.1037/022267

Terrill, M., & Danks, L. (2016). Cost overruns in transport infrastructure. Retrieved from Grattan Institute website: http://www.grattan.edu.au/

Toh, T. C., Ali, K. N., & Aliagha, G. U. (2011). Modeling construction cost factors in the Klang Valley area of Malaysia. ISBEIA 2011 - 2011 IEEE Symposium on Business, Engineering and Industrial Applications, 437–440. https://doi.org/10.1109/ISBEIA.2011.6088853

U.S. Department of Transportation (US DOT) (2015). Enhancing TSM&O Strategies Through Life Cycle Benefit/Cost Analysis. Authors: Salem, O., Chen, J., Salman, B. Washington, D.C.

US DOT, (RITA). (2010). Intellidrive Benefit-Cost Analysis. Produced by Volpe National Transportation Systems Center.

US FHWA. (2006). Guide to risk assessment and allocation for highway construction management. Retrieved from https://rosap.ntl.bts.gov/view/dot/50369

Vu, H. A., Wang, J., Min, L., Mai, S. H., & Nguyen, H. P. (2016). Research on Cost Overrun Risk of Construction Phase of Vietnam Highway International Contracting Project. Engineering, 08(03), 86–98. https://doi.org/10.4236/eng.2016.83011

Washington State Department of Transportation. (2014). Project Risk Management Guide. Engineering and Regional Operations. Retrieved December 30, 2019, from https://www.wsdot.wa.gov/publications/fulltext/CEVP/ProjectRiskManagementGuide.pdf.

White, G. E., & Ostwald, P. F. (1976). Life Cycle Costing. Management Accounting, 57(7), 39–42. https://doi.org/10.1111/j.1559-3584.1969.tb05481.x

Woodberry, O., Nicholson, A. E., Korb, K. B., & Pollino, C. (2004). Parameterising Bayesian networks. Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science). https://doi.org/10.1007/978-3-540-30549-1_108

WSP. (2015). Insights | WSP. Retrieved January 13, 2020, from https://www.wsp.com/en-CA/insights

Zhao, X., Hwang, B. G., & Yu, G. S. (2013). Identifying the critical risks in underground rail international construction joint ventures: Case study of Singapore. International Journal of Project Management, 31(4), 554–566. https://doi.org/10.1016/j.ijproman.2012.10.014

Žujo, V., Car-Pušic, D., & Brkan-Vejzović, A. (2010). Contracted price overrun as contracted construction time overrun function. Tehnicki Vjesnik, 17(1), 23–29.

# Appendices

# Appendix A

## *Appendix A. 1: Questionnaire Survey*

Carleton UNIVERSITY
Canada's Capital University

Date:
To:
From: Baraa Alfasi, Ph.D. Candidate, Carleton University, Ottawa, Canada
**Re: Request for Participation in an International Survey of Causes of Highway Infrastructure Cost Overruns**

I am a Ph.D. candidate at Carleton University (Ottawa) pursuing research in "Enhancing sustainability in highway infrastructure: risk analysis models". My studies are sponsored by the Saudi-Arabian government. A part of thesis research requires an international survey of causes of highway infrastructure cost overruns so that a detailed statistical study can be carried out with the use of Factor Analysis. The results are expected to contribute knowledge on reduction of incidence of cost overruns.

The context of the survey can be seen in the following description of overall analyses and modelling tasks: Life cycle cost variables and linkages; data acquisition on completed projects; cost overrun models; Monte Carlo method; Bayesian model; **Survey of factors responsible for cost overrun and Factor Analysis.**

Given that cost overruns in major highway projects is a subject of interest around the world, I would like to invite you to be a part of this research by completing a brief survey questionnaire. The survey questionnaire should take approximately 20 minutes to be completed. All information will be used for academic research purposes only and results will be shown in aggregated form.

There are no risks or personal benefits to participants in your organization. In accordance with the ethics protocol, the completion of the questionnaire is regarded as consent to participate. Upon request, contributed information can be withdrawn. The completed questionnaire files will be kept in a secure safety box for a period of three years and then deleted from the storage device. The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research (ID #110291). *This survey will comply with all clauses of Research Ethics. Should the participant have questions or concerns,* Dr. Bernadette Campbell, Chair, Carleton University Research Ethics Board-B can be contacted (by phone at 613-520-2600 ext. 4085 or via email at ethics@carleton.ca).

My thesis supervisor, Professor Ata Khan (Ph.D., P.Eng., F.CSCE, F. ITE) will be happy to answer questions regarding the overall thesis research or the survey part of research (ata.khan@carleton.ca; +613 520 2600 Extension 5786).

Sincerely,

Baraa Alfasi, PhD candidate, Department of Civil and Environmental Engineering Department, Carleton University, 1125 Colonel By Drive, Ottawa, Ontario, K1S5B6, Canada
E-mail: Baraa.Alfasi@carleton.ca

**Survey Questionnaire**

The questionnaire designed for the study is structured and multiple-choice type of questions are included. The questionnaire is divided into sections. Section 'A' comprises the background information of the respondents. Section 'B' is designed in relation to the purpose of the study. The questions are asked on a five-point Likert scale (very high, high, medium, low, very low).

We are requesting the Director/Head of the Department or a senior knowledgeable officer to complete the questionnaire. Following the completion of the questionnaire, please scan and send it as an email attachment. A hard copy can be mailed to my address noted above. If there are a number of knowledgeable persons in the organization, we would appreciate it if they could complete separate questionnaires. For statistical reasons, a high sample size is desirable.

## Part A General

**What is your current position?**

Director _____
Project Manager _____
Senior Project Engineer _____
Superintendent _____
Project Engineer _____
Field Engineer _____
Other _____

**Do/did you work with any of the following infrastructure projects?**

Highways _____
Bridges and tunnels _____
Other: _____

**How many years of experience do you have?**

1-5 years _____
6-10 years _____
More than 10 years _____
Other: _____

**Which region do you currently work in?**

Australia _____
New Zealand _____
Asia _____
Europe _____
U.K. _____
Middle East _____
Canada _____
U.S.A. _____
Mexico _____

**If you work for a funding agency (e.g. the World Bank), where are your projects located?**
Asia _____
Middle East _____
South America _____
Africa _____
Other: _____

**Which sector do you work for?**
Federal government _____
Provincial/State government _____
Municipal government _____
Private sector _____

If Private Sector (e.g. Consulting company, Contractor company), Client organization:
Federal government _____
Provincial/State government _____
Funding agency (e.g. World Bank) _____
Other: _____

**Have you worked in projects that have experienced cost overruns?**
Yes _____
No _____
Do not know _____
Other: _____

**How do you rate the occurrence of cost overruns in highway infrastructure projects?**
Usually _____
Sometimes _____
Rare _____
Do not know _____
Other: _____

## Part B Causes of Cost Overruns

PLEASE CHECK THE TYPE OF PROJECT FOR WHICH THE FOLLOWING RESPONSES ARE APPLICABLE. IF YOU HAVE EXPERIENCE WITH MORE THAN ONE TYPE OF PROJECT, PLEASE USE A SEPARATE FORM FOR EACH TYPE OF PROJECT.

**Project Type**

| Public | Private | Public-Private Partnership |
|---|---|---|
| | | |

**How would you rate each one of the following causes in contributing to cost overruns in infrastructure projects?**

**Policy**

**Changes in government funding policies**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Deal termination due to changes in law, government policy or protocols**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Change in regulations**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Design, Construction and scheduling**

**Complexity of the project (e.g. Project size, Project type, scope of work)**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Design changes during construction work**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Re-work due to the construction errors**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Unexpected technical problem**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Design errors that represent insufficient deliverables**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Changes by owner on the completion date of the project**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Scope changes by Owner during construction**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Delays related to owner or owner representative (e.g. stop work)**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Unrealistic project scheduling**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Acceleration to maintain schedule**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Delays in sending important documents to construction site (e.g. drawings, design changes)**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Type of construction contract (e.g. unit price contract)**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Unnecessary practices, specifications, procedures and documentation requirements forced onto the construction site workers**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Replacing unsatisfactory subcontractors from site by hiring new subcontractors**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Delay by subcontractor**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Site conditions/Environment**

**Poor site management**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Unexpected weather conditions**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Accidents due to poor site safety**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Materials and Equipment**

**Shortage of materials & equipment on site**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Damages in materials and equipment in transit to the construction site**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Late delivery of materials & equipment at the construction site**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Equipment Selection Changes**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Construction variations due to equipment selection**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Defective materials**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

| Labor, staff and vendor |
|---|

**Shortage of skilled labor**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Lack of staff at the time of construction**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**labor strikes & vendor strikes**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

| Permits and Approvals |
|---|

**Delays and approval of shop drawings and installation procedures**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Building Permit to the construction contractor**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Government/ Municipal Approvals**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

| Estimation / Budget and Financial | | | | |
|---|---|---|---|---|
| **Lack of expertise in setting the budget** | | | | |
| Very High | High | Medium | Low | Very Low |
| | | | | |
| Comments: | | | | |
| **The approved budget was too low** | | | | |
| Very High | High | Medium | Low | Very Low |
| | | | | |
| Comments: | | | | |
| **Absence of a detailed Estimate Plan** | | | | |
| Very High | High | Medium | Low | Very Low |
| | | | | |
| Comments: | | | | |
| **Changes in prices of items that have already been approved** | | | | |
| Very High | High | Medium | Low | Very Low |
| | | | | |
| Comments: | | | | |
| **Economic and financial factors** | | | | |
| Very High | High | Medium | Low | Very Low |
| | | | | |
| Comments: | | | | |
| **Inappropriate and inadequate procurement (e.g. payment terms, pricing)** | | | | |
| Very High | High | Medium | Low | Very Low |
| | | | | |
| Comments: | | | | |
| **Shortage of contingency and management reserve funds** | | | | |
| Very High | High | Medium | Low | Very Low |
| | | | | |
| Comments: | | | | |
| **Unaddressed overtime work or multiple shifts that was not included in the base estimate** | | | | |
| Very High | High | Medium | Low | Very Low |
| | | | | |
| Comments: | | | | |

**Bankruptcy of subcontractors and vendors during construction work**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Currency fluctuations**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Other Factors**

**Bad luck**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Lack of technical qualifications of the client**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Overly high expectations**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Poor communication and coordination between all parties**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Disputes between parties (designer, contractor, owner)**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Political Factors**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Technological risk**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Land acquisition issues within right-of-way**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Quality assurance and quality control**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

**Inexperienced Project Managers, Estimators and Planners**

| Very High | High | Medium | Low | Very Low |
|---|---|---|---|---|
| | | | | |

Comments:

# Appendix B

## *Appendix B. 1: SPSS Programming for Factor Analysis*

**Reliability Scale**

```
 /VARIABLES=V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
V20 V21 V22 V23 V24
   V25 V26 V27 V28 V29 V30 V31 V32 V33 V34 V35 V36 V37 V38 V39 V40 V41 V42 V43
V44 V45 V46 V47 V48 V49
   V50 V51 V52 V53
 /SCALE('ALL VARIABLES') ALL
 /MODEL=ALPHA.
```

**Version (1-1) Analysis 43 Variables**

```
FACTOR
 /VARIABLES V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V1
9 V20 V21 V22 V23 V24
   V25 V26 V27 V28 V29 V30 V31 V32 V33 V34 V35 V36 V37 V38 V39 V40 V41 V42 V43
 /MISSING LISTWISE
 /ANALYSIS V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V14 V15 V16 V17 V18 V19
V20 V21 V22 V23 V24
   V25 V26 V27 V28 V29 V30 V31 V32 V33 V34 V35 V36 V37 V38 V39 V40 V41 V42 V43
 /PRINT UNIVARIATE INITIAL CORRELATION DET KMO EXTRACTION ROTATION
 /FORMAT SORT BLANK(.5)
 /PLOT EIGEN
 /CRITERIA MINEIGEN(1) ITERATE(25)
 /EXTRACTION PC
 /CRITERIA ITERATE(25)
 /ROTATION VARIMAX
 /METHOD=CORRELATION.
```

**Version (1-2) Analysis 10 Variables**

```
 FACTOR
  /VARIABLES V44 V45 V46 V47 V48 V49 V50 V51 V52 V53
  /MISSING LISTWISE
  /ANALYSIS V44 V45 V46 V47 V48 V49 V50 V51 V52 V53
  /PRINT UNIVARIATE INITIAL CORRELATION SIG DET KMO EXTRACTION
 ROTATION
  /PLOT EIGEN
  /CRITERIA MINEIGEN(1) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25)
  /ROTATION VARIMAX
  /METHOD=CORRELATION.
```

**Analysis (2) Analysis 31 Variables**

```
FACTOR
  /VARIABLES V1 V2 V3 V4 V5 V6 V8 V9 V13 V14 V15 V16 V17 V18 V19 V21 V24 V26
V28 V31 V32 V34 V36 V38 V39 V40 V41 V42 V46 V52 V53
  /MISSING LISTWISE
  /ANALYSIS V1 V2 V3 V4 V5 V6 V8 V9 V13 V14 V15 V16 V17 V18 V19 V21 V24 V26
V28 V31 V32 V34 V36 V38 V39 V40 V41 V42 V46 V52 V53
  /PRINT INITIAL CORRELATION DET KMO AIC EXTRACTION ROTATION
/FORMAT SORT BLANK(.5)
  /PLOT EIGEN
  /CRITERIA MINEIGEN(1) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25)
  /ROTATION VARIMAX
  /METHOD=CORRELATION.
```

**Version (3) Analysis 27 Variables**

```
FACTOR
  /VARIABLES V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V15 V18 V19 V20 V28 V31 V33
V34 V35 V36 V37 V45 V47 V48 V51 V52 V53
  /MISSING LISTWISE
  /ANALYSIS V4 V5 V6 V7 V8 V9 V10 V11 V12 V13 V15 V18 V19 V20 V28 V31 V33 V34
V35 V36 V37 V45 V47 V48 V51 V52 V53
  /PRINT INITIAL CORRELATION DET KMO EXTRACTION ROTATION
  /FORMAT SORT BLANK(.50)
  /PLOT EIGEN
  /CRITERIA MINEIGEN(1) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25)
  /ROTATION VARIMAX
  /SAVE REG(ALL)
  /METHOD=CORRELATION.
```

## *Appendix B. 2: Factor Analysis Output – (Analysis Version 1)*

### Correlation Matrix

| | V1 | V2 | V3 | V4 | V5 | V6 | V7 |
|---|---|---|---|---|---|---|---|
| V1 | 1.000 | | | | | | |
| V2 | 0.681 | 1.000 | | | | | |
| V3 | 0.772 | 0.638 | 1.000 | | | | |
| V4 | 0.529 | 0.437 | 0.534 | 1.000 | | | |
| V5 | 0.475 | 0.375 | 0.414 | 0.725 | 1.000 | | |
| V6 | 0.515 | 0.418 | 0.446 | 0.406 | 0.375 | 1.000 | |
| V7 | 0.277 | 0.162 | 0.228 | 0.260 | 0.353 | 0.090 | 1.000 |
| V8 | 0.575 | 0.424 | 0.609 | 0.680 | 0.638 | 0.440 | 0.380 |
| V9 | 0.709 | 0.651 | 0.619 | 0.737 | 0.550 | 0.444 | 0.086 |
| V10 | 0.234 | 0.202 | 0.164 | 0.117 | 0.017 | 0.153 | -0.061 |
| V11 | 0.322 | 0.318 | 0.343 | 0.319 | 0.513 | 0.100 | 0.320 |
| V12 | 0.434 | 0.351 | 0.383 | 0.539 | 0.412 | 0.295 | 0.034 |
| V13 | 0.603 | 0.433 | 0.504 | 0.648 | 0.648 | 0.449 | 0.438 |
| V14 | 0.782 | 0.699 | 0.822 | 0.549 | 0.534 | 0.477 | 0.266 |
| V15 | 0.549 | 0.402 | 0.499 | 0.742 | 0.610 | 0.324 | 0.262 |
| V16 | 0.665 | 0.646 | 0.785 | 0.690 | 0.563 | 0.543 | 0.178 |
| V17 | 0.768 | 0.481 | 0.713 | 0.380 | 0.400 | 0.531 | 0.262 |
| V18 | 0.597 | 0.656 | 0.507 | 0.608 | 0.514 | 0.569 | 0.294 |
| V19 | 0.599 | 0.610 | 0.606 | 0.540 | 0.548 | 0.503 | 0.295 |
| V20 | 0.106 | 0.082 | 0.214 | 0.249 | 0.431 | 0.073 | 0.217 |
| V21 | 0.687 | 0.623 | 0.606 | 0.385 | 0.510 | 0.558 | 0.295 |
| V22 | 0.574 | 0.486 | 0.536 | 0.282 | 0.303 | 0.408 | 0.137 |
| V23 | 0.664 | 0.550 | 0.681 | 0.464 | 0.507 | 0.564 | 0.309 |
| V24 | 0.671 | 0.577 | 0.622 | 0.448 | 0.558 | 0.539 | 0.259 |
| V25 | 0.689 | 0.487 | 0.549 | 0.423 | 0.511 | 0.512 | 0.212 |
| V26 | 0.739 | 0.648 | 0.618 | 0.473 | 0.478 | 0.566 | 0.191 |
| V27 | 0.264 | 0.251 | 0.255 | 0.101 | 0.085 | 0.391 | -0.086 |
| V28 | 0.655 | 0.535 | 0.600 | 0.603 | 0.529 | 0.431 | 0.342 |
| V29 | 0.628 | 0.526 | 0.585 | 0.539 | 0.521 | 0.410 | 0.266 |
| V30 | 0.552 | 0.278 | 0.520 | 0.239 | 0.312 | 0.345 | 0.226 |
| V31 | 0.631 | 0.605 | 0.643 | 0.676 | 0.645 | 0.451 | 0.263 |
| V32 | 0.739 | 0.705 | 0.520 | 0.534 | 0.453 | 0.401 | 0.291 |
| V33 | 0.560 | 0.576 | 0.506 | 0.392 | 0.457 | 0.251 | 0.241 |
| V34 | 0.686 | 0.650 | 0.537 | 0.609 | 0.600 | 0.487 | 0.398 |
| V35 | 0.383 | 0.403 | 0.380 | 0.295 | 0.366 | 0.181 | 0.161 |
| V36 | 0.703 | 0.683 | 0.592 | 0.606 | 0.447 | 0.456 | 0.228 |
| V37 | 0.500 | 0.374 | 0.292 | 0.279 | 0.230 | 0.302 | 0.185 |
| V38 | 0.724 | 0.572 | 0.738 | 0.571 | 0.364 | 0.471 | 0.249 |
| V39 | 0.755 | 0.783 | 0.611 | 0.526 | 0.524 | 0.546 | 0.229 |
| V40 | 0.781 | 0.623 | 0.709 | 0.538 | 0.452 | 0.390 | 0.218 |
| V41 | 0.739 | 0.625 | 0.658 | 0.456 | 0.443 | 0.490 | 0.216 |
| V42 | 0.692 | 0.587 | 0.671 | 0.461 | 0.464 | 0.494 | 0.259 |
| V43 | 0.725 | 0.633 | 0.731 | 0.277 | 0.349 | 0.376 | 0.155 |

| | V8 | V9 | V10 | V11 | V12 | V13 | V14 |
|---|---|---|---|---|---|---|---|
| V8 | 1.000 | | | | | | |
| V9 | 0.606 | 1.000 | | | | | |
| V10 | 0.279 | 0.285 | 1.000 | | | | |
| V11 | 0.300 | 0.372 | -0.162 | 1.000 | | | |
| V12 | 0.412 | 0.537 | 0.118 | 0.104 | 1.000 | | |
| V13 | 0.743 | 0.534 | 0.198 | 0.389 | 0.489 | 1.000 | |
| V14 | 0.664 | 0.670 | 0.216 | 0.295 | 0.466 | 0.702 | 1.000 |
| V15 | 0.598 | 0.709 | 0.146 | 0.325 | 0.405 | 0.580 | 0.590 |
| V16 | 0.624 | 0.797 | 0.208 | 0.411 | 0.464 | 0.622 | 0.790 |
| V17 | 0.432 | 0.509 | 0.116 | 0.284 | 0.338 | 0.534 | 0.660 |
| V18 | 0.493 | 0.593 | 0.107 | 0.252 | 0.482 | 0.684 | 0.649 |
| V19 | 0.637 | 0.543 | 0.098 | 0.175 | 0.606 | 0.644 | 0.732 |
| V20 | 0.406 | 0.132 | 0.025 | 0.292 | 0.094 | 0.360 | 0.292 |
| V21 | 0.432 | 0.408 | 0.097 | 0.226 | 0.418 | 0.657 | 0.758 |
| V22 | 0.522 | 0.508 | 0.125 | 0.366 | 0.324 | 0.518 | 0.674 |
| V23 | 0.473 | 0.471 | 0.114 | 0.229 | 0.371 | 0.600 | 0.827 |
| V24 | 0.647 | 0.596 | 0.198 | 0.196 | 0.472 | 0.661 | 0.817 |
| V25 | 0.519 | 0.583 | 0.184 | 0.202 | 0.543 | 0.714 | 0.714 |
| V26 | 0.575 | 0.715 | 0.260 | 0.262 | 0.533 | 0.653 | 0.819 |
| V27 | 0.232 | 0.204 | -0.172 | 0.184 | 0.045 | 0.214 | 0.343 |
| V28 | 0.759 | 0.662 | 0.218 | 0.368 | 0.389 | 0.766 | 0.697 |
| V29 | 0.673 | 0.650 | 0.221 | 0.362 | 0.350 | 0.688 | 0.708 |
| V30 | 0.323 | 0.357 | 0.159 | 0.391 | 0.137 | 0.529 | 0.481 |
| V31 | 0.641 | 0.649 | 0.148 | 0.263 | 0.435 | 0.703 | 0.791 |
| V32 | 0.496 | 0.638 | 0.211 | 0.316 | 0.308 | 0.600 | 0.667 |
| V33 | 0.394 | 0.491 | 0.003 | 0.555 | 0.175 | 0.410 | 0.497 |
| V34 | 0.576 | 0.557 | 0.145 | 0.382 | 0.464 | 0.750 | 0.584 |
| V35 | 0.284 | 0.328 | -0.209 | 0.377 | 0.157 | 0.275 | 0.369 |
| V36 | 0.552 | 0.716 | 0.181 | 0.333 | 0.594 | 0.646 | 0.727 |
| V37 | 0.240 | 0.391 | 0.170 | 0.035 | 0.404 | 0.342 | 0.402 |
| V38 | 0.444 | 0.677 | 0.128 | 0.360 | 0.370 | 0.488 | 0.642 |
| V39 | 0.537 | 0.708 | 0.219 | 0.337 | 0.556 | 0.664 | 0.819 |
| V40 | 0.529 | 0.662 | 0.168 | 0.291 | 0.475 | 0.590 | 0.785 |
| V41 | 0.615 | 0.599 | 0.256 | 0.227 | 0.490 | 0.662 | 0.840 |
| V42 | 0.548 | 0.585 | 0.147 | 0.225 | 0.416 | 0.544 | 0.805 |
| V43 | 0.467 | 0.502 | 0.206 | 0.238 | 0.367 | 0.526 | 0.836 |

| | V15 | V16 | V17 | V18 | V19 | V20 | V21 |
|---|---|---|---|---|---|---|---|
| V15 | 1.000 | | | | | | |
| V16 | 0.715 | 1.000 | | | | | |
| V17 | 0.368 | 0.658 | 1.000 | | | | |
| V18 | 0.520 | 0.643 | 0.590 | 1.000 | | | |
| V19 | 0.569 | 0.630 | 0.569 | 0.694 | 1.000 | | |
| V20 | 0.266 | 0.255 | 0.195 | 0.203 | 0.280 | 1.000 | |
| V21 | 0.445 | 0.605 | 0.708 | 0.715 | 0.691 | 0.127 | 1.000 |
| V22 | 0.379 | 0.530 | 0.523 | 0.526 | 0.554 | 0.176 | 0.553 |
| V23 | 0.474 | 0.685 | 0.695 | 0.675 | 0.685 | 0.264 | 0.797 |
| V24 | 0.470 | 0.622 | 0.637 | 0.702 | 0.761 | 0.297 | 0.760 |
| V25 | 0.423 | 0.571 | 0.651 | 0.694 | 0.658 | 0.177 | 0.705 |
| V26 | 0.418 | 0.702 | 0.645 | 0.706 | 0.646 | 0.229 | 0.627 |
| V27 | 0.178 | 0.231 | 0.226 | 0.245 | 0.276 | -0.035 | 0.297 |
| V28 | 0.588 | 0.651 | 0.539 | 0.649 | 0.634 | 0.373 | 0.495 |
| V29 | 0.555 | 0.600 | 0.469 | 0.583 | 0.602 | 0.264 | 0.545 |
| V30 | 0.277 | 0.527 | 0.690 | 0.314 | 0.338 | 0.267 | 0.513 |
| V31 | 0.718 | 0.719 | 0.542 | 0.798 | 0.715 | 0.200 | 0.707 |
| V32 | 0.677 | 0.650 | 0.513 | 0.644 | 0.612 | 0.138 | 0.631 |
| V33 | 0.478 | 0.529 | 0.483 | 0.560 | 0.420 | 0.020 | 0.483 |
| V34 | 0.620 | 0.628 | 0.608 | 0.816 | 0.657 | 0.295 | 0.705 |
| V35 | 0.478 | 0.392 | 0.296 | 0.436 | 0.397 | 0.033 | 0.416 |
| V36 | 0.637 | 0.683 | 0.496 | 0.761 | 0.652 | 0.135 | 0.617 |
| V37 | 0.390 | 0.391 | 0.400 | 0.492 | 0.384 | 0.012 | 0.465 |
| V38 | 0.545 | 0.712 | 0.779 | 0.669 | 0.604 | 0.249 | 0.471 |
| V39 | 0.572 | 0.730 | 0.579 | 0.744 | 0.706 | 0.176 | 0.765 |
| V40 | 0.592 | 0.688 | 0.696 | 0.643 | 0.608 | 0.212 | 0.601 |
| V41 | 0.438 | 0.629 | 0.563 | 0.664 | 0.678 | 0.155 | 0.642 |
| V42 | 0.436 | 0.676 | 0.720 | 0.682 | 0.645 | 0.365 | 0.607 |
| V43 | 0.319 | 0.607 | 0.725 | 0.491 | 0.584 | 0.138 | 0.684 |

| | V22 | V23 | V24 | V25 | V26 | V27 | V28 |
|---|---|---|---|---|---|---|---|
| V22 | 1.000 | | | | | | |
| V23 | 0.586 | 1.000 | | | | | |
| V24 | 0.746 | 0.746 | 1.000 | | | | |
| V25 | 0.627 | 0.563 | 0.804 | 1.000 | | | |
| V26 | 0.675 | 0.704 | 0.781 | 0.814 | 1.000 | | |
| V27 | 0.555 | 0.301 | 0.352 | 0.268 | 0.292 | 1.000 | |
| V28 | 0.668 | 0.526 | 0.752 | 0.612 | 0.708 | 0.264 | 1.000 |
| V29 | 0.662 | 0.535 | 0.785 | 0.607 | 0.684 | 0.321 | 0.926 |
| V30 | 0.430 | 0.504 | 0.447 | 0.435 | 0.443 | 0.149 | 0.577 |
| V31 | 0.525 | 0.677 | 0.769 | 0.680 | 0.625 | 0.327 | 0.672 |
| V32 | 0.522 | 0.570 | 0.578 | 0.481 | 0.520 | 0.302 | 0.632 |
| V33 | 0.571 | 0.377 | 0.458 | 0.384 | 0.381 | 0.409 | 0.519 |
| V34 | 0.463 | 0.571 | 0.619 | 0.606 | 0.586 | 0.205 | 0.710 |
| V35 | 0.466 | 0.333 | 0.356 | 0.296 | 0.297 | 0.247 | 0.373 |
| V36 | 0.543 | 0.565 | 0.656 | 0.632 | 0.754 | 0.287 | 0.701 |
| V37 | 0.276 | 0.419 | 0.400 | 0.401 | 0.406 | 0.073 | 0.339 |
| V38 | 0.462 | 0.631 | 0.509 | 0.533 | 0.624 | 0.223 | 0.606 |
| V39 | 0.665 | 0.679 | 0.783 | 0.725 | 0.817 | 0.346 | 0.675 |
| V40 | 0.530 | 0.558 | 0.652 | 0.652 | 0.742 | 0.232 | 0.693 |
| V41 | 0.692 | 0.684 | 0.788 | 0.746 | 0.868 | 0.362 | 0.763 |
| V42 | 0.562 | 0.791 | 0.738 | 0.641 | 0.840 | 0.232 | 0.664 |
| V43 | 0.683 | 0.683 | 0.792 | 0.672 | 0.741 | 0.342 | 0.639 |

| | V29 | V30 | V31 | V32 | V33 | V34 | V35 |
|---|---|---|---|---|---|---|---|
| V29 | 1.000 | | | | | | |
| V30 | 0.589 | 1.000 | | | | | |
| V31 | 0.690 | 0.295 | 1.000 | | | | |
| V32 | 0.600 | 0.440 | 0.716 | 1.000 | | | |
| V33 | 0.494 | 0.318 | 0.659 | 0.677 | 1.000 | | |
| V34 | 0.613 | 0.490 | 0.729 | 0.732 | 0.642 | 1.000 | |
| V35 | 0.356 | 0.179 | 0.441 | 0.424 | 0.660 | 0.509 | 1.000 |
| V36 | 0.673 | 0.290 | 0.703 | 0.660 | 0.522 | 0.748 | 0.489 |
| V37 | 0.324 | 0.245 | 0.423 | 0.478 | 0.292 | 0.488 | 0.412 |
| V38 | 0.501 | 0.537 | 0.564 | 0.569 | 0.496 | 0.633 | 0.372 |
| V39 | 0.705 | 0.409 | 0.722 | 0.743 | 0.532 | 0.742 | 0.432 |
| V40 | 0.639 | 0.443 | 0.686 | 0.581 | 0.549 | 0.696 | 0.473 |
| V41 | 0.738 | 0.417 | 0.698 | 0.578 | 0.469 | 0.625 | 0.369 |
| V42 | 0.592 | 0.491 | 0.622 | 0.465 | 0.373 | 0.562 | 0.322 |
| V43 | 0.660 | 0.562 | 0.624 | 0.542 | 0.496 | 0.517 | 0.309 |

| | V36 | V37 | V38 | V39 | V40 | V41 | V42 | V43 |
|---|---|---|---|---|---|---|---|---|
| V36 | 1.000 | | | | | | | |
| V37 | 0.513 | 1.000 | | | | | | |
| V38 | 0.663 | 0.441 | 1.000 | | | | | |
| V39 | 0.870 | 0.509 | 0.593 | 1.000 | | | | |
| V40 | 0.833 | 0.477 | 0.709 | 0.766 | 1.000 | | | |
| V41 | 0.816 | 0.418 | 0.608 | 0.816 | 0.780 | 1.000 | | |
| V42 | 0.645 | 0.378 | 0.707 | 0.648 | 0.720 | 0.759 | 1.000 | |
| V43 | 0.635 | 0.305 | 0.555 | 0.731 | 0.778 | 0.827 | 0.722 | 1.000 |

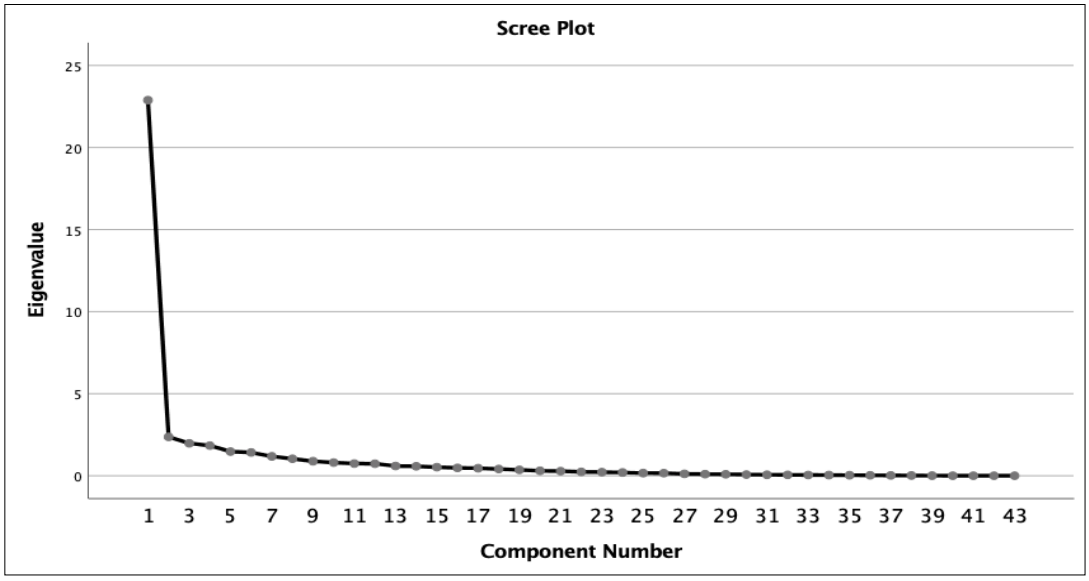
| **KMO and Bartlett's Test** | | |
|---|---|---|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.622 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 2933.042 |
| | df | 903 |
| | Sig. | 0.000 |

**Communalities**

| | Initial | Extraction |
|---|---|---|
| **V1** | 1.000 | 0.821 |
| **V2** | 1.000 | 0.664 |
| **V3** | 1.000 | 0.787 |
| **V4** | 1.000 | 0.876 |
| **V5** | 1.000 | 0.759 |
| **V6** | 1.000 | 0.731 |
| **V7** | 1.000 | 0.639 |
| **V8** | 1.000 | 0.787 |
| **V9** | 1.000 | 0.878 |
| **V10** | 1.000 | 0.786 |
| **V11** | 1.000 | 0.702 |
| **V12** | 1.000 | 0.683 |
| **V13** | 1.000 | 0.827 |
| **V14** | 1.000 | 0.867 |
| **V15** | 1.000 | 0.770 |
| **V16** | 1.000 | 0.852 |
| **V17** | 1.000 | 0.871 |
| **V18** | 1.000 | 0.796 |
| **V19** | 1.000 | 0.744 |
| **V20** | 1.000 | 0.600 |
| **V21** | 1.000 | 0.856 |
| **V22** | 1.000 | 0.796 |
| **V23** | 1.000 | 0.809 |
| **V24** | 1.000 | 0.886 |
| **V25** | 1.000 | 0.743 |
| **V26** | 1.000 | 0.865 |
| **V27** | 1.000 | 0.783 |
| **V28** | 1.000 | 0.865 |
| **V29** | 1.000 | 0.833 |
| **V30** | 1.000 | 0.762 |
| **V31** | 1.000 | 0.806 |
| **V32** | 1.000 | 0.812 |
| **V33** | 1.000 | 0.834 |
| **V34** | 1.000 | 0.834 |
| **V35** | 1.000 | 0.709 |
| **V36** | 1.000 | 0.860 |
| **V37** | 1.000 | 0.603 |
| **V38** | 1.000 | 0.810 |
| **V39** | 1.000 | 0.857 |
| **V40** | 1.000 | 0.832 |
| **V41** | 1.000 | 0.884 |
| **V42** | 1.000 | 0.820 |
| **V43** | 1.000 | 0.878 |
| Extraction Method: Principal Component Analysis. | | |

## Total Variance Explained

| Component | Initial Eigenvalues | | | Extraction Sums of Squared | | | Rotation Sums of Squared | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 22.885 | 53.220 | 53.220 | 22.885 | 53.220 | 53.220 | 10.153 | 23.612 | 23.612 |
| 2 | 2.366 | 5.503 | 58.724 | 2.366 | 5.503 | 58.724 | 6.102 | 14.191 | 37.803 |
| 3 | 1.979 | 4.603 | 63.327 | 1.979 | 4.603 | 63.327 | 5.146 | 11.968 | 49.770 |
| 4 | 1.838 | 4.274 | 67.601 | 1.838 | 4.274 | 67.601 | 3.427 | 7.970 | 57.740 |
| 5 | 1.473 | 3.427 | 71.028 | 1.473 | 3.427 | 71.028 | 3.224 | 7.497 | 65.237 |
| 6 | 1.419 | 3.299 | 74.327 | 1.419 | 3.299 | 74.327 | 2.722 | 6.331 | 71.568 |
| 7 | 1.180 | 2.745 | 77.072 | 1.180 | 2.745 | 77.072 | 1.936 | 4.502 | 76.070 |
| 8 | 1.037 | 2.412 | 79.483 | 1.037 | 2.412 | 79.483 | 1.468 | 3.413 | 79.483 |
| 9 | 0.886 | 2.060 | 81.543 | | | | | | |
| 10 | 0.805 | 1.872 | 83.415 | | | | | | |
| 11 | 0.744 | 1.730 | 85.145 | | | | | | |
| 12 | 0.735 | 1.710 | 86.855 | | | | | | |
| 13 | 0.596 | 1.385 | 88.240 | | | | | | |
| 14 | 0.578 | 1.345 | 89.586 | | | | | | |
| 15 | 0.523 | 1.217 | 90.803 | | | | | | |
| 16 | 0.481 | 1.119 | 91.922 | | | | | | |
| 17 | 0.458 | 1.065 | 92.987 | | | | | | |
| 18 | 0.409 | 0.952 | 93.939 | | | | | | |
| 19 | 0.359 | 0.836 | 94.774 | | | | | | |
| 20 | 0.303 | 0.704 | 95.479 | | | | | | |
| 21 | 0.281 | 0.653 | 96.131 | | | | | | |
| 22 | 0.238 | 0.553 | 96.685 | | | | | | |
| 23 | 0.222 | 0.516 | 97.201 | | | | | | |
| 24 | 0.199 | 0.463 | 97.664 | | | | | | |
| 25 | 0.164 | 0.381 | 98.045 | | | | | | |
| 26 | 0.157 | 0.364 | 98.410 | | | | | | |
| 27 | 0.113 | 0.264 | 98.673 | | | | | | |
| 28 | 0.100 | 0.232 | 98.905 | | | | | | |
| 29 | 0.088 | 0.204 | 99.109 | | | | | | |
| 30 | 0.075 | 0.175 | 99.284 | | | | | | |
| 31 | 0.059 | 0.137 | 99.422 | | | | | | |
| 32 | 0.054 | 0.125 | 99.547 | | | | | | |
| 33 | 0.048 | 0.111 | 99.658 | | | | | | |
| 34 | 0.041 | 0.094 | 99.753 | | | | | | |
| 35 | 0.033 | 0.076 | 99.829 | | | | | | |
| 36 | 0.026 | 0.061 | 99.890 | | | | | | |
| 37 | 0.019 | 0.044 | 99.934 | | | | | | |
| 38 | 0.014 | 0.033 | 99.968 | | | | | | |
| 39 | 0.007 | 0.017 | 99.985 | | | | | | |
| 40 | 0.003 | 0.007 | 99.992 | | | | | | |
| 41 | 0.002 | 0.005 | 99.998 | | | | | | |
| 42 | 0.001 | 0.002 | 100.000 | | | | | | |
| 43 | 0.000 | 0.000 | 100.000 | | | | | | |
| Extraction Method: Principal Component Analysis. | | | | | | | | | |




## Unrotated Component Matrix

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| **V1** | 0.854 | -0.132 | 0.054 | -0.023 | 0.246 | 0.076 | 0.057 | 0.034 |
| **V2** | 0.743 | -0.101 | 0.186 | -0.214 | 0.103 | 0.092 | 0.052 | -0.018 |
| **V3** | 0.789 | -0.103 | 0.032 | 0.131 | 0.247 | 0.174 | -0.209 | -0.024 |
| **V4** | 0.681 | 0.499 | -0.215 | -0.207 | 0.024 | 0.088 | -0.237 | 0.102 |
| **V5** | 0.651 | 0.503 | -0.128 | 0.110 | -0.118 | -0.135 | -0.145 | 0.021 |
| **V6** | 0.596 | -0.188 | -0.077 | -0.047 | -0.057 | -0.154 | -0.267 | 0.484 |
| **V7** | 0.331 | 0.364 | -0.022 | 0.328 | 0.113 | -0.386 | 0.355 | -0.042 |
| **V8** | 0.726 | 0.285 | -0.282 | 0.125 | -0.225 | 0.155 | 0.020 | 0.091 |
| **V9** | 0.780 | 0.147 | -0.092 | -0.259 | 0.095 | 0.375 | -0.147 | 0.007 |
| **V10** | 0.211 | -0.189 | -0.487 | -0.183 | 0.153 | 0.383 | 0.432 | 0.279 |
| **V11** | 0.406 | 0.444 | 0.368 | 0.370 | 0.071 | 0.187 | -0.061 | -0.150 |
| **V12** | 0.555 | 0.027 | -0.311 | -0.383 | -0.091 | -0.117 | -0.185 | -0.273 |
| **V13** | 0.794 | 0.226 | -0.219 | 0.145 | -0.156 | -0.128 | 0.163 | 0.093 |
| **V14** | 0.906 | -0.158 | -0.065 | 0.070 | -0.033 | 0.048 | -0.066 | -0.062 |
| **V15** | 0.691 | 0.464 | -0.013 | -0.215 | 0.037 | 0.114 | -0.069 | 0.109 |
| **V16** | 0.841 | 0.100 | -0.036 | -0.042 | 0.191 | 0.172 | -0.236 | 0.101 |
| **V17** | 0.755 | -0.239 | 0.045 | 0.242 | 0.381 | -0.131 | -0.123 | 0.077 |
| **V18** | 0.819 | 0.050 | 0.028 | -0.190 | -0.038 | -0.284 | 0.007 | 0.052 |
| **V19** | 0.803 | 0.005 | -0.104 | -0.073 | -0.148 | -0.226 | -0.097 | -0.013 |
| **V20** | 0.291 | 0.335 | -0.334 | 0.490 | -0.050 | -0.055 | -0.114 | -0.182 |
| **V21** | 0.793 | -0.189 | 0.095 | 0.029 | 0.003 | -0.392 | 0.059 | 0.157 |
| **V22** | 0.716 | -0.232 | 0.229 | 0.185 | -0.333 | 0.168 | 0.048 | -0.024 |
| **V23** | 0.794 | -0.191 | -0.036 | 0.151 | 0.073 | -0.269 | -0.167 | 0.118 |
| **V24** | 0.859 | -0.198 | -0.106 | 0.104 | -0.274 | -0.099 | 0.055 | -0.012 |
| **V25** | 0.791 | -0.206 | -0.147 | 0.000 | -0.157 | -0.152 | 0.028 | -0.072 |
| **V26** | 0.854 | -0.264 | -0.170 | -0.002 | -0.080 | 0.049 | -0.047 | -0.160 |
| **V27** | 0.352 | -0.238 | 0.459 | 0.084 | -0.475 | 0.126 | -0.249 | 0.286 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **V28** | 0.833 | 0.128 | -0.109 | 0.178 | -0.151 | 0.209 | 0.208 | -0.039 |
| **V29** | 0.801 | 0.046 | -0.057 | 0.160 | -0.231 | 0.239 | 0.225 | 0.005 |
| **V30** | 0.565 | -0.109 | 0.002 | 0.516 | 0.335 | 0.083 | 0.129 | 0.169 |
| **V31** | 0.850 | 0.142 | 0.035 | -0.149 | -0.150 | -0.078 | -0.002 | 0.108 |
| **V32** | 0.770 | 0.132 | 0.208 | -0.159 | 0.088 | 0.047 | 0.239 | 0.256 |
| **V33** | 0.640 | 0.199 | 0.591 | -0.017 | 0.001 | 0.126 | 0.125 | 0.056 |
| **V34** | 0.822 | 0.235 | 0.087 | -0.072 | 0.082 | -0.201 | 0.196 | 0.071 |
| **V35** | 0.496 | 0.214 | 0.586 | -0.124 | -0.032 | -0.099 | 0.023 | -0.215 |
| **V36** | 0.846 | 0.020 | 0.039 | -0.300 | -0.039 | 0.048 | 0.068 | -0.208 |
| **V37** | 0.511 | -0.062 | 0.056 | -0.374 | 0.242 | -0.277 | 0.240 | -0.050 |
| **V38** | 0.765 | -0.011 | 0.056 | 0.021 | 0.417 | 0.067 | -0.202 | -0.031 |
| **V39** | 0.890 | -0.111 | 0.046 | -0.178 | -0.087 | -0.013 | 0.091 | -0.053 |
| **V40** | 0.841 | -0.080 | 0.041 | -0.086 | 0.155 | 0.096 | 0.024 | -0.276 |
| **V41** | 0.858 | -0.259 | -0.063 | -0.022 | -0.171 | 0.074 | 0.103 | -0.174 |
| **V42** | 0.809 | -0.210 | -0.143 | 0.158 | 0.101 | -0.066 | -0.172 | -0.179 |
| **V43** | 0.789 | -0.416 | 0.060 | 0.182 | 0.004 | 0.112 | 0.074 | -0.162 |
| Extraction Method: Principal Component Analysis.<br>8 components extracted. | | | | | | | | |

**Unrotated Component Matrix (only loading values >0.5)**

| | **Component** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| **V14** | 0.906 | | | | | | | |
| **V39** | 0.890 | | | | | | | |
| **V24** | 0.859 | | | | | | | |
| **V41** | 0.858 | | | | | | | |
| **V26** | 0.854 | | | | | | | |
| **V1** | 0.854 | | | | | | | |
| **V31** | 0.850 | | | | | | | |
| **V36** | 0.846 | | | | | | | |
| **V16** | 0.841 | | | | | | | |
| **V40** | 0.841 | | | | | | | |
| **V28** | 0.833 | | | | | | | |
| **V34** | 0.822 | | | | | | | |
| **V18** | 0.819 | | | | | | | |
| **V42** | 0.809 | | | | | | | |
| **V19** | 0.803 | | | | | | | |
| **V29** | 0.801 | | | | | | | |
| **V13** | 0.794 | | | | | | | |
| **V23** | 0.794 | | | | | | | |
| **V21** | 0.793 | | | | | | | |
| **V25** | 0.791 | | | | | | | |
| **V43** | 0.789 | | | | | | | |
| **V3** | 0.789 | | | | | | | |
| **V9** | 0.780 | | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **V32** | 0.770 | | | | | | | |
| **V38** | 0.765 | | | | | | | |
| **V17** | 0.755 | | | | | | | |
| **V2** | 0.743 | | | | | | | |
| **V8** | 0.726 | | | | | | | |
| **V22** | 0.716 | | | | | | | |
| **V15** | 0.691 | | | | | | | |
| **V4** | 0.681 | | | | | | | |
| **V5** | 0.651 | 0.503 | | | | | | |
| **V33** | 0.640 | | 0.591 | | | | | |
| **V6** | 0.596 | | | | | | | |
| **V30** | 0.565 | | | 0.516 | | | | |
| **V12** | 0.555 | | | | | | | |
| **V37** | 0.511 | | | | | | | |
| **V11** | | | | | | | | |
| **V35** | | | 0.586 | | | | | |
| **V10** | | | | | | | | |
| **V20** | | | | | | | | |
| **V27** | | | | | | | | |
| **V7** | | | | | | | | |

## Rotated Component Matrix

| | **Component** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| **V41** | 0.829 | 0.218 | 0.242 | 0.167 | 0.192 | 0.069 | 0.085 | 0.118 |
| **V26** | 0.784 | 0.291 | 0.357 | 0.031 | 0.151 | 0.057 | 0.08 | 0.067 |
| **V24** | 0.776 | 0.224 | 0.212 | 0.085 | 0.175 | 0.249 | 0.294 | 0.057 |
| **V43** | 0.763 | -0.008 | 0.47 | 0.213 | 0.099 | 0.047 | 0.081 | 0.1 |
| **V25** | 0.706 | 0.232 | 0.233 | -0.002 | 0.262 | 0.182 | 0.186 | 0.02 |
| **V22** | 0.702 | 0.061 | 0.179 | 0.406 | -0.037 | 0.064 | 0.309 | 0.047 |
| **V14** | 0.694 | 0.333 | 0.428 | 0.156 | 0.147 | 0.131 | 0.158 | 0.056 |
| **V39** | 0.666 | 0.342 | 0.237 | 0.253 | 0.377 | 0.054 | 0.149 | 0.096 |
| **V42** | 0.642 | 0.255 | 0.541 | -0.006 | 0.122 | 0.162 | 0.045 | -0.086 |
| **V29** | 0.639 | 0.308 | 0.147 | 0.341 | -0.008 | 0.29 | 0.137 | 0.299 |
| **V36** | 0.62 | 0.456 | 0.173 | 0.289 | 0.387 | -0.008 | -0.045 | 0.047 |
| **V40** | 0.611 | 0.328 | 0.421 | 0.277 | 0.275 | 0.032 | -0.14 | 0.03 |
| **V28** | 0.61 | 0.377 | 0.2 | 0.319 | 0.01 | 0.36 | 0.062 | 0.274 |
| **V19** | 0.564 | 0.419 | 0.205 | 0.035 | 0.307 | 0.216 | 0.238 | -0.097 |
| **V2** | 0.451 | 0.284 | 0.328 | 0.343 | 0.358 | -0.093 | 0.08 | 0.109 |
| **V4** | 0.163 | 0.867 | 0.174 | 0.116 | 0.124 | 0.176 | 0.075 | 0.043 |
| **V15** | 0.163 | 0.73 | 0.152 | 0.317 | 0.213 | 0.158 | 0.077 | 0.102 |
| **V9** | 0.399 | 0.689 | 0.328 | 0.252 | 0.097 | -0.121 | -0.005 | 0.218 |
| **V5** | 0.231 | 0.631 | 0.118 | 0.159 | 0.067 | 0.479 | 0.136 | -0.128 |
| **V16** | 0.355 | 0.589 | 0.536 | 0.216 | 0.122 | 0.043 | 0.142 | 0.086 |
| **V8** | 0.455 | 0.574 | 0.103 | 0.128 | -0.073 | 0.38 | 0.164 | 0.218 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **V12** | 0.497 | 0.533 | 0.033 | -0.186 | 0.29 | -0.054 | -0.11 | -0.128 |
| **V31** | 0.479 | 0.509 | 0.158 | 0.263 | 0.33 | 0.184 | 0.282 | 0.042 |
| **V17** | 0.367 | 0.105 | 0.773 | 0.091 | 0.238 | 0.183 | 0.171 | -0.01 |
| **V38** | 0.295 | 0.385 | 0.694 | 0.21 | 0.216 | 0.037 | -0.011 | -0.007 |
| **V30** | 0.228 | -0.039 | 0.665 | 0.195 | -0.033 | 0.385 | 0.118 | 0.254 |
| **V3** | 0.438 | 0.336 | 0.651 | 0.22 | 0.047 | 0.033 | 0.073 | 0.034 |
| **V1** | 0.479 | 0.28 | 0.549 | 0.259 | 0.314 | 0.068 | 0.082 | 0.187 |
| **V23** | 0.484 | 0.214 | 0.526 | -0.003 | 0.269 | 0.232 | 0.343 | -0.097 |
| **V33** | 0.221 | 0.209 | 0.202 | 0.778 | 0.242 | 0.072 | 0.178 | -0.001 |
| **V35** | 0.219 | 0.171 | 0.068 | 0.648 | 0.334 | 0.039 | 0.009 | -0.306 |
| **V11** | 0.052 | 0.244 | 0.258 | 0.627 | -0.19 | 0.331 | -0.076 | -0.166 |
| **V32** | 0.273 | 0.351 | 0.252 | 0.458 | 0.428 | 0.14 | 0.236 | 0.289 |
| **V37** | 0.245 | 0.153 | 0.178 | 0.088 | 0.686 | 0.022 | -0.061 | 0.076 |
| **V18** | 0.46 | 0.406 | 0.215 | 0.147 | 0.504 | 0.195 | 0.238 | -0.057 |
| **V21** | 0.485 | 0.113 | 0.363 | 0.105 | 0.494 | 0.262 | 0.387 | -0.048 |
| **V34** | 0.338 | 0.396 | 0.246 | 0.327 | 0.488 | 0.367 | 0.129 | 0.079 |
| **V7** | 0.037 | 0.033 | 0.094 | 0.156 | 0.263 | 0.728 | -0.067 | -0.012 |
| **V20** | 0.186 | 0.275 | 0.163 | -0.083 | -0.305 | 0.579 | -0.111 | -0.123 |
| **V13** | 0.478 | 0.429 | 0.144 | 0.113 | 0.187 | 0.531 | 0.197 | 0.159 |
| **V27** | 0.339 | -0.001 | 0.022 | 0.388 | -0.147 | -0.204 | 0.66 | -0.137 |
| **V6** | 0.263 | 0.308 | 0.355 | -0.102 | 0.201 | 0.008 | 0.621 | 0.06 |
| **V10** | 0.174 | 0.12 | 0.085 | -0.199 | 0.084 | -0.057 | -0.063 | 0.825 |
| Extraction Method: Principal Component Analysis.<br>Rotation Method: Varimax with Kaiser Normalization. | | | | | | | | |

**Rotated Component Matrix (only loading values >0.5)**

| | **Component** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| **V41** | 0.829 | | | | | | | |
| **V26** | 0.784 | | | | | | | |
| **V24** | 0.776 | | | | | | | |
| **V43** | 0.763 | | | | | | | |
| **V25** | 0.706 | | | | | | | |
| **V22** | 0.702 | | | | | | | |
| **V14** | 0.694 | | | | | | | |
| **V39** | 0.666 | | | | | | | |
| **V42** | 0.642 | | 0.541 | | | | | |
| **V29** | 0.639 | | | | | | | |
| **V36** | 0.620 | | | | | | | |
| **V40** | 0.611 | | | | | | | |
| **V28** | 0.610 | | | | | | | |
| **V19** | 0.564 | | | | | | | |
| **V2** | | | | | | | | |
| **V4** | | 0.867 | | | | | | |
| **V15** | | 0.730 | | | | | | |
| **V9** | | 0.689 | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **V5** | | 0.631 | | | | | | |
| **V16** | | 0.589 | 0.536 | | | | | |
| **V8** | | 0.574 | | | | | | |
| **V12** | | 0.533 | | | | | | |
| **V31** | | 0.509 | | | | | | |
| **V17** | | | 0.773 | | | | | |
| **V38** | | | 0.694 | | | | | |
| **V30** | | | 0.665 | | | | | |
| **V3** | | | 0.651 | | | | | |
| **V1** | | | 0.549 | | | | | |
| **V23** | | | 0.526 | | | | | |
| **V33** | | | | 0.778 | | | | |
| **V35** | | | | 0.648 | | | | |
| **V11** | | | | 0.627 | | | | |
| **V32** | | | | | | | | |
| **V37** | | | | | 0.686 | | | |
| **V18** | | | | | 0.504 | | | |
| **V21** | | | | | | | | |
| **V34** | | | | | | | | |
| **V7** | | | | | | 0.728 | | |
| **V20** | | | | | | 0.579 | | |
| **V13** | | | | | | 0.531 | | |
| **V27** | | | | | | | 0.660 | |
| **V6** | | | | | | | 0.621 | |
| **V10** | | | | | | | | 0.825 |

## Factor Score

| FAC1_1 | FAC2_1 | FAC3_1 | FAC4_1 | FAC5_1 | FAC6_1 | FAC7_1 | FAC8_1 |
|---|---|---|---|---|---|---|---|
| 0.68038 | -0.50383 | 0.11114 | -0.32051 | -1.39344 | -2.36480 | -0.98751 | -1.09514 |
| 0.45883 | 0.59598 | 0.21650 | -0.60559 | 1.81621 | -3.69299 | 0.71782 | -1.51355 |
| -0.16348 | -0.16607 | 0.07406 | 0.99815 | 0.46012 | -0.94882 | -1.04080 | 0.92547 |
| 0.12312 | -1.04547 | -0.27155 | 0.61072 | 0.53381 | -0.80461 | 0.43214 | -0.42030 |
| 0.35211 | -0.55171 | 0.23342 | 0.43381 | -0.22398 | -0.77987 | -0.03531 | 0.81995 |
| -1.43556 | -0.56597 | 1.75206 | 0.39915 | 0.34739 | -1.02456 | -0.59919 | 0.40165 |
| 0.34462 | -0.99877 | 0.35711 | 0.96140 | 0.83795 | -0.10057 | -0.58233 | -0.91197 |
| -0.22997 | -0.64632 | 0.60129 | -0.53656 | -0.85037 | -0.60387 | 2.33183 | 1.41706 |
| 0.55725 | -0.20476 | -3.43576 | -0.91482 | 1.61819 | 0.05130 | -0.38623 | -1.90258 |
| 1.04704 | -2.16752 | 0.30214 | -0.72358 | 1.23251 | 0.13111 | 0.57151 | -0.47360 |
| 1.01898 | -1.35507 | -1.78096 | 1.76343 | 0.72357 | -0.79885 | -1.41872 | 2.42137 |
| 0.53489 | -0.03177 | 0.14973 | 0.19696 | 0.10680 | 0.32187 | 0.29834 | 0.33700 |
| -0.28444 | 0.52668 | 1.03472 | -0.18535 | 0.19092 | -0.14914 | -0.67495 | -0.04306 |
| -0.14093 | -0.18833 | -1.91796 | -0.99581 | 0.14814 | 0.24236 | -0.66374 | -0.72124 |
| 0.33370 | -1.09909 | 0.34629 | 1.06692 | 0.56240 | 1.06534 | 0.60501 | 0.66105 |
| -2.82876 | 1.33486 | -2.57102 | -0.56797 | -1.23494 | -0.93875 | -0.11889 | 0.47394 |
| 1.91186 | -1.36096 | -0.85329 | -2.39141 | -1.68044 | 1.29746 | 0.44726 | -0.69469 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| -1.07062 | -0.96696 | 2.32296 | 0.14112 | -1.35495 | -0.39746 | -1.29533 | -0.49221 |
| 0.34635 | -0.67437 | -2.28341 | 2.17318 | -2.42384 | -0.01534 | -0.67536 | 1.68740 |
| -0.18354 | -0.66807 | -1.21411 | -0.47905 | -0.53024 | 1.29853 | 1.28724 | -0.07456 |
| 0.50927 | -1.15880 | 0.73945 | -1.23617 | -1.24231 | -0.16475 | -0.51172 | -0.38445 |
| -0.02691 | -1.69422 | 0.54801 | -0.75920 | -1.21265 | 1.66761 | 1.01634 | 0.15419 |
| -0.16069 | 0.02151 | 1.40759 | 0.00715 | -0.33859 | -0.22980 | -0.98005 | -0.15811 |
| 1.23731 | -1.95642 | 0.53045 | -0.98726 | 0.86401 | 0.53450 | -0.65575 | -0.44146 |
| 0.65273 | -0.45968 | 0.67870 | -0.39567 | -1.90162 | -0.83112 | -1.30291 | -0.59063 |
| -3.65016 | -0.93193 | -0.59309 | 0.09760 | 0.77330 | 0.21842 | -0.22541 | -0.57767 |
| -3.55655 | -1.23143 | 0.55073 | -0.46073 | 0.62514 | 0.99240 | 0.50119 | -1.00506 |
| -0.86990 | 1.02500 | -0.24451 | -1.37089 | -1.90752 | -1.05730 | -2.13744 | 0.09473 |
| -0.64354 | -0.41020 | -0.31836 | 0.11310 | 0.71425 | -0.30866 | 1.17016 | 0.74060 |
| -0.61664 | -0.72235 | -1.38889 | -0.74391 | -0.15552 | -0.28479 | 0.84734 | 0.23534 |
| 0.20727 | -1.40587 | 0.77562 | 0.96600 | 0.87219 | -1.03036 | 0.53630 | 0.28567 |
| 0.24406 | 0.87193 | 0.10998 | 0.60973 | 0.21765 | 0.73813 | 0.28581 | -1.15312 |
| 0.17752 | 0.85961 | 0.07722 | 0.42369 | 0.32410 | 0.67131 | 0.24762 | 0.43163 |
| 0.81784 | 0.96216 | 0.34498 | -2.14218 | 0.10111 | -0.96793 | 0.92208 | 1.66994 |
| 0.23778 | 1.17959 | 0.33868 | -0.28408 | -1.15364 | 0.74334 | 0.93819 | 0.96256 |
| 0.13248 | 0.97460 | 0.29494 | -0.34452 | -0.01739 | 0.75870 | 0.56247 | 1.08177 |
| 0.54024 | 0.58476 | 0.02507 | 0.44330 | 0.96337 | 1.51110 | -3.13733 | 0.61791 |
| 0.24406 | 0.87193 | 0.10998 | 0.60973 | 0.21765 | 0.73813 | 0.28581 | -1.15312 |
| 0.74782 | 0.55501 | -0.24757 | 1.18522 | 0.14484 | 0.95480 | -1.33916 | -1.26428 |
| -0.25444 | 1.47231 | 0.76667 | -3.21456 | 1.68313 | 1.27587 | -1.29355 | 1.63178 |
| 0.33712 | 1.23781 | 0.16771 | 0.33895 | -0.15923 | -0.42749 | 0.49401 | 0.39560 |
| 0.24406 | 0.87193 | 0.10998 | 0.60973 | 0.21765 | 0.73813 | 0.28581 | -1.15312 |
| 0.24406 | 0.87193 | 0.10998 | 0.60973 | 0.21765 | 0.73813 | 0.28581 | -1.15312 |
| -0.46209 | 0.13606 | 0.58887 | 1.33468 | -0.06895 | 0.90758 | 0.60214 | 1.45651 |
| 0.24406 | 0.87193 | 0.10998 | 0.60973 | 0.21765 | 0.73813 | 0.28581 | -1.15312 |
| 0.49894 | 1.16997 | -0.07959 | -0.49697 | 0.41826 | -0.72481 | 0.88618 | 0.76739 |
| 0.06995 | 0.67761 | -0.00974 | 0.90012 | 0.89291 | -0.05574 | 0.67264 | -0.71573 |
| 0.40366 | 1.25013 | 0.20047 | 0.52499 | -0.26568 | -0.36067 | 0.53220 | -1.18916 |
| 0.33712 | 1.23781 | 0.16771 | 0.33895 | -0.15923 | -0.42749 | 0.49401 | 0.39560 |
| 0.03647 | 0.90169 | 0.38261 | -0.13220 | 1.03619 | 1.29442 | -1.51236 | 0.72907 |
| 0.31359 | 0.13910 | -0.15415 | -0.03325 | 1.72766 | -0.83526 | 0.87919 | 1.57796 |
| 0.50896 | 1.45512 | 0.24422 | 0.58543 | -1.40193 | -0.37603 | 0.90791 | -1.30837 |
| -0.11727 | 0.50894 | 0.48293 | 1.26959 | -1.13026 | 1.07317 | 1.24389 | -0.62968 |

## *Appendix B. 3: Factor Analysis Output – (Analysis Version 2)*

**Correlation Matrix**

| | V1 | V2 | V3 | V4 | V5 | V6 | V8 | V9 | V13 | V14 | V15 | V16 | V17 | V18 | V19 | V21 | V24 | V26 | V28 | V31 | V32 | V34 | V36 | V38 | V39 | V40 | V41 | V42 | V46 | V52 | V53 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| V1 | 1.000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| V2 | 0.681 | 1.000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| V3 | 0.772 | 0.638 | 1.000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| V4 | 0.529 | 0.437 | 0.534 | 1.000 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| V5 | 0.475 | 0.375 | 0.414 | 0.725 | 1.000 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| V6 | 0.515 | 0.418 | 0.446 | 0.406 | 0.375 | 1.000 | | | | | | | | | | | | | | | | | | | | | | | | | |
| V8 | 0.575 | 0.424 | 0.609 | 0.680 | 0.638 | 0.440 | 1.000 | | | | | | | | | | | | | | | | | | | | | | | | |
| V9 | 0.709 | 0.651 | 0.619 | 0.737 | 0.550 | 0.444 | 0.606 | 1.000 | | | | | | | | | | | | | | | | | | | | | | | |
| V13 | 0.603 | 0.433 | 0.504 | 0.648 | 0.648 | 0.449 | 0.743 | 0.534 | 1.000 | | | | | | | | | | | | | | | | | | | | | | |
| V14 | 0.782 | 0.699 | 0.822 | 0.549 | 0.534 | 0.477 | 0.664 | 0.670 | 0.702 | 1.000 | | | | | | | | | | | | | | | | | | | | | |
| V15 | 0.549 | 0.402 | 0.499 | 0.742 | 0.610 | 0.324 | 0.598 | 0.709 | 0.580 | 0.590 | 1.000 | | | | | | | | | | | | | | | | | | | | |
| V16 | 0.665 | 0.646 | 0.785 | 0.690 | 0.563 | 0.543 | 0.624 | 0.797 | 0.622 | 0.790 | 0.715 | 1.000 | | | | | | | | | | | | | | | | | | | |
| V17 | 0.768 | 0.481 | 0.713 | 0.380 | 0.400 | 0.531 | 0.432 | 0.509 | 0.534 | 0.660 | 0.368 | 0.658 | 1.000 | | | | | | | | | | | | | | | | | | |
| V18 | 0.597 | 0.656 | 0.507 | 0.608 | 0.514 | 0.569 | 0.493 | 0.593 | 0.684 | 0.649 | 0.520 | 0.643 | 0.590 | 1.000 | | | | | | | | | | | | | | | | | |
| V19 | 0.599 | 0.610 | 0.606 | 0.540 | 0.548 | 0.503 | 0.637 | 0.543 | 0.644 | 0.732 | 0.569 | 0.630 | 0.569 | 0.694 | 1.000 | | | | | | | | | | | | | | | | |
| V21 | 0.687 | 0.623 | 0.606 | 0.385 | 0.510 | 0.558 | 0.432 | 0.408 | 0.657 | 0.758 | 0.445 | 0.605 | 0.708 | 0.715 | 0.691 | 1.000 | | | | | | | | | | | | | | | |
| V24 | 0.671 | 0.577 | 0.622 | 0.448 | 0.558 | 0.539 | 0.647 | 0.596 | 0.661 | 0.817 | 0.470 | 0.622 | 0.637 | 0.702 | 0.761 | 0.760 | 1.000 | | | | | | | | | | | | | | |
| V26 | 0.739 | 0.648 | 0.618 | 0.473 | 0.478 | 0.566 | 0.575 | 0.715 | 0.653 | 0.819 | 0.418 | 0.702 | 0.645 | 0.706 | 0.646 | 0.627 | 0.781 | 1.000 | | | | | | | | | | | | | |
| V28 | 0.655 | 0.535 | 0.600 | 0.603 | 0.529 | 0.431 | 0.759 | 0.662 | 0.766 | 0.697 | 0.588 | 0.651 | 0.539 | 0.649 | 0.634 | 0.495 | 0.752 | 0.708 | 1.000 | | | | | | | | | | | | |
| V31 | 0.631 | 0.605 | 0.643 | 0.676 | 0.645 | 0.451 | 0.641 | 0.649 | 0.703 | 0.791 | 0.718 | 0.719 | 0.542 | 0.798 | 0.715 | 0.707 | 0.769 | 0.625 | 0.672 | 1.000 | | | | | | | | | | | |
| V32 | 0.739 | 0.705 | 0.520 | 0.534 | 0.453 | 0.401 | 0.496 | 0.638 | 0.600 | 0.667 | 0.677 | 0.650 | 0.513 | 0.644 | 0.612 | 0.631 | 0.578 | 0.520 | 0.632 | 0.716 | 1.000 | | | | | | | | | | |
| V34 | 0.686 | 0.650 | 0.537 | 0.609 | 0.600 | 0.487 | 0.576 | 0.557 | 0.750 | 0.584 | 0.620 | 0.628 | 0.608 | 0.816 | 0.657 | 0.705 | 0.619 | 0.586 | 0.710 | 0.729 | 0.732 | 1.000 | | | | | | | | | |
| V36 | 0.703 | 0.683 | 0.592 | 0.606 | 0.447 | 0.456 | 0.552 | 0.716 | 0.646 | 0.727 | 0.637 | 0.683 | 0.496 | 0.761 | 0.652 | 0.617 | 0.656 | 0.754 | 0.701 | 0.703 | 0.660 | 0.748 | 1.000 | | | | | | | | |
| V38 | 0.724 | 0.572 | 0.738 | 0.571 | 0.364 | 0.471 | 0.444 | 0.677 | 0.488 | 0.642 | 0.545 | 0.712 | 0.779 | 0.669 | 0.604 | 0.471 | 0.509 | 0.624 | 0.606 | 0.564 | 0.569 | 0.633 | 0.663 | 1.000 | | | | | | | |
| V39 | 0.755 | 0.783 | 0.611 | 0.526 | 0.524 | 0.546 | 0.537 | 0.708 | 0.664 | 0.819 | 0.572 | 0.730 | 0.579 | 0.744 | 0.706 | 0.765 | 0.783 | 0.817 | 0.675 | 0.722 | 0.743 | 0.742 | 0.870 | 0.593 | 1.000 | | | | | | |
| V40 | 0.781 | 0.623 | 0.709 | 0.538 | 0.452 | 0.390 | 0.529 | 0.662 | 0.590 | 0.785 | 0.592 | 0.688 | 0.696 | 0.643 | 0.608 | 0.601 | 0.652 | 0.742 | 0.693 | 0.686 | 0.581 | 0.696 | 0.833 | 0.709 | 0.766 | 1.000 | | | | | |
| V41 | 0.739 | 0.625 | 0.658 | 0.456 | 0.443 | 0.490 | 0.615 | 0.599 | 0.662 | 0.840 | 0.438 | 0.629 | 0.563 | 0.664 | 0.678 | 0.642 | 0.788 | 0.868 | 0.763 | 0.698 | 0.578 | 0.625 | 0.816 | 0.608 | 0.816 | 0.780 | 1.000 | | | | |
| V42 | 0.692 | 0.587 | 0.671 | 0.461 | 0.464 | 0.494 | 0.548 | 0.585 | 0.544 | 0.805 | 0.436 | 0.676 | 0.720 | 0.682 | 0.645 | 0.607 | 0.738 | 0.840 | 0.664 | 0.622 | 0.465 | 0.562 | 0.645 | 0.707 | 0.648 | 0.720 | 0.759 | 1.000 | | | |
| V46 | 0.578 | 0.492 | 0.403 | 0.630 | 0.450 | 0.376 | 0.561 | 0.598 | 0.643 | 0.583 | 0.544 | 0.484 | 0.380 | 0.796 | 0.587 | 0.463 | 0.619 | 0.624 | 0.694 | 0.744 | 0.605 | 0.699 | 0.835 | 0.563 | 0.666 | 0.687 | 0.724 | 0.596 | 1.000 | | |
| V52 | 0.598 | 0.564 | 0.518 | 0.454 | 0.695 | 0.508 | 0.515 | 0.460 | 0.563 | 0.626 | 0.500 | 0.574 | 0.545 | 0.652 | 0.650 | 0.776 | 0.722 | 0.560 | 0.545 | 0.698 | 0.576 | 0.702 | 0.630 | 0.439 | 0.709 | 0.562 | 0.628 | 0.558 | 0.548 | 1.000 | |
| V53 | 0.608 | 0.488 | 0.663 | 0.825 | 0.603 | 0.454 | 0.679 | 0.688 | 0.642 | 0.652 | 0.855 | 0.769 | 0.484 | 0.677 | 0.689 | 0.509 | 0.514 | 0.498 | 0.625 | 0.765 | 0.653 | 0.714 | 0.761 | 0.711 | 0.637 | 0.688 | 0.549 | 0.527 | 0.707 | 0.580 | 1.000 |

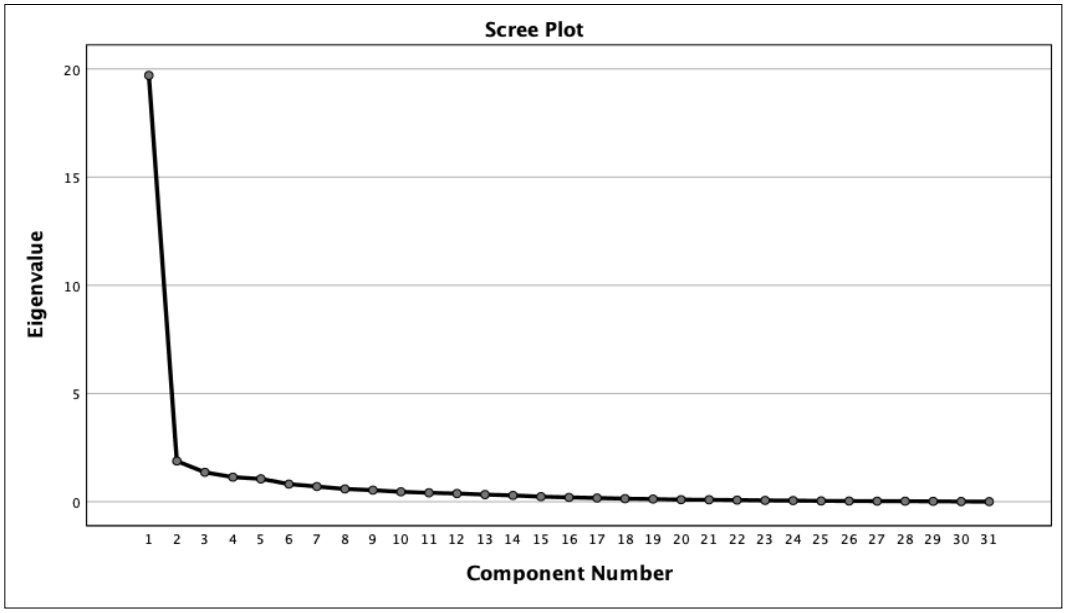
| KMO and Bartlett's Test | | |
|---|---|---|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.816 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 2189.718 |
| | df | 465 |
| | Sig. | 0.000 |

**Communalities**

| | Initial | Extraction |
|---|---|---|
| **V1** | 1.000 | 0.800 |
| **V2** | 1.000 | 0.714 |
| **V3** | 1.000 | 0.834 |
| **V4** | 1.000 | 0.855 |
| **V5** | 1.000 | 0.782 |
| **V6** | 1.000 | 0.483 |
| **V8** | 1.000 | 0.829 |
| **V9** | 1.000 | 0.783 |
| **V13** | 1.000 | 0.769 |
| **V14** | 1.000 | 0.860 |
| **V15** | 1.000 | 0.829 |
| **V16** | 1.000 | 0.839 |
| **V17** | 1.000 | 0.809 |
| **V18** | 1.000 | 0.797 |
| **V19** | 1.000 | 0.692 |
| **V21** | 1.000 | 0.907 |
| **V24** | 1.000 | 0.863 |
| **V26** | 1.000 | 0.879 |
| **V28** | 1.000 | 0.796 |
| **V31** | 1.000 | 0.808 |
| **V32** | 1.000 | 0.755 |
| **V34** | 1.000 | 0.812 |
| **V36** | 1.000 | 0.902 |
| **V38** | 1.000 | 0.807 |
| **V39** | 1.000 | 0.864 |
| **V40** | 1.000 | 0.799 |
| **V41** | 1.000 | 0.901 |
| **V42** | 1.000 | 0.800 |
| **V46** | 1.000 | 0.871 |
| **V52** | 1.000 | 0.793 |
| **V53** | 1.000 | 0.891 |

## Total Variance Explained

| Component | Initial Eigenvalues | | | Extraction Sums of Squared | | | Rotation Sums of Squared | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 19.699 | 63.545 | 63.545 | 19.699 | 63.545 | 63.545 | 5.777 | 18.635 | 18.635 |
| 2 | 1.879 | 6.061 | 69.605 | 1.879 | 6.061 | 69.605 | 5.586 | 18.02 | 36.655 |
| 3 | 1.358 | 4.38 | 73.986 | 1.358 | 4.38 | 73.986 | 4.682 | 15.103 | 51.758 |
| 4 | 1.133 | 3.654 | 77.64 | 1.133 | 3.654 | 77.64 | 4.578 | 14.769 | 66.527 |
| 5 | 1.055 | 3.403 | 81.043 | 1.055 | 3.403 | 81.043 | 4.5 | 14.516 | 81.043 |
| 6 | 0.813 | 2.624 | 83.667 | | | | | | |
| 7 | 0.701 | 2.262 | 85.928 | | | | | | |
| 8 | 0.59 | 1.903 | 87.831 | | | | | | |
| 9 | 0.532 | 1.717 | 89.548 | | | | | | |
| 10 | 0.457 | 1.473 | 91.021 | | | | | | |
| 11 | 0.412 | 1.328 | 92.349 | | | | | | |
| 12 | 0.377 | 1.216 | 93.565 | | | | | | |
| 13 | 0.329 | 1.062 | 94.628 | | | | | | |
| 14 | 0.289 | 0.931 | 95.559 | | | | | | |
| 15 | 0.233 | 0.751 | 96.31 | | | | | | |
| 16 | 0.199 | 0.642 | 96.952 | | | | | | |
| 17 | 0.171 | 0.55 | 97.502 | | | | | | |
| 18 | 0.14 | 0.452 | 97.954 | | | | | | |
| 19 | 0.123 | 0.396 | 98.35 | | | | | | |
| 20 | 0.095 | 0.306 | 98.656 | | | | | | |
| 21 | 0.09 | 0.29 | 98.946 | | | | | | |
| 22 | 0.075 | 0.243 | 99.189 | | | | | | |
| 23 | 0.059 | 0.189 | 99.378 | | | | | | |
| 24 | 0.051 | 0.164 | 99.542 | | | | | | |
| 25 | 0.035 | 0.113 | 99.655 | | | | | | |
| 26 | 0.032 | 0.102 | 99.757 | | | | | | |
| 27 | 0.026 | 0.084 | 99.841 | | | | | | |
| 28 | 0.022 | 0.072 | 99.913 | | | | | | |
| 29 | 0.016 | 0.051 | 99.964 | | | | | | |
| 30 | 0.008 | 0.025 | 99.989 | | | | | | |
| 31 | 0.004 | 0.011 | 100 | | | | | | |




**Unrotated Component Matrix (only loading values >0.5)**

| | Component | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| V14 | 0.893 | | | | |
| V39 | 0.886 | | | | |
| V31 | 0.867 | | | | |
| V36 | 0.866 | | | | |
| V16 | 0.847 | | | | |
| V41 | 0.846 | | | | |
| V1 | 0.845 | | | | |
| V40 | 0.843 | | | | |
| V26 | 0.838 | | | | |
| V18 | 0.838 | | | | |
| V24 | 0.836 | | | | |
| V34 | 0.834 | | | | |
| V28 | 0.822 | | | | |
| V53 | 0.818 | | | | |
| V19 | 0.810 | | | | |
| V42 | 0.800 | | | | |
| V9 | 0.794 | | | | |
| V13 | 0.793 | | | | |
| V3 | 0.779 | | | | |
| V21 | 0.778 | | | | |
| V32 | 0.775 | | | | |
| V38 | 0.769 | | | | |
| V46 | 0.769 | | | | |
| V52 | 0.755 | | | | |
| V2 | 0.745 | | | | |

| | | | | | |
|---|---|---|---|---|---|
| V8 | 0.738 | | | | |
| V17 | 0.731 | | | | |
| V4 | 0.726 | 0.547 | | | |
| V15 | 0.721 | 0.517 | | | |
| V5 | 0.667 | | | | |
| V6 | 0.605 | | | | |

**Rotated Component Matrix (only loading values >0.5)**

| | Component | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| V3 | 0.776 | | | | |
| V17 | 0.740 | | | | |
| V38 | 0.735 | | | | |
| V1 | 0.651 | | | | |
| V16 | 0.634 | 0.523 | | | |
| V42 | 0.579 | | | | 0.568 |
| V14 | 0.564 | | | | |
| V40 | 0.546 | | | | |
| V4 | | 0.837 | | | |
| V15 | | 0.785 | | | |
| V53 | | 0.748 | | | |
| V5 | | 0.683 | 0.509 | | |
| V8 | | 0.656 | | | 0.540 |
| V9 | 0.518 | 0.547 | | | |
| V13 | | 0.521 | | | |
| V31 | | 0.514 | | | |
| V21 | | | 0.799 | | |
| V52 | | | 0.738 | | |
| V6 | | | 0.515 | | |
| V19 | | | 0.507 | | |
| V36 | | | | 0.688 | |
| V46 | | | | 0.659 | 0.512 |
| V32 | | | | 0.603 | |
| V18 | | | | 0.577 | |
| V39 | | | | 0.572 | |
| V2 | | | | 0.570 | |
| V34 | | | | 0.555 | |
| V41 | | | | | 0.700 |
| V26 | | | | | 0.666 |
| V28 | | | | | 0.623 |
| V24 | | | 0.564 | | 0.599 |

## *Appendix B. 4: Factor Analysis Output – (Analysis Version 3)*

| | V4 | V5 | V6 | V7 | V8 | V9 | V10 | V11 | V12 | V13 | V15 | V18 | V19 | V20 | V28 | V31 | V33 | V34 | V35 | V36 | V37 | V45 | V47 | V48 | V51 | V52 | V53 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **V4** | 1.000 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **V5** | 0.725 | 1.000 | | | | | | | | | | | | | | | | | | | | | | | | | |
| **V6** | 0.406 | 0.375 | 1.000 | | | | | | | | | | | | | | | | | | | | | | | | |
| **V7** | 0.260 | 0.353 | 0.090 | 1.000 | | | | | | | | | | | | | | | | | | | | | | | |
| **V8** | 0.680 | 0.638 | 0.440 | 0.380 | 1.000 | | | | | | | | | | | | | | | | | | | | | | |
| **V9** | 0.737 | 0.550 | 0.444 | 0.086 | 0.606 | 1.000 | | | | | | | | | | | | | | | | | | | | | |
| **V10** | 0.117 | 0.017 | 0.153 | -0.061 | 0.279 | 0.285 | 1.000 | | | | | | | | | | | | | | | | | | | | |
| **V11** | 0.319 | 0.513 | 0.100 | 0.320 | 0.300 | 0.372 | -0.162 | 1.000 | | | | | | | | | | | | | | | | | | | |
| **V12** | 0.539 | 0.412 | 0.295 | 0.034 | 0.412 | 0.537 | 0.118 | 0.104 | 1.000 | | | | | | | | | | | | | | | | | | |
| **V13** | 0.648 | 0.648 | 0.449 | 0.438 | 0.743 | 0.534 | 0.198 | 0.389 | 0.489 | 1.000 | | | | | | | | | | | | | | | | | |
| **V15** | 0.742 | 0.610 | 0.324 | 0.262 | 0.598 | 0.709 | 0.146 | 0.325 | 0.405 | 0.580 | 1.000 | | | | | | | | | | | | | | | | |
| **V18** | 0.608 | 0.514 | 0.569 | 0.294 | 0.493 | 0.593 | 0.107 | 0.252 | 0.482 | 0.684 | 0.520 | 1.000 | | | | | | | | | | | | | | | |
| **V19** | 0.540 | 0.548 | 0.503 | 0.295 | 0.637 | 0.543 | 0.098 | 0.175 | 0.606 | 0.644 | 0.569 | 0.694 | 1.000 | | | | | | | | | | | | | | |
| **V20** | 0.249 | 0.431 | 0.073 | 0.217 | 0.406 | 0.132 | 0.025 | 0.292 | 0.094 | 0.360 | 0.266 | 0.203 | 0.280 | 1.000 | | | | | | | | | | | | | |
| **V28** | 0.603 | 0.529 | 0.431 | 0.342 | 0.759 | 0.662 | 0.218 | 0.368 | 0.389 | 0.766 | 0.588 | 0.649 | 0.634 | 0.373 | 1.000 | | | | | | | | | | | | |
| **V31** | 0.676 | 0.645 | 0.451 | 0.263 | 0.641 | 0.649 | 0.148 | 0.263 | 0.435 | 0.703 | 0.718 | 0.798 | 0.715 | 0.200 | 0.672 | 1.000 | | | | | | | | | | | |
| **V33** | 0.392 | 0.457 | 0.251 | 0.241 | 0.394 | 0.491 | 0.003 | 0.555 | 0.175 | 0.410 | 0.478 | 0.560 | 0.420 | 0.020 | 0.519 | 0.659 | 1.000 | | | | | | | | | | |
| **V34** | 0.609 | 0.600 | 0.487 | 0.398 | 0.576 | 0.557 | 0.145 | 0.382 | 0.464 | 0.750 | 0.620 | 0.816 | 0.657 | 0.295 | 0.710 | 0.729 | 0.642 | 1.000 | | | | | | | | | |
| **V35** | 0.295 | 0.366 | 0.181 | 0.161 | 0.284 | 0.328 | -0.209 | 0.377 | 0.157 | 0.275 | 0.478 | 0.436 | 0.397 | 0.033 | 0.373 | 0.441 | 0.660 | 0.509 | 1.000 | | | | | | | | |
| **V36** | 0.606 | 0.447 | 0.456 | 0.228 | 0.552 | 0.716 | 0.181 | 0.333 | 0.594 | 0.646 | 0.637 | 0.761 | 0.652 | 0.135 | 0.701 | 0.703 | 0.522 | 0.748 | 0.489 | 1.000 | | | | | | | |
| **V37** | 0.279 | 0.230 | 0.302 | 0.185 | 0.240 | 0.391 | 0.170 | 0.035 | 0.404 | 0.342 | 0.390 | 0.492 | 0.384 | 0.012 | 0.339 | 0.423 | 0.292 | 0.488 | 0.412 | 0.513 | 1.000 | | | | | | |
| **V45** | 0.581 | 0.540 | 0.380 | 0.358 | 0.685 | 0.573 | 0.183 | 0.269 | 0.515 | 0.791 | 0.666 | 0.693 | 0.656 | 0.315 | 0.688 | 0.768 | 0.543 | 0.727 | 0.375 | 0.761 | 0.394 | 1.000 | | | | | |
| **V47** | 0.659 | 0.534 | 0.332 | 0.275 | 0.639 | 0.509 | 0.244 | 0.218 | 0.607 | 0.654 | 0.654 | 0.598 | 0.567 | 0.226 | 0.609 | 0.701 | 0.503 | 0.683 | 0.318 | 0.671 | 0.493 | 0.766 | 1.000 | | | | |
| **V48** | 0.601 | 0.509 | 0.496 | 0.244 | 0.632 | 0.415 | 0.145 | 0.228 | 0.463 | 0.681 | 0.504 | 0.506 | 0.469 | 0.260 | 0.576 | 0.642 | 0.394 | 0.570 | 0.168 | 0.506 | 0.414 | 0.631 | 0.790 | 1.000 | | | |
| **V51** | 0.383 | 0.405 | 0.416 | 0.022 | 0.347 | 0.460 | -0.063 | 0.376 | 0.182 | 0.406 | 0.543 | 0.548 | 0.444 | 0.218 | 0.416 | 0.567 | 0.548 | 0.558 | 0.555 | 0.520 | 0.308 | 0.479 | 0.419 | 0.289 | 1.000 | | |
| **V52** | 0.454 | 0.695 | 0.508 | 0.190 | 0.515 | 0.460 | 0.041 | 0.373 | 0.383 | 0.563 | 0.500 | 0.652 | 0.650 | 0.345 | 0.545 | 0.698 | 0.470 | 0.702 | 0.375 | 0.630 | 0.398 | 0.629 | 0.554 | 0.449 | 0.560 | 1.000 | |
| **V53** | 0.825 | 0.603 | 0.454 | 0.319 | 0.679 | 0.688 | 0.118 | 0.335 | 0.482 | 0.642 | 0.855 | 0.677 | 0.689 | 0.286 | 0.625 | 0.765 | 0.514 | 0.714 | 0.485 | 0.761 | 0.444 | 0.694 | 0.746 | 0.620 | 0.593 | 0.580 | 1.000 |

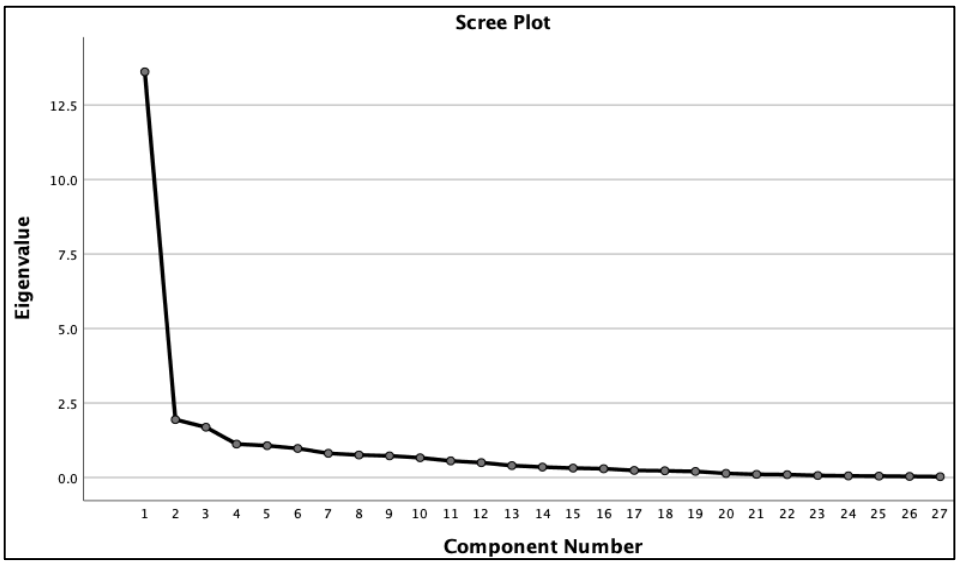
| KMO and Bartlett's Test | | |
|---|---|---|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.825 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 1251.089 |
| | df | 351 |
| | Sig. | 0.000 |

**Communalities**

| | Initial | Extraction |
|---|---|---|
| V4 | 1.000 | 0.794 |
| V5 | 1.000 | 0.734 |
| V6 | 1.000 | 0.617 |
| V7 | 1.000 | 0.772 |
| V8 | 1.000 | 0.766 |
| V9 | 1.000 | 0.812 |
| V10 | 1.000 | 0.448 |
| V11 | 1.000 | 0.679 |
| V12 | 1.000 | 0.521 |
| V13 | 1.000 | 0.805 |
| V15 | 1.000 | 0.764 |
| V18 | 1.000 | 0.794 |
| V19 | 1.000 | 0.691 |
| V20 | 1.000 | 0.636 |
| V28 | 1.000 | 0.690 |
| V31 | 1.000 | 0.783 |
| V33 | 1.000 | 0.740 |
| V34 | 1.000 | 0.815 |
| V35 | 1.000 | 0.753 |
| V36 | 1.000 | 0.777 |
| V37 | 1.000 | 0.604 |
| V45 | 1.000 | 0.761 |
| V47 | 1.000 | 0.760 |
| V48 | 1.000 | 0.622 |
| V51 | 1.000 | 0.720 |
| V52 | 1.000 | 0.761 |
| V53 | 1.000 | 0.810 |

## Total Variance Explained

| Component | Initial Eigenvalues | | | Extraction Sums of Squared | | | Rotation Sums of Squared | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| **1** | 13.613 | 50.417 | 50.417 | 13.613 | 50.417 | 50.417 | 6.703 | 24.824 | 24.824 |
| **2** | 1.943 | 7.198 | 57.614 | 1.943 | 7.198 | 57.614 | 5.235 | 19.387 | 44.211 |
| **3** | 1.687 | 6.247 | 63.862 | 1.687 | 6.247 | 63.862 | 3.576 | 13.244 | 57.455 |
| **4** | 1.120 | 4.149 | 68.010 | 1.120 | 4.149 | 68.010 | 2.018 | 7.472 | 64.927 |
| **5** | 1.069 | 3.959 | 71.969 | 1.069 | 3.959 | 71.969 | 1.901 | 7.042 | 71.969 |
| **6** | 0.973 | 3.604 | 75.573 | | | | | | |
| **7** | 0.810 | 3.001 | 78.574 | | | | | | |
| **8** | 0.756 | 2.800 | 81.375 | | | | | | |
| **9** | 0.726 | 2.688 | 84.062 | | | | | | |
| **10** | 0.663 | 2.454 | 86.516 | | | | | | |
| **11** | 0.557 | 2.064 | 88.580 | | | | | | |
| **12** | 0.498 | 1.846 | 90.426 | | | | | | |
| **13** | 0.398 | 1.476 | 91.901 | | | | | | |
| **14** | 0.349 | 1.294 | 93.196 | | | | | | |
| **15** | 0.315 | 1.168 | 94.364 | | | | | | |
| **16** | 0.293 | 1.084 | 95.448 | | | | | | |
| **17** | 0.237 | 0.879 | 96.327 | | | | | | |
| **18** | 0.224 | 0.830 | 97.157 | | | | | | |
| **19** | 0.201 | 0.746 | 97.904 | | | | | | |
| **20** | 0.136 | 0.505 | 98.408 | | | | | | |
| **21** | 0.105 | 0.389 | 98.797 | | | | | | |
| **22** | 0.095 | 0.354 | 99.150 | | | | | | |
| **23** | 0.066 | 0.245 | 99.395 | | | | | | |
| **24** | 0.054 | 0.201 | 99.596 | | | | | | |
| **25** | 0.045 | 0.167 | 99.763 | | | | | | |
| **26** | 0.036 | 0.133 | 99.897 | | | | | | |
| **27** | 0.028 | 0.103 | 100.000 | | | | | | |



**Un-rotated Component Matrix (only loading values >0.5)**

| | Component | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| V53 | 0.880 | | | | |
| V31 | 0.879 | | | | |
| V34 | 0.869 | | | | |
| V45 | 0.851 | | | | |
| V36 | 0.841 | | | | |
| V13 | 0.832 | | | | |
| V18 | 0.827 | | | | |
| V28 | 0.812 | | | | |
| V47 | 0.812 | | | | |
| V15 | 0.797 | | | | |
| V4 | 0.794 | | | | |
| V19 | 0.786 | | | | |
| V8 | 0.781 | | | | |
| V9 | 0.759 | | | | |
| V52 | 0.752 | | | | |
| V5 | 0.745 | | | | |
| V48 | 0.716 | | | | |
| V33 | 0.652 | | | | |
| V51 | 0.621 | | | | |
| V12 | 0.590 | | | | |
| V6 | 0.558 | | | | |
| V37 | 0.518 | | | | |
| V35 | 0.522 | 0.610 | | | |
| V10 | | -0.607 | | | |
| V11 | | 0.559 | | | |
| V20 | | | 0.637 | | |
| V7 | | | | 0.519 | |

**Rotated Component Matrix (only loading values >0.5)**

| | Component | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| V4 | 0.789 | | | | |
| V9 | 0.786 | | | | |
| V15 | 0.721 | | | | |
| V53 | 0.690 | | | | |
| V47 | 0.680 | | | | |
| V8 | 0.660 | | | | |
| V12 | 0.588 | | | | |

| | | | | | |
|---|---|---|---|---|---|
| V36 | 0.581 | 0.537 | | | |
| V28 | 0.565 | | | | |
| V45 | 0.563 | | | | |
| V31 | 0.562 | 0.547 | | | |
| V48 | 0.558 | | | | |
| V13 | 0.542 | | | | |
| V10 | 0.513 | | | | |
| V6 | | 0.733 | | | |
| V18 | | 0.719 | | | |
| V52 | | 0.716 | | | |
| V19 | | 0.662 | | | |
| V34 | | 0.603 | | | |
| V37 | | | | | |
| V35 | | | 0.802 | | |
| V33 | | | 0.745 | | |
| V11 | | | 0.649 | | |
| V51 | | 0.503 | 0.613 | | |
| V7 | | | | 0.841 | |
| V20 | | | | | 0.745 |
| V5 | | | | | 0.533 |

# Appendix C

## *Appendix C. 1: Logistic Regression Output – (Analysis Version 1)*

```
LOGISTIC REGRESSION VARIABLES COSTOVERRUN
  /METHOD=ENTER FAC1_1
  /PRINT=GOODFIT ITER(1) CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

| **Omnibus Tests of Model Coefficients** | | | | |
|---|---|---|---|---|
| | | Chi-square | df | Sig. |
| Step 1 | Step | 0.719 | 1 | 0.397 |
| | Block | 0.719 | 1 | 0.397 |
| | Model | 0.719 | 1 | 0.397 |

| **Model Summary** | | | |
|---|---|---|---|
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 71.828 | 0.013 | 0.018 |

| **Variables in the Equation** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| | | | | | | | | Lower | Upper |
| Step 1[a] | Factor 1 | -0.250 | 0.307 | 0.663 | 1 | 0.415 | 0.779 | 0.427 | 1.421 |
| | Constant | 0.272 | 0.279 | 0.946 | 1 | 0.331 | 1.312 | | |
| a. Variable(s) entered on step 1: Factor 1. | | | | | | | | | |

```
LOGISTIC REGRESSION VARIABLES COSTOVERRUN
  /METHOD=ENTER FAC1_1 FAC2_1
  /PRINT=GOODFIT ITER(1) CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

| **Omnibus Tests of Model Coefficients** | | | | |
|---|---|---|---|---|
| | | Chi-square | df | Sig. |
| Step 1 | Step | 20.854 | 2 | 0.000 |
| | Block | 20.854 | 2 | 0.000 |
| | Model | 20.854 | 2 | 0.000 |

| **Model Summary** | | | |
|---|---|---|---|
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 51.692 | 0.325 | 0.436 |

| **Variables in the Equation** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| | | | | | | | Lower | Upper |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Step 1[a] | Factor 1 | -0.425 | 0.352 | 1.454 | 1 | 0.228 | 0.654 | 0.328 | 1.304 |
| | Factor 2 | 1.597 | 0.445 | 12.866 | 1 | 0.000 | 4.938 | 2.063 | 11.818 |
| | Constant | 0.323 | 0.347 | 0.870 | 1 | 0.351 | 1.382 | | |
| a. Variable(s) entered on step 1: Factor 1, Factor 2. | | | | | | | | | |

```
LOGISTIC REGRESSION VARIABLES COSTOVERRUN
  /METHOD=ENTER FAC1_1 FAC2_1 FAC3_1
  /PRINT=GOODFIT ITER(1) CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

| **Omnibus Tests of Model Coefficients** | | | | |
|---|---|---|---|---|
| | | Chi-square | df | Sig. |
| Step 1 | Step | 21.240 | 3 | 0.000 |
| | Block | 21.240 | 3 | 0.000 |
| | Model | 21.240 | 3 | 0.000 |

| **Model Summary** | | | |
|---|---|---|---|
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 51.307 | 0.330 | 0.443 |

| **Variables in the Equation** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| | | | | | | | | Lower | Upper |
| Step 1[a] | Factor 1 | -0.467 | 0.367 | 1.614 | 1 | 0.204 | 0.627 | 0.305 | 1.288 |
| | Factor 2 | 1.633 | 0.457 | 12.736 | 1 | 0.000 | 5.118 | 2.088 | 12.546 |
| | Factor 3 | -0.205 | 0.334 | 0.378 | 1 | 0.539 | 0.814 | 0.423 | 1.567 |
| | Constant | 0.317 | 0.348 | 0.832 | 1 | 0.362 | 1.373 | | |
| a. Variable(s) entered on step 1: Factor 1, Factor 2, Factor 3. | | | | | | | | | |

```
LOGISTIC REGRESSION VARIABLES COSTOVERRUN
  /METHOD=ENTER FAC1_1 FAC2_1 FAC3_1 FAC4_1
  /PRINT=GOODFIT ITER(1) CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

| **Omnibus Tests of Model Coefficients** | | | | |
|---|---|---|---|---|
| | | Chi-square | df | Sig. |
| Step 1 | Step | 26.919 | 4 | 0.000 |
| | Block | 26.919 | 4 | 0.000 |
| | Model | 26.919 | 4 | 0.000 |

| **Model Summary** | | | |
|---|---|---|---|
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 45.627 | 0.398 | 0.534 |

| Variables in the Equation | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| | | | | | | | | Lower | Upper |
| Step 1[a] | Factor 1 | -0.626 | 0.406 | 2.382 | 1 | 0.123 | 0.535 | 0.241 | 1.184 |
| | Factor 2 | 1.964 | 0.545 | 12.999 | 1 | 0.000 | 7.126 | 2.450 | 20.724 |
| | Factor 3 | -0.248 | 0.349 | 0.504 | 1 | 0.478 | 0.780 | 0.393 | 1.548 |
| | Factor 4 | 0.917 | 0.404 | 5.145 | 1 | 0.023 | 2.502 | 1.133 | 5.525 |
| | Constant | 0.268 | 0.373 | 0.516 | 1 | 0.472 | 1.307 | | |
| a. Variable(s) entered on step 1: Factor 1, Factor 2, Factor 3, Factor 4. | | | | | | | | | |

```
LOGISTIC REGRESSION VARIABLES COSTOVERRUN
  /METHOD=ENTER FAC1_1 FAC2_1 FAC3_1 FAC4_1 FAC5_1
  /PRINT=GOODFIT ITER(1) CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

| Omnibus Tests of Model Coefficients | | | | |
|---|---|---|---|---|
| | | Chi-square | df | Sig. |
| Step 1 | Step | 27.012 | 5 | 0.000 |
| | Block | 27.012 | 5 | 0.000 |
| | Model | 27.012 | 5 | 0.000 |

| Model Summary | | | |
|---|---|---|---|
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 45.534 | 0.399 | 0.536 |

| Variables in the Equation | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| | | | | | | | | Lower | Upper |
| Step 1[a] | Factor 1 | -0.621 | 0.407 | 2.330 | 1 | 0.127 | 0.538 | 0.242 | 1.193 |
| | Factor 2 | 1.958 | 0.544 | 12.970 | 1 | 0.000 | 7.087 | 2.441 | 20.571 |
| | Factor 3 | -0.237 | 0.349 | 0.461 | 1 | 0.497 | 0.789 | 0.398 | 1.564 |
| | Factor 4 | 0.937 | 0.411 | 5.205 | 1 | 0.023 | 2.553 | 1.141 | 5.710 |
| | Factor 5 | 0.118 | 0.385 | 0.094 | 1 | 0.759 | 1.125 | 0.529 | 2.394 |
| | Constant | 0.247 | 0.380 | 0.423 | 1 | 0.516 | 1.280 | | |
| a. Variable(s) entered on step 1: Factor 1, Factor 2, Factor 3, Factor 4, Factor 5. | | | | | | | | | |

```
LOGISTIC REGRESSION VARIABLES COSTOVERRUN
  /METHOD=ENTER FAC1_1 FAC2_1 FAC3_1 FAC4_1 FAC5_1 FAC6_1
  /PRINT=GOODFIT ITER(1) CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

| Omnibus Tests of Model Coefficients | | | | |
|---|---|---|---|---|
| | | Chi-square | df | Sig. |
| Step 1 | Step | 34.185 | 6 | 0.000 |
| | Block | 34.185 | 6 | 0.000 |

| | | | | |
|---|---|---|---|---|
| | Model | 34.185 | 6 | 0.000 |

| Model Summary | | | |
|---|---|---|---|
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 38.361 | 0.475 | 0.638 |

| Variables in the Equation | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| | | | | | | | | Lower | Upper |
| Step 1[a] | Factor 1 | -0.416 | 0.383 | 1.180 | 1 | 0.277 | 0.660 | 0.312 | 1.397 |
| | Factor 2 | 1.968 | 0.569 | 11.965 | 1 | 0.001 | 7.158 | 2.347 | 21.834 |
| | Factor 3 | -0.121 | 0.347 | 0.123 | 1 | 0.726 | 0.886 | 0.449 | 1.747 |
| | Factor 4 | 0.938 | 0.461 | 4.143 | 1 | 0.042 | 2.554 | 1.035 | 6.300 |
| | Factor 5 | 0.171 | 0.409 | 0.174 | 1 | 0.676 | 1.186 | 0.532 | 2.643 |
| | Factor 6 | 1.168 | 0.521 | 5.026 | 1 | 0.025 | 3.214 | 1.158 | 8.921 |
| | Constant | 0.466 | 0.440 | 1.125 | 1 | 0.289 | 1.594 | | |
| a. Variable(s) entered on step 1: Factor 1, Factor 2, Factor 3, Factor 4, Factor 5, Factor 6. | | | | | | | | | |

```
LOGISTIC REGRESSION VARIABLES COSTOVERRUN
  /METHOD=ENTER FAC1_1 FAC2_1 FAC3_1 FAC4_1 FAC5_1 FAC6_1 FAC7_1
  /PRINT=GOODFIT ITER(1) CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

| Omnibus Tests of Model Coefficients | | | | |
|---|---|---|---|---|
| | | Chi-square | df | Sig. |
| Step 1 | Step | 36.052 | 7 | 0.000 |
| | Block | 36.052 | 7 | 0.000 |
| | Model | 36.052 | 7 | 0.000 |

| Model Summary | | | |
|---|---|---|---|
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 36.494 | 0.494 | 0.662 |

| Variables in the Equation | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| | | | | | | | | Lower | Upper |
| Step 1[a] | Factor 1 | -0.480 | 0.402 | 1.421 | 1 | 0.233 | 0.619 | 0.281 | 1.362 |
| | Factor 2 | 2.157 | 0.659 | 10.720 | 1 | 0.001 | 8.642 | 2.377 | 31.427 |
| | Factor 3 | -0.142 | 0.366 | 0.151 | 1 | 0.698 | 0.868 | 0.424 | 1.777 |
| | Factor 4 | 1.024 | 0.469 | 4.768 | 1 | 0.029 | 2.784 | 1.111 | 6.980 |
| | Factor 5 | 0.096 | 0.421 | 0.052 | 1 | 0.820 | 1.101 | 0.483 | 2.510 |
| | Factor 6 | 1.132 | 0.491 | 5.305 | 1 | 0.021 | 3.100 | 1.184 | 8.121 |
| | Factor 7 | 0.573 | 0.418 | 1.881 | 1 | 0.170 | 1.773 | 0.782 | 4.019 |
| | Constant | 0.509 | 0.458 | 1.233 | 1 | 0.267 | 1.663 | | |

| |
|---|
| a. Variable(s) entered on step 1: Factor 1, Factor 2, Factor 3, Factor 4, Factor 5, Factor 6, Factor 7. |

```
LOGISTIC REGRESSION VARIABLES COST OVERRUN
  /METHOD=ENTER FAC1_1 FAC2_1 FAC3_1 FAC4_1 FAC5_1 FAC6_1 FAC7_1 FAC8_1
  /PRINT=GOODFIT ITER(1) CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Initial -2 Log Likelihood: 72.546

| **Omnibus Tests of Model Coefficients** | | | | |
|---|---|---|---|---|
| | | Chi-square | df | Sig. |
| Step 1 | Step | 36.075 | 8 | 0.000 |
| | Block | 36.075 | 8 | 0.000 |
| | Model | 36.075 | 8 | 0.000 |

| **Model Summary** | | | |
|---|---|---|---|
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 36.472 | 0.494 | 0.662 |

| **Variables in the Equation** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| | | | | | | | | Lower | Upper |
| Step 1[a] | Factor 1 | -0.466 | 0.411 | 1.284 | 1 | 0.257 | 0.628 | 0.280 | 1.405 |
| | Factor 2 | 2.175 | 0.679 | 10.249 | 1 | 0.001 | 8.801 | 2.324 | 33.328 |
| | Factor 3 | -0.130 | 0.372 | 0.122 | 1 | 0.726 | 0.878 | 0.424 | 1.819 |
| | Factor 4 | 1.031 | 0.472 | 4.778 | 1 | 0.029 | 2.803 | 1.112 | 7.062 |
| | Factor 5 | 0.085 | 0.424 | 0.040 | 1 | 0.842 | 1.088 | 0.474 | 2.498 |
| | Factor 6 | 1.129 | 0.486 | 5.405 | 1 | 0.020 | 3.094 | 1.194 | 8.018 |
| | Factor 7 | 0.586 | 0.429 | 1.865 | 1 | 0.172 | 1.796 | 0.775 | 4.164 |
| | Factor 8 | -0.071 | 0.475 | 0.022 | 1 | 0.881 | 0.932 | 0.367 | 2.363 |
| | Constant | 0.538 | 0.503 | 1.143 | 1 | 0.285 | 1.713 | | |

a. Variable(s) entered on step 1: Factor 1, Factor 2, Factor 3, Factor 4, Factor 5, Factor 6, Factor 7, Factor 8.

## *Appendix C. 2: Logistic Regression Output – (Analysis Version 2)*

```
LOGISTIC REGRESSION VARIABLES COSTOVERRUN
  /METHOD=ENTER FAC1_1
  /PRINT=GOODFIT ITER(1) CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

| Omnibus Tests of Model Coefficients | | | | |
|---|---|---|---|---|
| | | Chi-square | df | Sig. |
| Step 1 | Step | 0.155 | 1 | 0.694 |
| | Block | 0.155 | 1 | 0.694 |
| | Model | 0.155 | 1 | 0.694 |

| Model Summary | | | |
|---|---|---|---|
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 72.391 | 0.003 | 0.004 |

| Variables in the Equation | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| | | | | | | | | Lower | Upper |
| Step 1[a] | Factor 1 | -0.111 | 0.283 | 0.153 | 1 | 0.695 | 0.895 | 0.514 | 1.560 |
| | Constant | 0.267 | 0.278 | 0.922 | 1 | 0.337 | 1.306 | | |
| a. Variable(s) entered on step 1: Factor 1. | | | | | | | | | |

```
LOGISTIC REGRESSION VARIABLES COSTOVERRUN
  /METHOD=ENTER FAC1_1 FAC234
  /PRINT=GOODFIT ITER(1) CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

| Omnibus Tests of Model Coefficients | | | | |
|---|---|---|---|---|
| | | Chi-square | df | Sig. |
| Step 1 | Step | 18.883 | 2 | 0.000 |
| | Block | 18.883 | 2 | 0.000 |
| | Model | 18.883 | 2 | 0.000 |

| Model Summary | | | |
|---|---|---|---|
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 53.664 | 0.300 | 0.402 |

| Variables in the Equation | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| | | | | | | | | Lower | Upper |
| Step 1[a] | Factor 1 | -0.165 | 0.322 | 0.264 | 1 | 0.607 | 0.848 | 0.451 | 1.592 |
| | Factor 2, 3, and 4 | 2.612 | 0.747 | 12.227 | 1 | 0.000 | 13.626 | 3.152 | 58.915 |
| | Constant | 0.238 | 0.339 | 0.493 | 1 | 0.483 | 1.269 | | |
| a. Variable(s) entered on step 1: Factor 1, Factor 2, 3, and 4. | | | | | | | | | |

```
LOGISTIC REGRESSION VARIABLES COSTOVERRUN
  /METHOD=ENTER FAC1_1 FAC234 FAC5_2
  /PRINT=GOODFIT ITER(1) CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

| **Omnibus Tests of Model Coefficients** | | | | |
|---|---|---|---|---|
| | | Chi-square | df | Sig. |
| Step 1 | Step | 19.872 | 3 | 0.000 |
| | Block | 19.872 | 3 | 0.000 |
| | Model | 19.872 | 3 | 0.000 |

| **Model Summary** | | | |
|---|---|---|---|
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 52.674 | 0.313 | 0.419 |

| **Variables in the Equation** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| | | | | | | | | Lower | Upper |
| Step 1 | Factor 1 | -0.158 | 0.320 | 0.244 | 1 | 0.621 | 0.854 | 0.456 | 1.599 |
| | Factor 2, 3, and 4 | 2.791 | 0.821 | 11.548 | 1 | 0.001 | 16.305 | 3.259 | 81.572 |
| | Factor 5 | -0.336 | 0.351 | 0.915 | 1 | 0.339 | 0.715 | 0.359 | 1.423 |
| | Constant | 0.178 | 0.352 | 0.256 | 1 | 0.613 | 1.195 | | |
| Variable(s) entered on step 1: Factor 1, Factor 2, 3, 4, Factor 5. | | | | | | | | | |

*Appendix C. 3: Logistic Regression Output – (Analysis Version 3)*

```
LOGISTIC REGRESSION VARIABLES COSTOVERRUN
  /METHOD=ENTER FAC1_1
  /PRINT=ITER(1)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

| Omnibus Tests of Model Coefficients | | | | |
|---|---|---|---|---|
| | | Chi-square | df | Sig. |
| Step 1 | Step | 7.743 | 1 | 0.005 |
| | Block | 7.743 | 1 | 0.005 |
| | Model | 7.743 | 1 | 0.005 |

| Model Summary | | | |
|---|---|---|---|
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 64.804 | 0.136 | 0.182 |

| Variables in the Equation | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| | | | | | | | | Lower | Upper |
| Step 1 | Factor 1 | 0.845 | 0.330 | 6.539 | 1 | 0.011 | 2.328 | 1.218 | 4.449 |
| | Constant | 0.296 | 0.299 | 0.976 | 1 | 0.323 | 1.344 | | |
| Variable(s) entered on step 1: Factor 1 | | | | | | | | | |

```
LOGISTIC REGRESSION VARIABLES COSTOVERRUN
  /METHOD=ENTER FAC1_1 FAC2_1
  /PRINT=ITER(1)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

| Omnibus Tests of Model Coefficients | | | | |
|---|---|---|---|---|
| | | Chi-square | df | Sig. |
| Step 1 | Step | 8.582 | 2 | 0.014 |
| | Block | 8.582 | 2 | 0.014 |
| | Model | 8.582 | 2 | 0.014 |

| Model Summary | | | |
|---|---|---|---|
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 63.964 | 0.149 | 0.201 |

| Variables in the Equation | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| | | | | | | | | Lower | Upper |
| Step 1 | Factor 1 | 0.866 | 0.336 | 6.641 | 1 | 0.010 | 2.377 | 1.230 | 4.594 |
| | Factor 2 | 0.280 | 0.307 | 0.833 | 1 | 0.362 | 1.324 | 0.725 | 2.418 |
| | Constant | 0.308 | 0.303 | 1.034 | 1 | 0.309 | 1.360 | | |
| Variable(s) entered on step 1: Factor 1, Factor 2 | | | | | | | | | |

```
LOGISTIC REGRESSION VARIABLES COSTOVERRUN
  /METHOD=ENTER FAC1_1 FAC2_1 FAC3_1
  /PRINT=ITER(1)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

| Omnibus Tests of Model Coefficients | | | | |
|---|---|---|---|---|
| | | Chi-square | df | Sig. |
| Step 1 | Step | 13.156 | 3 | 0.004 |
| | Block | 13.156 | 3 | 0.004 |
| | Model | 13.156 | 3 | 0.004 |

| Model Summary | | | |
|---|---|---|---|
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 59.391 | 0.220 | 0.295 |

| Variables in the Equation | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| | | | | | | | | Lower | Upper |
| Step 1 | Factor 1 | 0.890 | 0.342 | 6.760 | 1 | 0.009 | 2.435 | 1.245 | 4.763 |
| | Factor 2 | 0.256 | 0.311 | 0.674 | 1 | 0.412 | 1.291 | 0.701 | 2.377 |
| | Factor 3 | 0.670 | 0.322 | 4.343 | 1 | 0.037 | 1.955 | 1.041 | 3.671 |
| | Constant | 0.304 | 0.317 | 0.917 | 1 | 0.338 | 1.355 | | |
| Variable(s) entered on step 1: Factor 1, Factor 2, Factor 3 | | | | | | | | | |

```
LOGISTIC REGRESSION VARIABLES COSTOVERRUN
  /METHOD=ENTER FAC1_1 FAC2_1 FAC3_1 FAC4_1
  /PRINT=ITER(1)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

| Omnibus Tests of Model Coefficients | | | | |
|---|---|---|---|---|
| | | Chi-square | df | Sig. |
| Step 1 | Step | 14.609 | 4 | 0.006 |
| | Block | 14.609 | 4 | 0.006 |
| | Model | 14.609 | 4 | 0.006 |

| Model Summary | | | |
|---|---|---|---|
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 57.938 | 0.241 | 0.323 |

| Variables in the Equation | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| | | | | | | | | Lower | Upper |
| Step 1 | Factor 1 | 0.907 | 0.345 | 6.904 | 1 | 0.009 | 2.478 | 1.259 | 4.876 |
| | Factor 2 | 0.249 | 0.315 | 0.624 | 1 | 0.429 | 1.283 | 0.692 | 2.380 |
| | Factor 3 | 0.683 | 0.325 | 4.417 | 1 | 0.036 | 1.980 | 1.047 | 3.742 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Factor 4 | 0.395 | 0.334 | 1.402 | 1 | 0.236 | 1.484 | 0.772 | 2.854 |
| | Constant | 0.335 | 0.324 | 1.063 | 1 | 0.303 | 1.397 | | |
| Variable(s) entered on step 1: Factor 1, Factor 2, Factor 3, Factor 4 | | | | | | | | | |

```
LOGISTIC REGRESSION VARIABLES COSTOVERRUN
  /METHOD=ENTER FAC1_1 FAC2_1 FAC3_1 FAC4_1 FAC5_1
  /PRINT=ITER(1)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

| **Omnibus Tests of Model Coefficients** | | | | |
|---|---|---|---|---|
| | | Chi-square | df | Sig. |
| Step 1 | Step | 19.863 | 5 | 0.001 |
| | Block | 19.863 | 5 | 0.001 |
| | Model | 19.863 | 5 | 0.001 |

| **Model Summary** | | | |
|---|---|---|---|
| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
| 1 | 52.683 | 0.313 | 0.419 |

| **Variables in the Equation** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I.for EXP(B) | |
| | | | | | | | | Lower | Upper |
| Step 1 | Factor 1 | 0.960 | 0.360 | 7.107 | 1 | 0.008 | 2.611 | 1.289 | 5.289 |
| | Factor 2 | 0.286 | 0.331 | 0.750 | 1 | 0.387 | 1.332 | 0.696 | 2.547 |
| | Factor 3 | 0.780 | 0.359 | 4.730 | 1 | 0.030 | 2.181 | 1.080 | 4.406 |
| | Factor 4 | 0.366 | 0.348 | 1.103 | 1 | 0.294 | 1.442 | 0.728 | 2.853 |
| | Factor 5 | 0.771 | 0.359 | 4.622 | 1 | 0.032 | 2.162 | 1.070 | 4.367 |
| | Constant | 0.442 | 0.358 | 1.526 | 1 | 0.217 | 1.556 | | |
| Variable(s) entered on step 1: Factor 1, Factor 2, Factor 3, Factor 4, Factor 5 | | | | | | | | | |