Transportation Infrastructure Project Cost
Overrun Risk Analysis
*Risk Factor Analysis Models*

by
Qing Wu

B.Sc., Shanghai University of Finance & Economics, 1999

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Business Administration)

THE UNIVERSITY OF BRITISH COLUMBIA

July 2006

© Qing Wu, 2006
Abstract

The main purpose of this thesis is to analyze some common risk factors and to propose several useful analytical models for cost overrun risk analysis in transportation infrastructure investments. Probability models and regression models are proposed and partially (due to data insufficiency) applied using the VIHP (Vancouver Island Highway Project) data. The VIHP case study shows that cost overrun ratio increases as project sizes increase for small road/highway projects (budget < $0.25 million) and bridge/tunnel projects (budget >$0.85 million). However for road/highway projects with budgets over $0.25 million, cost overrun ratio decreases as project size increases. Using the VIHP database, results of a distribution fitting model and a Monte Carlo simulation model are compared. Compared with the distribution fitting model, the Monte Carlo simulation model is shown to underestimate both the upper bound value of project cost overrun ratio and slightly the probabilities of cost overrun. The distribution fitting model and regression model are shown to have close estimates of project costs and cost ranges at each confidence level.
# Table of Contents

Abstract .................................................................................................................. i

Table of Contents ................................................................................................. ii

List of Tables ........................................................................................................ v

List of Figures ......................................................................................................... vi

1. Introduction and Background ............................................................................. 1

2. Literature Review ............................................................................................... 2

   2.1. Introduction ................................................................................................ 2

   2.2. Definition of Risk and Risk Analysis ....................................................... 2

       2.2.1. Risk and Uncertainty ....................................................................... 2

       2.2.2. Risk Components and Definition ................................................... 2

   2.3. Necessity and Importance of Risk Analysis ............................................. 3

   2.4. Sources and Types of Risks ..................................................................... 4

   2.5. Methods of Risk Analysis ....................................................................... 9

       2.5.1. Qualitative versus Quantitative Risk Analysis ................................. 9

       2.5.2. Regression Models for Risk Analysis ............................................ 10

       2.5.3. The Decision Tree ......................................................................... 12

       2.5.4. Sensitivity Analysis ....................................................................... 13

       2.5.5. Probability Distributions Used in Project Risk Analysis .............. 13


   2.7. Conclusions on Literature Review ............................................................. 15

3. Risk Analysis Models and Discussions ............................................................ 16

   3.1. Introduction .............................................................................................. 16

   3.2. Probability Distributions for Project Cost Overrun Risks .................. 17

       3.2.1 Distribution Fitting Method vs. Monte Carlo Method .................... 18

       3.2.2 Probability Distribution for Project Cost Overrun Ratio .......... 19
3.2.2 Conditional Probability Distribution Model for Project Cost Overrun Ratio .......................................................................................................................... 20

3.3. The PPP Models .................................................................................................. 22

3.3.1. Public-Private Partnership (PPP or P3) ......................................................... 22

3.3.2 Risk Allocation in PPP ...................................................................................... 24

3.3.3 Conditions for PPP Success ........................................................................... 24

3.3.4 Comparisons between Public, PPP and Private Projects .............................. 25

3.3.5 PPP Cost Overrun Risk Analysis Models ....................................................... 26

3.4. Regression Model .............................................................................................. 31

3.4.1. Objective and Importance ........................................................................... 31

3.4.2 General Introduction of Risk Variables for the Model ................................. 31

3.4.3 Variable Specification and Analysis .............................................................. 33

3.4.4 Incorporation of Cost Overrun Risk into the Project Evaluation Process .... 38

3.5. Vancouver Island Highway Project Case Study .............................................. 40

3.5.1. The Database ............................................................................................... 40

3.5.2 Cost Overrun Ratio Distribution Fitting ....................................................... 41

3.5.3 Monte Carlo Simulation Model Results vs. Beta Fitting Model Results .... 44

3.5.4 Project Cost Overrun vs. Project Size in the VIHP ..................................... 46

3.5.5 A Numerical Example of Predicting Project Cost for a New Project Using a Regression Model and a Beta Fitting Model ........................................... 55

3.5.6. Multiple Regression Model Analysis ......................................................... 58

4. How the Project Managers Can Use these Models ........................................... 60

5. Summary and Conclusions ............................................................................... 61

References ............................................................................................................. 62

Appendix ................................................................................................................... 64

Appendix 1. Beta Distribution ................................................................................. 64

Appendix 2 Maximum Likelihood Estimation for Beta Distribution Shape Parameters ........................................................................................................... 65
List of Tables

Table 1. Akingtoyie's Survey Conclusions .......................................................... 6
Table 2. Main Project Risk Sources and Their Impacts ........................................ 7
Table 3. Main Findings in Four Project Cost Overrun Studies ............................... 14
Table 4. TRRL Report Conclusions ..................................................................... 14
Table 5. Aalborg University Report Conclusions for Different Project Type ........... 15
Table 6. Probabilities for Project Cost to Exceed Certain Values at Period n+1, with Beta
Distribution (4, 6, 2, 15) and Project Budget (G) of $10 million ........................ 22
Table 7. Scenarios for PPP Projects Cost Overrun Risk and Revenue Shortage Risk
Allocation in Private Sector’s Perspective ............................................................ 27
Table 8. Cost Overrun Risk Bearer Dummy Variable Values under Different Scenarios 33
Table 9. Cost and Cost Overrun Ratio Ranges of VIHP data ................................. 41
Table 10. Cost Overrun Ratio Descriptive Statistic Report ...................................... 42
Table 11. Beta Distribution Parameter Setting and Estimates ................................. 43
Table 12. Probabilities for Project Cost over Certain Values ................................... 44
Table 13. Beta Distribution Parameter Estimates Comparison for Cost Overrun Ratio
Variable between Monte Carlo Simulation Method and Beta Fitting Method .......... 45
Table 14. Cost Overrun Probability Estimation Comparison between Monte Carlo
Method and Beta Fitting Method ......................................................................... 45
Table 15. Expected Cost and Confidence Interval Estimation for a Project with Budget of
$1,200,000 ........................................................................................................... 56
Table 16. Project Cost Prediction Intervals at Different Confidence Levels for a Project
with Budget of $1,200,000 .................................................................................. 56
Table 17. Mean, Mode and Standard deviation Estimates ....................................... 57
Table 18. Project Type Variable Values for Each Project Type ............................... 59
Table 19. Cost Overrun Ratio Means and 95% Confidence Intervals at Each Project Stage
......................................................................................................................... 60
List of Figures

Figure 1. Risk, Uncertainty and Information Availability for Risk Events ......................................... 2
Figure 2. Main Risks for Transportation Project .................................................................................. 9
Figure 3. A Simple Decision Tree Example ....................................................................................... 12
Figure 4. Beta Distribution (4, 6, 2, 15) PDF for Project Cost at period n+1 .............................. 21
Figure 5. Cost Overrun Ratio Histograms for the VIHP data .......................................................... 42
Figure 6. Beta Distribution PDF Plot for Cost Overrun Ratio in the VIHP ........................................ 43
Figure 7. Beta Probability Plot for Cost Overrun Ratio Variable ...................................................... 43
Figure 8. Probability Density Function Comparison between Monte Carlo Simulation Method and Beta Fitting Method ........................................................................................................ 45
Figure 9. Scatter Plot of Cost vs. Budget for Road/Highway Projects ............................................ 47
Figure 10. Scatter Plot of Log (Cost) vs. Log (Budget) for Road/Highway Projects .................... 47
Figure 11. Project Cost vs. Budget Scatter Plot for Small R/H Projects .......................................... 48
Figure 12. Estimated Cost vs. Budget Model for Small Road/Highway Projects .......................... 49
Figure 13. Cost Increase for Each $10,000 Project Budget Increase for Small Road/Highway Projects ...................................................................................................................................... 49
Figure 14. Log (Cost) vs. Log (Budget) Scatter Plot for Medium Sized R/H Projects .................. 50
Figure 15. Cost vs. Budget for Medium Sized Road/Highway Projects ........................................... 50
Figure 16. Cost Increase for Each $50,000 Project Budget Increase for Medium Sized Road/Highway Projects ...................................................................................................................................... 51
Figure 17. Project Cost vs. Budget Scatter Plot for Large Road/Highway Projects ....................... 51
Figure 18. Cost vs. Budget for Large R/H Projects ......................................................................... 52
Figure 19. Cost Increase for Each $500,000 Project Budget Increase for Large Road/Highway Projects ...................................................................................................................................... 52
Figure 20. Cost vs. Budget for Bridge and Tunnel Projects ............................................................ 53
Figure 21. Log (Cost) vs. Log (Budget) for Bridge and Tunnel Projects ........................................... 53
Figure 22. Log (Cost) vs. Log (Budget) for Large Bridge and Tunnel Projects ............................... 54
Figure 23. Cost vs. Budget for Large Bridge and Tunnel Projects (Budgets over $850,000) ....... 54
Figure 24. Cost Increase for Each $500,000 Project Budget Increase for Bridge/Tunnel Projects with Budget over $850,000 ........................................................................................................ 55
Figure 25. Prediction Intervals at Each Prediction Confidence Level for a Project with Budget of $1,200,000 ...................................................................................................................................... 56
Figure 26. Comparisons on Project Cost Prediction and Confidence Intervals from the Regression and Beta Fitting Models respectively ................................................................. 58
1. Introduction and Background

Today, with globalization and urbanization, transportation infrastructure construction projects are carried out all over the world. According to Transportation Canada’s provincial transportation investment report, transportation investment in each province (including purchases of machinery and equipment) constitutes more than a quarter of total social investment in Newfoundland, Prince Edward Island and Nova Scotia between 1992 and 1995. In Quebec, Ontario and the western provinces, the corresponding ratio is about one fifth.

Transportation infrastructure construction projects are stated as a major contributor to economic growth and community well-being. This point is emphasized in the reports of Transportation Canada. Yet, we frequently hear reports of project failures. The abandonment of Mirabel Airport in Montreal, Canada, is such an example. Transportation project failures often cause millions or even billions of dollars in losses to the society, the costs of which are generally transferred to investors and taxpayers. Although Mirabel Airport is an extreme example, most transportation infrastructure projects are vulnerable to another widely existed risk which imposes magnificent costs on the society as well: cost overrun.

Cost overrun is the excess of actual project costs over budgeted project costs and may be caused by underestimation of costs at the planning stages or by the escalation of the original scope of the project. Cost overrun may not necessarily lead to project failure if the project can obtain sufficient revenue and benefits to cover its costs. However, the viability of the project, initially based on the underestimated costs, would be in question when cost overrun risk materializes. Projects with underestimated costs are chosen for their higher estimated NPV compared with other alternative projects. With cost underestimation, the choice becomes questionable.

The main objective of transportation infrastructure projects is to improve net social benefit. Net social benefit can be calculated by deducting a project’s social costs from its social benefits. However, with the large and irreversible investments required for transportation projects, neglecting or underestimating project risks creates large costs for society - exactly opposite to projects’ general objective and viability. Risk analysis is therefore crucial to transportation projects.

Underestimation of project risk may cause large social costs. On the other hand, overestimation of risk, leading to over-conservative designing, planning and budgeting, may lose potential social benefits. Reliable risk analysis is important to transportation project cost benefit analysis and management.

The main tasks of this thesis are to analyze cost overrun risk factors for transportation infrastructure projects, to propose models to quantify their influences on project cost overrun, and to fit proper probability distributions for project cost overrun risk variables. This work focuses on the development of models for transportation project cost overrun risk analysis, including probability distribution models, Public-Private-Partnership models, and risk factor regression analysis models. Vancouver Island Highway Project
case study will be an application example for probability distribution and regression analysis models. Project cost estimates using different analysis models are compared. Hopefully our analysis, models and conclusions will help project evaluators make better risk-adjusted cost estimation for transportation infrastructure projects.

2. Literature Review

2.1. Introduction
This literature review provides a general picture for the application of risk analysis in transportation project assessment and builds a theoretical base for our project cost overrun risk analysis models. It covers topics of risk definition, project risk sources, and risk evaluation methods. Section 2.2 will review definitions of risk and risk analysis; Section 2.3 explains the importance and necessity of risk analysis for transportation project assessment; Section 2.4 identifies the main risk sources for transportation projects; Section 2.5 discusses some widely used methods in risk analysis; Section 2.6 shows some interesting results and conclusions from several empirical studies of project risk.

2.2. Definition of Risk and Risk Analysis

2.2.1. Risk and Uncertainty
As Pouliquen (1970) states, risk analysis is a method for dealing with uncertainty. However, risk and uncertainty are not synonymous. Frame (2003) explains the difference is that when making decisions under conditions of risk, we know or can estimate the probability distribution of this risk event, whereas under conditions of uncertainty, we are unable to estimate the probability distribution. So Frame distinguishes risk and uncertainty according to information availability. Figure 1 roughly depicts this difference.

![Figure 1. Risk, Uncertainty and Information Availability for Risk Events](Sources: Frame (2003))

Referring to project risk and risk analysis instead of project uncertainty, we actually assume we know or can estimate the probability distributions for project risks. Project risk analysis should be based on these probability distributions.

2.2.2. Risk Components and Definition
In some cases, such as in financial markets, the risk often means the variability of return. Two stocks with the same expected return might be categorized differently with respect to risk, because of different degrees of variability of their returns. The riskier stock (with greater variability of return) may bring a higher return than the less risky stock in favorable scenarios. It may also bring a lower return than the less risky stock in unfavorable scenarios. Investors with different level of risk tolerance will have different preferences toward these two stocks. The risk-averse people would prefer the stock with
less return variability and the risk-takers would prefer the one with greater variability. Normally, investors behave to be risk averse. Similarly in project management, two projects may have equivalent expected net present values, but reasonable project sponsors and investors are assumed to prefer the one with less variability in expected NPV.

Although in some cases risks are simply interpreted as return variability, risk likelihood and risk impact are the two important faces of risk. Neglecting either of these two elements leads to an incomplete view of risk. As Frame (2003) points out, some risk events may have high impact but have such low likelihood of happening that we normally do not take them into concern. He gives the example of a comet hitting the earth.

Cooper and Chapman (1987) define risk as the “exposure to the possibility of economic or financial loss or gain, physical damage or injury, or delay, as a consequence of the uncertainty associated with pursuing a particular course of action”. And risk analysis is said to involve various approaches for dealing with the problems caused by uncertainty including the identification, evaluation, control and management of risks.

Sturk et al. (1996) defines risk as “expected consequence” or “expected loss of utility”. Based on this definition, they evaluate risks as:

\[
\text{Risk} = \text{Probability} \times \text{Consequence}
\]

Similar to Sturk et al., Jaafari (1999) defines risk as “the exposure to loss/gain, or the probability of occurrence of loss/gain multiplied by its respective magnitude”. For example, project cost overrun risk can be interpreted as the product of project cost overrun probability and the cost overrun magnitude. These formula and definition are rather for risk definition than for risk measurement. They are used to articulate the common view that risk has two components: likelihood and impact.

Every project faces the risk of failure. Project failure can be caused by cost overrun, benefit shortage, construction delay, and bad quality. And these risks may be traced further to more specific risk factors. Our interest here is what these risk factors are and how they affect project consequences.

2.3. Necessity and Importance of Risk Analysis

In cost benefit analysis of projects, the most common methods used today are based on net present value estimation. However, Pindyck (1991) argues that the NPV rule, "Invest in a project when the present value of its expected cash flows is at least as large as its cost” would not be reliable any more in scenarios with large irreversible investments and the option of postponing investments. Investment irreversibility, accompanied by large sunk costs, is common to transportation infrastructure projects. Pindyck finds that irreversibility makes a project much more sensitive to project risks. His study illustrates the unreliability of traditional cost benefit analysis, which simply relies on the NPV rule, and the necessity of analysis of project risks.

\[
\text{Project risks refer to the uncertain events that may cause project failure. These risks normally can be depicted by some random variables.}
\]
Flyvbjerg (2003) also argues that risk analysis is very important for project decision makers in order to obtain a clearer and more realistic view of possible outcomes. Risk analysis is fundamental for risk management. It provides project managers with a basis for identifying strategies to reduce risk, including: allocating risk among parties, transferring risks to professional risk management institutions, hedging positions in financial markets, and so on.

As Cooper and Chapman (1987) point out, risk analysis is crucially important when low-probability, high-consequence events are critical for decision making. These risk events are very likely to be neglected in the calculation of the net present value of projects. For examples, terrorist attack has very low probability of happening. However, without any preparation for its happening, the outcome would be an even larger disaster. As an example, New York City's preparation for disaster system helps to prevent the terrorists' attack on World Trade Center from thousands of additional life loss.

Risk analysis is important for better understanding project risks, risk effects and risk interaction, according to Cooper and Chapman's discussion. Risk analysis requires a conscious and systematic method to analyze the risks. During this process, project investors and evaluators can gain more complete and detailed knowledge about risk events, interactions of these risk events, and their relationships with project consequences.

Furthermore, based on knowledge of these risk events and relationships, project investors and designers can make better plans and designs to control potential loss, in case risk events are materialized. Thus risk analysis can help with better planning and responses for risk events; help and provide feedbacks to project design and planning; and help with better project construction and operation risks handling.

Cooper and Chapman conclude that risk analysis is especially important when projects require large capital investments; have unbalanced cash flows; require a large proportion of total investment before any returns are obtained; are subject to significant new technology; have unusual legal, insurance or contractual arrangements; are subject to important political, economic or financial parameters; face sensitive environmental or safety issues; or have stringent regulatory or licensing requirements. Transportation infrastructure projects obviously fit this description.

Although risk analysis can neither prevent nor accurately predict risks, it can facilitate superior preparedness and responses to risks. One example is a transportation project designed to meet the emergency safety requirement for minimizing life loss in case of fire.

2.4. Sources and Types of Risks

Ayyub (2003) discusses potential risk events for projects, including technological risks, economic climate risks, political risks, large project risks, and contractual or legal risks.

Technological risk refers to the fact that rapidly improving technology may make a project out-dated and abandoned. Technology planned for a project can require replacement by newer technology during the project construction period, adding to
project costs. The decision to replace the old technology or not should be based on whether the benefit of the new technology will at least compensate for the cost of replacing the old one. In some situations, there may be no option: the new technology will be necessary due to technological improvement throughout the industry or technological improvements of complementary equipment.

Project outcomes are influenced by social and domestic economic trends, called economic climate. The economic growth of the society has high impacts on traffic volume, financial market stability, and project performance.

Political risk refers to the circumstances under which projects are subjective to political factors. For example, foreign investors in some developing countries may face unfriendly local governments, or potential expropriation. Since large transportation projects often require approval and financial support from local and federal governments, conflicts between and within governments may bring about project failure.

Contract and legal risks arise from inappropriate responsibility division among contractors. During the whole project period, land using, payment dispute, and other legal issues might be raised unexpectedly. These contractual and legal issues may cause delay, suspension or even abandonment of the projects.

Unfortunately, Ayyub (2003) does not give theoretical analysis or quantified estimation for these risks’ impacts on project outcomes.

For transportation infrastructure projects specifically, Flyvbjerg (2003) emphasizes the financial risks, including: cost overrun, revenue shortage (lower than expected revenues, caused by insufficient traffic volume) and increased financing costs (exchange rate for cross-nation projects and interest rate). Inefficient project management, contractor conflicts and accidents may cause construction cost overrun. Benefit shortage is caused directly by insufficient users, probably due to poor project design or social economy changes. Given the huge amount of debt required for financing transportation projects, even a small change in interest rates can increase financing costs dramatically and affect the viability of the projects. Investors finance their investments in capital markets via bonds, stocks and other borrowing tools. Changes in interest rates influence the values of these financing tools and influence project financing costs. For those projects financed in international capital markets, exchange rates are another concern.

With regard to large projects, Jaafari (2001) suggests project risks include: market, political, technical, financing, environmental, cost estimate, operating, etc. Market risks refer to the adverse changes in labor, material or any other supply markets, which may increase costs or decrease benefits. Political risks are caused by policy or regulation changes from political sectors. Technical risks refer to situations in which new technologies create the possibility of not achieving budget, schedule or other targets. Financing risk is correlated with cost overrun, and benefit shortage. Environmental risks refer to the possibility of natural hazards. Cost estimate risk arises from estimated costs
that are insufficient compared with actual costs. Operating risks refer to the unexpected events which adversely affect project operation.

Based on their surveys on the Kuwait construction industry, Kartam et al. (2001) present their findings about Kuwait contractors’ perceptions of sources of risk. Financial factors are regarded as the most significant risks, followed by contractual and labor, material and equipment availability. Kartam et al. also find that sub-contracting is a powerful tool for minimizing project risks, as long as risks are properly allocated between the investors and contractors.

According to Akintoye (1997), construction project risks can be defined as variables causing variability in construction project costs, duration and quality. They identify environmental, design, financial, legal, political, construction and operation risks. According to their survey on UK project contractors and managers, there exist some differences between these two groups’ perceptions for individual risk source premium ranking. However, the two groups draw similar conclusions regarding the relative importance of these risk sources. They both regard financial and contractual risks as the most important types of risks to projects, consistent with Kuwait contractors’ views according to Kartam. Their views are summarized in Table 1.

Table 1. Akintoye's Survey Conclusions

<table>
<thead>
<tr>
<th>Risk Sources</th>
<th>Perception of Risk Premium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environmental (e.g. weather)</td>
<td>Low</td>
</tr>
<tr>
<td>Political, Social &amp; Economic (e.g. inflation)</td>
<td>Moderate</td>
</tr>
<tr>
<td>Contractual agreement (e.g. responsibilities)</td>
<td>High</td>
</tr>
<tr>
<td>Financial</td>
<td>High</td>
</tr>
<tr>
<td>Construction (productivity, injury, safety)</td>
<td>Moderate</td>
</tr>
<tr>
<td>Market/industry (availability of workload)</td>
<td>High</td>
</tr>
<tr>
<td>Company (corporate)</td>
<td>Moderate</td>
</tr>
<tr>
<td>Development in IT</td>
<td>Low</td>
</tr>
<tr>
<td>Project (design information)</td>
<td>High</td>
</tr>
</tbody>
</table>

(Source from Akintoye (1997))

According to this survey, both contractors and managers depend mainly on their intuition and subjective judgment to manage risks. About half of managers claim to be familiar with sensitivity analysis\(^2\), yet few managers use this technique in practice. Based on this survey, contractors and managers are said to doubt on the usage of quantitative risk analysis techniques in practice.

Shapira (1994) also finds that managers are quite ‘insensitive’ to the probability estimation for project outcomes. They seem to emphasize outcome values more than

\(^2\) Sensitivity analysis is a method of risk analysis that examines the influence of possible changes to project parameters on project NPV. We will discuss this method in detail in Section 5.
probabilities. Shapira reasons that this is due to managers' beliefs that risks can be controlled. Unfortunately, this belief may not necessarily be true. The backup insurance companies and other financial instruments may create the sense to project managers that risks are controllable. However, project managers neglect the fact that there always exist some risks that are impossible to be controlled, such as social economic trends, which could put insurance companies and financial markets into difficulty as well. Shapira finds that under unique, non-repeated decision conditions, managers neglect statistical analysis easily. This kind of managerial behavior has significant influence on the reliability of project cost benefit analysis.

Public-Private-Partnerships are now widely used in transportation infrastructure projects, as well as other projects requiring large investment which exceeds the means of private investors or governments alone. One example of a Public-Private-Partnership is "BOT" (build-operate-transfer), where private companies build and operate the infrastructure for a certain period then transfer its ownership to the public sector. The key risks of design, construction, cost overrun and financing are believed to be transferred to private companies. The popularity of PPP depends on the assumption that private sectors can manage certain project risk better than the public sector can.

According to Grimsey and Louis' (2002), risks of PPP projects differ little from those of other projects and can be evaluated and managed similarly. The main risks for PPP projects are said to be caused by the complexity of the PPP arrangements themselves, in terms of complicated documentation, financing, sub-contracting and so on. In PPP projects, the success of documentation depends on interactions between different contractors and investors. It requires additional effort and time to reach agreements on the issues raised by different contractors. The assignment of responsibilities and rights among the parties and the complicated contract issues become key sources of risk to the projects, in addition to the risk factors present in any single project sponsor scenarios. We discuss PPP project risks in details in the PPP model section.

The above studies provide a general portrait of major risk factors in transportation projects. They help identify risk variables for our models. Table 2 provides a summary of them.

Table 2. Main Project Risk Sources and Their Impacts

<table>
<thead>
<tr>
<th>Risk Factors</th>
<th>Major impacts on Project</th>
<th>Management Perception</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial Risk</td>
<td>Benefit deficit</td>
<td>High</td>
</tr>
<tr>
<td>Economics Risk</td>
<td>Benefit deficit and Cost overrun</td>
<td>High</td>
</tr>
<tr>
<td>Construction Risk</td>
<td>Benefit deficit and Cost overrun</td>
<td>High</td>
</tr>
<tr>
<td>Contractual Risk</td>
<td>Cost overrun</td>
<td>High</td>
</tr>
<tr>
<td>Environment Risk</td>
<td>Cost overrun</td>
<td>Moderate</td>
</tr>
<tr>
<td>Funding Risk</td>
<td>Project termination</td>
<td>High</td>
</tr>
<tr>
<td>Project Design Risk</td>
<td>Benefit deficit, Cost overrun and Project abandonment</td>
<td>High</td>
</tr>
</tbody>
</table>
One important issue here is whether these risks are measurable. Due to the uniqueness of each project, exact project risk probability distributions for individual projects are almost impossible to obtain. Previous project data could be used to obtain approximate probability distributions for these risk factors. Based on the probability distributions, we can estimate the likelihoods of project risks for individual projects. The risk analysis method section in the Literature Review provides a more detailed discussion of risk measurement.

The risk factors discussed above are mostly in project investors and contractors’ perceptions. Since project cost benefit analysis is made from society’s perspective, externalities may be another significant concern for project risk sources. For example, the rapid transit line connecting Vancouver and Richmond (RAV) may expose residents along its route to noise and safety issues, which may lead to conflicts or even law suits between the project operators and the residents. Externality risks should be included in project risk analysis for RAV line project evaluation.

Identification of project risk factors and their relationships are presented in Figure 2. This structure will be used as the basis for our regression analysis.

As reflected in Figure 2, transportation project failure\(^3\) can be caused by negative project NPV, funding risk, externality risk and design risk. Funding risk refers to the possibility that investors may fail to provide further funding. Externality risk materializes when large negative externality of the project causes legal suits from the affected community, which may lead to construction delay, suspension or even project abandonment. Poor project designs not only influence costs and benefits, but only affect quality adversely, making projects vulnerable to other risks such as natural hazards.

\textit{Negative project NPV} is directly caused by cost overrun and benefit shortage. Cost overrun can be caused by construction risk, project design risk, environmental risk, contractual risk and economic risk. Design, financial, economic and construction risks are the major contributors to project benefit shortage.

These risk factors are somehow correlated. For example, economic risk may influence financial and construction risk, due to their sensitivities to social economic changes. Environmental and contractual risks may increase construction risk. Changes in weather and natural hazards may cause construction delays, safety issues and so on. Contractual disputes may cease project construction, causing project construction delay and productivity or quality issues.

\(^3\) The words in italics are corresponding to the terms in Figure 2.
2.5. Methods of Risk Analysis

2.5.1. Qualitative versus Quantitative Risk Analysis

According to the methods used for risk assessment, risk analysis can be categorized as either qualitative or quantitative. Qualitative analysis assesses variable probabilities and consequences using subjective judgments, expertise opinions and experiences. Quantitative analysis is based on probabilistic and statistical methods. To contrast these two: When one declares some risk events are very likely to occur, he/she is doing qualitative risk analysis. When one demonstrates that the probability for a specific risk event to happen is about 60%, which is based on statistical analysis of historical data, he is doing quantitative risk analysis.

As Ayyub (2003) concludes, qualitative analysis may be sufficient to identify the risks for a system, while quantitative risk analysis examines the system in greater details for risk assessment. Frame (2003) argues that narrative statements in qualitative risk analysis are normally unclear and subject to varied interpretations by different people. Furthermore, they are often not testable. Therefore, it is difficult to judge whether their conclusions are wrong or not. Quantitative risk analysis, on the other hand, generally gives clear and testable conclusions. According to Ayyub and Frame’s arguments, quantitative risk analysis is important for giving project evaluators clear and detailed views of project risks, which cannot be replaced by qualitative risk analysis.
Based on a survey of project managers, Lyons (2004) finds that the most commonly used tools for risk identification in practice are brainstorming, the case-based approach and the checklist. His finding is consistent with several other researchers' findings, as discussed in this Literature Review, such as Akintoye's. Intuition, judgment and experience are the most frequently used risk assessment techniques in practice. Qualitative risk assessments are said to dominate quantitative ones. However, according to this survey, there is no significant factor to limit implementation of quantitative risk analysis in practice. It is said that all these factors, including: cost concerns, difficulty in seeing benefits, human/organizational resistance, lack of accepted industry model for risk analysis, lack of dedicated resources, lack of expertise in the techniques, lack of familiarity with the techniques, lack of information, and lack of time, were shown to be low or moderately relevant to managers' reluctance to do quantitative risk analysis.

One agreement from the above papers and survey results is that project managers normally do not make quantitative risk analysis in practice. They seem to rely on qualitative risk analysis only. This is not because they lack the techniques of quantitative risk analysis, but probably due to their lack of belief in its value. Despite the importance to use expertise experiences and judgments in projects, absolute dependence on subjective judgments may bring severe problems. First, it is questionable about the qualification and definition for the expertise. Who can be titled as an expertise? What are the criteria? Even for a person spending his lifetime on project assessment, his experience is still quite limited, compared with the huge project database used to achieve statistical analysis. And his judgments and estimates may not be accurate, even if they are not mixed with his personal bias. Second, absolute dependence on subjective analysis may easily raise moral hazard problems in project evaluation. One example is the overestimation for future project user numbers and project benefits, which can be found now and then in project cost benefit analysis to balance the project budgets. The project evaluators are normally employed by the project sponsors. They are very likely to be influenced for the project sponsors’ interest concerns. And this challenges the objectiveness of their analysis and evaluation. Although quantitative risk analysis does not eliminate this problem, it can help control it. Reliable and easily interpretable quantitative risk analysis is necessary for improving project risk analysis.

### 2.5.2. Regression Models for Risk Analysis

Regression models are widely used to analyze the influences of risk factors. Not only are they effective for explaining risk events and quantifying the impacts of risk factors, they are also easily understood by people.

The Center for Operation Excellence at the University of British Columbia undertook a project to analyze the risk of oil leakage for an oil shipping company. Using historical data, Poisson regression model is successfully used to analyze the relationships between potential risk factors and oil leaking risk. Some interesting and unexpected conclusions about the risk factors’ influences on oil leaking risk are concluded from that regression model, which shows the existence of subjective misunderstanding or bias in previous qualitative risk analysis.
Flyvbjerg (2004) uses simple linear regression models to analyze the relationships between transportation project cost overrun and three potential risk factors respectively: length of project implementation phase, project size and project ownership.

The length of implementation phase is defined as the period from the project construction decision to completion of construction. Flyvbjerg treats projects as outliers if their implementation phases are 13 years long or even longer. The estimated regression model for project cost overrun and project length is \( \Delta C = 0.4 + 4.64 \times T \), where \( \Delta C \) is the cost escalation (in %) and \( T \) is the length (years) of the implementation phase. The 95% confidence interval for the slope is (2.10, 7.17). According to this estimated model, for every additional year of project implementation, the project cost escalation is expected to increase by 4.64%. The influences of implementation phase length on cost escalation are similar for all modes: rail, fixed-link (bridge and tunnel) and road projects.

Using project budgets to represent project sizes, Flyvbjerg looks into the impacts of project size on project cost. Flyvbjerg finds that for bridge and tunnel projects, larger projects tend to have larger cost escalations; however, for rail and road projects, this relationship is not significant. For all project types, the data do not support the assumption that a bigger project has a larger risk of cost escalation.

In Flyvbjerg's model, project ownership is categorized into private, state-owned enterprise, and other public ownership. The data do not support the assumption that public ownership is problematic and private ownership is a main source of efficiency in controlling cost escalation. State-owned enterprises show the poorest performance with an average cost overrun of 110%. Privately owned fixed links have an average cost overrun of 34%. Other public ownership shows the best performance with an average cost overrun of “only” 23%. Flyvbjerg argues that the main problem in relation to cost overrun may not be public versus private ownership but a certain kind of public ownership, namely state-owned enterprises, which lack not only the transparency and public control that public sectors would implement but also the competence the private sector would bring.

Flyvbjerg's models test the assumed impacts of these three potential risk factors on project cost overrun. Regression models help to provide intuitive and accurate understandings for these impacts. For example, he concludes that the average cost escalation would be about 4.64% for each additional project year, instead of the vague and subjective assumption that project delay may cause cost overrun.

However, Flyvbjerg investigates only the individual impact of each risk factor on project cost escalation and neglects the combined impact of several risk factors on project cost. He does not discuss the relationships between the risk factors. However, the relationships between these risk factors may influence the estimates of their individual impacts on project cost. He uses simple linear regression models for his study on the impacts of risk factors. He does not explain the viability of using simple linear regression models.
An advantage of using a regression model for risk analysis is its intuitive appeal. The uncertainties projects face can be related to risk factors with economical meanings. For example, uncertainty about interest rate changes can be represented by the sensitivity of project performance on interest rate changes, using the regression coefficient estimate for the interest rate change variable estimated in the regression model.

2.5.3. The Decision Tree
A decision tree is another risk analysis tool widely used in practice. Cooper and Chapman (1987) give a simple example for this method. For each single cost risk factor, we can give the probability of its happening. For its happening case, the risk event can be grouped, for example, Minor, Modest, Major or Maximum, with their respective probability. For each group of risk events, its impact on the project can be further grouped into sub-groups, such as None, Negligible, Significant, and Catastrophic, based on probability estimation. The outcome under each condition can be calculated based on this structure.

Figure 3. A Simple Decision Tree Example

Boardman and Greenberg (1996) carry out decision tree analysis in two basic stages. First, they use a diagram, called a decision tree, to specify the logical structure of the decision problem with sequences of decisions and their contingencies. Second, work backward from final outcomes to the initial decision, calculating expected net benefit values under each condition and finding out the branch with largest net benefit value.

As an example, for the decision tree in Figure 3, the expected NPV of “Taking Project” would be \( P_1 \times NPV_1 + P_2 \times NPV_2 \). The expected NPV of “Not Taking Project” would be the result of Do-Nothing strategy. If \( P_1 \times NPV_1 + P_2 \times NPV_2 \) is greater than the quantified value of Do-Nothing strategy, we would take the project, otherwise, we would suspend or abandon the project.

Let \( P_1 = 0.6, P_2 = 0.4, NPV_1 = -2, NPV_2 = 2 \), and quantified Do-Nothing result is \(-0.2\), the expected NPV of “Taking Project” is \(-0.4\), less than the quantified result of Do-Nothing strategy, \(-0.2\). Therefore, we should not carry on with this project.

—

4 Quantified result of Do-Nothing Strategy is often a negative value and not necessarily equal to 0. For example, without necessary road expansion, traffic congestion would lead to extra costs to society making the NPV of “not taking” the road expansion project would be negative.
The decision tree method is very intuitive for risk analysis. However, it is generally used under conditions of discrete probability distributions and may not applicable to analyze risk factors with continuous and correlated probability distributions.

2.5.4. Sensitivity Analysis
As shown in the surveys on contractors and managers in section 2.4, sensitivity analysis may be the most frequently used risk analysis method in project evaluations. In many cases, it is the only risk analysis tool used in project cost benefit analysis. Partial sensitivity analysis takes uncertainties into cost benefit analysis by varying cost and benefit estimation within a certain presumed ranges of variability. Range of expected project net present value can be estimated.

Boardman and Greenberg (1996) mention worst-and-best-case analysis, checking whether combinations of reasonable assumptions reverse the sign of net present value. Worst-and-best analysis is used to obtain the most conservative and most optimistic estimates for project net present values. For example, project evaluator may have conservative estimates for three independent risk factors. Using these conservative estimates for risk variables, which is the worst-case scenario, the worst-case estimate can be obtained. Similarly, the best-case estimate can be calculated by using optimistic estimates of risk variables. Worst-and-best case analysis gives a range of project performance estimation. Neither partial sensitivity analysis nor worst-and-best case analysis takes into account the probability distributions of risk factors and project outcomes. They provide only the range of net present values for changes in risk factors.

2.5.5. Probability Distributions Used in Project Risk Analysis
For choosing proper probability distribution, Pouliquen (1970) argues that the target of risk analysis is not to find the real distribution but rather the distribution best representing "the judgment of an appraisal team". If this statement is true, a probability distribution would be valid as long as it reflects the project team's experience and understanding of project. For example, the project appraisal team may think one project risk factor follows a normal distribution based on their knowledge and experience. Then in risk analysis, this risk factor can be assumed to follow the normal distribution. This argument is obviously based on the strong assumption on the appraisal team's reliability and credibility.

Two well-known probability distributions are used widely in project risk analysis: the normal distribution and the beta distribution.

Frame (2003) concludes that the normal distribution is generally used when routine processes are involved. For example, the duration of each bus trip is likely to follow the normal distribution for its repeated and routine schedules. However, the normal distribution is not generally useful when unique events are involved, for example, a

---

5 See Appendix 1 for introduction of Beta distribution.
transportation project using new technology. The normal distribution is not feasible to those variables with upper or lower bounds, or has a PDF (probability density function) shape different from the normal distribution’s bell shape.

Regnier (2005) states that the probability distribution of project activity time is widely represented by the beta distribution due to desirable properties of Beta distribution.

- The Beta distribution has finite limits. Many real-world random variables have lower and upper bounds. For example, project duration has a lower bound that is greater than zero. The model should reflect this property of these random variables.
- The Beta distribution can be asymmetric. Many variables in reality are right or left skewed to some unlikely but significant result.
- The Beta distribution is flexible and can reflect different shapes, including U and inverted-U shapes.

### 2.6. Empirical Studies on Magnitude of Cost Overrun Risk for Transportation Projects

Flyvbjerg (2003) describes four studies comparing budgeted and actual costs for transportation projects. His main findings are listed in Tables 3, 4 and 5.

Table 3. Main Findings in Four Project Cost Overrun Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Size</th>
<th>Average Cost Overrun</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auditor-General of Sweden¹</td>
<td>15</td>
<td>86%</td>
<td>2% - 182%</td>
</tr>
<tr>
<td>US Department of Transportation²</td>
<td>-</td>
<td>61%</td>
<td>-10% - 106%</td>
</tr>
<tr>
<td>TRRL UK³</td>
<td>13</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Aalborg University, Denmark⁴</td>
<td>258</td>
<td>28%</td>
<td>-</td>
</tr>
</tbody>
</table>

1. The study by Auditor-General of Sweden study covers rail and road projects. For rail projects, the cost overrun average was 17%, ranging from -14% to 74%. This study might underestimate the real cost overrun, since two-thirds of its sample projects were still carried on.

2. Rail projects only

3. TRRL: Transportation and Road Research Laboratory

Table 4. TRRL Report Conclusions

<table>
<thead>
<tr>
<th>Cost Overrun</th>
<th># of Metro Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10%-20%</td>
<td>4</td>
</tr>
<tr>
<td>20%-50%</td>
<td>3</td>
</tr>
<tr>
<td>50%-100%</td>
<td>4</td>
</tr>
<tr>
<td>100%-500%</td>
<td>2</td>
</tr>
</tbody>
</table>

(Sources: Flyvbjerg (2003))

4. Covers bridge, tunnel, road and rail projects, located in 20 countries, and completed between 1927 and 1998. Its project cost overrun mean and standard deviation estimates for each project type are listed in Table 5.
Table 5. Aalborg University Report Conclusions for Different Project Type

<table>
<thead>
<tr>
<th>Project Type</th>
<th>Average Cost Overrun</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rail</td>
<td>45%</td>
<td>38</td>
</tr>
<tr>
<td>Tunnels &amp; Bridges</td>
<td>34%</td>
<td>62</td>
</tr>
<tr>
<td>Road</td>
<td>20%</td>
<td>30</td>
</tr>
</tbody>
</table>

(Sources: Flyvbjerg (2003))

These four studies have a common conclusion: cost overrun occurs commonly among transportation projects. According to the Aalborg University report, standard deviations of cost overrun for their sample projects are large. Assuming project cost overrun follows the normal distribution, the 95% confidence interval for cost overrun estimation for rail projects would be −30% to 120%. For tunnel and bridge projects, the confidence interval would be wider, due to their larger standard deviations. The high likelihood of cost overrun in transportation projects increases the importance of cost overrun risk analysis in transportation project evaluation.

Another interesting finding in the Aalborg university report is that project cost overrun has not improved over the past seventy years. There seems to be no element of learning in transportation project cost estimation and management, despite improved knowledge, experience and technology. However, due to the varied locations of the sample projects, it is not clear whether this conclusion is valid in all the states or just some of them. The Aalborg university report finds that cost underestimation and overrun are more pronounced in developing countries than in developed countries. The possibility exists therefore that this no-learning characteristic comes from including more developing country observations in recent years. Further analysis is necessary to draw a conclusion.

2.7. Conclusions on Literature Review

Risk analysis is important to transportation infrastructure projects because of their large irreversible investments and high sensitivity to risks. Although managers and contractors almost rely only on qualitative risk analysis in practice, quantitative risk analysis is essential for better project risk assessment and management.

Transportation projects encounter a long list of risks including economic, financial, safety, contractual, regulatory risks and natural hazard. Here we focus on the risk of cost overrun. Despite our focus on project cost overrun risk, cost overrun is not necessarily a problem itself for transportation projects. It is a deficit in project net present value caused by cost overrun that will lead to failure of transportation projects.

Unfortunately, current literature does not provide probability distribution analysis for transportation project risk factors either from empirical studies or in theoretical analysis. This impedes the application of quantitative risk analysis to project cost benefit analysis. Furthermore, although works on project risk factor identification have been done, few theoretical studies have been conducted to build models to quantify the impacts of these risk factors on project consequences. In this thesis, we attempt to make up for these missing points in project cost overrun risk analysis.
3. Risk Analysis Models and Discussions

3.1. Introduction
From the Literature Review, we know that risk is composed of likelihood and impacts. Neglecting either one will lead to an incomplete understanding of project risks. The Literature Review also discusses some methods for project risk analysis, including regression models, decision trees, and sensitivity analyses. Each method has advantages and disadvantages.

Both the decision tree and the sensitivity analysis methods are case-based. For each individual project, we use a decision tree and sensitivity analysis to analyze project risks. Their dependence on information from previous projects is mostly brought by project evaluators' subjective experiences and views. Quantitative analysis on risk probability and impacts are necessary to support these kinds of case-based methods, not only for better using the information from historical data, but also for alleviation of subjective bias in risk analysis.

It would be no use to take historical data into account if a project is absolutely unique, which means it is being carried out with absolutely new technology, design, and materials in an absolutely new environment, under new policy and so on. However, project uniqueness does not mean projects are absolutely unique, rather that each project has some unique features. That is, projects may share some common risk factors, while each is also affected by some unique risk factors. In section 2.4 of Literature Review, we discuss some common project risk factors. Understanding the impacts of these common risk factors on project outcomes will help to better manage these risks and to better estimate project outcomes.

Regression models using historical data are useful to test and quantify the impacts of these common risk factors on project outcomes. For example, using regression analysis, project evaluators can conclude that project cost is expected to increase 4% for each one-year increase in the length of the project construction phase, instead of the subjective and obscure statement "project delay may increase project cost". In addition, this impact estimate of risk will help project evaluators obtain more accurate project cost benefit analysis, assist project decision-makers make better decisions and strategy and improve project managers' risk management strategy. Incorporating project risk estimates into project evaluation can make cost and benefit estimation more realistic. Project risk estimates also provide a quantitative reference to project decision makers and managers. These estimates can be different from project evaluators' subjective expectations on project risks. This difference actually yields some useful information from historical data for project evaluators' reference. In Section 3.4.4, we will discuss incorporating project risk estimates into project evaluation in details.

As discussed in the Literature Review, when we analyze risks, we cannot neglect either the probability of the risk event or the impact of the risk event. Accordingly we will propose two data-based models to analyze project cost overrun risk. The probability
distribution model can be used to estimate the probability of project cost overrun risk. The regression model would be used to estimate impacts of project risk factors on project cost.

We will introduce probability distribution models to analyze project cost overrun risk variables in Section 3.2. In Section 3.3, we discuss Private Public Partnership (PPP) in transportation projects and propose a PPP cost overrun risk analysis model. In Section 3.4, we propose a multiple linear regression model to analyze the impacts of risk factors on transportation project cost overrun risk. The Vancouver Island Highway Project (VIHP) case study in Section 3.5 provides examples of applications of probability distribution analysis and regression analysis to analyze transportation project cost overrun risk.

3.2. Probability Distributions for Project Cost Overrun Risks
Project evaluators may be interested in the expected cost overrun ratio\(^6\) for a new project as well as the probability of the project cost overrun ratio exceeding a specific value. Probability distribution analysis can provide estimates of mean and standard deviation and confidence intervals for project cost overrun risk variables. Project evaluators normally estimate project cost and set project budgets using their knowledge and experience from similar past projects. However, knowledge and experience are subjective and difficult to be quantified accurately. Moreover, individual project evaluator’s experience and knowledge may be biased and may not reflect the reality of the whole industry. Well-founded statistical models are needed to analyze valid historical data and provide reliable information and estimates for project costs.

Both of the two probability distribution models in this section are based on historical data. An important requirement for properly using these two models is that the database for fitting these distributions should include only those projects with similar construction, economic and political conditions. To estimate the costs for a new project, project evaluators should use probability distribution models fitted to data from projects with similar characteristics, such as:

- construction and management technologies
- size
- economic and political conditions
- private investment structures (for example, PPP) and public-private risk sharing/arrangements

We treat cost overrun risk variables as random variables, which are influenced by some random risk factor and build probability distribution models for them.

---

\(^6\) The project cost overrun ratio is calculated as project cost/project budget. Project cost is what project investors actually pay for the project; project budget is the estimated project cost before projects start. Projects have cost overrun problem if project cost overrun ratio is greater than one; otherwise project costs are under control.
3.2.1 Distribution Fitting Method vs. Monte Carlo Method

Distribution fitting is an analytical method which uses empirical data to find the 'best' fit probability distribution for variables. In many statistical analysis tools, distribution fitting functions are available for various distribution types, including beta, gamma, exponential, and others. The general steps of distribution fitting are:

1. Using valid data, plot the histogram of the target variable.
2. Based on this histogram and experience, determine the distribution type most clearly reflected.
3. Use distribution fitting functions to estimate distribution parameters.

Some software, such as the Input Analyzer in Arena, may automatically choose the 'best' fitting distribution model, based on the sample data pattern. However, project evaluators should test the validity of this 'best' fit distribution model with their own judgment. These automatically fitted distributions may not be realistic.

The Monte Carlo method is based on simulation techniques. The general steps of Monte Carlo simulation are the following:

1. Make assumptions on the probability distributions of the factor variables
2. Use computer to generate a large number of 'random draws' from the assumed probability distribution of each factor variable
3. Use these 'randomly' drawn values of the factor variables to estimate the target variable
4. Use these target variable estimates to calculate mean, standard deviation, etc. for the target variable

The Monte Carlo simulation method is especially useful when the target variable is influenced by many complicated and correlated factor variables or/and when data are not available. However, there are some potential problems in the implementation of Monte Carlo simulation in transportation project evaluation.

First, Monte Carlo simulation results can be biased due to project evaluators' subjective views on project risk factors. Since the whole Monte Carlo simulation model is based on project evaluators' prior assumptions on project risk factors, the results of Monte Carlo simulation are possibly biased. Similarly, Monte Carlo simulation results can be intentionally made up, by adjusting the prior assumptions.

Second, the 'random draws' from a computer are not really random. To date, computers have not been able to generate real random numbers. These 'random draws' may lose some information from the original data. In the VIHP case study, we provide comparisons between results from Beta Fitting model and Monte Carlo Simulation model. The simulated probability distribution of cost overrun ratio in the Monte Carlo Simulation model has a smaller range than that of the Beta Fitting model. The discrepancy between these two model results is due to the inability of the computer to provide truly random draws. This may not be a severe problem in some cases. However, for project cost overrun ratio variable, the bounds of the values, especially the maximum value of the project cost overrun ratio, are what interest project sponsors most.
Underestimation of the value of the maximum cost overrun would cause problems. In VIHP case study, we will show that cost overrun probability estimates in Monte Carlo simulation model are slightly lower than those in Beta Fitting model.

Third, when a system becomes more complicated, Monte Carlo simulation may be difficult to be implemented. It is common that when the system gets complicated in real cases, simulation techniques may give results, which cannot be validated by real data. Often project evaluators/analysts’ countermeasure to this situation is to ‘adjust’ model assumptions to ‘make’ the model validated. Due to the large time and expertise efforts required to examine a complicated simulation model, the validity of a simulation model would hardly be challenged. However, the validity and reliability of a Monte Carlo Simulation model is a big question.

However, we are not arguing that Monte Carlo Simulation would not be applicable in transportation project evaluation. The Distribution Fitting method cannot be a full replacement for Monte Carlo Simulation, especially when valid data are not available. The advantage of the Distributing Fitting method is that it is based on real data and therefore its results are expected to be more objective and realistic. We recommend using Distribution Fitting method to estimate the distributions of factor variables to be used in a Monte Carlo method. It would help alleviate the bias problem in Monte Carlo Simulation Models. And we would recommend using Distribution Fitting method in project evaluation when valid data are available, due to its better fitting the reality, its relative simplicity in application, and the possibility for its results to be audited and examined.

In this thesis, we focus on Beta Distribution Fitting (Beta Fitting), since our target variables, transportation project cost and project cost overrun ratio, are expected to follow Beta distributions. The models in this thesis can be extended to analyze other variables following various distributions.

### 3.2.2 Probability Distribution for Project Cost Overrun Ratio

We use the Beta distribution\(^7\) to fit project cost overrun ratio data, due to its nice features, such as its finite upper and lower bounds and flexible shape. A project cost overrun ratio has finite minimum and maximum values as do Beta distributions. When project construction and operation progress as scheduled, actual project cost should be reasonably controlled; project cost overrun ratio should be reasonably close to the industry norm. However, if unexpected situations arise and cause, for example, project delays or safety or quality issues, the actual project costs may increase and the cost overrun ratio may deviate unfavorably from the expected.

So transportation project cost overrun ratio histogram is expected to have a tail to the right, an upper bound, and a lower bound greater than 0, due to the fact that project real cost is always larger than 0. Probability distributions with infinite bounds (for example, normal distribution) are incapable to depict these features for project cost overrun ratio variable. Nor would those distributions with inflexible shapes different from the cost overrun ratio data pattern be applicable, for example, the exponential distribution.

\(^7\) See Appendix 1 for Beta distribution introduction
The histogram pattern of the project cost overrun ratio provides the basis for our assumption that we can use the Beta distribution to reflect the transportation project cost overrun ratio variable. With valid and adequate data, we can fit a beta distribution to transportation project cost overrun ratio. Based on the fitted distribution, we can estimate probabilities of project costs exceeding a certain level and the confidence intervals for these probability estimations. In VIHP case study, we have a numerical example using this model.

3.2.2 Conditional Probability Distribution Model for Project Cost Overrun Ratio

Transportation projects have several phases: evaluation phase, design, construction, operation, etc. Moreover, during the construction phase, project managers may evaluate project cost control at several sub-stages. The conditional probability distribution can help the project evaluator use information from earlier stages to better estimate project cost overrun probability. Based on the estimated cost for the next stage, the project manager can adjust and schedule construction to control costs.

Model Assumptions

- The project can be divided into several periods or stages. Project costs at these periods or stages are independent of one another. Absolute independence between costs at different project stages may not be exactly true. However this does not make this assumption unrealistic. Independent costs of different project stages mean that project cost at one stage would not be quite different no matter whether project cost at other stages high or low. Take RAV line project as an example, RAV line is composed with several sections of high speed rail, some of which are underground while others are on the ground or aboveground. In contrast to construction costs of on the ground sections and above ground sections, construction costs for the underground section are likely to have different structure and sources of risk. We assume project evaluators are able to divide projects into some evaluation phases with reasonably independent costs.

- Project costs at each period or stage follow a Beta distribution. Since we made the assumption from the earlier model that project cost overrun ratio can be fitted with a Beta distribution and since the cost overrun ratio is calculated as actual project cost divided by project budgeted cost, which is constant, project cost can be assumed to follow Beta distribution as well.

The Model

Suppose we are at the end of project evaluation period \( n \), and we know the accumulated project cost from earlier \( n \) periods is \( \sum_{i=1}^{n} C_i \). We are interested in the probability for project cost exceeding project budget at period \( n+1 \). Assume cost at period \( n+1 \) follows a Beta distribution \( (P, Q, A, B) \), probability for project cost to exceed budget at period \( n+1 \) is:

---

8 Absolute independence may not exist at all in some philosophers' view.
9 See Appendix 1 for Beta Distribution introduction. P and Q are the shape parameters; A and B are the value parameters.
\[
\Pr(\sum_{i=1}^{n+1} C_i > G | \sum_{i=1}^{n} C_i) = \Pr \left[ C_{n+1} > (G - \sum_{i=1}^{n} C_i) | \sum_{i=1}^{n} C_i \right] = \begin{cases} 
1 & \text{if } \sum_{i=1}^{n} C_i \geq G \\
\int_{-\infty}^{B(P,Q)} f(x|P,Q,A,B) \, dx & \text{if } \sum_{i=1}^{n} C_i < G 
\end{cases}
\]

\( C_i \) = project cost at each period \( i \)

\( G \) = project budget

\( f(x|P,Q,A,B) = \frac{(x-A)^{p-1}(B-x)^{q-1}}{B(P,Q)(B-A)^{p+q-1}} \), \( A \leq x \leq B; \ P, Q > 0 \)

\( B(P,Q) = \int_{0}^{1} x^{p-1}(1-x)^{q-1} \, dx \)

This model can also be used to calculate the probability of the project cost being between specific value bounds. As an example, the probability of a project cost being between 1.2 times and 1.5 times of budget is:

\[
\Pr(1.2G < \sum_{i=1}^{n+1} C_i < 1.5G | \sum_{i=1}^{n} C_i) = \Pr \left[ (1.2G - \sum_{i=1}^{n} C_i) < C_i < (1.5G - \sum_{i=1}^{n} C_i) | \sum_{i=1}^{n} C_i \right] = \int_{1.2G-\sum_{i=1}^{n} C_i}^{1.5G-\sum_{i=1}^{n} C_i} f(x|P,Q,A,B) \, dx
\]

Project evaluators can use this method to estimate probabilities of project costs exceeding various special values, for example, 2 times project budget.

**A Numerical Example**

Say a project has a budget \( G \) of $10 million and we know accumulated project costs up to the end of project period \( n \) are about 40% of project budget. We assume project costs in period \( n+1 \) follow a Beta Distribution \((4, 6, 2, 15)\). Figure 4 shows the probability density function for this Beta distribution.

Figure 4. Beta Distribution \((4, 6, 2, 15)\) PDF for Project Cost at period \( n+1 \)

With \( G = 10 \), \( P = 4 \), \( Q = 6 \), \( A = 2 \) and \( B = 15 \), the probability for project cost to exceed budget at the end of period \( n+1 \) can be estimated as:

---

\(^{10}\) P=4, Q=6, A=2, B=15
The probability of project cost exceeding 1.5 times budget at the end of period \( n+1 \) is:

\[
\Pr(\sum_{i=1}^{n+1} C_i > 1.5G | \sum_{i=1}^{n} C_i = 40\%G) = \Pr(C_{n+1} > 1.1G) = \int_{1.1}^{6} f(x | 4, 6, 2, 15) dx = 2.9\%
\]

The probability of the project cost exceeding budget, but not exceeding 1.5 times budget at the end of period \( n+1 \) is:

\[
\Pr(G < \sum_{i=1}^{n+1} C_i < 1.5G | \sum_{i=1}^{n} C_i = 40\%G) = \Pr(0.6G < C_{n+1} < 1.1G) = \int_{0.6}^{1.1} f(x | 4, 6, 2, 15) dx = 68.3\%
\]

The following shows probabilities of project costs exceeding certain values at the end of period \( n+1 \), given various project cumulative cost values for the previous \( n \) periods.

Table 6. Probabilities for Project Cost to Exceed Certain Values at Period \( n+1 \), with Beta Distribution (4, 6, 2, 15) and Project Budget (G) of $10 million

<table>
<thead>
<tr>
<th>Cumulative Project Cost at end of period</th>
<th>For project cost at the end of period ( n+1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( &gt;0.5G )</td>
</tr>
<tr>
<td>0.1G</td>
<td>96.3%</td>
</tr>
<tr>
<td>0.2G</td>
<td>99.7%</td>
</tr>
<tr>
<td>0.4G</td>
<td>100%</td>
</tr>
<tr>
<td>0.5G</td>
<td>100%</td>
</tr>
<tr>
<td>0.8G</td>
<td>100%</td>
</tr>
</tbody>
</table>

The advantage of this conditional probability model is that it makes use of cost information from previous periods or stages of the project. This method would give more reliable project cost overrun probability estimates than our first model, in which project costs are not divided and analyzed at stages.

The key requirements for this conditional probability model are project stage division, validity of model assumptions and reliability of the project database. Regulated standards and rules are necessary for project managers to properly define project stages; Cost records and report systems should be set in each project; Project cost information at each stage should be collected and built into a database at transportation research institutions, for example, at risk management departments of provincial Transport Canada offices.

### 3.3. The PPP Models

#### 3.3.1. Public-Private Partnership (PPP or P3)

Grimsey and Lewis (2004) define PPPs as “arrangements whereby private parties participate in, or provide support for, the provision of infrastructure”. A PPP project results in a contract, based on which a private entity provides public infrastructure-based services.
PPP means more than the statement that the private sector finances public infrastructure. Grimsey and Lewis view the essence of PPP as the public sector buying “a stream of services under specified terms and conditions”. Compared with simple contracting-out, PPP requires close partnerships between the parties, as well as more complicated and specific contracts identifying each party’s risk, responsibilities and rights.

PPP is expected to improve project efficiency by involving the private sector entity into project management, revenue or demand management and cost control. Compared to typical public project management, the private sector can manage projects better by setting clear targets or goals at each stage of the project, clarifying the responsibilities of managers, and making managers accountable for their responsibilities.

The private entity is expected to be more proficient at revenue or demand analysis and other marketing research work. Marketing information will improve project design, construction and operation to better meet user demands. The private sector is expected to bring “market disciplines into public service” (Grimsey and Lewis, 2004).

PPP is expected to bring cost effectiveness. Instead of contracting out project design, building, operation and maintenance to different contractors, PPP often uses one contractor. As Grimsey and Lewis point out, “bundling” of a long-term partnership compels the project contractor to manage projects more thoughtfully and carefully starting at the design stage, which leads to project cost effectiveness.

There are many types of PPPs. They can be BOT (Build Operate Transfer), BOO (Build Own Operate), joint ventures of public and private sectors, and others. Different types of PPPs may reflect different risk-sharing relationships between public and private partners. Simply grouping projects into “private” and “public” may be misleading, since what really bring about project cost effectiveness are the risk and responsibility shared by private sector, which drive the private sector to manage project in a more efficient way.

There is another fact increasing the importance to investigate project risk sharing structures between partners. A reasonable private entity would try to avoid risks that are not fully compensated for. Although one PPP project may be classified as BOT, in which case project risks are considered to be primarily on private partner’s side, the public sector may nevertheless take the majority of project revenue or demand risk. The RAV (Richmond-Airport-Vancouver) project may be one example. The private entity will build and operate this rapid transit line and transfer it to public sector after a certain period. However according to the contract, TransLink will collect the fares and pay the private entity a certain amount of fees every period, which are based on the pre-assumed RAV line ridership numbers. This term almost exempts the private sector from revenue shortage risk. The public sector will face an even worse situation if RAV Line ridership forecasts turn out to be over estimated.
3.3.2 Risk Allocation in PPP

Should all project risks be transferred to private sector to induce its cost saving incentive? This depends on the ability of the private sector to manage project risks. As Grimsey and Lewis argue, if the private sector is not the better manager of project risks, the risk premium it demands would outweigh the benefits to the public sector. So, they conclude, “optimum, rather than maximum, risk transfer is the objective of the PPP arrangement”. Here the optimum refers to the scenario when risks are assigned to the parties who can best manage them. Optimal risk allocation of transportation projects requires minimizing both probability and impacts of risks at least cost. One typical case is when the public sector transfers the majority of design, construction, and operation risk to the private sector, including cost overrun risk, construction delay risk, and so on. The public sector retains the risk of land use, policy, environment, and so on.

On the government’s side, the failure of a transportation infrastructure project would be unlikely to influence the “credibility” of the government. Governments would still be regarded with little default risk. The project risks are actually born by the taxpayers and other members of the society.

The private sector normally sets a subsidiary limited company for the project and project risks are transferred to this limited company. If the project fails, this subsidiary limited company bears all the financial loss and may declare bankruptcy, while the private sector may still survive. In this way, the project risks, namely transferred to the private project sector, are actually transferred to the investors, financiers and other creditors of that subsidiary limited company.

Although the private sector can transfer majority of project risks to a limited company, they do bear certain “credibility” risk. Failure of a project would adversely influence credibility of companies responsible for it. And diminished company credibility would further influence the future borrowing rates and financing costs of the private sector. The private sector will therefore have incentives to manage project risks efficiently unless it can always be a monopoly, which is unlikely in developed and open markets today. So we can assume proper involvement of private sector would improve project risk management.

3.3.3 Conditions for PPP Success

Grimsey and Lewis (2004) quote six key ‘drivers’ for successful projects in one report from Arthur Andersen:

- Risk transfer
- Long-term contracting
- Output-based project specification
- Competition
- Performance measurement

- Private sector management skills
In addition, they recommend centralized monitoring of project performance, communication among government departments, use of ongoing benchmarks, and careful assessment of the risk transferred to the private sector.

Grimsey and Lewis list another four main aspects of successful projects from one study\textsuperscript{12} of the National Audit Office: clear definition of objectives, proper procurement processes, high-quality bids, and ensuring final deal make sense.

From these studies, we can find some key conditions for PPP project success:
1. Clear identification of risk and responsibility sharing among each partner
2. Output-based performance measurements at each project stage
3. Competitive bids
4. Long-term partnership between public and private sectors
5. Cost-effective risk allocation so that risks are allocated to the party who can best manage them

These conditions can be instructive to all transportation infrastructure projects. The key issue for these conditions is to make sure allocating public and private sector the tasks they are best with and obliging both of them to be responsible for their tasks.

3.3.4 Comparisons between Public, PPP and Private Projects
The conditions for PPP project success show that private involvement alone cannot ensure project success. The key for PPP project success is that the parties in projects are taking what they can best cope with and are responsible for their tasks indeed. In an ideal world, these conditions may be achieved no matter whether private sector would be a partner or not. However in real world, competitiveness and other market contributors must be implemented into this system to achieve these conditions. Incorporation of private capital is expected to bring up better project management and other market contributors to transportation infrastructure projects.

Should all transportation projects be purely privatized then? In other words, should government leave transportation infrastructure projects to the market completely? The answer should be “no”.

First, transportation infrastructure projects often require huge amounts of investments, which can hardly be funded by single private companies. Even if private companies can raise such large amounts of money, the capital costs and the project financing risk may be high.

Second, private companies make investment decisions for the sake of their own profitability, not for the society’s well-being. However, major benefits of transportation infrastructure projects, such as economic improvement, alleviation of congestion, and so on, are in society’s interest. For their own profits concern, the private companies might

either abandon some projects or collect high tolls from users, both of which would depress project social benefits and worsen transportation problems such as congestion and safety.

Third, absolute privatization may cause land-use problems. As an example, in some developing countries, the profitability of toll ways has attracted the building of additional toll ways in the same area and connecting the same cities. Redundancy of highways causes not only profit loss to private companies but also land waste. The “invisible hand” of the market might bring balance finally. However transportation construction is quite different from many other industries such as pencil production. One redundant highway may cause millions of loss to the society, or even more. Project assessment and approval, auditing, monitoring and regulation should be important tasks of governments.

Fourth, because of their high requirements of investments, land-use and environmental impact, transportation infrastructure projects bear high political and environment risks, both in risk possibility and impacts. Once these risks materialize, the project is likely to fail, since individual private sector can hardly manage these risks. For example, delay of government approval for land use or changes to environmental standards may cause transportation projects to be postponed, delayed, or even abandoned. Involvement of public sector will at least minimize or alleviate these risks.

However, it does not mean privatization is not feasible for transportation projects ever. Private projects can be successful but should not be out of the monitor of governments.

As a cooperative relationship between governments and the private sector, is PPP the best choice for transportation projects? It depends. The complicated contracts and legal issues of PPP lead to extra costs, such as negotiation costs and contracting costs. For a small transportation project, a PPP would be too costly. Simple contracting-out would usually be a better choice. For large transportation projects which require high management skills and cooperation between the sectors, PPP may be a good choice. PPP should be used only when benefits of PPP to the project outweigh its costs.

3.3.5 PPP Cost Overrun Risk Analysis Models

3.3.5.1 PPP Cost Overrun Risk and Revenue Shortage Risk Allocation Scenarios
The success of PPP project asks for optimal risk allocations between public and private sectors. For example, the private sector takes design and construction risks and the public sector takes environmental, political and land-use risks. Even under this optimal risk allocation, different cost overrun risk and revenue shortage risk allocation between public and private sectors may cause different results to projects. Here we define the cost overrun risk taker as the one who will reinvest for increased project cost in addition to project budget; The revenue shortage risk taker is the one exposed to revenue or demand variation.

For PPP projects, the private sector may bear cost overrun risk and revenue shortage risk or leave them to the public sector. There are four scenarios in general for PPP project cost
overrun risk and revenue shortage risk allocation, in perspective of whether private sector takes these two risks. Cost efficiency in PPP projects is generally assumed to be ensured by the private sector's cost control management. So we focus on discussing and comparing the incentives of the private sector to control project cost in these scenarios.

Project benefits to the private sector can be more than its project revenue share. As we have discussed, the private sector may want to improve its reputation or credibility in the industry for future profits. Unfortunately this aspect of benefit depends on the viewpoint and evaluations of individual private company, and is too difficult to be included in the following scenario comparison, which discusses private sector in general. So we assume private sectors in these scenarios are pursuing profit maximization. Benefits other than revenues are not included in the following discussion.

Table 7. Scenarios for PPP Projects Cost Overrun Risk and Revenue Shortage Risk Allocation in Private Sector's Perspective

<table>
<thead>
<tr>
<th>Scenario Index</th>
<th>Cost Overrun Risk</th>
<th>Revenue Shortage Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Not Responsible</td>
<td>Not Responsible</td>
</tr>
<tr>
<td>II</td>
<td>Not Responsible</td>
<td>Responsible</td>
</tr>
<tr>
<td>III</td>
<td>Responsible</td>
<td>Not Responsible</td>
</tr>
<tr>
<td>IV</td>
<td>Responsible</td>
<td>Responsible</td>
</tr>
</tbody>
</table>

First, we compare the effects of scenarios I and II on project costs. In these two scenarios, the public sector takes the cost overrun risk and will invest the additional capital if project costs exceed budgets.

- In scenario I, the public sector also takes revenue shortage risk, which means the private sector does not pay the increased costs and receives the fixed revenue. With fixed revenues, the private sector will try to lower costs so that its profit is maximized. As long as the private sector expects possibility of controlling project costs under budget to pay less for the project, it will have an incentive to lower project costs.

- In scenario II, the revenue of the private sector depends on future demands arising from this project. In this case, project cost increases may improve project design and quality and, therefore, future demand and revenue. To maximize its profit, the private sector may increase project cost to improve project quality as long as the expected marginal revenue is greater than the marginal cost. There is no guarantee that the private sector would keep project costs as low as possible in this scenario.

For project cost control, the risk allocation in scenario I may be better at driving the private sector to reduce project cost than in scenario II. However, it would be too cursory to conclude that scenario I would be a better choice for project risk allocation. The cost efficiency in scenario I is likely to be obtained in the prices of project quality decline and may cause the public sector, or even the whole project, to incur losses of benefits.

In scenarios III and IV, the private sector is responsible for project cost overrun risk, which means when project costs exceed budget, private sector will be asked to pay at least part of the additional.
Compared with scenario I, private sector in scenario III would not change its cost control behavior much. As long as the private sector faces fixed revenue, it will always have an incentive to lower its costs so that its profit would be maximized.

In scenario IV, the private sector will still try to find the optimal cost input, with which the marginal benefit of the private sector equals its marginal cost and its profit is maximized, as in scenario II. However, scenarios II and IV are likely to result in different optimal cost input amounts, since the revenue and cost functions of the private section under these two scenarios are not the same.

According to the analysis of these four scenarios, it seems that when private sector is offered fixed benefits, its cost control behavior is irrelevant to the project cost overrun risk allocation. However, if the project budget in scenario I is ‘delusively’ underestimated at the project assessment and evaluation stage and private sector knows it, scenarios I and III may induce different cost management behavior in the private sector. If the private sector knows or expects the project budget to be underestimated, in scenario I, since private sector is not required to re-invest the increased costs and expects little possibility to control the costs under budget, private sector may not have as much incentive to control project cost as in scenario III.

Compared with revenue conditional on real project demand, fixed revenue may be better at inducing private sector to control costs. However, this scenario analysis is from the perspective of the private sector’s profit and helps to understand private sector’s cost control behaviors under these scenarios. The better choice for cost control in this scenario analysis may not be the better choice from the perspective of the social benefit. Corresponding regulations are necessary to make sure project cost reductions do not damage public or social interests. For example, more detailed and clearly defined project quality standards may prevent the private sector decreasing project quality to lower costs.

This PPP risk allocation scenario analysis provides a general picture of how the involvement of the private sector in project cost overrun risk and revenue shortage risk can influence project cost control. The scenarios simply classify private sector involvement in these two project risks as “Responsible” (involved) and “Not Responsible” (not involved). In the following model, we use the proportion of private investment to represent private sector involvement in transportation projects and analyze its impacts on project cost.

3.3.5.2. PPP Cost Overrun Risk Analysis Model: A Necessary Condition for Cost- Efficient Private Involvement in Transportation Project

As we discuss in earlier sections, when properly monitored and used, private sector involvement in transportation projects can achieve better revenue and cost management. In this model, we use the proportion of private investment to represent private sector involvement under the assumption that the more the private sector invests, the greater the private sector involvement. Our objective is to find some necessary conditions for making sure private involvement will help control project costs in the scenario under which private sector benefit is not fixed. In this model, project benefits to private sector include revenue and other benefits, such as raising company reputation and future opportunities.
Model Assumptions
1. For a project with \$I (in current dollar value) investment budget, the private sector invests \(a\) (in %) and the public sector invests \(1-a\) of total investment, including project budget and increased cost if there is project cost overrun.
2. To private sector, the value of expected project cost overrun ratio is \(E\gamma\).
3. \(\gamma\) is assumed to be a function of \(a\), which means the project cost overrun ratio would be influenced by how much the private sector is involved in the project. This is based on our assumption that more investment may strengthen the power of the private sector management in projects and make private sector more responsible, which will induce the private sector to better control project costs. As we have discussed, private sector’s involvement in transportation projects may improve project cost control, as long as some conditions, such as optimal project risk allocation, are satisfied. We have concluded that neither purely privatization nor no private involvement of transportation projects should not be the optimal way to control project cost. So, we assume that with an increased private investment proportion, expected project cost overrun ratio will first decrease, and then increase, which is \(\frac{\partial^2 E\gamma(a)}{\partial a^2} > 0\).
4. The private sector will make a return of \(R\) (in current $ value). As we have discussed, involvement of private sector may bring market advantages to projects and improve project demand/revenue management. So, we assume \(R\) could be expressed as a function of \(a\). Since reasonable private companies will not invest without being paid, we assume that the larger the private investment proportion, the greater the return to the private sector, which means \(\frac{\partial R(a)}{\partial a} \geq 0\). However, it is unlikely for private sector to obtain an increasing marginal benefit for its investment in transportation projects. We assume that when private investment proportion increases, marginal private benefit to investment proportion will decrease, \(\frac{\partial^2 R(a)}{\partial a^2} < 0\). When \(a\) increases, \(\frac{\partial R(a)}{\partial a}\) decreases until it reaches zero.

According to assumptions 3 and 4, for increasing \(a\), private sector benefit would increase at a decreasing rate, while private sector cost would first decrease and then increase. So, we expect project profit to the private sector would first increase and then decrease as \(a\) increases, potentially subjective to a maximum.

For the private company, its target is to maximize its profits.

\[
\max_a R(a) - \alpha I \times \gamma(a), \quad 0 < \alpha < 1, \gamma(a) > 0
\]

The First Order Condition for private company profit maximization function would be:

\[
\frac{\partial R(a)}{\partial a} - I \times \gamma(a) - \alpha I \frac{\partial \gamma(a)}{\partial a} = 0,
\]

---

\(\gamma = \text{Project Cost/ Project Budget}; \text{both cost and budget are in current } $\text{ value.}\)
Since $\alpha l > 0$, 
When $\frac{\partial R(\alpha)}{\partial \alpha} > I \times E[\gamma(\alpha)]$, $\frac{\partial E\gamma(\alpha)}{\partial \alpha} > 0$. .......................... (4.2) 
When $\frac{\partial R(\alpha)}{\partial \alpha} < I \times E[\gamma(\alpha)]$, $\frac{\partial E\gamma(\alpha)}{\partial \alpha} < 0$ .......................... (4.3) 

Intuitively, $\frac{\partial R(\alpha)}{\partial \alpha}$ refers to the marginal benefit of the private investment proportion. $I \times E\gamma(\alpha)$ is the expected project cost for a certain proportion of private investment. $\frac{\partial E\gamma(\alpha)}{\partial \alpha}$ is the expected marginal cost overrun ratio of the proportion of private investment.

According to Equation 4.2, for every percentage that the proportion of private sector investment increases, if the private company can achieve marginal benefits greater than the value of expected project costs, more involvement of private company will increase expected cost overrun ratio. One possible intuitive explanation for equation 4.2 is that when the marginal benefit of private investment proportion exceeds expected project cost, the private company may not be induced to control project costs, since increasing project costs may bring more profit to the private sector. In this case, more involvement of private sector will increase expected project cost overrun ratio.

Once provided with marginal benefit greater than expected project cost, private sector may be over-paid for the risks they actually take. Private sector may have an incentive to increase its investment proportion to gain more profit, as long as its marginal benefit is larger than expected project cost. Moreover, private sector may take risks that they may not be able to manage. So, project cost overrun risk, in terms of the expected project cost overrun ratio, may increase as private sector investment proportion increases.

Now we have one necessary condition for cost-efficient private sector involvement in transportation projects, which is that the marginal benefit of private investment proportion to the private sector should not exceed the expected project cost. One important point is that the marginal benefit to the private sector could be much larger than the marginal revenue. For a newly launched company, for example, an important benefit may be the building of its reputation in the industry and future opportunities. Such project benefits are hidden from the contract. Yet they are important considerations for investment and project management strategies including project cost control strategies. The public sector should investigate the marginal benefits the private sector expects from projects and pay special attention to those associated with large potential marginal benefits. Terms related to this condition should be included in PPP contracts to ensure that the involvement of private sector indeed exerts positive influence on project cost control.
3.4. Regression Model

3.4.1. Objective and Importance

Despite their uniqueness, transportation projects are influenced by common risk factors such as social interest rate changes and project delays. We propose this regression model to test and quantify the impacts of these transportation risk factors on the project cost overrun ratio.

Quantifying the sensitivities of the project cost to these risk factors helps project managers to obtain a better view of the impacts of these risk factors, to improve their ability to estimate the likelihood of cost overrun risk and to implement proper risk management strategies at early stages.

Furthermore, quantification of the sensitivity of project costs to these risk factors is meaningful to governments. Governments are generally the major investors in transportation infrastructure projects. Improved understanding on how risk factors, such as interest rate changes affect project costs help governments improve their economic development and fiscal strategies. For example, provided with significant project cost sensitivity estimate to the interest rate fluctuation, the governments can adjust interest rates to favorable levels to control transportation project costs.

Since transportation projects are often financed using local, provincial or federal taxes, improved estimation and management of project cost overrun risk are beneficial to the society in general.

Probability distribution modeling methods, such as Distribution Fitting, cannot provide quantitative analysis of the impacts of risk factors on project cost. A regression model, designed to analyze the impacts of factor variables on the target variable, can be used as a supplementary tool for project cost overrun risk analysis.

3.4.2 General Introduction of Risk Variables for the Model

This regression model is based on risk source analysis in the literature review. The model uses historical data from transportation projects. In addition to their expected effects on project cost, the following risk variables are chosen in concerns of mathematical modeling possibility and data collection convenience. In this section, we discuss these risk variables in general. Detailed risk variable specification and analysis are discussed in the next section.

The dependent variable is the project cost overrun ratio, calculated as project cost divided by project budget\(^{14}\).

The independent variables are:

1. \textit{Average annual interest rate change ratio over construction period}

\(^{14}\) Both project cost and budget are in current $ values.
The interest rate change ratio is calculated by dividing next-year interest rate by current-year interest rate. This variable is used to test the impact of annual interest rate changes on project costs. Changes in the interest rate may increase project costs, especially if project financing is subjective to floating interest rates.

2. Annual interest rate variability

Standard deviations of annual interest rates during the project period are used to quantify the impact of annual interest rate fluctuations on project cost overrun risk. In contrast to the annual interest rate change ratio, whose log values show negative and positive interest rate changes, the standard deviation of interest rate during the project period is always positive and depicts interest rate volatility only.

3. Project construction duration

Project construction duration is the period from the time the project decision is made to the time construction is finished and the project is ready to use. Project construction duration influences other project risk factors. For example, construction delay may increase the chance that input material prices will increase.

4. Percentage Investment of total project funding by private sector

We assume that more participation by the private sector in projects may improve project cost control at first, and that this effect will diminish after private investment proportion reaches a certain level. Here, the assumptions are the same as they were in the PPP cost overrun risk analysis model.

5. Percentage Investment of total project funding by federal government

The federal governments generally use tax revenue for project funding and are assumed to be less sensitive to individual project results than private companies and local governments. This variable is used to estimate project cost sensitivity to investments by federal governments in transportation projects.

Transportation construction projects may have both public and private investors. The public investors may include local, provincial and/or federal governments. Here we assume a federal government would not fund the project without local government investment, which means the investment ratio of local government would be greater than zero. So, the sum of investment proportions for private sector and federal government would always be less than one.

6. Cost overrun risk bearer

Private sector is generally regarded as better at risk management. Project capital costs are expected to be better controlled when the private sector is responsible for them. Cost overrun risk bearer variable is used to examine whether this assumption is true and to quantify its impact on project cost overrun risk. We assume that it is the investors responsible for project increased cost who bear project cost overrun risk. There are three scenarios for cost overrun risk bearer: cost overrun risk is born by private sector alone, public sector alone or both public and private sector. Two dummy variables are used to analyze these three scenarios.

---

15 In Flyvbjerg's model, this variable is called project implementation phase.
Table 8. Cost Overrun Risk Bearer Dummy Variable Values under Different Scenarios

<table>
<thead>
<tr>
<th>Cost Overrun Risk Bearer</th>
<th>Dummy Variable Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenarios</td>
<td>Private risk Bearer</td>
</tr>
<tr>
<td>Private</td>
<td>1</td>
</tr>
<tr>
<td>Public</td>
<td>0</td>
</tr>
<tr>
<td>Public and Private</td>
<td>0</td>
</tr>
</tbody>
</table>

Transportation project cost overruns are often caused by project cost underestimation. One reason for cost underestimation may be project uncertainties such as delays. This kind of cost underestimation is caused by variations of project risk factors including the risk variables in our regression model.

Another reason for cost underestimation is “biased” or “delusive” behavior of project evaluators whose activities are influenced by project sponsors and their interests. As discussed in Section 2.4, Literature Review, investors may regard their own investment, instead of those from all investors, as project costs. So, investors may benefit from projects even though the project net present value is negative. Employed by project sponsors to carry out project evaluation, project evaluators are likely to consider the interests of project sponsors to be their first concern and project costs may be underestimated. This kind of cost underestimation is due to human behavior and is difficult to analyze in a regression model. Objective numerical estimates from probability distribution and regression analyses can be helpful to minimizing this problem. For example, if every one year delay in project construction is expected to increase project costs by 4%, the project evaluators would have difficulty assuming that project costs will not change when projects are delayed.

3.4.3 Variable Specification and Analysis

When we discuss project cost overrun risk, we mean both likelihood and magnitude of project cost overrun. However, what most interest project evaluators and investors may not be whether cost overrun will occur, but how large the overrun ratio will be. Because of their large investment requirements, transportation infrastructure projects can incur millions of dollars in extra costs with just moderate cost overrun ratios. So, our regression model considers the size of the cost overrun in transportation infrastructure industry.

The projects of transportation modes (road, highway, rail, bridge, tunnel, etc.) may not have the same cost sensitivities to risk factors. For example, for road projects, one-year construction delay may not have the same effect on project cost overrun ratio as the same delay in a bridge or tunnel project. Bridge and tunnel projects generally face higher construction risks and technical requirements. The cost sensitivity to construction delay is likely to be larger for bridge or tunnel projects than for road projects.

Different project size may lead to different estimated regression models due to their differences in technologies, cost structures and investment structures. For example, a 1% increase in the private investment ratio for a project with a budget of $10,000 will have different implications to the private investors from the implications of a similar increase
to a project with a budget of $1 billion. It affects the incentives of private investors to control project costs.

So, a regression model should use data from projects which are similar in mode, type and size.

**Project cost overrun ratio variable**
The project cost overrun ratio should always be positive. If we treat this variable as the dependent variable in the regression model and use the estimated model for the prediction of cost overrun ratio, we may get negative predicted values of the project cost overrun ratio. To avoid this problem, we use log values of project cost overrun ratio variable as the dependent variable for our regression model. When project costs exceed project budget, the log value of the cost overrun ratio will be positive; otherwise, it will be negative.

**Project construction duration variable**
According to Flyvbjerg’s model and data, the project duration variable is shown to be linearly correlated with cost escalation variable. However, the present value of project cost is calculated as 

\[
P(V) = \sum_{i=0}^{T_i} \frac{C_{i,t}}{(1+r)^t} 
\]

- \(T_i\) = Number of project construction years for project i
- \(t\) = Project construction year index from construction starting time (year 0) to finish (year \(T\))
- \(r\) = Project discount rate
- \(i\) = Project index assigning a number to each project in database

According to this project cost present value calculation formula, project duration is likely to have a non-linear relationship with project cost. As discussed in the conditional cost overrun probability model, projects can be divided into several evaluation stages. For each stage, the project may have different cost structures and marginal cost increases per year. For example, the cost structure and marginal cost increase per year in the construction period of a road bed may not necessarily be the same as those of a road facility (for example, street plate, road painting, traffic light, etc.). So, we assume a positive, non-linear relationship between project cost overrun ratio and project duration.

**Interest Rate Variables**
The average annual interest rate change over the construction period is

\[
\frac{\sum_{i=1}^{T_i} r_{i,t+1}/r_{i,t}}{T_i} 
\]

- \(i\) = Project index assigning a numerical number to each project in database
- \(T_i\) = Number of project construction years for project i
- \(t\) = Project construction year index from construction starting time (year 0) to finish (year \(T\))
- \(r_{i,t}\) = The interest rate for project i at year t
This formula shows not only the magnitude of interest rate changes over the construction year but also the direction (increase or decrease). A value which is greater than one indicates an increase. As we have seen in section 5.4, Literature Review, unfavorable changes in interest rates will cause project financing costs increase.

The annual interest rate variability variable can be estimated by the standard deviation of annual interest rates during project periods.

Two main factors affecting interest rates are inflation and government monetary policy change. Inflation increases goods prices over the economy, including the capital goods, which are priced in interest rates; Governments may use interest rates as a tool to achieve goals of monetary or fiscal policy. For example, government may lower interest rate to encourage investments.

According to the project cost present value calculation formula above, changes in interest rates will cause non-linear changes in the project cost present value, no matter whether such changes are caused by inflation or by government monetary policy. Nominal project costs at each time period are discounted by project discount rate powered with the number of time periods. A one-percent increase in the discount rate would have a different effect on the PV of project cost in period 1 from that in period 5. If the interest rate change is caused by inflation, both interest rates and project costs will change, since inflation will change the prices of project inputs.

In our models, interest rate change variables are assumed to be non-linearly correlated with the project cost overrun variable. Log values of the variables are used to test these relationships. The interest rate change variables are expected to be positively correlated with the project cost overrun ratio.

**Private investment percentage variable**

Instead of using ownership dummy variables, which are used in Flyvbjerg’s model, our regression model is interested in the effects of respective proportions of private sector and public sector investment on project cost overrun. As discussed, classifying project investment structure simply as private or as public may lead to misleading conclusions, since what really influence project cost control are the risk sharing structure and obligations of each party.

It is reasonable to expect that a larger percentage of private investment would increase the influence of the private sector on transportation projects and so improve project efficiency. A larger percentage of investment entitles private companies to have more control over project decisions and management. With larger investments, private companies have more incentive to actively participate in projects. So greater involvement of private sectors would help control cost overrun risk. However, as we have assumed in the PPP cost overrun risk analysis model, when the private investment proportion reaches a certain level, further increase in private investment may not bring further cost efficiency. The private sector may take more risks than it can handle well, for which it will seek large risk premiums. Moreover, improper risk allocation will increase project cost
overrun risk. So, we assume that when private investment reaches a particular level, further private investment would increase project cost overrun ratio. We assume a non-linear relationship between private investment proportion and project cost overrun variable.

Federal government investment percentage variable
The involvement of a federal government in transportation projects is an interesting topic. The key question is how local governments and other project sponsors view the role of federal government. Although federal government funding would be considered as a part of project total cost in project cost benefit analysis, local governments or private sponsors may regard this funding as a subsidy from the federal government for local transportation improvement. To obtain this subsidy, the local government may be induced to over-optimistically estimate project benefits to get it approved and funded.

Normally, investors are regarded as risk averse, which means facing a game with a half chance to double his property and a half chance to lose his property, an investor would choose not to play the game. However, in the case with federal government funding, local governments seem to be no more risk-averse. Suppose there are two projects with the same total investment amounts and same expected net present values, and project A has a higher probability of project failure than project B. The present values of investments and benefits (in Millions of dollars) by each investor for each project are:

\[ I_{A,L} = 3, \quad I_{A,P} = 4, \quad I_{A,F} = 3 \]
\[ I_{B,L} = 6, \quad I_{B,P} = 4 \]
\[ R_{A,L} = R_{B,L} = 7 \]

A = Project index for project A
B = Project index for project B
L = Investor index for local government
P = Investor index for private sector
F = Investor index for federal government
I = Investment
R = Project Benefit

Which project will the local government prefer? The answer would highly probably be project A despite its higher failure probability. Does this mean local government is risk seeking, instead of risk averse? We would suggest that the expected net present values of these two projects in society’s view are not what they are in the local government’s view. With $3 million of federal government investment in project A, the “real” cost to local government is $3 million. Using this “real” cost as its project cost, project A has a higher expected net present value to the local government.

\[ NPV_{A,L} = 7 - 3 = 4 \]
\[ NPV_{B,L} = 7 - 6 = 1 \]

If project A’s net benefit is large enough to compensate for its higher risk, local government would prefer project A to project B. As long as the risk premium\(^{16}\) that local

---

\(^{16}\) Risk premium is the compensation investor would ask for taking more risk.
government obtains for project A is less than $NPV_{A,L} - NPV_{B,L} = \$4 - \$1 = \$3$ million, the local government will choose project A.

Furthermore, the federal government often does not have a reliable and complete inspection and management system for the projects, which makes financial abuses and poor management possible. We suggest that the greater involvement of the federal government would increase cost overrun risk.

Similar to the relationship between the proportion of private investment and project cost overrun, federal government participation in the project may not be linearly related to project cost overrun. It is unlikely that increasing federal government investment from 10% to 11% will increase project cost overrun at the same proportion as increasing it from 80% to 81%. We assume a non-linear relationship between the proportion of federal government investment and project cost overrun.

**Project cost overrun risk bearer variable**

Two dummy variables are used for the project cost overrun risk bearer. One is private cost overrun risk bearer, which equals one when the private sector bears capital cost overrun risk completely and zero otherwise. Similarly, the public-private cost overrun risk bearer variable equals one when both the public and private sectors bear capital cost overrun risk and zero otherwise. See Table 8 for the three cost overrun risk bearer scenarios. The key issue here is to define the risk bearer. The cost overrun risk taker is the party who is responsible for making additional investment when project costs exceed budget. These two dummy variables are helpful for examining the assumption that project costs can be better controlled when cost overrun risk is properly allocated. The PPP contracts provide information on private and public investment ratios in the budget and increased cost when project cost exceeds budget.

So, the regression model for analyzing the impacts of these risk factors would be:

$$
\log Y = \beta_0 + \beta_1 \log X_1 + \beta_2 \log X_2 + \beta_3 \log X_3 + \beta_4 \log X_4 + \beta_5 \log X_5 + \beta_6 X_6 + \beta_7 X_7
$$

- $Y = \frac{\text{ActualCost}}{\text{Cost Budget}}$
- $X_1 = \text{Average annual interest rate change over construction period}$
- $X_2 = \text{Annual interest rate variability}$
- $X_3 = \text{Project construction duration}$
- $X_4 = \text{Investment ratio of private sector in total project funding}$
- $X_5 = \text{Investment ratio of federal government in total project funding}$
- $X_6 = \text{Private cost overrun risk bearer}$
- $X_7 = \text{Public Private cost overrun risk bearer}$
- $\beta_0 = \text{Constant term}$

We have discussed on the non-linear relationships between the dependent variables and the numerical independent variables. To test and estimate these non-linear relationships, we use Log values of these numerical independent values in this regression model.
\[ \beta_i = \text{Regression coefficients of each independent variable} \]

We do not have the data to test and estimate this regression model. However, data used in this regression model are not difficult for project evaluators to collect. This model can be used in practice to test and quantify the impacts of these risk factors on project cost overrun. Furthermore, it would help the project evaluator predict the cost overrun ratio for a new project.

3.4.4 Incorporation of Cost Overrun Risk into the Project Evaluation Process

First of all, our question is: Can we use the conclusions from our regression model for the prediction of a new transportation project cost overrun risk?

The regression model we propose tries to test and estimate effects of some common risk factors on transportation project cost overrun. This is based on our understanding that despite the uniqueness of each transportation project, transportation project costs are influenced by common risk factors. This regression model is very likely to have an \( R^2 \) value less than 0.5, which means the regression model explains less than half of the variance of the project cost overrun ratio. However, this does not mean the models are meaningless for project cost overrun risk analysis. Low \( R^2 \) values show that project cost overrun is influenced by some other unknown factors or is due to the uniqueness of projects. The seven independent variables in our regression model definitely cannot represent all risk factors for project cost overrun. We use them for modeling because we assume they have influences on project cost overrun and their data are relatively easy to gather and analyze. We are interested in estimating the effects of these common risk factors on project costs. We call them common risk factors not only because they are risk factors shared by all transportation projects but also because their data are relatively easy to collect. So, estimations of their influence have empirical meaning to a new project sharing these common risk factors.

Projects of different types use different technologies and have different cost structures. Even different sized projects of the same type may use different technologies. So, we suggest using data from similar sizes and types of projects. If data are available, it would be interesting to add some dummy variables to the regression model to test and estimate the influences of project technologies and cost structures on project cost. For example, to test and estimate the influence of technology A on project cost overrun ratio, we can set a dummy variable \( T_A \), where

\[ T_A = \begin{cases} 0 & \text{if the project does not use this technology} \\ 1 & \text{if the project uses this technology} \end{cases} \]

We can use the regression model to analyze technology A’s influence on project cost overrun ratio.

Another factor that may make these regression models infeasible is the macroeconomic and policy environment. Flyvbjerg’s study shows that the average transportation cost overrun ratio seems to be different among areas that have different macroeconomic and policy environments. So, data for these regression models should be specific to location. In Canada, provincial governments control most transportation projects and have project
cost tracking systems. We assume the economic and political environment is reasonably stable between successive transportation projects in the same province. Using data from projects of similar size and type done in the same specific province may exclude the majority of project cost overrun variation caused by technology, cost structure and economic and political factors when regression models use data from similar sized projects, of the same project type, and in the same province. The residuals of these regression models can be assumed to represent some random variation attributable to project uniqueness.

It would be reasonable for us to assume that the influences of these common risk factors on project cost overrun are not very different from each other on projects with the same technological, financial, economic, political and geographic characteristics. This is the basic assumption for using the estimated regression model to predict the cost overrun ratio for a new project. Actually, this assumption is used in practice when project evaluators use their experiences from similar projects to estimate project costs and benefits for a new project.

However, there is no guarantee that the predicted value of the project cost overrun ratio will be exactly equal to the real project cost overrun ratio. As discussed, the uniqueness of project may cause large variation in the project cost overrun ratio. So, we need some statistical estimation in a probability range for it.

**Prediction Interval**

Based on the regression model, we estimate the prediction interval for the dependent variable. The prediction interval gives the range of dependent variable for a particular set of independent variable values, within which there is a certain probability (95%, for example) that the dependent variable value will be, based on the fit of current data. As Wooldridge (2000) states, the prediction interval is a confidence interval for a predicted value.

A prediction interval for cost overrun for a new project can be computed as:

\[
\hat{Y}_{\text{new}} \pm t_{\alpha/2, n-p} \cdot s_{\text{new}}
\]

\[
s_{\text{new}}^2 = \frac{MSE \left[ 1 + \frac{1}{n} + \left( X_{\text{new}} - \bar{X} \right)^2 / \sum_{i=1}^{n} (X_i - \bar{X})^2 \right]}
\]

\[
\hat{Y}_{\text{new}} = \text{Predicted value of dependent variable for a new project}
\]

\[
MSE = \text{Mean Squared Error of the regression model}
\]

\[
s_{\text{new}}^2 = \text{Variance of the prediction error}
\]

\[
X_{\text{new}} = \text{The independent variable values of a new project}
\]

\[
\bar{X} = \text{Mean values of independent variables for regression model observations}
\]

---

18 In contrast, the 95% confidence interval is the variable range that has a 95% chance of containing the true regression line, which provides a useful way to assess the quality of prediction. Prediction interval is generally wider than confidence interval at the same probability level. This is because prediction interval not only includes the uncertainties or noises from current data, but also those from the new data.
In project evaluation, the net present value of the project is the key measurement.

\[ NPV = PV(B) - PV(C) = \sum_{t=0}^{k} B_t / (1+r)^t - \sum_{t=0}^{k} C_t / (1+r)^t \]

Instead of using a single value estimate of PV(C), we use the prediction interval for project cost overrun from our model to calculate a range for PV(C) with a certain confidence level. So, the NPV formula would become:

\[ NPV = PV(B) - PV(C) = PV(B) - \left[ \hat{Y}_{\text{low}} \pm t_{\alpha/2, n-p} \cdot s_{\text{low}} \right] \times \text{Budget} \]

For example, from the regression model, we may get the 95% prediction interval for the project cost overrun to be 20%-100%. Let the present value of the project budget be $100 million, we can say with 95% confidence that the expected present value of project costs will be between $120 million and $200 million. In words, the likelihood that project costs to be beyond this range is 5%. We can calculate a 95% prediction interval for project net present value using this project cost range.

Similarly, we can obtain 95% prediction interval for project net benefit, say $140 million to $220 million. If we can assume project benefits and project costs are independent of each other, we can find a 95% prediction interval for this project NPV, which is -$60 million to $100 million.

The question is whether we can make the assumption that benefit and cost are independent. Cost and benefit estimation in project evaluation are based mainly on the project design, including quality and safety issues. As long as the project design is not changed and there are no large fluctuations in macroeconomic environment, changes in project costs would have little impact on project benefits. So, it can be assumed that benefit and cost estimates in project NPV estimation are independent of each other as long as the social economy is reasonably stable and project design is not changed. Similar estimations can be applied to other project evaluation parameter such as cost-benefit ratio.

### 3.5. Vancouver Island Highway Project Case Study

#### 3.5.1. The Database

The Vancouver Island Highway Project (VIHP) involved upgrading the existing highways between Swartz Bay and Nanaimo, building a new parkway in Nanaimo and constructing a new inland highway from Parksville to Campbell River. The project included improvements to approximately 174 kilometers of highway and the construction of 146 kilometers of new highway. The VIHP was announced in 1993 and was substantially complete by 2003. This database is provided by Transport Canada office in British Columbia.
In this database, the VIHP is broken down into 163 independent projects, excluding the projects not in initial VIHP plan\(^{19}\) and two other projects with extremely high cost overrun ratios\(^{20}\). The projects consist of road and highway construction projects in different areas and individual bridge and tunnel projects. We assume that the project costs are independent of one another.

There are 127 road and highway construction projects and 36 bridge and tunnel projects in VIHP database, including their budget and cost information. Unfortunately construction beginning and ending times for each project are not available. Nor we have the investment and private involvement information. Due to inefficient data, our proposed multiple regression model can not be applied to the VIHP data. The cost overrun ratio probability distribution model and the simple regression model will be used to analyze cost overrun risk for the VIHP projects.

The costs and budgets are in 1993-dollar values. Cost overrun ratios are calculated as dividing project cost by project budget. There are 104 road/highway projects and 29 bridge/tunnel projects, each costing over $1 million. Table 9 gives aggregate cost data ranges for road/highway projects and for bridge/tunnel projects separately.

### Table 9. Cost and Cost Overrun Ratio Ranges of VIHP data

<table>
<thead>
<tr>
<th>Number of Projects with Cost Overrun</th>
<th>Number of Projects with Cost Overrun Ratio &gt; 1.5</th>
<th>Cost Min.</th>
<th>Cost Max.</th>
<th>Overrun Min.</th>
<th>Overrun Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road &amp; Highway Projects</td>
<td>104</td>
<td>3</td>
<td>$7,559</td>
<td>$88,076,460</td>
<td>0.48</td>
</tr>
<tr>
<td>Bridge &amp; Tunnel Projects</td>
<td>29</td>
<td>2</td>
<td>$7,592</td>
<td>$18,349,840</td>
<td>0.80</td>
</tr>
</tbody>
</table>

### 3.5.2 Cost Overrun Ratio Distribution Fitting

We discuss the probability distribution of the cost overrun ratio variable in the probability distribution model section. According to our analysis, the Beta Distribution fits our cost overrun ratio data. Figure 5 shows histograms of the cost overrun ratio for road/highway construction projects and for bridge/tunnel projects using the VIHP data.

\(^{19}\) These projects had zero budgets.
\(^{20}\) One is a road construction project, whose cost was about 14 times its budget; the other is a highway construction project, whose cost was about 42 times its budget.
According to the histograms, cost overrun ratio data are centered just higher than one, where project costs equal project budget, with a few observations reflecting high cost overrun ratios and skewing the histogram to right. These histograms confirm our expectation that most projects control their real costs reasonably well; however, in some cases, real costs can be much higher than budgets.

As shown in Table 10, the respective descriptive statistics for road/highway projects and for bridge/tunnel projects are close to one another.

<table>
<thead>
<tr>
<th></th>
<th>Road/Highway Projects</th>
<th>Bridge/Tunnel Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>1.059</td>
<td>1.052</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>1.055</td>
<td>1.055</td>
</tr>
<tr>
<td><strong>Mode</strong></td>
<td>1.045</td>
<td>1.059</td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>0.27</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Bridge and tunnel projects are generally expected to face more complicated technical and construction conditions, which make their costs more difficult to control. However, in the VIHP, the average cost overrun ratio and standard deviation of bridge and tunnel projects are even lower than those of road and highway projects. One possible reason may be that the bridge and tunnel projects in VIHP are relatively small, with costs mostly below $8 million. These bridge and tunnel projects are mostly supplementary projects connecting road and highway in VIHP.

In our distribution fitting model, we do not separate road and highway projects from bridge and tunnel projects because of their close descriptive statistics and insufficient data for bridge and tunnel projects. However, if data were sufficient, we would recommend setting different probability models for each transportation project type separately. We set zero as the Beta distribution lower bound and 2.35 as the upper bound.

21 See Table10 for number of projects with cost overrun ratio larger than 1.5.
bound to fit the VIHP data range, whose maximum cost overrun ratio data point is just lower than 2.35.

Distribution shape parameters, P and Q, are calculated using maximum likelihood estimation (MLE) method. MLE is an analytical maximization procedure. Likelihood refers to the probability of obtaining sample data, given a particular distribution model. The values of the model parameters that maximize the likelihood of certain sample data are the Maximum Likelihood Estimates.

Table 11. Beta Distribution Parameter Setting and Estimates

| Minimum (A) | 0 |
| Maximum (B) | 2.35 |
| P MLE | 9.02 |
| Q MLE | 11.05 |

Note: A and B values are set manually, based on cost overrun ratio data range of VIHP; P and Q values are estimated using MLE method in NCSS.

Figure 6. Beta Distribution PDF Plot for Cost Overrun Ratio in the VIHP

![Beta Distribution PDF Plot](image6)

Figure 7. Beta Probability Plot for Cost Overrun Ratio Variable

![Beta Probability Plot](image7)

22 For statistical analysis in this case study, we use NCSS, which requires setting Beta distribution value parameters manually. The upper bound should be not much larger than the maximum value in the sample. Otherwise it may result in a beta distribution not fitting the sample data well. The Beta distribution value parameters can also be calculated using MLE.

23 See Appendix 2
According to the beta probability plot in Figure 7, most data fit this beta distribution well, while some observations deviated from model estimates. The model would be good at estimating probability for project cost overrun ratio to exceed a certain value, which is less than 1.25. However when estimating the probability for project cost overrun ratio exceeding some value over 1.25, the model is very likely to give a conservative estimate. It means model probability estimate for a project to have a cost overrun ratio exceeding, for example, 1.5, may be an underestimation and should be used as a conservative estimate for project cost overrun risk. This conservative project cost overrun ratio estimate is caused by the statistical methods used for fitting distribution, instead of data pattern. In Beta distribution fitting, project data are divided into quartiles according to their values. The statistical Beta fitting model focuses on maximizing the likelihood of obtaining data from Beta quartile 25% to 75%. For VIHP database, 75% Beta quartile is at the point with cost overrun ratio equal to 1.25. The cost overrun probability estimations over this point are likely to be underestimated.

<table>
<thead>
<tr>
<th>For Cost Overrun Ratio</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 0.2</td>
<td>99.99%</td>
</tr>
<tr>
<td>&gt; 0.4</td>
<td>99.79%</td>
</tr>
<tr>
<td>&gt; 0.6</td>
<td>96.71%</td>
</tr>
<tr>
<td>&gt; 0.8</td>
<td>83.68%</td>
</tr>
<tr>
<td>&gt; 1.0</td>
<td><strong>57.82%</strong></td>
</tr>
<tr>
<td>&gt; 1.2</td>
<td>28.84%</td>
</tr>
<tr>
<td>&gt; 1.4</td>
<td>9.34%</td>
</tr>
<tr>
<td>&gt; 1.6</td>
<td>1.68%</td>
</tr>
<tr>
<td>&gt; 1.8</td>
<td>0.12%</td>
</tr>
</tbody>
</table>

According to this fitted beta distribution, the probability of road and highway construction project costs exceeding their budgets is 57.82%; the probability is 0.12% that a road or highway project cost exceeding 1.8 times of budget. However according to the Beta probability plot in Figure 7, this estimate may be underestimated. This model would be better to be used as providing conservative estimates of cost overrun for road, highway, and small bridge or tunnel construction projects.

### 3.5.3 Monte Carlo Simulation Model Results vs. Beta Fitting Model Results

Using Monte Carlo simulation\(^{24}\), we can get an estimated beta distribution (6.99, 8.91, 0.15, 2.22) for the project cost overrun ratio variable in the VIHP. Table 13 presents comparisons on the beta distribution parameter estimates by both a simulation tool (Arena) and a statistical tool (NCSS). Figure 8 provides graphical comparison on PDF patterns between Monte Carlo simulation and Beta Fitting method.

\(^{24}\) Using R, 5000 data points are generated for the estimated Beta distribution (9.02, 11.05, 0, 2.35) fitted with VIHP data. Then we use this generated new database to fit a beta distribution in Arena. Please read Appendix 4 for analysis summary report.
Table 13. Beta Distribution Parameter Estimates Comparison for Cost Overrun Ratio Variable between Monte Carlo Simulation Method and Beta Fitting Method

<table>
<thead>
<tr>
<th></th>
<th>Monte Carlo Simulation</th>
<th>Beta Fitting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum (A)</td>
<td>0.04</td>
<td>0</td>
</tr>
<tr>
<td>Maximum (B)</td>
<td>2.00</td>
<td>2.35</td>
</tr>
<tr>
<td>Shape Parameter1 (P)</td>
<td>7.18</td>
<td>9.02</td>
</tr>
<tr>
<td>Shape Parameter2 (Q)</td>
<td>6.73</td>
<td>11.05</td>
</tr>
</tbody>
</table>

Figure 8. Probability Density Function Comparison between Monte Carlo Simulation Method and Beta Fitting Method

In Figure 8, the distribution PDF from the Monte Carlo simulation model has a smaller range (higher minimum value A and lower maximum value B), a higher mode, as compared with the Beta Fitted distribution PDF. As discussed in the Probability Distribution Model section, this may be due to that the computer cannot generate truly random values.

Despite the slight differences in the shapes of their PDFs, the probability distribution simulated by Monte Carlo method gives similar estimates on project cost overrun probabilities. Table 14 gives the numerical comparison.

Table 14. Cost Overrun Probability Estimation Comparison between Monte Carlo Method and Beta Fitting Method

<table>
<thead>
<tr>
<th>For Project Cost Overrun Ratio &gt;</th>
<th>Monte Carlo Simulation</th>
<th>Beta Fitting Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>98.7%</td>
<td>99.0%</td>
</tr>
<tr>
<td>0.75</td>
<td>87.6%</td>
<td>88.2%</td>
</tr>
<tr>
<td>1</td>
<td>57.9%</td>
<td>57.9%</td>
</tr>
<tr>
<td>1.25</td>
<td>22.8%</td>
<td>22.8%</td>
</tr>
<tr>
<td>1.5</td>
<td>3.7%</td>
<td>4.3%</td>
</tr>
<tr>
<td>1.75</td>
<td>0.1%</td>
<td>0.3%</td>
</tr>
<tr>
<td>2</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>2.25</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

45
According to Table 14 and as compared to the Beta Fitting model, the Monte Carlo simulation model gives equal or slightly lower estimates of probabilities of project cost overrun. We have discussed that the Beta Fitting Model provides understated cost overrun probability estimates for cost overrun ratios greater than 1.25 so that its estimation should be regarded as a lower bound. Unfortunately, Monte Carlo Simulation seems not to be a tool for providing upper-adjusted estimation on project cost overrun probabilities. On the contrary, it seems to further underestimate project cost overrun probabilities as shown in Table 14. Project cost overrun probability estimation using Monte Carlo simulation is likely to give underestimated estimates.

The Beta Fitting Method and the Monte Carlo Method provide project cost overrun probability estimates. However, they can not explain why overrun may occur and how those cost overrun causes would influence project costs. They are not feasible tools if project managers want to examine the exact factors for project cost overrun in order to control project costs. For this purpose, we use regression models to test and analyze the impacts of risk factors on project cost. Unfortunately, we have only cost and budget data for each project in the VIHP database. In the following sections, we use project budget to represent project size and analyze its impact on project cost.

3.5.4 Project Cost Overrun vs. Project Size in the VIHP

Large transportation projects attract attention for their huge funding requirements. If the costs of large projects are more likely to be out of control, project sponsors should put extra efforts on controlling the associated risks.

Project size can be represented by project construction time, budget and cost. Unfortunately, the VIHP database provides only project budget and cost information. In Flyvbjerg’s study, project budget is used as the variable to represent project size. Project budget alone may not be a good indicator of project size, when underestimation of cost is purposely made at the project evaluation stage. Even without ‘delusive’ human behavior, project evaluators may overestimate or underestimate real project size without full information about construction conditions, such as the conditions of rock layers in tunnel projects. On the other hand, project cost alone may also provide misleading information about project size, if legal, environmental or political risks materialize. For example, new environmental regulations requiring design or construction changes may cause project cost overrun. In this regression analysis, we assume there is no ‘delusive’ human behavior and project evaluators are able to reasonably estimate project cost. Based on these assumptions, we use project budget to represent project size.

Road/Highway (R/H) Projects

First of all, we look at the scatter plot showing how project costs change as project budget increases. In Figure 9, project cost equals to project budget on the straight blue line; the green dots are the data points from the VIHP database. According to this figure, costs for larger projects deviate more from the Line. Although not obvious, the data show a non-linear relationship between project cost and budget. As discussed earlier, we use log values of project cost and budget and test for a linear relationship between them.
Using 1.05 as the log base\(^{25}\), we test the linear relationship between log values of project cost and budget. The model is:

\[
\text{Log}Y = \beta_0 + \beta_1 \text{Log}X + \varepsilon
\]

\(Y\) = Project cost  
\(X\) = Project budget  
\(\beta_0\) = Constant term  
\(\beta_1\) = Regression coefficient  
\(\varepsilon\) = Error term

Figure 10 shows non-constant cost variance for different project size. For log (Budget) values below 250, the VIHP data fit the model well. For log(Budget) values between 250 and 290, VIHP data show high variances; For log(Budget) values above 290, VIHP data

\(^{25}\) Log bases from 1.01 to 1.5 do not cause much difference in data pattern for project cost and budget. Log bases above 1.5 generate too few data points for the regression model. So we choose 1.05 as the log base.
show reasonably constant variances. Based on this data pattern, we divide VIHP road and highway into three groups: small (project budget lower than $250,000), medium (project budget between $250,000 and $1 million), large (project budget over $1 million). Unfortunately, we do not have engineering and technical backgrounds for these projects and cannot explain whether this division is due to project technology or other differences related to engineering.

Small Road/Highway Projects

Figure 11. Project Cost vs. Budget Scatter Plot for Small R/H Projects

According to linear regression assumption tests\(^2^6\), the model assumptions, such as linearity, homoskedasticity and so on, are satisfied. The ANOVA table shows that the model is statistically significant (P value = 0.0000). We estimate the model for small road and highway projects to be:

\[
\text{Log(Cost)} = -0.982 + 1.006 \text{Log(Budget)}
\]

Therefore approximately, \(\text{Cost} = 0.953 \text{Budget}^{1.006}\), as shown in Figure12.

\(^2^6\) See regression analysis report in Appendix 3
The cost increase associated with each $10,000 budget increase increases as budget increases. This is shown in Figure 13. The costs of larger project costs are more likely to go out of control than the costs of smaller road and highway projects.

Figure 13. Cost Increase for Each $10,000 Project Budget Increase for Small Road/Highway Projects
Medium Sized Road/Highway Project

Figure 14. Log (Cost) vs. Log (Budget) Scatter Plot for Medium Sized R/H Projects

The linear regression assumption tests of this model show that the assumptions are satisfied. The ANOVA table shows that the model is statistical significant (P value= 0.04). We estimate the model for medium sized road and highway projects to be:

\[ \log(\text{Cost}) = 105.068 + 0.62 \log(\text{Budget}) \]

And therefore \( \text{Cost} = 168.39 \cdot \text{Budget}^{0.62} \). Figure 15 shows how project costs increase with project budget for medium-sized road and highway projects.

Figure 15. Cost vs. Budget for Medium Sized Road/Highway Projects

As shown in Figure 16, the cost increase associated with each $50,000 budget increase decreases as budget increases. The larger projects are subject to less cost overrun risk than are smaller road and highway projects.
Large Road/Highway Project

Model assumption reports show constant variance assumption and linearity assumption are satisfied; however the residuals do not follow a normal distribution. The residual histogram shows a lightly skewed pattern. The violation of the normality assumption may influence the accuracy of the F-ratio of the model. However, the normality assumption "plays no role in the unbiasedness of OLS" (Wooldridge, 2000). Therefore it is not necessary to discard the t-statistics or the F ratio. In the ANOVA table, we find extremely low P-values. Therefore we would assume the violation of the normality assumption does not invalidate this model. The estimated model is:

\[ \text{Ordinary Least Square method used in linear regression analysis} \]
\[ \log(Cost) = 1.657 + 0.995 \log(Budget) \]

And therefore \( \text{Cost} = 1.084 \cdot \text{Budget}^{0.995} \), as shown in Figure 18.

Figure 18. Cost vs. Budget for Large R/H Projects

Figure 19 shows that the increase in cost associated with each $500,000 increase in project budget decreases as budget increases. Similar to medium-sized road and highway projects, large road and highway projects face decreasing cost overrun risk as project size increases.

Figure 19. Cost Increase for Each $500,000 Project Budget Increase for Large Road/Highway Projects

These regression models show that project costs are non-linearly correlated with project budget. According to this analysis, when project budget increases, marginal project cost increases for small projects, and decreases for medium and large projects. For small projects with budgets less than $250,000, project evaluators are more likely to lose control of costs the larger the project budget. However, for projects with budgets over $250,000, project costs are better controlled. One possible explanation for this project cost control pattern is that project evaluators may tend to be more careful when...
estimating the costs of large projects. Another possible reason may be that when project costs are out of control on large projects, project managers may separate some components of the projects and title them as new projects. In the VIHP database, we see that some projects with zero initial budgets were added to the project list during the VIHP construction periods. Although we exclude these projects from our regression analysis, they may deliver some hidden information about the costs of large projects.

**Bridge and Tunnel Project Regression Models**

As we did in Road/Highway regression analysis, we look at the cost budget scatter plot for bridge and tunnel projects first. Although not obviously, the data reflect a non-linear relationship between project cost and budget for bridge and tunnel projects, as we found for road and highway projects. We take log values of project cost and budget and test for a linear relationship between these two new variables.

Figure 20. Cost vs. Budget for Bridge and Tunnel Projects

![Cost vs Budget for B/T Projects](image)

Figure 21. Log (Cost) vs. Log (Budget) for Bridge and Tunnel Projects

![Log (Cost) vs. Log (Budget) for Bridge and Tunnel Projects](image)

Y Axis: Log values of Cost for Bridge/Tunnel Projects
X Axis: Log values of Budget for Bridge/Tunnel Projects

This scatter plot in Figure 21 shows that four data points with log (Budget) values below 280 deviate more from the predicted than the other observations do. When we remove these four
small bridge and tunnel projects (with budget below $850,000) out and run a linear regression, we get a new scatter plot for log values of project cost over budget, as follows.

Figure 22. Log (Cost) vs. Log (Budget) for Large Bridge and Tunnel Projects

The model assumption reports show that the constant variance and linearity assumptions are valid; but again the residuals seem not to have a normal distribution. The histogram of the residuals shows a tail to right. As was true for large road and highway projects, the P-value for this model is negligible. We assume that the violation of the normality assumption does not invalidate this model. The model is estimated to be:

\[
\text{Log}(\text{Cost}) = -0.983 + 1.003 \text{Log}(\text{Budget})
\]

And therefore \( \text{Cost} = 0.953 \text{Budget}^{1.003} \), as shown in Figure 23.

Figure 23. Cost vs. Budget for Large Bridge and Tunnel Projects (Budgets over $850,000)

Figure 24 shows that the marginal project cost increases when budget increases for bridge and tunnel project with budget greater than $850,000. Project costs are more difficult to control or more easily underestimated for large bridge and tunnel projects.
Using the VIHP database, our regression models test the linear relationships between the log values of project cost and project budget for each project type. The models give different conclusions from Flyvbjerg’s, in which project budget is shown to have a linear relationship with the project cost overrun ratio for bridge and tunnel projects.

Since project cost overrun can be caused by various risk factors, as we discuss in the risk source analysis in the Literature Review and in the multiple regression model section, our models that are based on the VIHP database may not be applicable for all projects worldwide. We expect that they could be applied to projects with similar political, environmental, technology and financial characteristics. However, the analytical methods and techniques used in modeling the probability distributions and regression models can be used to analyze transportation project cost overrun risk for any project type in any part of the world. Furthermore, these models can be applied to other transportation project risk analyses such as revenue shortage risk analysis.

3.5.5 A Numerical Example of Predicting Project Cost for a New Project Using a Regression Model and a Beta Fitting Model

Using the estimated regression models, we can predict expected cost for a new project. However, project evaluators may find a prediction interval to be more useful for project analysis than the single estimated value of expected cost. What follows is an example of a prediction interval in project cost benefit analysis.

According to the estimated model for medium sized road and highway projects, a project with $1,200,000 budget ($\log_{10.05}\text{Budget} = 286.8986$) has an expected $\log_{10}\text{Cost}$ value that is estimated to be 287.1212 ($\log_{10.05}\text{Cost} = 1.657 + 0.995\log_{10.05}\text{Budget}$). The mean squared error (MSE) for this model is 3.57, the average log project budget value is 319.86, number of observations (n) is 100 and the number of the independent variable (p) is 2. $ \sum (X_i - \bar{X})^2 = 43643.51$. So
The 95% prediction interval for the log project cost value is (282.7457, 291.4967)\(^{28}\), corresponding to the project cost interval ($979,905, $1,501,793). With 95% confidence, we can conclude that the project costs for a road or highway project with a budget of $1,200,000 would be between $979,905 and $1,501,793.

Table 15. Expected Cost and Confidence Interval Estimation for a Project with Budget of $1,200,000

<table>
<thead>
<tr>
<th>Project Budget</th>
<th>Model Estimated Project Cost</th>
<th>95% Prediction Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1,200,000</td>
<td>$1,213,098</td>
<td>$979,905 - $1,501,793</td>
</tr>
</tbody>
</table>

We also present Table 16, project cost prediction intervals for various confidence levels. Figure 25 provides a graphic view of how the prediction interval changes for different confidence levels for this project. When the confidence level decreases, the prediction interval shrinks.

Table 16. Project Cost Prediction Intervals at Different Confidence Levels for a Project with Budget of $1,200,000

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Lower Prediction Bound</th>
<th>Upper Prediction Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%</td>
<td>$979,905</td>
<td>$1,501,793</td>
</tr>
<tr>
<td>80%</td>
<td>$1,038,162</td>
<td>$1,417,519</td>
</tr>
<tr>
<td>60%</td>
<td>$1,074,847</td>
<td>$1,369,138</td>
</tr>
<tr>
<td>50%</td>
<td>$1,088,342</td>
<td>$1,352,161</td>
</tr>
<tr>
<td>30%</td>
<td>$1,110,835</td>
<td>$1,324,782</td>
</tr>
<tr>
<td>10%</td>
<td>$1,129,812</td>
<td>$1,302,530</td>
</tr>
</tbody>
</table>

Figure 25. Prediction Intervals at Each Prediction Confidence Level for a Project with Budget of $1,200,000

\[ s_{\text{new}}^2 = MSE \left[ 1 + 1/n + (X_{\text{new}} - \bar{X})^2 / \sum (X_i - \bar{X})^2 \right] \]
\[ = 3.57 \left[ 1 + 1/100 + (286.8986 - 319.86)^2 / 43643.54 \right] \]
\[ = 3.6946 \]

\[^{28}\] \( \hat{Y}_{\text{new}} \pm t_{\alpha/2, n-p} \cdot s_{\text{new}} = 287.1212 \pm 2.276 \times \sqrt{3.6946} \)
We have shown how to use regression models to test and estimate the impact of a risk factor on project costs; how to use the estimated regression model to predict the costs of a new project; and how to determine the prediction interval for project cost at a certain confidence level. In the following section, we compare the estimates from this regression model with those from the Beta Fitting model.

**Regression Analysis vs. Beta Fitting Method**

To make a fair comparison between the project cost overrun risk estimates from these two methods, we must ensure that the databases used are identical. The regression model in this numerical example uses the data for large road and highway projects whose budgets are greater than $1 million. We use these data to fit a Beta distribution model also. Using the techniques shown in Section 3.5.2, we obtain a Beta distribution \((67.44, 88.28, 0, 2.35)\). Table 17 gives the estimated mean, mode and standard deviation for the cost overrun ratio variable of this Beta distribution model.

<table>
<thead>
<tr>
<th>Table 17. Mean, Mode and Standard deviation Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta Distribution ((67.44, 88.28, 0, 2.35))</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Mode</td>
</tr>
<tr>
<td>Standard Deviation</td>
</tr>
</tbody>
</table>

In this Beta fitted model, the project cost for a project with budget of $1.2 million is expected to be $1,221,600\(^{29}\), $8,502 higher than the regression model estimate. The difference ratio is about 0.71%\(^{30}\).

In regression models, we use prediction intervals to estimate with certain confidence levels the range within which the real project cost will fall. In probability models, we use confidence intervals to estimate with certain confidence levels the range within which the real project cost will fall. It would be interesting to examine the difference between these two. Figure 26 gives a graphical comparison between the estimated prediction interval from the regression model and the estimated confidence interval from the Beta fitted model for cost estimation of this new project.

\(^{29}\) Expected Cost = Budget*Mean Project Cost Overrun Ratio = $1.2*1.018=$1,2216 million

\(^{30}\) Difference ratio = Difference in Cost Estimates/ Budget = $8,502/ $1,200,000 = 0.71%
Figure 26. Comparisons on Project Cost Prediction and Confidence Intervals from the Regression and Beta Fitting Models respectively

According to Figure 26, the prediction and confidence intervals for project cost using Regression and Beta Fitting models, respectively, give estimate ranges close to one another, especially at lower confidence levels. At confidence levels above 75%, the confidence intervals from the Beta Fitting model are lower than the prediction intervals from the regression model. While at confidence levels below 40%, the confidence intervals from the Beta Fitting model are higher than those from the regression model.

These results show that, everything else the same, the regression model and the beta fitting model yield similar cost interval estimates for road and highway projects with budgets over $1 million. The close estimates for these two intervals in this case might be due to that the cost overrun ratios of VIHP projects are centered at about 1.02. Therefore both the regression model and the beta fitting model have close estimated values for project cost overrun ratio.

3.5.6. Multiple Regression Model Analysis

In the previous sections, we analyze the relationship between project budget and cost overrun ratio for each project type. It would be interesting to see the effect of using more information from the data in the regression model.

First, we can add a project type variable as an independent variable in addition to project budget. In the previous sections, projects are grouped into small road/highway, medium-sized road/highway, large road/highway, and bridge/tunnel projects. In the VIHP, the cost overrun ratios for small road/highway projects have larger variations than those for other project types. So, we categorize the VIHP projects into two groups: small road/highway projects and other projects. We add one dummy variable to this multiple regression model as an independent variable: small road/ highway projects. Table 18 gives the dummy variable values for the previous four project types:

31 We tested other dummy variables in different grouping. However none of those is statistically significant.
Table 18. Project Type Variable Values for Each Project Type

<table>
<thead>
<tr>
<th>Project Types</th>
<th>Small R/H Project Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small R/H Projects</td>
<td>1</td>
</tr>
<tr>
<td>Other Projects</td>
<td>0</td>
</tr>
</tbody>
</table>

* ‘Other Projects’ include medium and large road and highway projects and all bridge and tunnel projects.

Second, we add another independent variable, the cost overrun ratio frequency. For each observation, we calculate its cost overrun ratio and then the frequency of this cost overrun ratio value happens among cost overrun ratios of all the observations. For example, there are 163 projects in VIHP database, 10 of which have cost overrun ratios of 1.03. So the cost overrun ratio frequency for each of these 10 projects is 10/163.

So, this multiple regression model can be written as:

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \]

\[ Y = \log \text{(Cost)} \]
\[ X_1 = \log \text{(Budget)} \]
\[ X_2 = \text{Cost Overrun Ratio Frequency} \]
\[ X_3 = 1 \text{ for small road and highway projects} \]
\[ = 0 \text{ for other projects} \]

According to the statistical tests\(^{32}\), these three independent variables are all significantly correlated to the independent variable, project cost. The estimated model is:

\[ \log \text{(Cost)} = 9.66 + 0.97 \log \text{(Budget)} - 4.11 \text{Frequency} - 2.75 \text{SmallR/H} \]

The estimated regression coefficient for Log (Budget) is 0.97, which indicates that project costs increase with larger budget, everything else the same. However, marginal project cost decreases as budget increases (\( \beta_1 = 0.97 < 1 \)), which means project costs are better controlled in larger projects, everything else the same.

The estimated regression coefficient for the Small Road/Highway project variable is \(-2.75\), indicating that small road and highway projects have lower costs than other projects. This conclusion also implies that small bridge and tunnel projects may have larger project cost overrun than small road and highway projects, since small bridge and tunnel projects are included in the “other projects” group.

The negative estimate for the regression coefficient for cost overrun ratio frequency indicates that for projects of a certain type and with a certain budget, an increase in project cost overrun ratio frequency is associated with a decrease in project costs. The increase in project cost overrun ratio frequency means that the project cost overrun ratio approaches its mode in the cost overrun ratio histogram. The negative sign of this regression coefficient indicates that projects with lower costs are more likely to have cost overrun ratios closer to the mode. One possible reason for this is most of VIHP projects are in small sizes.

\(^{32}\) See Appendix 5 for statistical analysis reports
This model cannot be used for project cost prediction, since the project cost overrun ratio frequency data are not available at the project evaluation stage. However, it provides some interesting conclusions on the relationships among the cost, budget, type and cost overrun ratio frequency of a project. For example, all else the same, project costs are better controlled in larger projects.

4. How the Project Managers Can Use these Models

For a new project, project managers can use the estimated probability or regression models that are based on historical project data with similar conditions (such as similar political, economical, environmental, cost structure and investment structure) to estimate project costs and to analyze the impacts of risk factors.

In some cases, the new project might use new technologies or have some new features (for example, new construction designs), which makes it incapable to use historical data fitted models directly. However, uniqueness of the new projects is unlikely to be absolutely unique. These new projects should incorporate some technologies or features used in previous projects. In these cases, we can split the new project into two parts: stages that can be analyzed using historical data and stages that are new. For example, RAV line is a new project launched in BC and we may not be able to find enough historical data of similar projects in similar sizes. One alternative can be splitting RAV line construction into stages as: underground rail construction, on-the-ground rail construction, and station construction (managers may set different project stages based on practical situations and needs). When splitting other urban rail projects, such as the Sky Train projects, into similar stages, project managers can use probability models or regression models to estimate costs and to analyze cost overrun risks for RAV line at each stage.

Suppose the probability distribution fitting model or the regression model shows the cost overrun ratio means and confidence intervals as listed in Table 19.

<table>
<thead>
<tr>
<th>Project Stages</th>
<th>Cost Overrun Ratio Mean</th>
<th>Cost Overrun Ratio CI (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over ground stage</td>
<td>1.05</td>
<td>0.9 ~ 1.5</td>
</tr>
<tr>
<td>Under ground stage</td>
<td>1.5</td>
<td>1.0 ~ 2.5</td>
</tr>
<tr>
<td>Station Building stage</td>
<td>1.07</td>
<td>1.0 ~ 1.5</td>
</tr>
</tbody>
</table>

For a new urban rail project with $10 million budget for over ground stage, $50 million budget for under ground stage and $10 million budget for station building stage. The expected costs would be $96.2 million ($10*1.05 + $50*1.5 + $10*1.07) in total for this new project. And project cost confidence interval at 95% level is between $69 million ($10*0.9 + $50*1.1 + $10*1) and $155 million ($10*1.5 + $50*2.5 + $10*1.5).

Another possible solution is: specifying the exact features that make the new project unique, using historical data to analyze the impacts of these features on project costs and using these estimates to analyze cost overrun risk for the new project. For example, a new project might use one technology that has never been used in this kind of projects, but in some other projects. Project managers can first use regression models to analyze the impacts of this technology on project costs in previous projects. And use this estimate to
analyze the cost overrun risk for the new project. For example, if the regression analysis shows that the involvement of a certain technology in project may increase project cost overrun ratio by 5%. For the cost overrun ratio of a new project using this technology is likely to be 5% larger than that of a similar project without using this technology.

After all, projects have at least some features that can be found correspondingly in previous projects. Regression models can help test and quantifying the impacts of these features on project costs and therefore help cost estimation and cost overrun risk analysis for a new project. We suggest instead of relying solely on personal experiences, project managers can improve the reliability and accuracy of cost overrun risk analysis by using quantitative analysis models.

5. Summary and Conclusions
Based on project risk source analysis, this thesis focuses on proposing risk analysis models to analyze cost overrun risk for transportation infrastructure projects.

Making use of historical data in addition to experiences of project evaluators, the probability distribution models and regression models are used in project cost benefit analysis and improve on project evaluators’ cost estimation in terms of reliability. Based on private sector behavior analysis, the PPP scenario analysis and PPP cost overrun risk analysis models determine some conditions for the success of PPP projects, which may help the public sector design successful PPP contracts.

Our point is that project cost overrun risk analysis should be conducted by relying on feasible historical data. Current despise on quantitative project risk analysis and solely dependence on the experiences of project evaluators should be changed, since no individual’s experiences can be comparable to the information we can draw from a valid project database.

As we discuss in the Risk Definition section of the Literature Review, risk comprises both the probability of a risk event and its potential impact. Failing to consider either of these has the potential to create severe misunderstanding of risks. The probability distribution model in this thesis provides probability estimation for project cost overrun risks and the regression model provides estimates of the impacts of the risk factor variables, as well as predictions of project costs.

The probability distribution model and the regression model in this thesis are both data based, intuitive and easily implemented in practical project cost risk analysis. The primary purpose of this thesis is to develop models that provide better project cost estimation without complicated technical and modeling requirements.
References


Pindyck, Robert S. *Irreversibility, Uncertainty and Investment*, Journal of Economic Literature, 29, 1110-1148, 1991


Regnier, Eva, *Activity Completion Times in PERT and Scheduling Network Simulation, Part II*, DRMI Newsletter, April 8, 2005


63
Appendix

Appendix 1. Beta Distribution

Beta probability distribution has some nice properties:

- has finite limits, which fit the lower and upper bounds for many real-world random variables.
- can be asymmetric. Many variables in reality are right or left skewed to some unlikely but significant result.
- flexible and can be in different shapes.

The probability density function of beta distribution is:

\[ f(x) = \frac{(x - A)^{p-1}(B - x)^{q-1}}{B(p, q)(B - A)^{p+q-1}}, A \leq x \leq B; p, q > 0 \]

P and Q are the shape parameters. Shape parameters define the shape of a distribution. Beta distributions with different P and Q values have different probability density functions. Some classes of distributions have pre-defined shapes. For example, all normal distribution are always bell shaped. The following figure shows different Beta probability density functions for different P and Q values.

(Sources from http://www.itl.nist.gov/div898/handbook/eda/section3/eda366h.htm)

A and B are the lower and upper bounds, respectively, of the distribution, and B(p, q) is the beta function.

\[ B(p, q) = \int_0^1 x^{p-1}(1 - x)^{q-1} \, dx \]

The beta function can take on different shapes depending on the values of the two parameters:
• \( P < 1, Q < 1 \) is U-shaped
• \( P < 1, Q \geq 1 \) or \( P = 1, Q > 1 \) is strictly decreasing
• \( P = Q = 1 \) is the uniform distribution
• \( P = 1, Q < 1 \) or \( P > 1, Q \leq 1 \) is strictly increasing.
• \( P > 1, Q > 1 \) is unimodal, which has a single local maximum value
• If \( P = Q > 1 \), the density function is symmetric about \( 1/2 \).

The mean and variance are calculated as follows:

\[
\mu = A + (B - A) \frac{P}{P + Q}
\]

\[
\sigma^2 = (B - A)^2 \frac{P \times Q}{(P + Q)^2(P + Q + 1)}
\]

\[
Mode = A + (B - A) \frac{P - 1}{P + Q - 2}, P > 1, Q > 1
\]

**Appendix 2 Maximum Likelihood Estimation for Beta Distribution Shape Parameters**

We can use the following equations to estimate \( P \) and \( Q \):

\[
\psi(P) - \psi(P + Q) = \frac{1}{n} \sum_{i=1}^{n} \log \left( \frac{\gamma_i - A}{B - A} \right)
\]

\[
\psi(Q) - \psi(P + Q) = \frac{1}{n} \sum_{i=1}^{n} \log \left( \frac{B - \gamma_i}{B - A} \right)
\]

\( n \) = Number of observations

\( \gamma_i \) = Cost overrun ratio for project \( i \)

\( \psi(x) \) is the digamma function of \( x \), which is given by the logarithmic derivative of the gamma function.

\[
\psi(x) = -\frac{d}{dx} \ln \Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}
\]

\( \Gamma(x) \) is the gamma function, which is defined as:

\[
\Gamma(x) = \int_{0}^{\infty} t^{x-1}e^{-t}dt \quad x > 0
\]

\[
\Gamma(x) = \frac{\Gamma(x+1)}{x} \quad x < 0
\]

**Appendix 3 Regression Analysis Reports for the VIHP**

• Small Road and Highway Project Regression Model

**Tests of Assumptions Section**

<table>
<thead>
<tr>
<th>Assumption/Test</th>
<th>Test Value</th>
<th>Prob Level</th>
<th>Reasonable at the 0.2000 Level of Significance?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Is the Assumption Reasonable at the 0.2000 Level of Significance?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Value</td>
<td>Prob Level</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Residuals follow Normal Distribution?
- Shapiro Wilk: 0.9433, 0.559802, Yes
- Anderson Darling: 0.3338, 0.508876, Yes
- D'Agostino Skewness: 0.2396, 0.810625, Yes
- D'Agostino Kurtosis: -0.6424, 0.520586, Yes
- D'Agostino Omnibus: 0.4701, 0.790512, Yes

Constant Residual Variance?
- Modified Levene Test: 1.0542, 0.331348, Yes

Relationship is a Straight Line?
- Lack of Linear Fit F(8, 1) Test: 0.5023, 0.804051, Yes

**Estimated Model**
\[-0.981527612769659 + (1.00551872149618) \times \text{Log}_\text{Small}_\text{R/H Project Budget}\]

**Analysis of Variance Section**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Ratio</th>
<th>Prob Level</th>
<th>Power (5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>569546.3</td>
<td>569546.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>1</td>
<td>4796.69</td>
<td>4796.69</td>
<td>4300.9663</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Error</td>
<td>9</td>
<td>10.03733</td>
<td>1.115259</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj. Total</td>
<td>10</td>
<td>4806.727</td>
<td>480.6727</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td>574353</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(s = \text{Square Root}(1.115259) = 1.056058\)

- Medium Sized R/H Project Regression Model

**Tests of Assumptions Section**

<table>
<thead>
<tr>
<th>Assumption/Test</th>
<th>Test Value</th>
<th>Prob Level</th>
<th>Is the Assumption Reasonable at the 0.2000 Level of Significance?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residuals follow Normal Distribution?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shapiro Wilk</td>
<td>0.9817</td>
<td>0.986787</td>
<td>Yes</td>
</tr>
<tr>
<td>Anderson Darling</td>
<td>0.2089</td>
<td>0.863874</td>
<td>Yes</td>
</tr>
<tr>
<td>D'Agostino Skewness</td>
<td>0.1667</td>
<td>0.867590</td>
<td>Yes</td>
</tr>
<tr>
<td>D'Agostino Kurtosis</td>
<td>0.7604</td>
<td>0.447033</td>
<td>Yes</td>
</tr>
<tr>
<td>D'Agostino Omnibus</td>
<td>0.6060</td>
<td>0.738614</td>
<td>Yes</td>
</tr>
</tbody>
</table>

| Constant Residual Variance?             |            |            |                                                               |
| Modified Levene Test                    | 0.1020     | 0.755472   | Yes                                                          |

| Relationship is a Straight Line?        |            |            |                                                               |
| Lack of Linear Fit F(9, 2) Test         | 0.4364     | 0.843076   | Yes                                                          |
Estimated Model
\((105.068296795956) + (0.619730185497456) \times \text{Log\_Medium\_R\_H\_Project\_Budget}\)

Analysis of Variance Section

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Ratio</th>
<th>Prob Level</th>
<th>Power (5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>965603.8</td>
<td>965603.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>1</td>
<td>350.3859</td>
<td>350.3859</td>
<td>5.4068</td>
<td>0.0402</td>
<td>0.5637</td>
</tr>
<tr>
<td>Error</td>
<td>11</td>
<td>712.8448</td>
<td>64.80408</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj. Total</td>
<td>12</td>
<td>1063.231</td>
<td>88.60256</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>13</td>
<td>966667</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(s = \text{Square Root}(64.80408) = 8.050098\)

Large R/H Project Regression Model

Tests of Assumptions Section

<table>
<thead>
<tr>
<th>Assumption/Test</th>
<th>Test Value</th>
<th>Prob Level</th>
<th>Is the Assumption Reasonable at the 0.2000 Level of Significance?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residuals follow Normal Distribution?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shapiro Wilk</td>
<td>0.8199</td>
<td>0.000000</td>
<td>No</td>
</tr>
<tr>
<td>Anderson Darling</td>
<td>7.1029</td>
<td>0.000000</td>
<td>No</td>
</tr>
<tr>
<td>D'Agostino Skewness</td>
<td>2.4162</td>
<td>0.015682</td>
<td>No</td>
</tr>
<tr>
<td>D'Agostino Kurtosis</td>
<td>4.3707</td>
<td>0.000012</td>
<td>No</td>
</tr>
<tr>
<td>D'Agostino Omnibus</td>
<td>24.9410</td>
<td>0.000004</td>
<td>No</td>
</tr>
</tbody>
</table>

Constant Residual Variance?

Modified Levene Test 0.0436 0.835035 Yes

Relationship is a Straight Line?

Lack of Linear Fit F(51, 47) Test 0.9606 0.557187 Yes

Estimated Model
\((1.65721935221895) + (0.99547546003812) \times \text{Log\_Large\_R\_H\_Project\_Budget}\)

Analysis of Variance Section

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Ratio</th>
<th>Prob Level</th>
<th>Power (5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>1.024448E+07</td>
<td>1.024448E+07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>1</td>
<td>35932.66</td>
<td>35932.66</td>
<td>10065.5253</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Residual Histogram Plot

Histogram of Residuals of Log_large_R_H_Project_Cost

Bridge and Tunnel Regression Model

Tests of Assumptions Section

<table>
<thead>
<tr>
<th>Assumption/Test</th>
<th>Test</th>
<th>Prob Level</th>
<th>Reasonable at the 0.2000 Level of Significance?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residuals follow Normal Distribution?</td>
<td>Shapiro Wilk, Anderson Darling, D'Agostino Skewness, D'Agostino Kurtosis, D'Agostino Omnibus</td>
<td>0.6985, 4.6111, 2.4304, 3.1520, 15.8418</td>
<td>No, No, No, No, No</td>
</tr>
<tr>
<td>Constant Residual Variance?</td>
<td>Modified Levene Test</td>
<td>1.1781</td>
<td>Yes</td>
</tr>
<tr>
<td>Relationship is a Straight Line?</td>
<td>Lack of Linear Fit F(22, 8) Test</td>
<td>0.9236</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Estimated Model

(-.98304086868152081) + (1.0033133225583) * (Log__large_B_T_Project_Budget)

Analysis of Variance Section

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Ratio</th>
<th>Prob Level</th>
<th>Power (5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>2999413</td>
<td>2999413</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>1</td>
<td>10899.37</td>
<td>10899.37</td>
<td>36947.4954</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Error</td>
<td>30</td>
<td>8.849884</td>
<td>0.2949961</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Adj. Total 31 10908.22 351.878
Total 32 3010321

$s = \text{Square Root}(0.2949961) = 0.5431355$

Histogram of Residuals of Log_{large_BJTJ^cject_Cost}

Appendix 4. Distribution Summary in Arena
Distribution: Beta
Expression: $0.04 + 1.96 \times \text{BETA}(7.18, 6.73)$
Square Error: 0.000274
Chi Square Test
Number of intervals = 16
Degrees of freedom = 13
Test Statistic = 147
Corresponding p-value < 0.005

Data Summary
Number of Data Points = 17049
Min Data Value = 0.215
Max Data Value = 1.89
Sample Mean = 1.05
Sample Std Dev = 0.254

Appendix 5. Multiple Regression Report
Run Summary Section
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable</td>
<td>log_cost</td>
</tr>
<tr>
<td>Number Ind. Variables</td>
<td>3</td>
</tr>
<tr>
<td>Weight Variable</td>
<td>None</td>
</tr>
<tr>
<td>R2</td>
<td>0.9930</td>
</tr>
<tr>
<td>Adj R2</td>
<td>0.9929</td>
</tr>
<tr>
<td>Coefficient of Variation</td>
<td>0.0095</td>
</tr>
<tr>
<td>Mean Square Error</td>
<td>8.421432</td>
</tr>
<tr>
<td>Square Root of MSE</td>
<td>2.90197</td>
</tr>
<tr>
<td>Completion</td>
<td></td>
</tr>
<tr>
<td>Parameter</td>
<td>Value</td>
</tr>
<tr>
<td>Rows Processed</td>
<td>163</td>
</tr>
<tr>
<td>Rows Filtered Out</td>
<td>0</td>
</tr>
<tr>
<td>Rows with X's Missing</td>
<td>0</td>
</tr>
<tr>
<td>Rows with Weight Missing</td>
<td>0</td>
</tr>
<tr>
<td>Rows with Y Missing</td>
<td>0</td>
</tr>
<tr>
<td>Rows Used in Estimation</td>
<td>163</td>
</tr>
<tr>
<td>Sum of Weights</td>
<td>163.000</td>
</tr>
<tr>
<td>Completion Status</td>
<td>Normal</td>
</tr>
</tbody>
</table>

69
Ave Abs Pct Error 0.529

**Regression Equation Section**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Regression Coefficient</th>
<th>Standard Error</th>
<th>T-Value to test</th>
<th>Prob Level</th>
<th>Reject H0 at 5%?</th>
<th>Power of Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>9.6613</td>
<td>2.6249</td>
<td>3.681</td>
<td>0.0003</td>
<td>Yes</td>
<td>0.9553</td>
</tr>
<tr>
<td>(C39=1)</td>
<td>-2.7525</td>
<td>1.1504</td>
<td>-2.393</td>
<td>0.0179</td>
<td>Yes</td>
<td>0.6621</td>
</tr>
<tr>
<td>Cost_Overrun_Ratio_Frequency</td>
<td>-4.1109</td>
<td>1.9515</td>
<td>-2.106</td>
<td>0.0367</td>
<td>Yes</td>
<td>0.5532</td>
</tr>
<tr>
<td>log_budget</td>
<td>0.9723</td>
<td>0.0083</td>
<td>117.217</td>
<td>0.0000</td>
<td>Yes</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Estimated Model

9.66126733146649-2.7525117238345*(C39=1)-4.11086840238232*Cost_Overrun_Ratio_Frequency+.972297221530233*log_budget

**Regression Coefficient Section**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Regression Coefficient</th>
<th>Standard Error</th>
<th>95% C.L. Lower</th>
<th>95% C.L. Upper</th>
<th>Standardized Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>9.6613</td>
<td>2.6249</td>
<td>4.4771</td>
<td>14.8455</td>
<td>0.0000</td>
</tr>
<tr>
<td>(C39=1)</td>
<td>-2.7525</td>
<td>1.1504</td>
<td>-5.0246</td>
<td>-0.4804</td>
<td>-0.0202</td>
</tr>
<tr>
<td>Cost_Overrun_Ratio_Frequency</td>
<td>-4.1109</td>
<td>1.9515</td>
<td>-7.9651</td>
<td>-0.2566</td>
<td>-0.0141</td>
</tr>
<tr>
<td>log_budget</td>
<td>0.9723</td>
<td>0.0083</td>
<td>0.9559</td>
<td>0.9887</td>
<td>0.9840</td>
</tr>
</tbody>
</table>

Note: The T-Value used to calculate these confidence limits was 1.975.

**Analysis of Variance Section**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>R2</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Ratio</th>
<th>Prob Level</th>
<th>Power (5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>1.528971E+07</td>
<td>1.528971E+07</td>
<td></td>
<td></td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Model</td>
<td>3</td>
<td>0.9930</td>
<td>189704.8</td>
<td>63234.95</td>
<td>7508.811</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Error</td>
<td>159</td>
<td>0.0070</td>
<td>1339.008</td>
<td>8.421432</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total(Adjusted)</td>
<td>162</td>
<td>1.0000</td>
<td>191043.8</td>
<td>1179.283</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Analysis of Variance Detail Section**

<table>
<thead>
<tr>
<th>Term</th>
<th>DF</th>
<th>R2</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Ratio</th>
<th>Prob Level</th>
<th>Power (5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>1.528971E+07</td>
<td>1.528971E+07</td>
<td></td>
<td></td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Model</td>
<td>3</td>
<td>0.9930</td>
<td>189704.8</td>
<td>63234.95</td>
<td>7508.811</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>C39</td>
<td>1</td>
<td>0.0003</td>
<td>48.20848</td>
<td>48.20848</td>
<td>5.724</td>
<td>0.0179</td>
<td>0.6621</td>
</tr>
<tr>
<td>Cost_Overrun_Ratio_Frequency</td>
<td>1</td>
<td>0.0002</td>
<td>37.36802</td>
<td>37.36802</td>
<td>4.437</td>
<td>0.0367</td>
<td>0.5532</td>
</tr>
<tr>
<td>log_budget</td>
<td>1</td>
<td>0.6057</td>
<td>115708.7</td>
<td>115708.7</td>
<td>13739.785</td>
<td>0.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Error</td>
<td>159</td>
<td>0.0070</td>
<td>1339.008</td>
<td>8.421432</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total(Adjusted)</td>
<td>162</td>
<td>1.0000</td>
<td>191043.8</td>
<td>1179.283</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>