UBC P3 Project

Working Paper
2006 - 05

Cost overruns Risk Analysis in Transportation Infrastructure Investments

Joseph Berechman
CA Chair Professor
Sauder School of Business and SCARP
University of British Columbia

and

Qing Wu
Graduate, Operation and Logistics Division
Sauder School of Business
University of British Columbia

December 7, 2006

Production of this working paper has been made possible through a financial contribution from Infrastructure Canada. The views expressed herein do not necessarily represent the views of the Government of Canada.

Phelps Centre for the Study of Government and Business
Sauder School of Business
University of British Columbia
2053 Main Mall, Vancouver, BC V6T 1Z2
Tel : 604 822 8399 or e-mail: phelps_centre@sauder.ubc.ca
Web: http://csgb.ubc.ca/p3_about.html

Last Revised: December 7, 2006

Cost overruns Risk Analysis in Transportation Infrastructure Investments

by

Joseph Berechman[1] and Qing Wu[2]

Abstract

Cost overruns have been shown to occur in a significant number of all transportation investment projects. To date, the risk of cost overruns has not been adequately examined in the literature relative to causes and approaches to quantitatively estimate its magnitude prior to the project's implementation. This paper proposes several methods for assessing cost overruns risk in transportation infrastructure projects. Probability distribution fitting methods, regression analysis and simulation models are developed. Using a database, made up of observations on a large number of highway projects, these methods are further used to empirically estimate costs overrun risk. The key conclusion is that these methods can provide realistic risk estimates, thereby reduce subjective biases in project cost benefit analysis.

Acknowledgment: For financial support of this study we wish to thank the UBC P3 Project, Phelps Centre for the Study of Government and Business, Sauder School of Business, University of British Columbia, Vancouver Canada.

[1] CN Chair Professor, Sauder School of Business and SCARP, University of British Columbia Vancouver V6T 1Z2, Canada (berechman@sauder.ubc.ca) or i9400@post.tau.ac.il

[2] Graduate, Operation and Logistic Division, Sauder School of Business, University of British Columbia Vancouver Canada (qingwu.c@gmail.com)

1. Introduction

Consider the following facts. Of 258 large-scale transportation investment projects in five continents, 9 out of 10 projects had cost overruns. All projects (road, fixed links and rail), on the average, experienced costs overruns of 28%, while rail projects, on the average, had cost overruns of 45%, (Flyvbjerg, 2004). A significant proportion of all surveyed passenger rail projects in the US had a significant over-estimation of actual demand levels and under-estimation of actual costs (Pickrell, 1989). Of 128 highway projects carried out in Vancouver Island 104 had some costs overruns, while out of 36 tunnels and bridges projects, carried out in the same region, 29 had cost overruns (see below). Why so many transportation investment projects are subject to significant risks of costs overruns or over-estimation of actual demand? How can these risks be assessed *ex ante* and what tools are available to this end? These are the key questions that this paper will try to answer. In particular, the paper will focus on costs overrun risk, which as the evidence show is a pervasive phenomenon in transportation infrastructure investments.

Cost overruns is defined as the excess of actual project costs over budgeted costs. Costs overrun may be caused by underestimation of costs at the planning stages or by the escalation of costs during implementation due to unforeseen events, changes in the scope of the project or by poor management. Costs overruns may not necessarily lead to project failure if the project can obtain sufficient funding to cover its excess costs. However, the economic viability of the project, which was assessed using the erroneously estimated costs, would be different if the risk of cost overruns was built-in the evaluation analysis.

Against this background this paper first analyzes the risk factors, which underlie costs overruns in transportation infrastructure projects. Subsequently, it proposes methods for estimating the overall risk likelihood of costs overruns, given the particular nature of the project. Subsequently, these methods are applied to a real-world database of highway transportation investments. It is shown that the use of such methods can improve the *ex ante* risk's analysis of transportation infrastructure projects.

The design of the paper is as follows. In Section 2, we examine various sources of risk in transportation investment projects. Section 3 presents results, reported in the literature of costs overruns in transportation projects. In Section 4 we introduce the methods that we have developed for assessing costs overruns. These include a Distribution Fitting Model (DFM), a regression model and a Monte Carlo simulation model. Empirical application of these models is in Section 5. Section 6 presents a numerical example of the use of these methods for cost overruns assessment. Summary and conclusions are in Section 7.

2. Sources of Risk in Transportation Investment Projects

We define risk of cost overruns as the probability that a given project will experience actual costs, which exceed its projected budget by a given factor. Since the computation of probabilities of future events requires that these would be indeed random events, one might ask about the "randomness" of events in transportation infrastructure investments.

Thus, we first discuss potential causes for costs overrun risks and highlight their random nature, which in turn, provides the rationale for risk analysis of transportation projects. Based on available literature several main categories of cost overruns risk factors are identified (Ayyub, 2003; Jaafari, 2001; Flyvbjerg et al, 2003). These are:

Technological risk: It refers to the fact that technology planned for a given project may need to be modified or replaced by a newer one as either the costs or the benefits of the new technology outperform those of the older one. In some situations, there may be a need for the new technology due to unexpected difficulties during construction. Technologies related to burrowing often fall within this category, with substantial impact on the actual costs of a project. In general, disregarding the possibility of dishonesty in the project's planning phase, technological risks are mainly due to unsystematic random technological complexities.

Construction risks: Large-scale transportation projects quite often are subject to unexpected construction snags, which range from bad weather, unexpected and random geo-technical events, equipment breakdowns, undelivered raw materials, unknown presence of other infrastructures (e.g., sewage lines) or unexpected soil problems. All of these imply construction delays, which in turn, affect the costs of projects.

General economic and financial risks: Unexpected changes in real interest rates, or in exchange rates or in unemployment rates may have considerable impacts on the actual costs of a project. Rising interest rates will affect the debt service cost component of a capital project. Shortages of skilled labor, which often characterize periods of rapid economic growth, are likely to have consequential impacts on labor costs.

Regulatory risks: These risk types stem from unexpected changes in regulation of externalities, for example, changes in emission and other environmental standards, which might take place during project implementation.

Organizational and project management risks: Often unpredictably, projects lose critical staff during the time of construction. In addition, projects may suffer from poor management, which may also yield to external pressure from interest groups to change the project's scope. As a result, delays appear along with rising costs.

Political risks: Political risk refers to unforeseen circumstances where a new government fails to keep commitments made by a previous one, or due to a budget crisis, it fails to provide already promised capital. Foreign investors in some developing countries may face unfriendly local governments or the risk of potential expropriation. Since large transportation projects often require approval and financial support from local and federal governments, conflicts between and within governments may cause project's delays, thus additional costs.

Contractual or legal risks: Contract and legal risks arise from inappropriate division of responsibility among contractors. During the project's planning and implementation periods, issues related to securing the rights of way, payments and other legal disputes

might appear unexpectedly. Issues of contract enforceability and post contractual disagreements may also cause project suspensions and delays with substantial effects on costs.

This categorization serves to show the wide range of risk factors, which underlie costs overruns as well as their random nature. Thus, in what follows we treat costs overruns as a random variable for which we will fit a probability distribution models.

How prevalent is the use of quantitative risk analysis methods in actual project planning and management? A survey by Akintoye (1997), has shown that both contractors and project managers mainly rely on their intuition and subjective judgment to manage risks. About half of the managers surveyed claimed to be familiar with sensitivity analysis, yet few actually have used this technique in practice. By and large, contractors and managers expressed doubts on the usefulness and practicality of quantitative risk analysis techniques.

Similarly, Shapira (1994) found that managers are quite uninterested in using probabilistic techniques to assess project's outcomes. He also found that under unique, non-repeated decision conditions, managers tend to neglect statistical analysis all too easily. Shapira reasons that this is due to managers' confidence in their ability to control risk, though as experience shows this kind of behavior many times have led to actual costs significantly exceeds planned budget and at times, to project's total failure.

Table 1 provides a summary of demand and cost risk factors and their impact on costs overruns and demand deficits.

Table 1: Main project risk sources and their impacts

Risk Factors	Major impacts on Project	Management Perception of Risk
Financial Risk	Benefit deficit and Costs overrun	High
Economics Risk	Benefit deficit and Costs overrun	High
Construction Risk	Costs overrun	High
Contractual Risk	Costs overrun	High
Environment Risk	Costs overrun	Moderate
Funding Risk	Project termination	High
Project Design Risk	Benefit deficit, Costs overrun and Project abandonment	High
Political Risk	Project delay or abandonment	Not Available

A key question here is whether these risks are measurable. That is, it might be argued that since each project, to an extent, is unique it is impossible to obtain probability distributions of its future events. We shall return to this question below.

3. Findings Related to Costs Overruns in Transportation Projects

In a seminal study Flyvbjerg et al (2003, 2004) used a simple linear regression models to analyze the relationships between transportation project cost overruns and three potential risk factors. These were: length of project implementation period, project size and project ownership. Their key finding is that for every additional year of project implementation, the project cost escalation is expected to increase by 4.64%. This effect holds for all facility types: rail, bridges and tunnels, and highways.

They have further found that for bridges and tunnels, larger projects tend to have larger costs escalations, though for rail and road projects, these relationships are not significant. For their entire database the assumption that bigger projects are associated with larger risks of costs overruns, was not validated.

Regarding the impact of ownership on project's costs, these authors have categorized ownership into private, state-owned enterprise, and other public ownership. Their analysis shows that state-owned enterprises have the poorest performance record with average cost overruns of 110%. Privately owned fixed links have average cost overruns of 34%. But "other public ownership" shows the best performance with average cost overruns of "only" 23%. Flyvbjerg et al (2004) argue that the main problem relative to cost overruns may not be public versus private ownership but a certain kind of public ownership, namely state-owned enterprises, which lack not only transparency and public control but also the competence of managing complicated projects that the private sector brings.

Still another interesting finding is that project costs overruns have not improved over the past several decades. There seems to be no element of learning in transportation projects' costs estimation and management, despite improved knowledge, experience and technology. The study also reports that cost underestimation and overruns are more pronounced in developing countries than in developed countries.

A "soft" risk assessment approach is to enumerate all risk factors, emanating from a given project and subsequently obtain an idea about their significance. The key problems with this approach are first that many risk factors tend to correlate so that treating each of these separately may result in erroneous estimation of the project's overall risk. Second, since different managers may conceive of risk differently, their estimates can vary widely. Third, in order to enable the comparison of alternative courses of actions we need an objective quantitative risk measure. For these reasons we prefer the use of costs overrun as the principal risk indicator of a project.

4. Methods for Assessing the Risk of Costs Overrun in Transportation Projects

We begin by defining a Cost overruns Ratio (COR) as the proportion of the project's real costs from the project's planned budget. Project's real costs are the total funds that the state or private investors have actually paid for the project's construction. The project's

budget represents the planned costs assigned to the project prior to its commencement, and which were used for the project's cost-benefit analysis. Cost overrun is then the case where this ratio is greater than one[3]. Our objective is to compute the probability of the project's COR exceeding a specific value, which we define as the *project's risk level*. To that end we need to determine a probability distribution function; subsequently, we can derive estimates of the distribution's moments as well as a confidence interval for a particular project's costs overrun risk level.

4.1 Probability Distributions Fitting Models of Costs Overrun: The Beta Distribution Model

For reasons, which are explained below, we have selected the Beta distribution for the analysis of the risk level of a project. It is defined as: $Beta(P, Q, A, B)$ where P and Q are shape parameters, and A and B are scale parameters, reflecting, respectively, the minimum and maximum values of the distribution's variable (i.e., COR). We use this distribution to fit project costs overrun ratio data. Appendix A provides a discussion of the Beta distribution.

The reason for using this distribution, in addition to its convenient mathematical properties, is the fact that the cost overrun ratio of a project has finite minimum and maximum values. A histogram of project's cost overruns ratio is expected to have a tail to the right, an upper bound, and a lower bound greater than zero, since project's real costs are always positive. Probability distributions with infinite bounds or with inflexible shapes are incapable of capturing these features as, for example, are the exponential or the normal distributions. With adequate data, we can fit a Beta distribution and then estimate the probability of the project's costs exceeding a certain level. We can further estimate confidence intervals for these probability estimates. This is done in Section 5.

The general steps of a distribution fitting method are:

1. Using the database, plot histogram of the target variable(s) is plotted
2. Based on this histogram (and experience), determine that distribution, which "best" fits it
3. Use a distribution-fitting function to estimate relevant parameters

Statistical software distribution fitting functions is available for various distribution types (Beta, Gamma, Exponential, and others)[4].

---

[3] The reason for not defining cost overruns as the difference between project's real costs and planned budget is to avoid obtaining negative values, which in the log regression model would be undefined.

[4] In this study we have used the statistical software package NCSS (see also Appendix B) (http://www.ncss.com/index.htm).

4.2 Regression Analysis of Project's Cost overruns Ratio

The main objective of this model is to explain the variability in Costs Overrun Ratios (COR) by a set of independent variables. We use a log model so that when $COR > 1$, $\log(COR) > 0$; Otherwise, it is negative. key explanatory variables are:

1. Project's Present Value of Costs: $PV(C_i) = \sum_{t=0}^{Ti} C_{i,t} / (1 + r_i)^t$ (1)

In (1) T_i = Number of construction years for the completion of project i; t = Project construction year index, ($t = 0,...,T$); r = Project discount rate; i = Project index ($i = 1,...,N$) where N is the number of projects in the database.

2. Interest Rate: The average annual interest rate change over the construction period:

$$\frac{\sum_{t=1}^{T_i} r_{i,t+1} / r_{i,t}}{T_i} \qquad (2)$$

In (2), $r_{i,t}$ = the interest rate for project i at year t. It is assumed that project's costs are affected by the actual interest payments, which in turn, are affected by the prevailing interest rate at time t. The annual interest rate variability variable can be estimated by the standard deviation of annual interest rates during project periods. As shown by (2) the actual interest rate has a non-linear impact on the project's cost present value.

3. Percentage of Private and Public Investments in Capital Costs: Instead of using ownership dummy variables, (as in Flyvbjerg's et al model), we assume that changes in COR are affected by the respective proportions of the private and public sectors investments in the project's capital costs. We assert that a larger percentage of private investment would improve the project's efficiency and management responsibility, hence reduces COR.

4. Percent of Federal Government Investment: Many transportation investment projects are planned, carried out and financed by local and state governments. Yet, local and state authorities prefer federal capital funding since it does not require raising local taxes or debt finance. On the other hand, federal funding may encourage the acceptance of riskier large-scale capital investment. Thus, we hypothesize that COR is positively related to the percentage of federal government funding and that that these relationships are non-linearly related.

5. Project's Cost overruns Risk Bearer: A key question in risk analysis is who actually bears the risk of costs overruns? That is, which party is responsible for making additional investments when the COR exceeds 1? This question is particularly acute when a project is implemented as Public Private Partnership (PPP), for example, using BOT. Two

dummy variables (assuming 0 or 1 values) are then used to indicate the bearer of the project's COR risk, as in Table 2.

Table 2: Cost overruns risk bearer dummy variable values under different scenarios

Cost overruns Risk Bearer Scenarios	Dummy Variable Values	
	Private risk Bearer	Public and Private Risk Bearer
Private	1	0
Public	0	0
Public and Private	0	1

6. Project's Duration: We hypothesize that the duration of a project (period of construction) has non-linear and increasing relationships with project's costs. Thus, we assume a positive relationship between the project's COR and its duration.

7. Project's budget: the budget reported for the construction of project i. We hypothesize that COR is positively associated with the size of the budget.

Given the above, the regression model is:

$$LogY_i = \beta_0 + \sum_k^K \beta_k LogX_{i,k} + e_i \qquad (3)$$

Y_i = Costs Overrun Ratio of project i

X_1 = Average annual interest rate change over construction period

X_2 = Annual interest rate change

X_3 = Project construction duration

X_4 = Investment ratio of private sector in total project funding

X_5 = Investment ratio of federal government in total project funding

X_6 = Private COR bearing

X_7 = Public-Private COR risk bearing (a dummy variable)

X_8 = Projects' budget

X_9 = Project's duration

e_i = Error terms

K = number of explanatory variables

The dummy variables X_6 and X_7 have the value of 1 if both sectors bear the risk; and the value of 0 if only one sector does so.

Flyvbjerg's et al (2003, 2004) argued that the average transportation COR varies among areas due to differences in macroeconomic conditions and public policy. Since these external factors are common to all projects in the VIHP database we assume that the error

terms of the regression model represent some random variation attributed to the unique circumstances of specific projects.

4.3 The Simulation Model

The development of the simulation model follows these steps:

1. Make assumptions on the probability distributions of the studied variables
2. Generate a large number of "random numbers" from the assumed probability distribution of each variable
3. Use these "randomly" drawn values to estimate the parameters of the studied variable
4. Use these parameters to calculate moments (e.g., the mean and standard deviation)

The key issues in a simulation model are the assumed distributions of the variables and their relationships. In our specific analysis, we will use the results of the Beta fitting and the regression models for exploring these features.

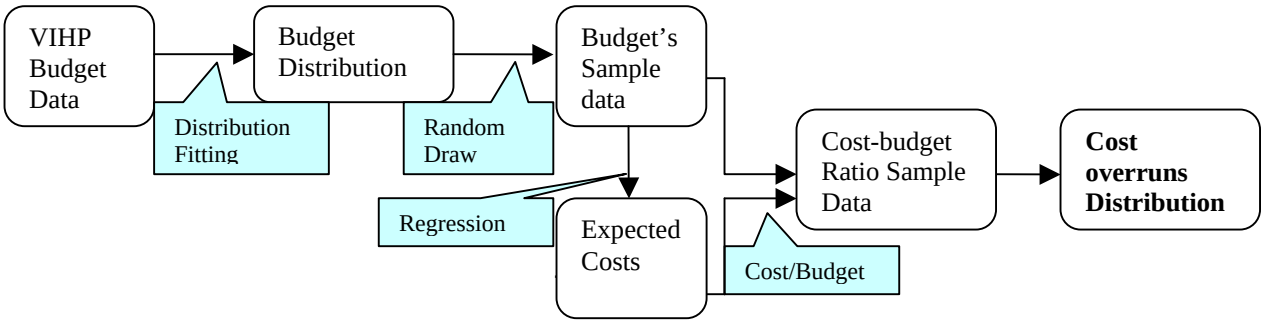
The simulation method is useful when a variable is influenced by many correlated factors and when specific data on these factors are not available. However, there are some potential problems in the implementation of simulation models for COR risk analysis. First, simulation results can be biased since the simulation model is based on prior assumptions on the project's risk factors. Second, computer "random draws" are not really random since to date computers are unable to generate real random numbers. Third, when a system gets quite complicated the results of a simulation analysis cannot easily be validated by empirical data. In brief, the validity and reliability results from a simulation model can be problematic.

We propose two simulation models. General steps of the first model are:

1. Randomly draw sample data from the fitted distribution for the projects' budget
2. Using the regression estimates calculate corresponding expected project costs for each randomly drawn data point
3. Based on (1) and (3) generate COR
4. Fit the Beta distribution for a project's COR, using the sample data

Figure 1 shows these steps.

Figure 1: Simulation Model I

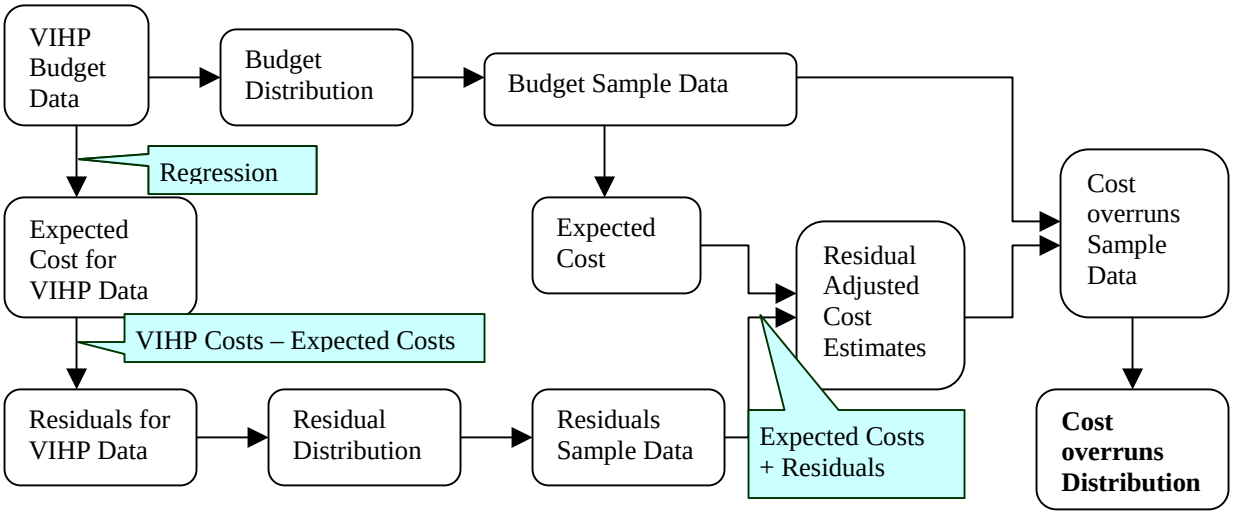


The general steps of the second simulation model are:

1. Find the 'best fit' distribution for the budget variable, using VIHP database
2. Randomly draw sample data from the fitted distribution for the projects' budget
3. Using the regression estimates calculate corresponding expected project costs for each randomly drawn data point
4. Find the 'best fit' distribution for the regression residuals
5. Randomly draw sample data from the fitted distribution for the residual[5]
6. Add residual estimates to the expected project costs and get the projects costs estimates
7. Generate the COR values
8. Fit the Beta distribution for project's COR using the sample data

These steps are shown in Figure 2.

Figure 2: Simulation Model II



5. Application: The Vancouver Island Highway Projects (VIHP)

5.1. The Database

The Vancouver Island Highway Project (VIHP) is a set of highway investments, which includes approximately 174 kilometers of highway improvements and 146 kilometers of new construction. The VIHP was announced in 1993 and was substantially completed by

[5] The residuals are calculated by subtracting actual costs from their expected costs, which are calculated using the VIHP budget data and the regression model. They reflect the 'noise' between expected costs and actual costs.

2003. It is composed of 163 independent projects[6] of which 127 were highway construction projects and 36 bridge and tunnel projects. Since these investments were carried out in various parts of the Vancouver Island and at different times, they were assumed to be independent of each other relative to planning and costs. Finally, many of these projects are carried out as Build and Transfer so they can be regarded as PPP type investments.

The actual costs and budgets for each of these projects are in 1993-dollar prices. Costs overrun ratios are calculated by dividing the project's actual cost by its budget. It might argued though that the difference between a project's actual costs and its budget merely reflect changes in the scope of the project or in quantities need during construction and hence, should not be regarded as cost overruns. Given the prevalence of the cost overruns phenomenon and the magnitude of COR (see Table 3 below), such a view raises an important question, namely if the majority of all projects are likely to experience some costs escalation during construction, why they were not factored into the original project's budget? Is it because in so doing it might have affected the NPV of a project and, as a result, the decision to construct it? It might also be argued that since the extra costs need to be funded, under an overall government's budget constraint some other projects would not be implemented and the loss of their benefits can be regarded as the price of having cost overruns. In brief, projects' budgets, if properly done, should reflect likely costs escalation. Assuming that to be the case, in this study the difference between stated budget and actual total costs is treated as cost overruns.

There are 104 highway projects and 29 bridge and tunnel projects in VIHP database. Table 3 provides basic statistics for these projects.

Table 3: Cost and cost overruns ratio ranges of VIHP data

	Number of Projects with Cost overruns	Number of Projects with Cost overruns Ratio > 1.5	Cost		Cost overruns Ratio	
			Min.	Max.	Min.	Max.
Road & Highway Projects	104 (N=127)	3	$7,559	$88,076,460	0.48	2.34
Bridge & Tunnel Projects	29 (N=36)	2	$7,592	$18,349,840	0.80	1.90

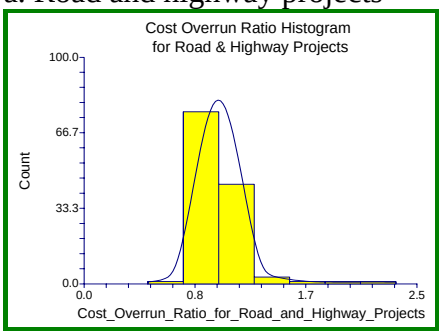
Source: Transport Canada, office in British Columbia

Given the VIHP database, Figure 3 shows histograms of the Costs Overrun Ratios (COR) highway construction and for the bridge and tunnel projects.

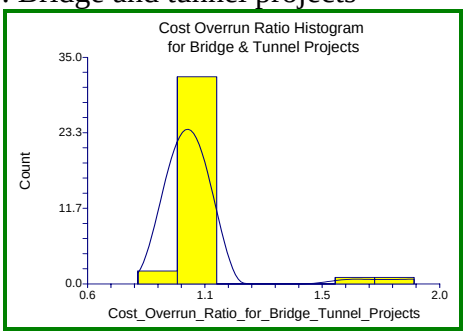
[6] Two projects with extremely high cost overruns ratios (14 and 42 times the budget) were excluded of the database.

Figure 3: Cost overruns ratio histograms for the VIHP database

a. Road and highway projects



b. Bridge and tunnel projects



The horizontal axis shows COR values, while the vertical axis frequency of occurrence. Apparently, for all project types COR observations are centered around a value little greater than one, though a sizeable proportion of all projects have COR much greater than 1. Table 4 provides descriptive COR statistics for highway and for bridge and tunnel projects.

Table 4: COR descriptive statistic

	Road/Highway Projects	Bridge/Tunnel Projects
Mean	1.059	1.052
Median	1.055	1.055
Mode	1.045	1.059
Standard Deviation	0.27	0.23

5.2 COR Beta Distribution Fitting Model

In this analysis we do not separate the projects by type (highways and bridges and tunnels). For the beta distribution we set the parameters as in Table 5.

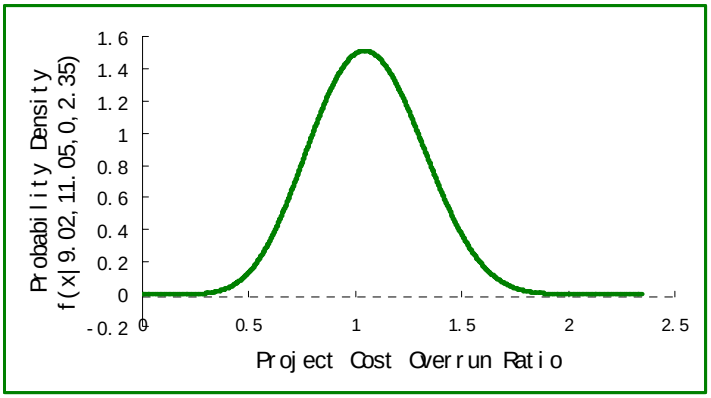
Table 5: Beta Parameters

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

Note: P and Q values are estimated using Maximum Likelihood Estimation (MLE).

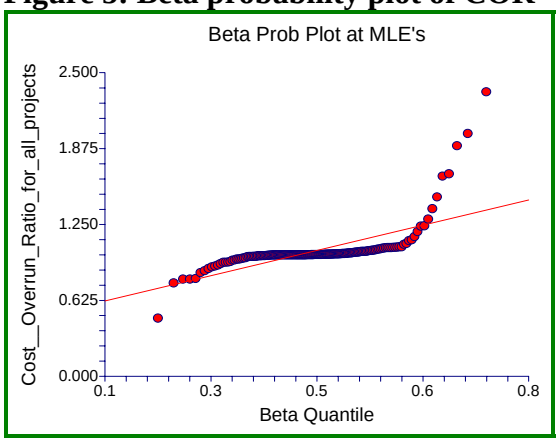
Using VIHP database the objective is to find the ‘best fit’ distribution for project’s COR. Figure 4 show the result of the Beta fitting analysis.

Figure 4: Beta distribution PDF plot for cost overruns ratio in the VIHP



Although this PDF plot is visually similar to a normal distribution PDF, it is rather asymmetric, with minimum value of 0, and a maximum value of 2.35. Figure 5 shows a plot of the Beta probability COR values.

Figure 5: Beta probability plot of COR



Note: Beta Quintiles are defined in Appendix B

As evident from Figure 5 the model estimates quite well the probability for project costs overrun ratio less than 1.25, but less well for higher values. Thus, the model's probability estimate for a project having COR >1.25 is likely to be an underestimate of the true probability[7]. Table 6 shows the probability distribution of COR.

[7] One explanation for these results is due to the statistical method used for fitting the Beta distribution. That is, in Beta distribution fitting, the COR observations are divided into Quintiles (see Appendix B). Subsequently, the statistical Beta fitting model attempts to maximize the likelihood of obtaining data from Beta Quintiles 25% to 75%. For VIHP database, 75% Beta Quintiles is at the point with cost overruns ratio equal to1.25. As a result, the cost overruns probability estimations beyond this point are likely to be underestimated.

Table 6: Probabilities for Project COR

For Cost overruns Ratio	Probability
> 0.2	99.99%
> 0.4	99.79%
> 0.6	96.71%
> 0.8	83.68%
> 1.0	**57.82%**
> 1.2	28.84%
> 1.4	9.34%
> 1.6	1.68%
> 1.8	0.12%

As seen from this table, in the VIHP database the probability of road and highway construction projects' costs exceeding their budgets (COR>1) is 57.82%, while the probability of COR> 1.8 is 0.12%.

While the Beta Fitting Method can estimate probabilities of costs overruns, it cannot explain why overruns may occur and which factors are responsible for them. For this purpose we next estimate the regression model discussed above.

5.3 Regressing Project's Cost overruns vs. Project's Budget

The VIHP database did not contain enough information on some important variables including the proportions of private and public sector investments and the actual time-length of specific projects. Therefore, we prefer to display here only limited results from the regression analysis. As the first step we tested for the relationships between projects' budget and actual costs[8], the model is:

$$LogY = \beta_0 + \beta_1 LogX + \varepsilon \qquad (5)$$

Y = Project's costs
X = Project budget
ε = error

The scatter plot of log values of project costs vs. budget shows non-constant cost variance for different project sizes. We have therefore divided the VIHP database into three groups: "small projects" (budget < $250,000), "medium projects" (budget = $250,000 - $1 million), and "large projects" (budget > $1 million). For small projects, the estimated model is:

$$Log(Cost) = -0.982 + 1.006 Log(Budget), \text{ or: } Cost = 0.953 Budget^{1.006}.$$

For medium-sized projects, the estimated model is:

[8] We used a log base of 1.05, which is the man of the VIHP database (see Table 4).

$Log(Cost) = 105.068 + 0.62 Log(Budget)$, or: $Cost = 168.39 Budget^{0.62}$.

For large road and highway projects, the estimated model is:

$Log(Cost) = 1.657 + 0.995 Log(Budget)$, or: $Cost = 1.084 Budget^{0.995}$

No obvious explanation can be given to these results. Apparently, costs tend to increase more rapidly with budget for smaller projects than for larger ones, but the pattern is inconsistent.

For bridge and tunnel projects, the estimated model is[9]:

$Log(Cost) = -0.983 + 1.003 Log(Budget)$ or, $Cost = 0.953 Budget^{1.003}$.

5.4 The Simulation Model

As explained in Section 4.3, the simulation model takes as inputs the results of the Beta and regression models. Using the above regression model results we have built two simulation models. The difference between them is that the second model takes into account also the regression residuals. The fitted distribution parameter estimates of the above two simulation models are shown in Table 7 for the 4 project types. For these comparisons, the Beta minimum and maximum values are set as: value is set: $A = 0$ and $B = 2.35$, respectively.

Table 7: Comparison of the Estimated Distribution Moments for the Simulation Models

Distribution Parameters	**Simulation Model I**				**Simulation Model II**			
	Small R&H	Medium R&H	Large R&H	B&T	Small R&H	Medium R&H	Large R&H	B&T
P	18482	16.79	21062	113339	14.78	3.17	14.86	40.48
Q	24135	18.82	28322	153774	22.62	4.63	20.39	55.63
Mean	1.019	1.108	1.002	0.997	0.929	0.955	0.990	0.990
Mode	1.019	1.107	1.002	0.997	0.924	0.936	0.987	0.989
Standard Deviation	0.0056	0.194	0.0052	0.0022	0.185	0.389	0.193	0.118

Legend: R&H = Roads and Highways; B&T= Bridges and Tunnels

[9] Four projects had log (Budget) values less than 280, largely deviating from those of all other projects in the database. They were therefore removed from the sample.

Table 7 shows that the moment estimates of simulation model I have higher means and smaller standard deviations than those of model II. The differences between the means might be due to the non-positive means of the residuals. In Model II, we derived the residuals as the difference between project's budget and the project's costs (as estimated from the regression model). But these residuals do not necessarily have zero means, which, in turn, drive down the means of Model II. The difference in the standard deviations between the two models is attributed to the use of these residuals.

Next we compare the results of the simulation models with those of Beta fitting model and the regression analysis, using a numerical example.

6. Predicting Project Costs for a New Project: A Numerical Example

Using the VIHP database for projects with budgets over $1.2 million, in this section we calculate the probability of COR using the 3 models: beta distribution, regression and simulation.

6.1 The Beta Fitting Model

We estimate a Beta distribution with the following parameters A=0; B=2.35; P=67.44; Q=88.28. Table 8 shows the estimated mean, mode and standard deviation for COR.

Table 8: Predicted Mean, Mode and Standard Deviation Estimates

	Beta Distribution
Mean	1.018
Mode	1.016
Standard Deviation	0.093

6.2 Regression Model

The objective is to compute a prediction interval for COR for projects with budget > $1,200,000 (N=100 observations).

$$s_{Ynew}^2 = MSE\left[1 + 1/n + (X_{new} - \overline{X})^2 / \sum (X_i - \overline{X})^2\right] = 3.6946$$

The 95% prediction interval for the log project cost value is (282.7457; 291.4967) [10], which corresponds to the project's cost interval of {$979,905; $1,501,793}. Stated

[10] Based on the estimated regression model for large road and highway projects, $Log(Cost) = 1.657 + 0.995 Log(Budget)$, we have expected Log(cost) value of 287.1212. Their Mean Squared Error (MSE) for this model is 3.57, the average log project budget

alternatively, with 95% confidence, the project costs for a highway project with a budget of $1,200,000 would be in the range $979,905; $1,501,793. Table 9 summarizes these results.

Table 9: Regression results for expected cost and confidence interval for a project with budget of $1,200,000

Project Budget	Model Estimated Project Cost	95% Prediction Interval	
		Lower	Upper
$1,200,000	$1,213,098	$979,905	$1,501,793

6.3 Comparison of the Three Models' COR Results: A Numerical Example

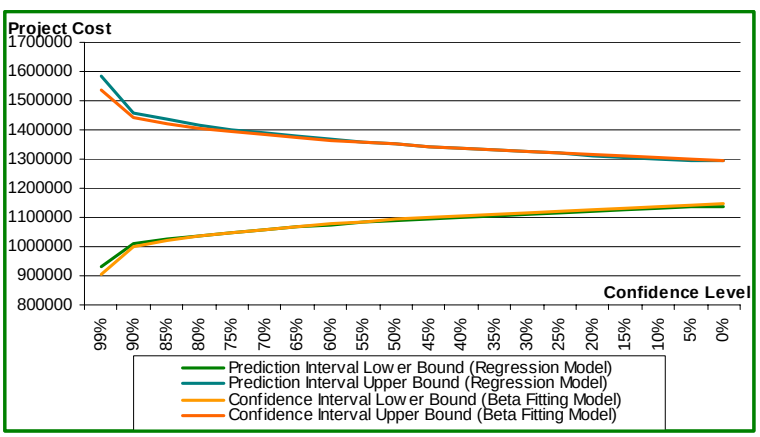
Table 10 compares the expected cost estimates[11] for a project with a budget of $1,200,000.

Table 10: Comparison of Expected Costs Estimates (Project's Budget $1,200,000)

	Regression Model	Beta Distribution	Simulation Model I	Simulation Model II
Expected Costs	$1,213,098	$1,221,600	$1,202400	$1,188,000

From Table 10, the difference between the results from the regression model and those from the other two models is: 0.71%, 0.9% and 2.1%, respectively[12].

Figure 6: Comparison of Prediction and Confidence Interval Estimates from the Regression and Beta Fitting Models



---

value is 319.86, number of observations (n) is 100 and the number of the independent variable (p) is 2. Thus, $\widehat{Y}_{new} \pm t_{\alpha/2,n-p} \cdot s_{Ynew} = 287.1212 \pm 2.276 \times \sqrt{3.6946}$.

[11] Expected Cost = Budget $\times$ Mean Project COR = $\$1.2 \times 1.018 = \$1.2216$ million

[12] Difference ratio = Difference in Cost Estimates/ Budget = ($1,213,098-$1,221,600)/ $1,200,000 = 0.71%

Figure 6 provides 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 the $1,200,000 project It shows that, everything else being the same, the regression model and the beta fitting model reach similar cost interval estimates for road and highway projects with budgets over $1 million. This can be explained by the COR of the VIHP database, which center at about 1.02 (see Figure 3), and which are used by both models.

6.4 Comparison of the Three Models COR Estimates by Projects of Size

Table 11 compares the COR estimates between the four models, for different project types.

Table 11: Comparison of Projects' COR Estimates

Project Type	**Beta Distribution**	**Regression Model**	**Simulation**	
			Model I	**Model II**
Small R&H Projects	1.071	$0.953*Budget^{0.006}$	1.019	0.929
Medium R&H Projects	1.076	$168.39*Budget^{-0.38}$	1.108	0.955
Large R&H Projects	1.018	$1.084*Budget^{-0.005}$	1.002	0.99
B&T Projects	1.023	$0.953*Budget^{0.003}$	0.997	0.99

Legend: R&H = Road and highway projects; B&T = Bridge and tunnel projects

In Table 11, except for the simulation Model I of medium-sized road and highway projects, the COR estimates of the Beta distribution model are higher than those of the simulation models[13].

7. Summary and Conclusions

Given the prevalence of cost overruns in transportation infrastructure investments, it is pertinent to ask how these risks can be assessed prior to project's implementation. The main objective of this paper is to analyze this question relative to underlying risk sources, quantitative risk assessment methods and application. Following a discussion of the main risk factors in transportation investments, we define our key risk indicator, Cost Overrun Ratio (COR), as the ratio of the project's actual costs to its planned budget. We then present three risk estimation models: a distribution fitting model, a regression model and

[13] For medium road and highway projects the sample size N=10. Hence the reliability of this result is uncertain.

a simulation model. Subsequently, we applied these models to a database composed of highway, tunnels and bridges projects, which were implemented in Vancouver Island, Canada, between the years 1993 and 2003.

Key results from the analysis relate to probability intervals estimates of the predicted COR, given the projects' budget and type (small, medium and large roads and highways, or bridges and tunnels). The major conclusion, therefore, is that these methods provide reasonable estimates for the risk of COR, and are readily applicable to most common transportation investments PPP included.

Two caveats to the above analysis are warranted. First, the risk of costs overrun, estimated at the planning stage of the project, may be different than that estimated at the implementation phase. As more information becomes available, the estimation of the risk of COR becomes more accurate. Hence, risk analysis should be a continuous process, which encompasses the project's entire life cycle from the initial planning stage, through construction, operation and maintenance, major upgrading and maturity.

A second caveat relates to the use of the methods presented in this paper for *ex ante* risk analysis of some unique projects. As was shown, the application of these methods is predicated on historical data, describing projects with similar attributes such as project type (highways, rail), construction conditions, economic environment and investment make-up (private, public and PPP). Yet, some projects may have unique characteristics, mainly due to the use of advanced technologies, new design or sheer magnitude (the cost of a one mile of a new subway line ranges from $200-$1 billion, depending on location and type of burrowing). Under such conditions, historical data may be of little use for estimating the risk of COR for projects of this nature.

This qualification, notwithstanding, rarely a new project is absolutely so unique that it becomes inapt for risk analysis on the basis of historical data. And even if this indeed is the case, it might be possible to divide the new project into two components: those, which are amenable to the kind of risk analysis shown above and those that are not. For example, the Canada Line (also known as the RAV line) is a new investment project in Vancouver, BC, which connects the city of Richmond, the Airport and downtown Vancouver using above ground and underground passenger rail line. At least the above ground and station construction components, which constitute a substantial portion of the entire investment, can be subjected to risk analysis, which is based on historical data. The underground component, which has no precedent in BC, may require a different risk analysis approach. This issue is the focus of further research on the subject.

With these caveats in mind, the main conclusion from this paper is that managers can vastly improve risk analysis of newly planned transportation infrastructure investments by using methods such as those expounded in this paper. Reliance solely on personal experiences and intuition might result in extensive costs overruns like those reported in the literature.

Finally, it has been frequently been argued that the private sector is more efficient in planning and executing transportation infrastructure projects. In this regard an interesting question for future research is whether projects done by the private sector exhibit lower rates of cost overruns. If so, policies favoring PPP type projects should be pursued.

References

Akintoye, Akintola S & Malcolm J MacLeod, *Risk analysis and management in construction*, International Journal of Project Management Vol. 15, No. 1, pp. 31-38, 1997

Ayyub, Bilal M. *Risk Analysis in Engineering and Economics*, Chapman & Hall/CRC, 2003

Boardman, Anthony E & Greenberg, David H. *Cost-benefit Analysis: Concepts and Practice*, Prentice Hall, 1996

Chapman, Chris, *Project risk analysis and management--PRAM the generic process*, *International Journal of Project Management* Vol. 15, No. 5, pp. 273-281, 1997

Cho, Hyo-Nam et al. *A risk assessment methodology for incorporating uncertainties using fuzzy concepts*, Reliability Engineering and System Safety 78 (2002) 173–183, 2002

Cooper, Dale & Chapman, Chris, *Risk Analysis for Large Projects: Models, Methods & Cases*, John Wiley & Sons, 1987

Flyvbjerg, Bent, et al. *Megaprojects and Risks: An Anatomy of Ambition*, Cambridge, 2003

Flyvbjerg, Bent, et al. *What Causes Cost overruns in Transport Infrastructure Projects*, Transport Reviews, Vol. 24, No. 1, 3–18, January 2004

Frame, J. Davidson *Managing Risk in Organizations: A guide for managers*, Jossey-Bass, 2003

Grimsey, Darrin & Lewis M. K, *Evaluate the Risks of Public Private Partnerships for Infrastructure Projects*, International Journal of Project Management, Vol 20, pp. 107-118, 2002

Grimsey, Darrin & Lewis, M. K. *Public Private Partnerships: The Worldwide Revolution in Infrastructure Provision and Project Finance*, Edward Elgar, 2004

Jaafari, Ali, *Management of risks, uncertainties and opportunities on projects: time for a fundamental shift*, International Journal of Project Management 19 (2001) 89-101, 2001

Kartam, Nabil A & Kartam, Saied A, *Risk and its Management in the Kuwait Construction Industry: a Contractor's Perspective*, International Journal of Project Management, vol.19, pp. 325-335, 2001

Lyons, Terry & Skitmore, Martin, *Project risk management in the Queensland engineering construction industry: a survey*, International Journal of Project Management 22 (2004) 51–61, 2004

Ministry of Employment and Investment, Province of British Columbia, *Risk Analysis for Public Private Partnership*, 1995

NCHRP, 2006, "Guidance for Estimation and Management for Highway Projects During Planning. Programming and preconstruction", Final Report, Project 8-49,

Pickrell d., 1989, "Urban Rail Transit Projects" Forecast versus Actual Ridership and costs", Cambridge Massachusetts, US Department of Transportation, Transportation Systems Center

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

Pouliquen, Louis Y, 1970, *Risk Analysis in Project Appraisal*, World Bank Staff Occasional Papers Number Eleven, John Hopkins Press,

Public Financial Management, Inc. *Final Report Intergovernmental Financial Risks of New Rail Start Projects*, DTUM60-87-C-71343 Task 4, February 7, 1989

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

Shapira, Zur, *Risk Taking: A Managerial Perspective*, Russell Sage Foundation, 1994

Sturk, R., et al. *Risk and Decision Analysis for Large Underground Projects, as Applied to the Stockholm Ring Road Tunnels, Planning and Development*, S0886-7798 (96) 00019-3, 1996

Titarenko B P, 1997, *Robust technology' in risk management*, International Journal of Project Management Vol. 15, No. 1, pp. 11-14,

Vermeulen, Erik M. et al. *Analyzing risk and performance using the multi-factor concept*, European Journal of Operational Research 93 (1996) 173-184, 1996

Williams, Terry, *A classified bibliography of recent research relating to project risk management*, European Journal of Operational Research 85 (1995) 18-38, 1995

Wooldridge, Jefferey M, 2000, *Introductory Econometrics: A Modern Approach*, Thomson South-western,

Ministry of Employment and Investment, Province of British Columbia, *Risk Analysis for Public Private Partnership*, 1995

Appendix A: The Beta Distribution

A Beta probability function has the following properties:

- It has finite limits on the lower and upper bounds of the studied variables.
- It can be asymmetric. In reality many variables are right or left skewed
- It is a flexible distribution that and can assume different shapes.

The probability density function of beta distribution is:

$$f(x) = \frac{(x-A)^{P-1}(B-x)^{Q-1}}{B(P,Q)(B-A)^{P+Q-1}}, A \le x \le B; P, Q > 0 \qquad \text{(A1)}$$

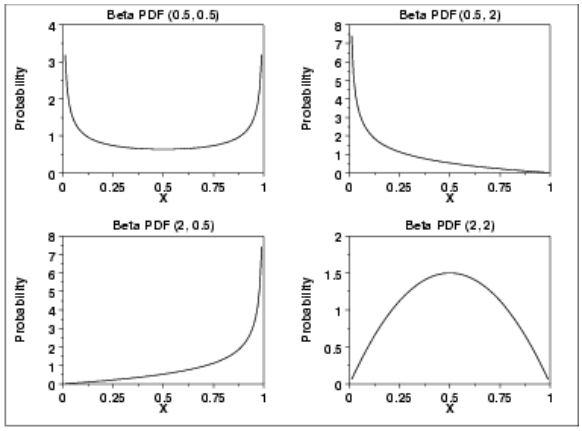
In this paper, x , is the event of COR accepting a certain value. In (A1) A and B are the distribution's lower and upper bounds, respectively. P and Q are the shape parameters. Beta distributions with different P and Q values have different probability density functions. $B(P,Q)$ is the beta function, i.e.,

$$B(P,Q) = \int_0^1 x^{P-1}(1-x)^{Q-1} dx \qquad \text{(A2)}$$

The function can have different shapes depending on the values of the P and Q parameters:

- $P < 1, Q < 1$, Beta is U-shaped
- $P < 1, Q \ge 1$ or $P = 1, Q > 1$, Beta is strictly decreasing
- $P = Q = 1$, Beta has a uniform distribution
- $P = 1, Q < 1$ or , Beta is strictly increasing
- $P > 1, Q > 1$, Beta is unimodal, which has a single local maximum
- If $P = Q$, the density function is symmetric about 1/2.

The following figure shows different Beta probability density functions for different P and Q values.



The mean, variance and mode are calculated as follows:

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

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

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

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

NCSS, the statistical package used in this study requires that we manually set the Beta distribution parameters. 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.

Appendix B: Computation of Quintiles in Distribution Probability Plot in NCSS

Let us assume that we have a set of numbers x1, x2, ..., xn and we wish to visually study whether the normality assumption is reasonable. The basic method is:

1. Sort the xi's from smallest to largest. Represent the sorted set of numbers as x(1), x(2), ..., x(n) . Hence, x(1) is the minimum and x(n) is the maximum of these data.

2. Define n empirical quantiles, p1, p2, ..., pn, where pi = i/n. These are similar to percentiles. For example, if n = 5 the pi's would be .2, .4, .6, .8, 1.0. The p2 value of .4 is interpreted as meaning that this is the 40th percentile.

3. Find a set of numbers, z1, z2, ..., zn, that would be expected from data that exactly follows the normal distribution. For example, z2 is the number that we would expect if we obtained 5 values from a normal distribution, sorted them, and selected the second from the lowest. These are called the quantiles.

4. Construct a scatter plot with the pairs x(1) and z1, x(2) and z2, and so on. If the xi's came from a normal distribution, we would anticipate that the plotted points will fall along a straight line. The degree of non-normality is suggested by the amount of curvature in the plot.