1-1-2010

A competitive bidding decision-making model considering correlation

Hesam Bahman-Bijari
Ryerson University

A COMPETITIVE BIDDING DECISION-MAKING MODEL CONSIDERING CORRELATION

By

Hesam Bahman-Bijari

Bachelor of Engineering, Civil Engineering,

Islamic Azad University of Mashad, Iran, Sep/2007

A thesis

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Applied Science

in the Program of

Civil Engineering

Toronto, Ontario, Canada, 2010

I hereby declare that I am the sole author of this thesis or dissertation.

I authorized Ryerson University to lend this thesis or dissertation to other institutions or individuals for the purpose of scholarly research.

Signature ....

I further authorize Ryerson University to reproduce this thesis or dissertation by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Signature ....

A COMPETITIVE BIDDING DECISION-MAKING MODEL

CONSIDERING CORRELATION

Hesam Bahman-Bijari

Master of Applied Science

Department of Civil Engineering, Ryerson University
2010

Abstract

Many general contractors obtain a majority of projects based on the low-bid award system. A major objective of a competitive bidding model is to determine an optimum markup size so as to maximize the contractor's long term profit. The new bidding model with explicit consideration of correlation is proposed since this important parameter is not considered in existing bidding models. In this study, the existence of a positive correlation coefficient between any two competitors' bid ratios was demonstrated. After that, a new competitive bidding model was proposed, and a statistical method in a Bayesian framework was developed. The significance of correlation on probability of winning and optimum markup decisions was investigated. For an illustration purpose, a case study of a bidding Data Set from actual projects was conducted. It has been found that as correlation increases, the probability of winning will increase, and hence an increased optimum markup can be used. In comparison with Friedman and Gates models, the proposed model with consideration of correlation coefficient derives different value of optimum markup which is closer to the real situation of the construction market since the correlation coefficient among competitors' bid ratios is considered.

Acknowledgments

First and foremost, I would like to thank my supervisor professor Arnold Yuan for not only providing me with the opportunity to undertake this research but also his supportive guidelines and insights.

I would also like to thank all supportive team of civil engineering department at Ryerson University to provide a productive research environment.

I appreciate my family for their continuously supports and understanding during the completion of this project.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background

Nowadays different types of project delivery systems are used in construction industry. Depending on the nature of a project and the project owner's in-house design and management capacity, the owner may choose design-bid-build (DBB), construction management (CM), design-build-operate-maintain (DBOM), or other innovative contracting methods such as private-public partnership (Halpin, 2006). In public sector, however, the DBB system based on competitive bidding is still the most commonly used project delivery system.

For the DBB mode, the project owner, or client, first acquires a consultant (depending on the type of project, this can be an architect or a civil engineer; can be called consultant in general). The consultant starts with a preliminary design. After a few rounds of interactions with the client, the consultancy services culminate in the final design of project and bid documents. The client then puts an advertisement for tendering; sometime it sends out invitations for bidding to those major contractors to which it has maintained a close relationship. When a general contractor (G.C.) receives the invitation, it needs to decide whether to bid or not. If yes, then the next step is to purchase the bid documents, study them, negotiate with and acquire necessary subcontractors, estimate the cost of the project, and then submit the bid with a properly determined markup size adding upon the estimated project cost. After bids are received, the client will evaluate the bids and determine the winning bidders based on merits of the bid. Thereafter, a construction contract is entered and signed, and then the construction will commence. Several criteria can be used for evaluation of the bids, for example, bid price, quality, schedule, performance specification, historical project records, or a combination of them. However, the lowest price is the commonly used criterion for bid evaluation, although quite often a prequalification process is added to screen out those contractors that have had very poor performance records and client relationships.

In the competitive bidding environment in which the bid price is the major factor, determination of the markup size is a very important, strategic task for the general contractors. As Park and Chapin (1992) have rightfully put, “the markup should be low enough to win the job, and yet high enough to earn a profit from the project.” Many factors affect the markup decision. First, the purpose of participating the bidding is an important one. In a typical situation, the G.C. participates to win the job in order to maximize its overall long-term profit. In some other situation, barely for a strategic purpose, the G.C. may enter the bidding game to just to lower the winning bid price and thus minimize its competitors’ profit. Sometimes, the G.C. submits a bid in order just to show its interest and existence in the market (the so-called discretionary bid). Obviously, with different bidding purposes, the markup decisions would be very different. The second important factor that affects the markup size is the nature of project. In particular, the complexity of and uncertainties involved in the project. The more complex and uncertain is a project, the greater the markup should be to countermeasure the cost contingency. Finally, the market competitiveness is also a very important factor that needs to be considered when furnishing the final bid price. The market competitiveness also includes many intriguing factors such as overall economical condition, health of local construction industry, supply chain of human and material resources, and more specifically, the number of potential bidders that are interested in winning the project. To sum up, the markup decision is important and yet delicate.

This study focuses on the markup decision in competitive bidding in which the bidder of the lowest bid price is the winner among all competitors.

1.2 Motivation

Competitive bidding modelling is a classical problem in both operations research and construction management. The first discussion of the problem, as R.M.Stark (1979) pointed out, can be traced back to 1944 by D.Emblen his PhD dissertation dealing with competitive bidding technique for securities, but the most cited work in competitive bidding modelling goes to Friedman (1956). Since then there have been many studies devoted to this problem, resulting in several well-known mathematical models: Friedman model, Gates model (1967), Carr model (1982), and Skitmore model (1994), just name a few. With the wide applications of computers in construction project management, many decision support systems by using artificial intelligence, neural network and/or fuzzy logic have also been developed later. These models and tools attacked the problem from different aspects. Unfortunately, the intensive research has not yet been accompanied by wide applications of those models in practice. According to several surveys conducted over the past two decades, the percentage of using formal statistical and optimization tools in markup sizing is still in a very low level of 10% (Ahmad, *et al.*, 1988; Fayek, *et al.*, 1999; Hegazi and Moselhi, 1995).

One of the reasons for the non-responsiveness of the industry to the research is that the competitive bidding model might not be realistic enough. Although most of the mathematical models suggest studying historical bidding data to characterize the competitors' bidding behaviours, none of those models considered the statistical correlation relationship among the data which will be discussed in Chapter 2. This observation motivates the study of the effect of correlation on markup sizing, and integrates it into the competitive bidding model.

1.3 Objectives of the Research

The objective of the study is to develop an innovative probabilistic model for markup decision making with consideration of the statistical correlation among bid data. Particularly, rationale behind the fact that why correlation must exist between any pairs of competitors' bid data from both mechanistic and empirical points of view should be investigated. If the correlation does exist indeed, then an innovative model is developed which considers the effects of correlation on the decision of markup size. Moreover, considering the common feature of the historical bid data that often involves missing data requires a development of a new effective and efficient statistical method to characterize and estimate the correlations.

1.4 Research Methodology

The methodology used in this study commences with analysing the overall bidding procedure, decisions involved therein, bid components and their uncertainty factors, which are followed by a literature review on existing bidding models. The common problem of the existing models is identified to be the lack of consideration of correlation among bids. The rationale behind the existence of correlation among bid ratios is then explored and for a new statistical method for estimating the correlation with consideration of missing data is proposed. In order to investigate the significance of correlation coefficient on different components of bidding models, a sensitivity analysis is performed, in which the major existing models are also evaluated in comparison with the proposed model. A case study is used to illustrate the proposed model and statistical methods and to verify the assumptions that have been made for the model.

1.5 Scope of the Research

The scope of this study is limited to the lowest-price bid award system in construction procurement through competitive bidding. Although several other objectives and criteria can be used, sometimes simultaneously, for the markup decision making, this study considers only the long-term profit as the sole objective of the markup decision. That is, the markup size is determined so as to maximize the bidder's long-term profit. Finally, the study belongs to a decision-theoretic approach. It is not the intention of this research to expand it to a game-theoretic study, although it is not impossible in principle.

1.6 Thesis Outline

The rest of the study is organized as follows. Chapter 2 serves as a literature review, in which decisions involved in bidding procedure and uncertainty factors of bid components will be introduced and various competitive bidding models proposed by Friedman, Gates, Carr are reviewed. In Chapter 3, a mechanistic argument for existence of correlation among the bid data is provided at first. Then a statistical approach to estimating the correlation coefficient among bids is developed in Section 3.3 based on Bayesian method. A new probabilistic bidding model is proposed with consideration of statistical correlation among bids in Chapter 4. Since the new model involves a multi-dimensional integral for the probability of winning, the computational techniques for evaluating the integral are also discussed. Using the new model, the effects of correlation on bid decisions are also investigated through a sensitivity analysis. In Chapter 5 a case study of an early published bidding data set is conducted and results from the Friedman, Gates, and the new model are compared. Chapter 6 summarizes the study with major conclusions drawn and recommendations for future researches made.

A list of mathematical notations and acronyms can be found in the Appendix C. As a general rule of notations in the study, a Greek letter is used to denote a parameter, an uppercase Latin letter to denote a random variable, and the corresponding lower case Latin letter to denote an observation of the random variable. The word “Contractor” is used to mean the decision maker, “Competitor” to mean the bidder that the Contractor needs to compete in a project, and “bidders” to mean all the bidders that participate in a project tendering.

Chapter 2 Literature Review

This chapter review a bidding procedure and major decisions involved in bidding process. After that, uncertainty factors involved in bidding is going to introduce. Friedman, Gates and Carr models of optimum markup decision making will be explained in this chapter.

2.1 Decisions Involved in Competitive Bidding

The general contractor (G.C.) should make different decisions in competitive bidding procedure. This section mainly focuses on bidding procedure and decisions involved in bidding procedure. The first decision is whether or not to bid and the second which is the concern of this study is a bid price. These decisions are complicated as consequences of both are uncertain and lots of uncertainty factors are involved (Shash, 1995).

2.1.1 Bidding Procedure and Cost Estimating

A competitor should follow a procedure for competitive bidding in order to advance the efficiency of bidding preparation as shown in Fig.2.1.

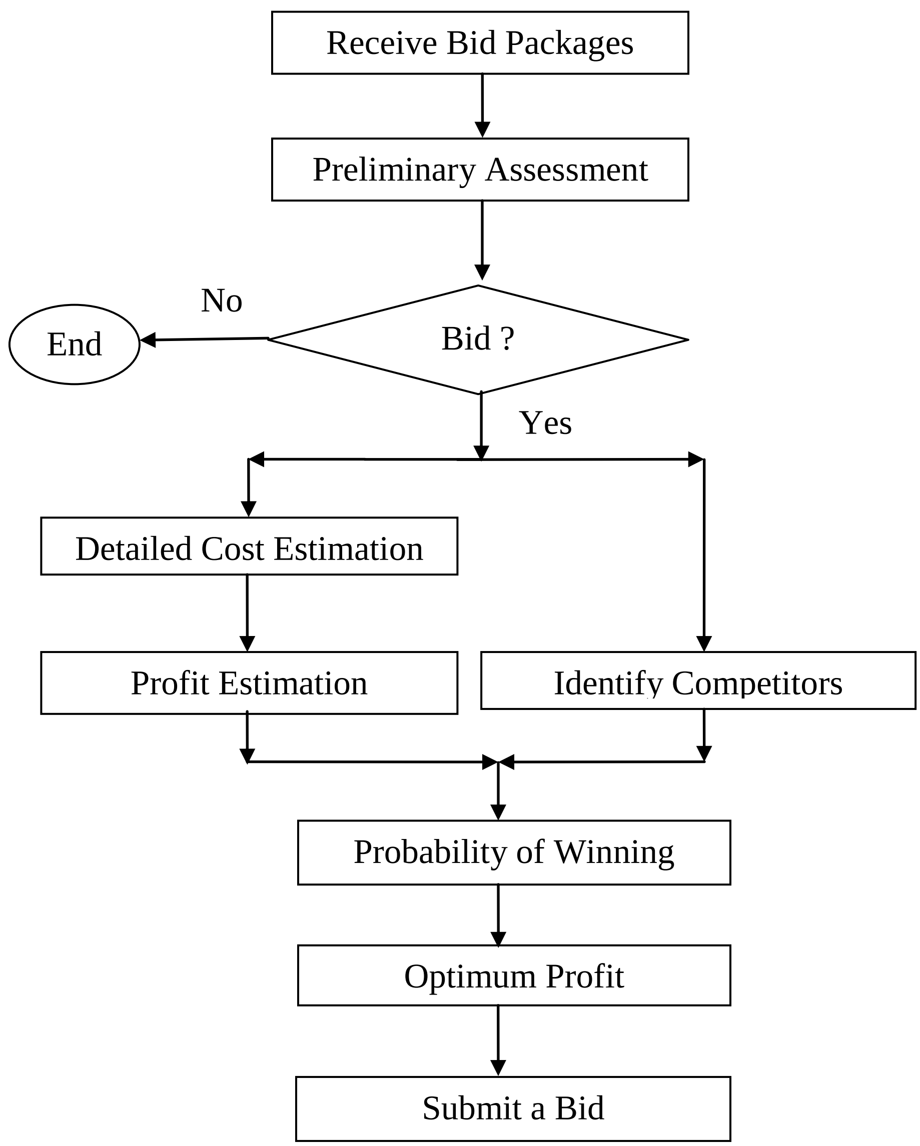

Fig.2.1: Bidding Procedure

As the process of bidding preparation has a considerable cost, mainly the cost of estimator teams which have a significant non-refundable fees, preliminary analysis stage attempts to identify an unprofitable tendering situations named the decision to bid or not. For bidding situations, the estimated cost of construction and probability of winning for specific bids should be conducted. The decision of bid price should be made and submitted at tendering date (Ward & Chapman, 1988).

Components of cost estimating in tendering based on released plan packages are included direct and indirect costs. As shown in Fig.2.2, direct cost is mainly included cost of materials, labors and equipments and indirect cost include project and general overhead.

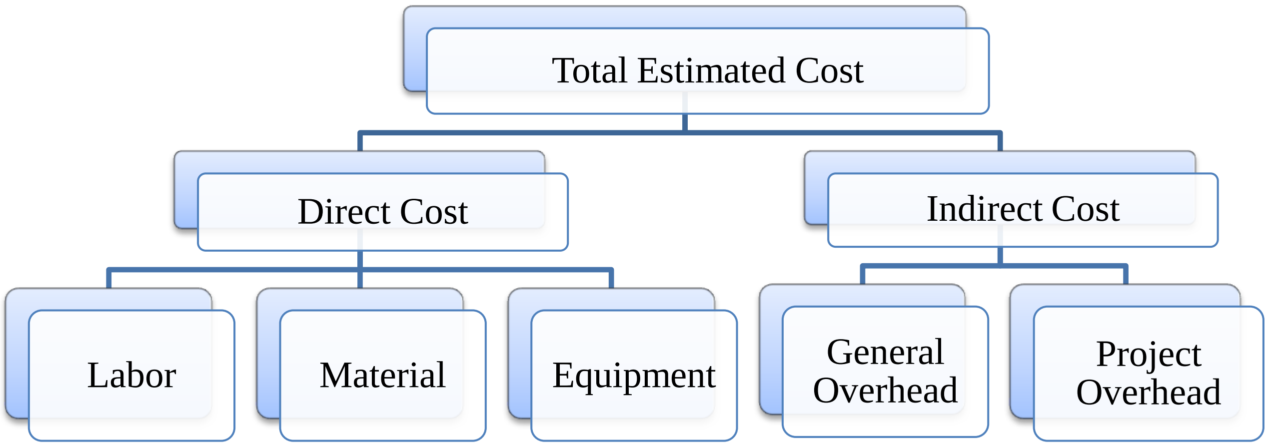

Fig.2.2: Estimated Cost Components

Direct cost of a project based on released plan packages which is available for general contractors can be done based on quantity take offs and unit prices. Detailed cost estimation is a method in which most of contractors (92%) have been used to estimate direct cost of a project (Hegazi & Moselhi, 1995).

2.1.2 Bid or Not Bid

In order to bid for a project, expectations of a general contractor should be satisfied which is different from one competitor to another. As a process of finalizing bid packages has a considerable non-refundable estimation cost for the general contractor, therefore this decision should be considered carefully by the bidding staff after invitation of contractors for a released project. There are several factors that influence this decision such as project type, job mix, degree of financial risk, location, bidding climate, chance of winning and reputation of the client. The AHP (Analytic Hierarchy Process) method is one of the decision making tools which can be used by weighting different criteria based on competitors' bidding policies in order to make a decision to bid or not for different released project (Holm, 2005; Moselhi & Hegazy, 1993; Chua & Li, 2000).

Fig.2.3 shows top 10 factors influencing decision to bid based on a survey of bidding practices of Canadian civil engineering construction contractors (Fayek *et al.*, 1999).

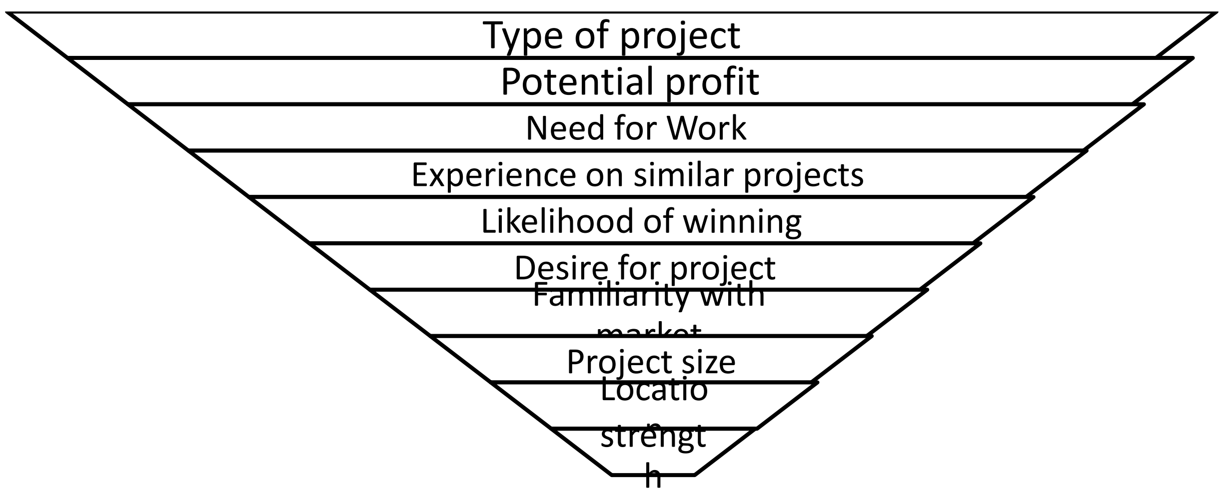

Fig.2.3: Top 10 Factors Influencing Bid Decision

If the general contractor desires to submit a bid for a released project, then the second major decision which is a bid price decision has to be made. In this study, the focus is the case in which a contractor passed a first major decision as to bid for a project, therefore, a second major decision which is the decision of the bid price should be conducted by the general contractor.

2.1.3 Markup Size

After a general contractor decide to bid on a project, a bid price which consists of two major parts, total estimated cost and markup, should be performed at this stage. The direct cost of material, labor and equipment through quantity takeoffs and unit prices should be estimated and added to the indirect costs of over head, financial interests, contingency and tax (Wallwork, 1999). In addition to the total estimated cost, profit should be added to the total estimated cost to get a bid price (*b*). As a result of this fact that contractors mostly use the same methods of cost estimation and share the same pool of material resources for the project, therefore total estimated cost of different contractors for the same project is close to each other, therefore a critical part of a bid price selection which affects a winner of a project (The lowest submitted bid price) is markup which should be added as a profit to the total estimated cost of a project. Therefore, the second contractors' major decision which is bid price decision turns into the markup size decision making. As contractor's definitions of markup are different from one to another, the following graph shows a portion of contractors with different definitions of markup based on a survey done by Hegazi and Moselhi (1995). As shown in the following figure, markup has different definitions among general contractors, however, the majority of contractors define markup as a profit, and therefore this study follows the same definition which has been used among the majority of contractors for markup which is the profit.

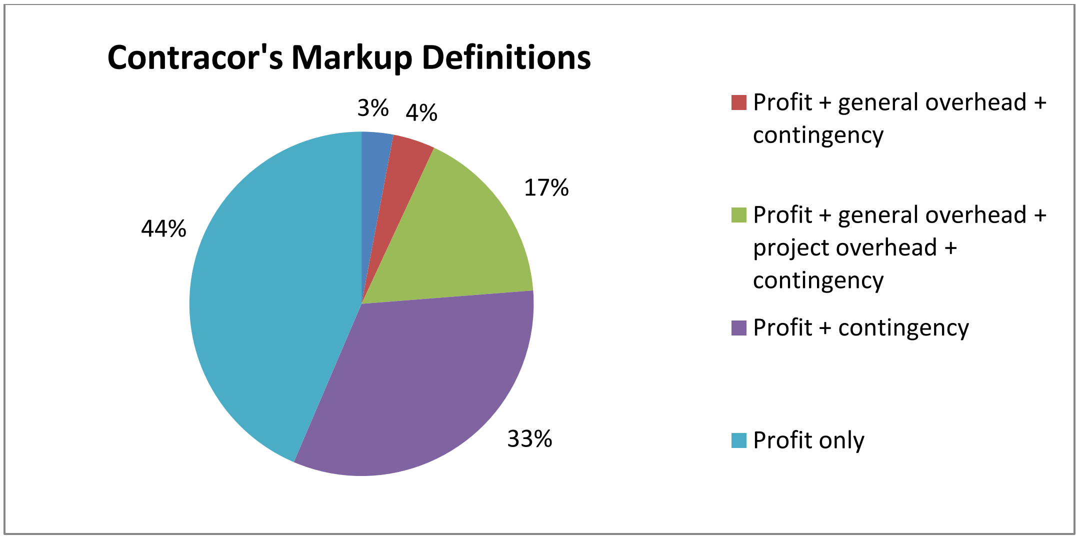

Fig.2.4: Different Markup Definitions

The most efficient factors which have affected markup decisions are degree of risk, difficulty of the project and risk of investment. The following figure shows top 10 factors which influences a markup rate decision (Shash 1992; Ahmed 1988).

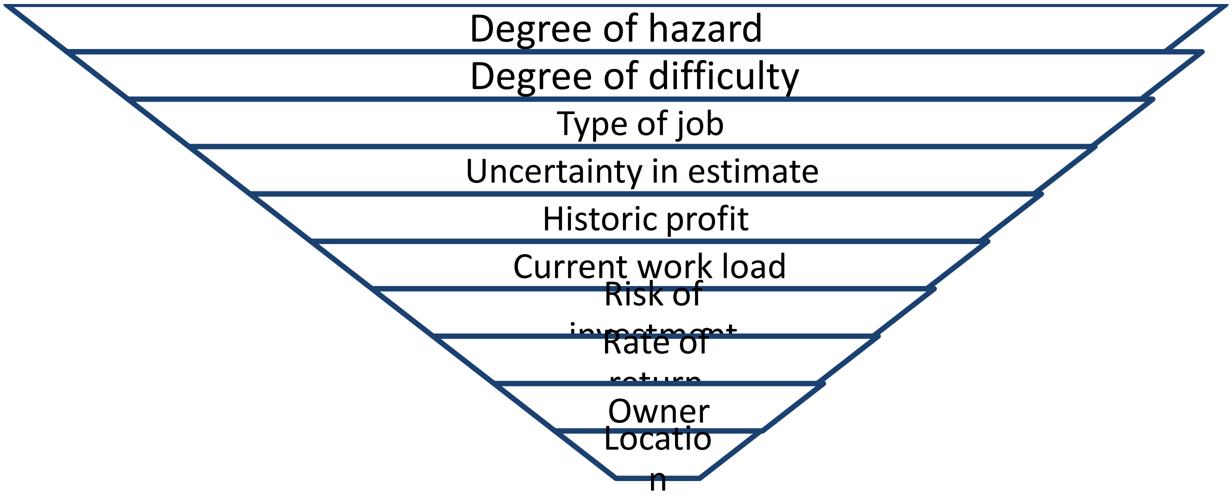

Fig.2.5: Factors Influencing Markup Rate Decision

As the major concern of this study is determination of Contractor bid price (Markup rate selection), therefore, uncertainty factors of bid components and associated models should be considered in following sections.

2.2 Uncertainties Involved in Bidding

As the main target of this study is involved in bidding procedure, therefore identification of uncertainties which are involved in bidding procedure requires to be investigated. In this section, bid components associated uncertainty factors introduced.

2.2.1 Cost Estimating Uncertainty Factors

Uncertainties associated with cost of a project can be reviewed in two stages, first is a pre-construction during a planning stage (Estimated cost uncertainties) and the second one is during construction stage (Actual cost uncertainties). Uncertainty factors associated to a direct estimated cost are quantity takeoff errors which can be caused by estimator or plans scale errors and construction cost uncertainties can be caused by unpredicted conditions on site and equipment's breakdown. The indirect cost estimating errors have been produced by factors such as inflation, regulation changes and management interferences. All of these abovementioned uncertainties result in the project's financial risk and contingency is one of the key factors for general contractors in order to reduce these financial risks (AbouRizk & Halpin, 1994; Holm, 2005; King & Mercer, 1988). As construction industry subjected to the variety of factors which leads uncertainties in cost estimating during planning and construction stages, therefore the traditional deterministic approaches improved toward the probabilistic methodologies to resolve limitations of deterministic cost estimating. The range-estimating methodology will result in an appropriate cost estimating which include the level of confidence for the cost of a project and as a result of this fact the estimator can predict the revelation of project financial risks (Back *et al.*, 2000; King and Mercer, 1988). In order to find an appropriate probability density function for each cost items, the numerical simulation methods like Monte Carlo simulation can be used to fit the selected distribution to the sample data. The statistical methods for parameter estimating based on an assigned distribution for a data set are maximum likelihood, least square minimization and moment matching method. Since markup is the most efficient bid component as an indicator of the winner of the project rather than estimated cost, therefore the primary focus of this study is on details of optimum markup decision making models (AbouRizk & Halpin, 1994; Schexnayder *et al.*, 2005).

2.2.2 Markup Uncertainty Factors

Markup has a major impact on winning of a project rather than estimated cost based on a fact that estimated cost is mostly the same among competitors, therefore, factors which leads uncertainties in markup decision making should be considered carefully. This decision can be those bid prices which have optimum balance between a bid price that is as practically low as possible to win the job and at the same time as practically high as possible to make maximum profit in long run for a general contractor (Shash, 1992).

Major markup uncertainty factors are competitiveness (number of competitors), overall construction industry economy, project type and size of a project (based on dollar amount). These factors should be considered to suggest an optimum markup to the Contractor which maximizes a long term profit of Contractor.

As the focus of this study is on markup decision making models therefore mathematical models such as Friedman, Gates and Carr are going to be introduced in following sections. In this study the focus is on first two models (Friedman and Gates) and these two models will be evaluated and results will be compared with proposed model in order to illustrate the effects of consideration of correlation coefficient among bid ratios in proposed model which has not been considered in existing bidding models.

2.3 Friedman Model

Friedman's model is one of the first mathematical based bidding models which was developed in 1956. This model is generated for cases which contractors are invited to submit their bids in order to win a job based on the lowest submitted bid and performs a method which has the purpose of maximizing an expected profit, *EP*. The expected profit can be defined as the multiplication of two values, probability of winning (P_w) over competitors and markup (m), $EP = P_w \times m$.

The markup rate which maximizes an expected profit should be calculated. Determination of probability of winning, P_w, is the most important part in calculation of *EP* and the optimum markup.

If Contractor has enough submitted historical bidding data from previous projects of our competitors and its own estimated cost for those projects (as Contractor participate on those projects) then the "bidding pattern" of our competitors can be derived. Assume a situation in which competitor *A* competes against Contractor. Since enough historical bidding data of competiror *A* and estimated costs of Contractor is available for previous projects, therefore, the ratios of competitor *A's* bid prices to Contractor cost estimates (c_0) on those previous projects they have participated can be derived, $X_A = B_A/c_0$, known as a bid ratio. The estimated cost of the Contractor, C_0, and competitor *A's* bid prices, B_A, of previous projects in which they both participated is known for the Contractor as the Contractor has recorded its own estimated cost and competitor bid prices from those previous projects which they participated.

Competitor *A* bid ratios follow a specific distribution which is a probability density function (PDF) of X_A values known as bidding pattern of competitor *A*. Fig.2.6 is used for an illustration purpose for competitor *A*'s bidding pattern.

The probability of being the lowest bid by submitting the bid price, b_0, with a bid ratio of $X=b_0/c_0$, is the area on the right side of the bid ratio probability distribution curve of the competitor *A* as shown in Fig.2.6. If Contractor knows the identity of competitors who are going to submit their bods, then competitors' bidding pattern curves derived and hence the probability of winning over each of them equals the areas on the right hand side (hatched area) of the bid ratio $X=b_0/c_0$ under each competitor's bidding distribution curve.

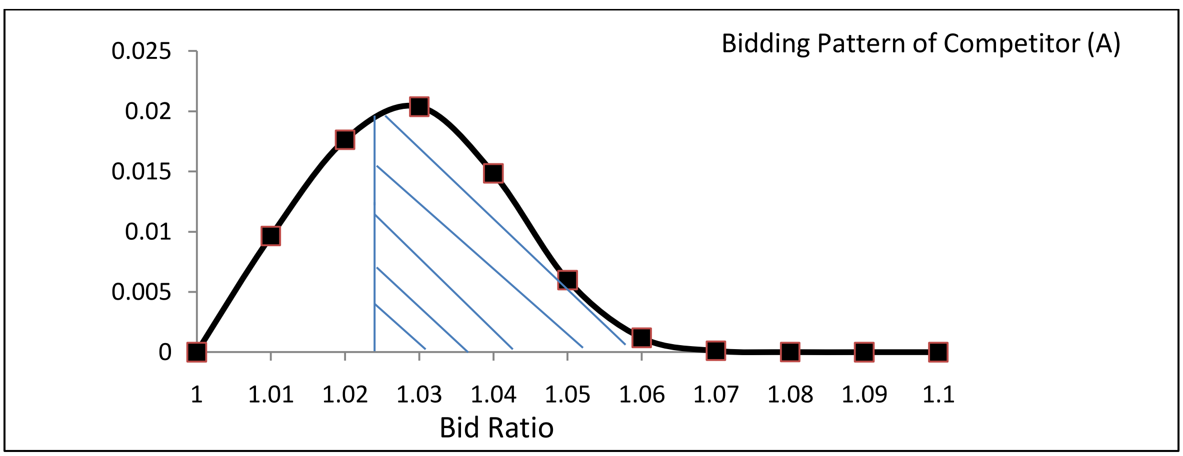

Fig.2.6: Competitor *A* Bidding Pattern Curve

Fig.2.7 shows the case when Contractor is going to bid against two competitors *A* and *B*. The same as the previous case the bidding pattern of competitor *B* can also identified and based on Friedman's assumption of independency among bid ratios of *k* competitors, $X_i=B_i/C_0$, (i=1...k) the probability of winning over all K competitors is the product of the probabilities of winning over each of them, which known as the general Friedman's model and can be shown as follows:

$$P_w = P\left[\frac{B_1}{c_0} > b_0/c_0\right] P\left[\frac{B_2}{c_0} > b_0/c_0\right] \dots P\left[\frac{B_K}{c_0} > b_0/c_0\right] \tag{2.1}$$

$$= \prod_{i=1}^{K} P\left[\frac{B_i}{c_0} > b_0\ /c_0\right] = \prod_{i=1}^{K} \int_{b_0}^{\infty} f\left(\frac{B_i}{C_0}\right) dx$$

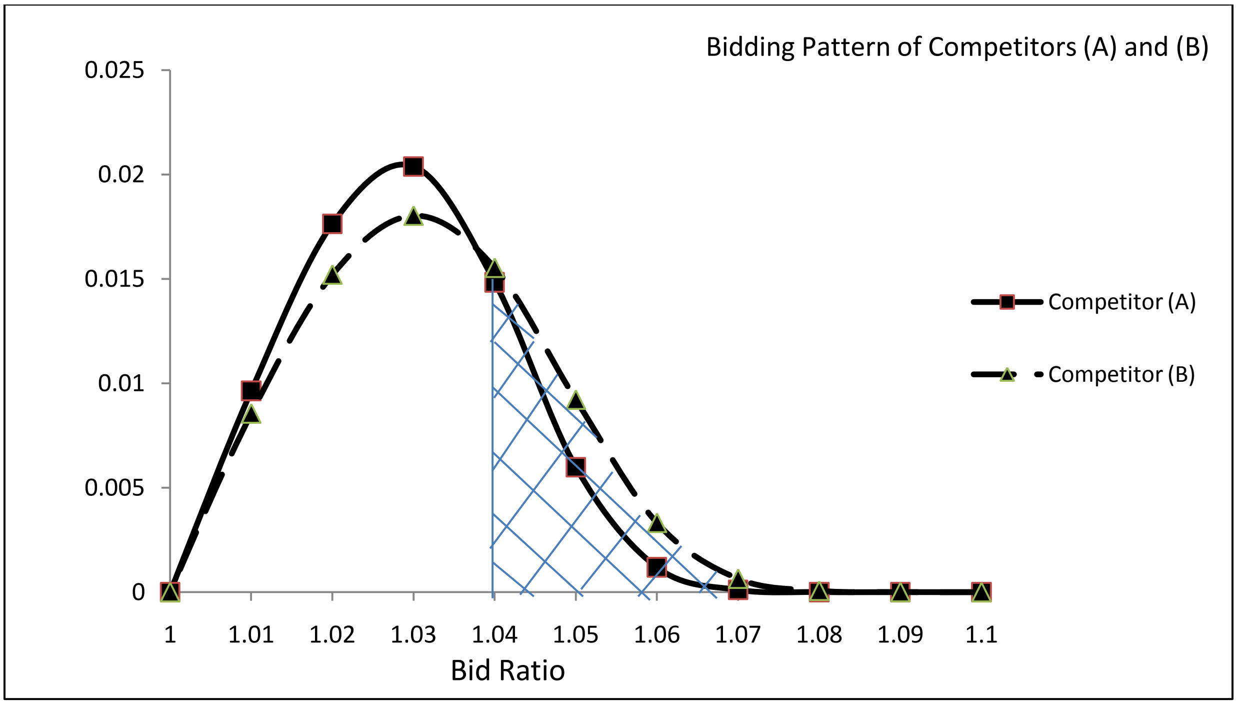

Fig.2.7: Probabilities of Winning Over Competitor *A* and *B*

As the model's purpose is to maximize the expected profit and since the value of probability of winning for different markup rates is available then the expected profit curve can be obtained and by visual observation of expected profit curve, the optimum matkup which maximizes an expected profit can be determined (Friedman, 1956).

2.4 Gates Model

In 1967, Gate's model has been published in order to find an optimum markup rate for a contractor which maximizes a value of expected profit based on statistical and numerical analysis.

In this model the author has obtained the relation between markup and probability of being the lowest bid among other competitors' bids based on historical bid prices of competitors. This model derived a probability of winning over each individual competitor, P_{wi}, for different markup rates. An overall probability of winning over *k* competitors, P_w, known as "All-Bidders-Known strategy", has been introduced as follows:

$$P_w = \frac{1}{1 + \sum_{i=1}^{k} \frac{1 - P_{wi}}{P_{wi}}} \tag{2.2}$$

Expected profit, *EP*, can be calculated by multiplication of P_w for specific markup rates and markup rates. As a result, relation between expected profit, *EP*, and markup, *m*, can be conducted and by maximization of expected value, *EP*, in terms of markup, *m*, value of optimum markup, m_{opt}, can be estimated.

The major difference between Friedman and Gates model comes from the method they conduct a probability of winning over all competitors which will be investigated in Chapters 4 and 5.

2.5 Carr Model

In 1982 Robert Carr introduced a model which is applicable in cases which distributions of contractors' estimated cost for a specific facility, C_i, and opponent's bids can be estimated. The estimated costs of competitors are distributed about the following value

$$\bar{C} = \frac{\sum_{i=1}^{k} C_i}{k} \tag{2.3}$$

in which $\bar{C}$ is the average of estimated costs of k competitors for a specific project. Therefore the standardized estimated cost of contractor *i* for that project, $\acute{C_\iota}$, can be calculated as follows

$$\acute{C_\iota} = \frac{C_i}{\bar{C}} \tag{2.4}$$

However, the only values that contractor *i* knows is his own estimated cost, C_i, and his competitor's bid price for the same project, B_j, therefore the contractor can compare the competitor's bids to his estimated cost as follows

$$\left(\frac{B}{C}\right)_{ij} = \frac{B_j}{C_i} = FBC_j \frac{C_j}{C_i} \tag{2.5}$$

As can be seen, the *B/C* random variable depends on competitor's *j* bid and cost of contractor *i*. As the ratio of $\left(\frac{B}{C}\right)_{ij}$ is for a specific project therefore this value for *N* project in which both contractors *i* and *j* have been attended is distributed about

$$MBC_{ij} = \frac{\sum_{z=1}^{N} \left(\frac{B}{C}\right)_{ijz}}{N} \tag{2.6}$$

where MBC_{ij} is the mean of *B/C* ratio of competitor's *j* bid to the estimated cost of contractor *i*. In addition, for the case of all competitors against contractor *i* the mean of B/C ratio has distributed around

$$MBC_i = \frac{\sum_{z=1}^{N} \sum_{j=1}^{n_z} \left(\frac{B}{C}\right)_{ijz}}{\sum_{z=1}^{N} n_z} \tag{2.7}$$

where MBC_i is the mean of *B/C* ratio of all competitors against contractor *i* for *N* projects which contractor *i* participates.

As a result of above calculation, standardized distributions of estimated cost of contractor *i*, bids of competitor *j* and the ratio of $(B/C)_{ij}$ can be derived. The probability that $\left(\frac{B}{C}\right)_{ij}$ will exceed *b* as the value of standardized estimated cost of contractor *i*, $\acute{C}_i$, is not known, given by:

$$P_w = P\left[\left(\frac{B}{C}\right)_{ij} > b\right] = \int_{-\infty}^{\infty} f\left(\frac{B_j}{C_i}\right) dx = \int_{-\infty}^{\infty} f(\acute{C}_i)\, P(\acute{B}_j > b\acute{C}_i) dx \tag{2.8}$$

$$= \int_{-\infty}^{\infty} f(\acute{C}_i) \int_{b\acute{C}_i}^{\infty} f(\acute{B}_j) dx dx$$

Which is the area on the right side of the value of *b* under $f\left(\frac{B_j}{C_i}\right)$.

As the lowest bid price is the winner of the project, therefore the lowest $\left(\frac{B}{C}\right)_{ij}$, which is called LBC_{ij} , should exceed the value of *b*. This probability is known as the general Carr bidding model and can be written as follows

$$P_w = P\big(LBC_{ij} > b\big) = \int_{-\infty}^{\infty} f(\acute{C}_\imath)\left\{\prod_{j=1}^{k}\left[\int_{bC_i}^{\infty} f(\acute{B}_J)\,dx\right]\right\}dx \qquad (2.9)$$

In this model $f(\acute{B}_J)$ and $f(\acute{C}_\imath)$ are independent. Although this model is not limited to assumptions such Friedman and Gate models do, but in order to observe the function of standardized estimated cost of contractor i, $f(\acute{C}_\imath)$, it has some difficulties as the contractor i, doesn't have the information of competitor's cost estimation based on the bidding history of competitors (Carr, 1982).

2.6 Skitmore Model

Skitmore utilized a multivariate approach in bidding model and in this model the Contractor can incorporate data of all of projects in which competitors participated in order to estimate the parameters of distribution. Skitmore model used multivariate approach in order to make a better use of available data in order to estimate parameters.

If, $f(y_1 y_2 y_3 \cdots y_n)$ represents the joint probability density function of y_i , $i = 1,2, \dots, n$ which are log transformed bids, $y_i{\sim}LN(\mu_i, {\sigma^2}_i)$, for a specific project treated as continuous random variables then the probability of winning based on skitmore model expressed as follows

$$\Pr(y_1 < y_i) = \int_{y_1=-\infty}^{\infty}\int_{y_2=y_1}^{\infty}\int_{y_3=y_1}^{\infty}\cdots\int_{y_n=y_1}^{\infty} f(y_1 y_2 y_3 \cdots y_n)\,dy_n \cdots dy_3 dy_2 dy_1 \quad (2.10)$$

Skitmore's approach uses an assumption of independency among all variables and therefore the probability of winning can be shown as follows:

$$\Pr(y_1 < y_i) = \int_{-\infty}^{\infty} f_1(y_1) \prod_{i=2}^{n} \left\{ \int_{y_i=y_1}^{\infty} f_i(y_i) dy_i \right\} dy_1 \tag{2.11}$$

As it can be recognized from Skitmore model, although he used a multivariate approach in his proposed model, but he did not consider the correlation coefficient as a parameter of interest since he made an assumption of independency (Skitmore, 1994).

As shown through Sections 2.3 to 2.6, none of the existing models consider the existence of the correlation coefficient among bid ratios in their proposed models for calculation of the probability of winning over competitors, therefore in order to fulfill this lack of consideration of correlation coefficient among bid ratios this study is going to propose a new model with consideration of correlation coefficient. In this study, the focus is on two major bidding models, Friedman and Gates, and the result of these two models will be compared with proposed model by parametric study in Section 4.3 and case study in Chapter 5.

Chapter 3 Correlation among Bid Ratios and Estimation Techniques

In this chapter, first the concept of correlation in probability and statistics is reviewed in Section 3.1 and then in Section 3.2 the rationale behind an existence of correlation among bid ratios is developed. A Bayesian statistical method for estimating the correlation coefficient among bid ratios from historical bidding data is proposed in Section 3.3.2.

3.1 Concept of Statistical Correlation

Correlation is an important concept in probability and statistics to measure statistical dependence relationship between two random variables. Two random variables are called positively correlated if one variable increases statistically when the other variable increases. The term 'increase statistically' here means that there is no guarantee that the former variable will increase for certainty; rather, it only says that the first variable will increase on average with the increase in the second one. On the other hand, the two random variables are called negatively correlated if one decreases statistically (or on average) when the other increases. In the bidding context, two competitors would be called positively correlated if one competitor would often decrease its bid when the other competitor lowered its bid.

In probability, the correlation relationship between two random variables is measured by a dimensionless quantity called correlation coefficient. Specifically, the correlation coefficient of random variables X and Y is defined as

$$\rho = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}} \tag{3.1}$$

where $\mathrm{COV}(X, Y)$ denotes the covariance of the two random variables, and $\mathrm{var}(X)$ the variance of X. Recall that $\mathrm{var}(X) = \mathrm{E}[X - \mathrm{E}(X)]^2$ and $\mathrm{cov}(X, Y) = \mathrm{E}\big[\big(X - \mathrm{E}(X)\big)\big(Y - \mathrm{E}(Y)\big)\big]$. The correlation coefficient scales the covariance by the standard deviation of each variable. Consequently, the correlation is a dimensionless quantity that can be used to compare relationships between pairs of variables.

The correlation coefficient defined above is also called Pearson's correlation or Pearson's rho (ρ) in statistics, to differentiate other measures of correlation, for example Pearman's tau (τ), a rank-based measure of correlation (Montgomery & Runger, 2003; Rodgers & Nicewander, 1988). When a random sample of the variables, $(x_i, y_i, i = 1, \ldots, n)$, is available, the Pearson's rho is estimated as

$$\rho = \frac{1}{s_X s_Y}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \tag{3.2}$$

in which $\bar{x}$ and s_X are the sample mean and sample standard deviation of X, respectively; and the same notations are used for Y. As it can be readily shown, the Pearson correlation ranges from -1 to +1. It is +1 for a perfect, increasing linear relationship, and -1 for a perfect, decreasing linear relationship. When the correlation coefficient is zero, the two random variables X and Y are called uncorrelated.

3.2 Rationale behind Positive Correlations among Competitors' Bid Ratios

The bid ratio X_i of Competitor i is defined by the competitor's bid price, B_i, divided by the Contractor's cost estimate, C_0; in mathematical notations,

$$X_i = B_i / C_0 \tag{3.3}$$

The fact that there is a positive correlation between a pair of competitors' bid ratios will be shown in two steps. First, it will be explained that under the assumption of the competitors' bid prices being statistically independent, the bid ratios are positively correlated due to the common random Contractor's cost estimate, C_0. Then at the second step illustrate through cost estimating procedures that competitors' bids are often also positively correlated, with which the correlation coefficient between the bid ratios can be even higher.

First consider the correlation between X_i and X_j, the bid ratios of Competitors i and j, under the assumption that B_i, B_j and C_0 are independent one another. For the ease of presentation, further consider that bid prices and estimated cost are both following a lognormal distribution (This assumption will be investigated later in Case study). That is

$$B_i \sim LN\left(\mu_i, \sigma_i^2\right), \text{or } \log(B_i) \sim N\left(\mu_i, \sigma_i^2\right) \tag{3.4}$$

$$B_j \sim LN\left(\mu_j, \sigma_j^2\right), \text{or } \log\left(B_j\right) \sim N\left(\mu_j, \sigma_j^2\right) \tag{3.5}$$

and

$$C_0 \sim LN(\mu_0, \sigma_0^2), \text{or } \log(C_0) \sim N(\mu_0, \sigma_0^2) \tag{3.6}$$

The notation of $X \sim LN(\mu, \sigma^2)$ is used to denote that the random variable X follows a lognormal distribution with parameters μ and σ^2. The probability density function of a lognormal distribution is expressed as

$$f_X(x;\mu,\sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right) \tag{3.7}$$

Recall that the logarithm of a lognormal random variable is normally (Gaussian) distributed. Since a sum or subtraction of two normal random variables is still normally distributed, a product or division of two lognormal variables is still a lognormal variable. Therefore,

$$X_i \sim LN(\mu_i - \mu_0, \sigma_i^2 + \sigma_0^2) \text{ and } X_j \sim LN(\mu_j - \mu_0, \sigma_j^2 + \sigma_0^2) \tag{3.8}$$

The covariance of the logarithm of these two variables is

$$\mathrm{cov}[\log(X_i), \log(X_j)] = \mathrm{E}[(\log(C_0) - \mu_0)^2] = \sigma_0^2 \tag{3.9}$$

Hence, the correlation coefficient between log (X_i) and log (X_j) is

$$\rho_{ij} = \frac{\sigma_0^2}{\sqrt{(\sigma_i^2 + \sigma_0^2)(\sigma_j^2 + \sigma_0^2)}} \tag{3.10}$$

which is strictly positive. By further assumption of $\sigma_i^2 = \sigma_j^2 = \sigma_1^2$, then

$$\rho_{ij} = \frac{\sigma_0^2}{\sigma_1^2 + \sigma_0^2} = \frac{1}{1 + \gamma^2} \tag{3.11}$$

in which $\gamma = \sigma_1/\sigma_0$.

Clearly, the correlation coefficient is strictly positive and its value depends on the variability in the bids and cost estimate. When the standard deviation of the bids (σ_1) is comparable to that of the contractor's cost estimate (σ_0), the correlation coefficient will be around 0.5. Only if the variability of the competitors' bids is much greater than that of cost estimate (hence $\gamma \gg 1$), the two bid ratios will then be close to be uncorrelated. As a numerical example, when γ equals 1.1, then ρ_{ij} equals 0.45.

Next consider the case which is closer to the reality; that is, the examination of the actual relationship of the bid prices and cost estimates in practice, and study its impact on the correlation among the bid ratios.

The bid price of a contractor, say B_i of Contractor i, consists of two major components: estimated cost (C_i) and markup (M_i). The estimated cost is further divided into direct cost (C_{Di}) and indirect cost (C_{Ii}), each of which consists of different subcomponents. Specifically, the direct cost includes four major subcomponents: material cost, labour cost and equipment cost of the work performed by the general contractor itself, and subcontract cost of the work done by subcontractors. The variability of the first three cost components are influenced by two major factors: quantity and unit price of the work. For a same project for which the design drawings and specifications are ready before tendering, the quantity takeoffs of different contractors should be of little difference. The unit price used for pricing the direct cost depends on material and equipment suppliers, construction methods the general contractor has selected, and management efficiency and productivity of the general contractor. Despite of sometimes substantial differences in construction methods as well as productivity between two general contractors, the uncertainties in material and equipment prices are general and they will affect any contractors in the same market. As a result, the material, labour and equipment costs should be positively correlated. Similarly, in a market economy, the general contractor is free to choose the subcontractors, and vice versa. So the subcontract costs of two general contractors usually have a positive correlation. Putting all of them together, therefore, the total direct costs of two different general contractors should be, at least weakly, positively correlated. The degree of correlation depends on the complexity of and percentage of subcontracting in the project. When the project is very complex, different contractors might use very different construction methods, resulting in very weak correlation in the direct cost estimates. On the other hand, when the project is of fairly routine nature and the subcontracting ratio is relatively high, most of the jobs are done by subcontractors. In this case, the direct costs would be strongly correlated. By no means should the direct costs be negatively correlated. Fig.3.1 explains that general contractors (G.C.) share the construction market as shown by thick arrows and illustrates that general contractors utilize the same sources for their cost estimation.

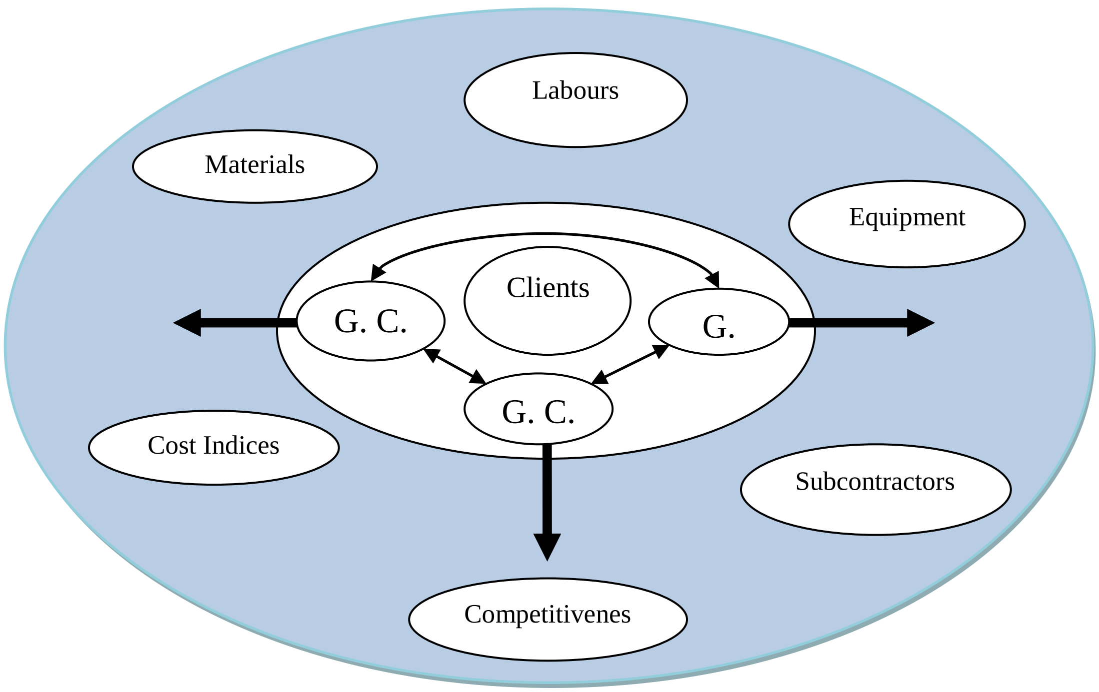

Fig.3.1: The Overall Construction Market Composition

The indirect costs of a project include mainly project overhead, general or company overhead, and financial costs (e.g., interest rates). Although the company overhead is largely dependent upon organizational structure of the general contractors, a common factor for variability of all of the indirect cost components is the total project duration, which is subject to both uncertainty and client's requirement of a certain deadline. Meanwhile, some of the indirect costs (e.g., cost incurred for safety measures) are also subject to a same set of municipal, and provincial and federal regulations. All of these contribute a positive correlation in the indirect cost estimates among different general contractors.

The last component of the bid price is the markup. As a strategic decision, the markup rates can be very different for different companies. The markup size depends on competitiveness (which can be measured by the number of participants in a tender), needs for jobs, and nature and strategic value of the projects to the contractors (Shash 1992; Ahmed 1988). Therefore, without actual data it is very hard to determine if the markup rates of different contractors are positively or negatively correlated or even just uncorrelated at all.

To sum up, the direct and indirect costs are positively correlated, while the markup seems to be anywhere in the range of correlation. However, the markup usually is a small portion of the total bid price. Consequently, the bid prices most likely are positively correlated. In rare cases when the overall construction industry is gloomy and some of the contractors needs job badly, might the bid prices appear a very weak positive correlation.

With the same assumption of lognormal distributions used previously, now assume further that $\log(B_i)$ and $\log(B_j)$ are correlated to each other with a coefficient r_1 and that $\log(B_i)$ and $\log(B_j)$ both are correlated to $\log(C_0)$ with a coefficient r_0. Note that $B_i = C_i M_i$. Hence the correlation coefficient r_0 represents the correlation between C_0 and C_i, as well as the correlation between C_0 and M_i, whereas r_1 represents the correlation between C_i and C_j, C_i and M_j, C_j and M_i, and M_i and M_j. If all of these components are positively correlated, then $r_1 > r_0$. If the markups are considered negatively correlated to each other, however, the previous relationship between two coefficients might not be true. Nevertheless, for both cases, the correlation between the logarithms of the two bid ratios X_i and X_j can be derived as follows.

Let $Z_1 = \log(X_i) = \log(B_i) - \log(C_0)$, and $Z_2 = \log(X_j) = \log(B_j) - \log(C_0)$. Then the variances of Z_1 is

$$\mathrm{var}[Z_1] = \mathrm{E}[\{(\log(B_i) - \log(C_0)) - (\mu_i - \mu_0)\}^2] \tag{3.12}$$

Therefore $\mathrm{var}[Z_1]$ can be simplified as

$$\begin{aligned} \mathrm{var}[Z_1] &= \mathrm{E}[(\log(B_i) - \mu_i)^2] - 2\mathrm{E}[(\log(B_i) - \mu_i)(\log(C_0) - \mu_0)] + \mathrm{E}[(\log(C_0) - \mu_0)^2] \\ &= \sigma_i^2 - 2r_0\sigma_i\sigma_0 + \sigma_0^2 \end{aligned} \tag{3.13}$$

Similarly,

$$\mathrm{var}[Z_2] = \sigma_j^2 - 2r_0\sigma_j\sigma_0 + \sigma_0^2 \tag{3.14}$$

For the covariance,

$$\begin{aligned} \mathrm{cov}[Z_1, Z_2] &= \mathrm{E}\big[(\mathrm{Z}_1 - (\mu_i - \mu_0))\big(\mathrm{Z}_2 - (\mu_j - \mu_0)\big)\big] \\ &= \mathrm{E}\big[(\log(B_i) - \mu_i)\big(\log(B_j) - \mu_j\big)\big] - \mathrm{E}[(\log(B_i) - \mu_i)((\log(C_0) - \mu_0)] \\ &\quad - \mathrm{E}\big[\big(\log(B_j) - \mu_j\big)((\log(C_0) - \mu_0)\big] \\ &\quad + \mathrm{E}[(\log(C_0) - \mu_0)^2] \\ &= r_1\sigma_i\sigma_j - r_0\sigma_0(\sigma_i + \sigma_j) + \sigma_0^2 \end{aligned} \tag{3.15}$$

Therefore, the correlation between the bid ratios is expressed as

$$\rho_{ij} = \frac{r_1\sigma_i\sigma_j - r_0\sigma_0(\sigma_i + \sigma_j) + \sigma_0^2}{\sqrt{(\sigma_i^2 - 2r_0\sigma_i\sigma_0 + \sigma_0^2)(\sigma_j^2 - 2r_0\sigma_j\sigma_0 + \sigma_0^2)}} \tag{3.16}$$

Again, when $\sigma_i^2 = \sigma_j^2 = \sigma_1^2$, it reduces to

$$\rho_{ij} = \frac{r_1\sigma_1^2 - 2r_0\sigma_1\sigma_0 + \sigma_0^2}{\sigma_1^2 - 2r_0\sigma_1\sigma_0 + \sigma_0^2} = \frac{r_1\gamma^2 - 2r_0\gamma + 1}{\gamma^2 - 2r_0\gamma + 1} \tag{3.17}$$

in which $\gamma = \sigma_1/\sigma_0$. It can be shown that when $r_1 > r_0 > 0$, ρ_{ij} is always positive and it is also greater than the value under the independence assumption of bid prices; however, this is not true if $r_1 < r_0$, as shown in Fig.3.2. For one case of the figure, $r_0 = 0.4$, which is less than r_1 shown in the range of the x-axis, the correlation between the bid ratios ρ_{ij} is higher than 0.45, which is the value corresponding to the case when $r_0 = r_1 = 0$ and is shown as the horizontal broken line. However, for the other case of $r_0 = 0.6$, ρ_{ij} is less than 0.45 when $r_1 < 0.6$. As discussed previously, r_1 is less than r_0 only if the markup rates are negatively correlated, or in other words, when the overall construction market is gloomy and the contractors badly need jobs to uphold a healthy account balance.

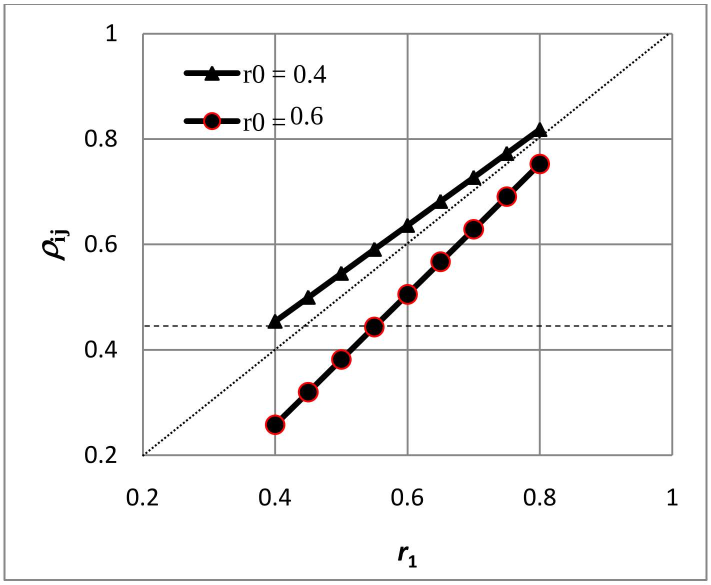

Fig.3.2: Effects of r_1 and r_0 on Correlation of Bid Ratios

3.3 Statistical Methods for Estimating Correlation

In elementary statistics, estimation of correlation coefficient from a sample of paired data $(x_i, y_i, i = 1, \ldots, n)$ is a fairly straightforward task as defined in Equation (3.2). The standard approach is to estimate the sample covariance, and then the correlation coefficient can be obtained by dividing the sample covariance by the sample standard deviations of the two variables. Unfortunately, the situation for the bidding data is more complex than this. Two major problems prevent one from using the abovementioned standard approach.

First, the estimation of correlation coefficients is very sensitive to outliers. An accurate estimation of the correlation coefficient usually requires a fairly large number of paired data, which are usually not available in historical bidding records. One approach to this limitation is to adopt a Bayesian perspective. The Bayesian method provides a nice machinery to include prior information and knowledge in statistical inferences and therefore can be used in small sample problems to improve the reliability of inferences.

The second problem, which is even more important, is: Rarely could one have a complete set of bidding records that allows an investigator to directly use the abovementioned estimation procedure. Quite often, all the competitors of interest did not participate in the historical projects. As a result, some bid ratios was not available. As shown in Table 3.1 below, some projects all of the three competitors participated and therefore the three bid ratios were available. For some projects, only two or even just one of them presented. In statistics, this kind of data is called missing data. An intuitive approach to dealing with the missing data is to neglect the records that are not complete. However, this is not practical, as the data size will reduce drastically if one uses only the complete records. As seen from Table 3.1, there is only one out of the seven projects has the complete records of the three competitors, and three projects have the complete records for Competitors 1 and 2. Therefore, an effective and efficient statistical method is desirable.

Table 3.1: Bidding Data with Missing Values

Project Number	Competitor 1	Competitor 2	Competitor 3
1	x_{11}	x_{21}	x_{31}
2	--	x_{22}	x_{32}
3	x_{13}	x_{23}	--
4	x_{14}	--	x_{34}
5	x_{15}	x_{25}	--
6	x_{16}	--	--
7	--	x_{27}	--
$\vdots$	$\vdots$	$\vdots$	$\vdots$

In this section, statistical methods for estimating the correlation of missing bidding data are discussed. Specifically, the positiveness of correlation, as discussed in the previous section, is introduced as prior knowledge in the parameter estimation in the Bayesian framework. Note that the Bayesian method discussed below is not the only choice. Alternative approaches include, for example, the maximum likelihood method, the EM (expectation-maximization) method, multiple imputation, and data augmentation (Little and Rubin, 2002). Since the likelihood function plays a central role in all of the methods, therefore the discussion of the likelihood function of the correlation coefficient for a missing data is introduced in following section.

3.3.1 Likelihood Function of Correlation for Missing Data

In general, a likelihood function is a function of the parameter or parameters of interest from a set of observed data. It is assumed that the data include information of the parameter(s) through a certain probabilistic model. Specifically, suppose that x_1, x_2,...,x_n are the observed values of a random variable Z which follows a probability distribution with a parameter θ, $f\,(z;\,\theta)$. Then the likelihood function is defined as

$$L(\theta|\boldsymbol{z}) = f(z_1, z_2, \dots, z_n; \theta) \tag{3.18}$$

in which $L(\theta|\boldsymbol{z})$ denotes the likelihood function of θ on the data $\boldsymbol{z} = (z_1, z_2, \dots, z_n)$, and $f(z_1, z_2, \dots, z_n; \theta)$ the joint probability density function of the data. Assume further that the data are independent random sample of Z. Then the likelihood function is reduced to the product of the marginal density function evaluated at the data, or

$$L(\theta|\boldsymbol{z}) = \prod_{i=1}^{n} f(z_i; \theta) \tag{3.19}$$

In essence, the likelihood function is just the probability density function of the data expressed in terms of the parameter. This concept of the likelihood function was developed in earlier 1920s by R. A. Fisher who proposed to estimate the parameters by maximizing the likelihood function, i.e., the maximum likelihood method (Casella and Berger, 1990; Montgomery, 2000).

A maximum likelihood estimate of the parameter, usually denoted by $\hat{\theta}$, is a value of θ that maximizes the log-likelihood function. Note that here the logarithm of the likelihood, usually denoted by the lower case $l(\theta|\boldsymbol{z})$, instead of likelihood function itself, is used for maximization. This is mainly for the sake of computational ease.

When the data are subject to missing values and the missing data are missing at random, the likelihood function can be defined based on only the observed values. Particularly, let $Z = (Z_{\text{obs}}, Z_{\text{mis}})$ denote the data that would occur in the absence of missing values, in which Z_{obs} denotes the observed data and Z_{mis} the missing data. Let $f(Z_{\text{obs}}, Z_{\text{mis}})$ denote the joint PDF of Z_{obs} and Z_{mis}. Then the likelihood function of the parameter is just the marginal distribution of the observed data; i.e.,

$$L(\theta|Z_{\text{obs}}) = \int f(Z_{\text{obs}}, Z_{\text{mis}}; \theta) dZ_{\text{mis}} \tag{3.20}$$

The parameter can then be estimated by maximizing the log-likelihood based on the observed data. Note that the discussion here is based on the 'missing at random' assumption, which holds true for the case of competitive bidding. For more detailed discussion of the missing mechanisms, readers are referred to Little and Rubin (2002, Section 6.2).

The maximum likelihood method can be used to estimate the correlation coefficients among bid ratios of competitors, provided that complete data sets of the bid ratios are available. For this, consider two competitors, say 1 and 2, of which the bid ratios of n projects all are known. Take the same assumption as made in Section 3.2 that the two bid ratios follow a bivariate lognormal distribution. Then the logarithms of the bid ratios follow a bivariate normal distribution with five parameters: $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$. Instead of using the raw bid ratio data, $z_{ij} = \log(x_{ij})$, $i = 1, \dots, n$ and $j = 1,2$, can be used to construct the likelihood function. The log-likelihood function of the parameter vector $\boldsymbol{\theta}$ is thus expressed as

$$l(\boldsymbol{\theta}|\boldsymbol{z}) = \sum_{i=1}^{n} \log f(z_{i1}, z_{i2}; \boldsymbol{\theta}) = -\frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{z}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}) \qquad (3.21)$$

where $f(z_{i1}, z_{i2}; \boldsymbol{\theta})$ is the PDF of the bivariate normal distribution; $\boldsymbol{z}_i = (z_{i1}, z_{i2})^T$; $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$; and $\boldsymbol{\Sigma}$ is the 2×2 covariance matrix, which is written as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \qquad (3.22)$$

In which $\sigma_{12} = \rho\sigma_1\sigma_2$ and $\sigma_{jj} = \sigma_j^2$ for $j = 1, 2$. Maximization of the log-likelihood function in Eq. (3.21) leads to the maximum likelihood estimates (*m.l.e.*) as the follows:

$$\hat{\mu}_j = \bar{z}_j = n^{-1}\sum_{i=1}^{n} z_{ij}\,; \;\; \hat{\sigma}_{jk} = s_{jk} = n^{-1}\sum_{i=1}^{n}(z_{ij} - \hat{\mu}_j)(z_{ik} - \hat{\mu}_k) \qquad (3.23)$$

for $j, k = 1,2$. The *m.l.e.* for the correlation coefficient is

$$\hat{\rho} = \hat{\sigma}_{12}/\sqrt{\hat{\sigma}_{11}\hat{\sigma}_{22}} \qquad (3.24)$$

Little and Rubin (2002) discussed a special case of missing data for the bivariate normal distribution. In particular, the missing data occurs only in one of the variables, say z_2. So the observed data include a complete set of paired data (z_{i1}, z_{i2}) for $i = 1, \dots, r$ and for the remaining $(n - r)$ cases only z_{i1} for $i = r + 1, \dots, n$. It has been shown by using the factorization technicque that the estimation of correlation coefficient can be shown as follows (Little and Rubin, 2002; pp.135-137):

$$\hat{\rho} = \frac{s_{12}}{\sqrt{s_{11}s_{22}}}(\hat{\sigma}_{11}/s_{11})^{1/2}(s_{22}/\hat{\sigma}_{22})^{1/2} \tag{3.25}$$

in which

$$s_{jk} = r^{-1}\sum_{i=1}^{r}\left(z_{ij}-\bar{z}_j\right)(z_{ik}-\bar{z}_k) \text{ for } j,k = 1,2 \tag{3.26}$$

$$\hat{\sigma}_{11} = n^{-1}\sum_{i=1}^{n} z_{i1} \tag{3.27}$$

$$\hat{\sigma}_{22} = s_{22} + \left(\frac{s_{12}}{s_{11}}\right)^2 (\hat{\sigma}_{11} - s_{11}) \tag{3.28}$$

Note that if only the complete data are considered, the correlation coefficient equals $s_{12}/\sqrt{s_{11}\, s_{22}}$. With consideration of the missing data in Z_2, this estimation should be adjusted by the factor of $(\hat{\sigma}_{11}/s_{11})^{1/2}(s_{22}/\hat{\sigma}_{22})^{1/2}$. When the missing percentage is very small, the impact of missing data on the correlation is also minor. However, in bidding data, the missing percentage is usually fairly considerable. Neglecting the missing data would therefore lead to a biased estimation for the correlation coefficient.

In the general case of missing data as shown in Table 3.1, however, the analytical solution for the correlation coefficient is hard to obtain. In this study, the general missing data in a bivaraite normal distribution situation is considered. Since the major concern is the uncertainty of the correlation coefficient, the likelihood function for this parameter on the observed bid ratio pairs is then

$$L(\rho|Z_{\text{obs}}) = \text{const} \times \prod_{i=1}^{r} f(z_{i1}, z_{i2}|\rho) \prod_{i=r+1}^{q_1} f(z_{i1}|\rho) \prod_{i=r+1}^{q_2} f(z_{i2}|\rho) \tag{3.29}$$

where Z_{obs} is the observed bidding data in the logarithm form, including three parts: a complete set of r pairs (z_{i1}, z_{i2}) for $i = 1, \ldots, r$, q_1 observations of Z_1 only, and q_2 observations of Z_2 only. Note that $n = r + q_1 + q_2$. Since the last two components of Equaltion (3.29) are marginal distributions and independent of ρ, the likelihood function can be simplified as follows

$$L(\rho|Z_{\text{obs}}) \propto \left(\frac{1}{\sqrt{1-\rho^{2}}}\right)^{r} \times \exp\left\{-\frac{1}{2(1-\rho^2)} \sum_{i=1}^{r} \left[\frac{\left(z_{i1} - \hat{\mu}_{z_1}\right)^2}{\hat{\sigma}_{z_1}{}^2} - 2\rho \frac{\left(z_{i1} - \hat{\mu}_{z_1}\right)\left(z_{i2} - \hat{\mu}_{z_2}\right)}{\hat{\sigma}_{z_1}\hat{\sigma}_{z_2}} + \frac{\left(z_{i2} - \hat{\mu}_{z_2}\right)^2}{\hat{\sigma}_{z_2}{}^2}\right]\right\} \tag{3.30}$$

in which $\hat{\mu}_{z_1}$, $\hat{\mu}_{z_2}$, $\hat{\sigma}_{z_1}$ and $\hat{\sigma}_{z_2}$ are estimates of the mean and variance based on $(r + q_1)$ and $r + q_2$ observations of Z_1 and Z_2, respectively.

3.3.2 Bayesian Method for Estimation of Correlation

As mentioned in Section 3.3, the bidding data is usually scarce. The occurrence of missing data results in even less information for the correlation coefficient. Bayesian method supplements further information in the statistical estimation through a prior distribution, denoted by $\pi(\theta)$, based on author's understanding of the correlation coefficient. Indeed, as discussed in Section 3.2, the correlation coefficient is usually positive because of the common denominator in the bid ratios as well as the fact that the competitors share the same industry market. With consideration of the range of the correlation coefficient, this prior knowledge can be represented by a probability distribution spanned over the interval from 0 to 1, for example, a beta distribution

(Gross, 1995). In this study, the standard uniform distribution is an appropriate choice for the correlation coefficient as it provides the least prior information (or non-informative), i.e.,

$$\pi(\rho) = \begin{cases} 1, & \text{for } 0 \leq \rho \leq 1 \\ 0, & \text{otherwise} \end{cases} \tag{3.31}$$

From Bayesian perspective, all unknowns, including the parameter ρ, are treated as random variables, and the estimation of the correlation coefficient is represented by a posterior distribution, denoted by $p(\rho)$. Based on Bayes' theorem, the posterior distribution of ρ after observing the data Z_{obs} is

$$p(\rho|Z_{\text{obs}}) = \frac{\pi(\rho)L(\rho|Z_{\text{obs}})}{f(Z_{\text{obs}})} \tag{3.32}$$

Since the denominator is independent of the parameter and can be considered as a normalization factor of the posterior distribution, the posterior distribution is proportional to the product of the prior distribution and the likelihood function of the parameter; i.e.,

$$p(\rho|Z_{\text{obs}}) \propto \pi(\rho)L(\rho|Z_{\text{obs}}) \tag{3.33}$$

Point estimate of ρ can be obtained as measures of the center of the posterior distribution, such as posterior mean, median, or mode. In this study, the posterior mean is used as an estimation of correlation coefficient among bid ratios:

$$E_p[\rho] = \frac{\int_0^1 \rho\, L(\rho)\, \pi(\rho)\, d\rho}{\int_0^1 L(\rho)\, \pi(\rho)\, d\rho} = \frac{\int_0^1 \rho\, L(\rho)\, d\rho}{\int_0^1 L(\rho)\, d\rho} \tag{3.34}$$

Since $\pi(\rho)$ is a uniform distribution over [0, 1], the term $\pi(\rho)$ can be eliminated and hence the second equality in the above equation. For the integration, numerical quadrature techniques can be used.

Chapter 4 Proposed Bidding Model

4.1 The Model

In the competitive bidding setting, the data that the contractor has prior to bid opening are its own estimated cost (c_0) of the project under tendering, and some historical records of its own cost estimates ($C_{0i},\ i = 1, \dots, n$) and competitors' bidding prices ($B_{ki}, k = 1, \dots, K$) for previous projects they have participated. The key to an optimal markup decision making is the characterization of the potential competitors' bidding behaviour. This is done by studying the historical bidding records (B_{ki}) with the Contractor's own cost estimates (C_{0i}) as the reference point.

As discussed in Chapter 2, debates on competitive bidding modeling over the past four to five decades focused on the treatment of uncertainties in cost estimate, albeit an intense and yet fruitless debate over Gates' mysterious formula for probability of winning. Not very long after Friedman published his seminal paper, a few researchers (e.g., Capen et al., 1971; Fuerst, 1976 and 1977; Ioannou 1988) started to realize the importance of the uncertainties in cost estimate and its impact on markup size decision. Later Carr (1982) developed a so-called 'general bidding model' with an explicit consideration of cost estimate uncertainty. However, he considered the uncertainty of cost estimate only in calculation of the probability of winning, and not in the expected profit. More importantly, he adopted a set of very strong assumptions before he could decompose the variance of cost estimate from the variance of the bid ratio, the latter being the only directly estimable quantities. Particularly, he assumed that the variance of cost estimate equals the variance of bid. This amounts to assuming that the markup is a fixed, deterministic value. Obviously this is not true.

Some other researchers attempted to directly include the uncertainty of cost estimate in the markup decision model. For example, a commonly used model for the expected profit is expressed as

$$\psi(m) = \int (mc_0 - c)\Pr(B_1 > mc_0, B_2 > mc_0, \dots, B_n > mc_0)\, f_{C_0}(c_0)dc_0 \tag{4.1}$$

in which $\psi(m)$ is the expected profit as a function of the markup factor m; and c is the actual project cost. Since the actual project cost is unknown prior to the project, it is usually replaced by the mean value of the cost estimate c_0. Unfortunately, this model has only theoretical value. It is not practically operational. The reason is simple: there is no data for one to establish the probability distributions of the competitors' bid prices themselves. The only operational approach to developing the probability distributions of the competitors' bids is through the modeling of the bid ratios. Friedman and Gates took this approach, but they both ignored the important factor of correlation among the bid ratios. The novelty of the new model proposed next is the explicit consideration of the correlation effects through a multivariate distribution.

Assumptions associated with the proposed model are as follows:

1) The bid ratios, X_k $(k = 1, \dots, K)$, follow a multivariate lognormal distribution.
2) The competitorss' bidding behaviour is stationary; that is, the competitors will bid for the current project in the same manner as they used to.
3) The Contractor has an unbiased cost estimate.
4) The number of competitors and their identity is known for the project under tendering.

Similar to other competitive bidding models, the objective of the proposed bidding model is to determine an optimal markup factor (m) so as to maximize the Contractor's profit in the long run. Therefore, the objective function is the expected profit expressed as

$$\psi(m) = (mc_0 - c)P_w(m) \tag{4.2}$$

where $P_w(m)$ denotes the probability of winning when the Contractor bids at a markup factor of m on top of the cost estimate c_0. With the assumption of unbiased estimate, i.e., $c = c_0$, the normalized expected profit can be used as

$$\psi(m) = (m - 1)P_w(m) \tag{4.3}$$

Therefore, the key issue of the model is the probability of winning. First recall that a multivariate normal (or Gaussian) distribution has a joint probability density function (PDF) expressed as

$$f(\boldsymbol{z}) = (2\pi)^{-\frac{K}{2}}(\det(\boldsymbol{\Sigma}))^{-\frac{1}{2}}\exp\left\{-\frac{(\boldsymbol{z}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})^T}{2}\right\} \tag{4.4}$$

in which $\boldsymbol{z} = (z_1, z_2, \dots, z_K)^T$ is a K-dimensional column vector; $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and covariance matrix, respectively, of the random vector $\boldsymbol{Z} = (Z_1, Z_2, \dots, Z_K)^T$; and $\det(\cdot)$ denotes the determinant operator of a matrix. In short hand, $\boldsymbol{Z} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes for the multivariate normal random vector with mean $\boldsymbol{\mu}$ and covariance$\boldsymbol{\Sigma}$. The covariance matrix is expressed as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1k}\sigma_1\sigma_k \\ & \sigma_2^2 & \cdots & \rho_{2k}\sigma_2\sigma_k \\ & & \ddots & \vdots \\ sym. & & & \sigma_k^2 \end{bmatrix} \tag{4.5}$$

A multivariate lognormal distribution has a joint PDF as

$$f(\boldsymbol{x}) = (2\pi)^{-\frac{K}{2}}(\det(\boldsymbol{\Sigma}))^{-\frac{1}{2}}\left(\prod_{k=1}^{K} x_k\right)^{-1} \exp\left\{-\frac{(\log(\boldsymbol{x}) - \boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\log(\boldsymbol{x}) - \boldsymbol{\mu})^T}{2}\right\} \tag{4.6}$$

It can be readily shown that $\log(\boldsymbol{X}) \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In other words, the logarithm of a multivariate-lognormally vector follows a multivariate normal distribution.

The probability of the Contractor winning the K contractors at the bid price of $b_0 = mc_0$, or equivalently, the probability of b_0 being the lowest bid among all (K+1) bidders including the Contractor itself is

$$P_w(m) = \Pr\left(\bigcap_{k=1}^{K} B_k > b_0\right) = \Pr\left(\bigcap_{k=1}^{K} X_k > m\right) \tag{4.7}$$

Unlike many previous models, the joint events represented by $\bigcap_{k=1}^{K}(X_i > m)$ are not independent. Under the assumption of the multivariate lognormal distribution for X_k, the probability of winning can be evaluated by a multivariate normal distribution function through the abovementioned logarithm transformation. That is,

$$P_w(m) = \int_{y_1}^{\infty} \cdots \int_{y_K}^{\infty} \phi_K(\boldsymbol{z};\, \boldsymbol{\rho}) dz_n \cdots dz_1 = \Phi_K(-\boldsymbol{y};\, \boldsymbol{\rho}) \tag{4.8}$$

in which $y_k = (\log(m) - \mu_k)/\sigma_k$; and $\phi_K(\cdot; \boldsymbol{\rho})$ and $\Phi_K(\cdot; \boldsymbol{\rho})$ denote the joint PDF and CDF of a standard multivariate normal distribution with correlation matrix of

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1k} \\ & 1 & \cdots & \rho_{2k} \\ & & \ddots & \vdots \\ sym. & & & 1 \end{bmatrix} \tag{4.9}$$

As can be observed, the probability of winning over K competitors now becomes a multiple integration of a joint multivariate probability density function, instead of a product of univariate distribution functions in Friedman model, or a so-called 'empirical' formula as proposed by Gates.

Evaluation of the multivariate normal distribution function is not an intractable task nowadays. In fact, for a low-dimension case where K is no more than 10, the commonly used mathematical computer software MATLAB can easily find the solution by using the recently introduced function 'mvncdf'(Matlab Help Documentation). For a high-dimension problem, this function is not efficient. In this case, some other special-purpose algorithms are ready. For example, the product of conditional marginals (PCM) approach developed can be used (Pandey 1998, Yuan and Pandey, 2006). In a very special case where all the correlation coefficients in the correlation matrix are equal, i.e., $\rho_{ij} = \rho$ for $i, j = 1, \dots, K$, the probability of winning can be reduced to the following univariate integral:

$$P_w = \int_{-\infty}^{+\infty} \phi(u) \times \prod_{k=1}^{K} \Phi\left(-\frac{y_k + u\sqrt{\rho}}{\sqrt{1-\rho}}\right) du \tag{4.10}$$

in which ϕ and Φ are the standard normal PDF and CDF, respectively.

4.2 Implementation of the Proposed Model and Statistical Method

The detailed steps for the implementation of the proposed model and statistical method can be summarized below:

1) Congregate Competitor's bid prices B_k and Contractor estimated cost C_0 from historical available bid data set.

2) Calculate Bid ratios, $X_k = B_k / C_0$, of competitors.

3) Transform bid ratios to $\mathrm{z} = \log(\boldsymbol{X})$.

4) Calculate an estimation of μ and σ based on observed transformed bid ratios for each competitor.

4) Calculate likelihood function, $L(\rho|\boldsymbol{z})$, based on Equation (3.30).

5) Calculate ρ based on Equation (3.34).

6) Calculate covariance matrix based on Equation (4.5) by using estimation of standard deviation and correlation coefficient from steps 4 and 5.

7) Calculate probability of winning based on Equation (4.8) by using Matlab code.

8) Calculate expected profit for different markup rates by multiplication of probability of winning and markup and draw a curve of expected profit versus markup.

9) Observe an optimum markup which maximizes an expected profit

The above steps will be followed in chapter 5 for a case study.

4.3 Effects of Correlation in Optimal Markup

In order to investigate significance of the correlation of bid ratios on the markup decision making, a parametric study is conducted here. The proposed model is programmed in MATLAB 7.4.0 (R2007a). It is assumed that the bid ratios of the K competitors all follow the same lognormal distribution with parameters $\mu = 0.05$ and $\sigma = 0.03$. Also assume that each pair of the bid ratios between any two competitors have the same correlation coefficient, i.e., $\rho_{ij} = \rho$ for

$i, j = 1, \dots, K$. By varying the correlation coefficient from 0 (no correlation) to 0.95 (very high correlation), the effects of the correlation on the probability of winning and the optimal markup is going to be investigated. The author also likes to study how the effects of the correlation would change at different numbers of competitors. Since Carr's model has an assumption that the variance of cost estimate equals the variance of bid, therefore based on Equation (3.11) the correlation coefficient among bid ratios in Carr model is 0.5. Since the main aim of this study is the study of Friedman and Gates model, therefore these two models are shown in following figures, but the position of Carr model is around the case which correlation coefficient is 0.5.

Let fix the number of competitor to $K = 5$. Probability of winning versus percentage markup for different value of correlation coefficient is shown in Fig.4.1. As can be understood from this graph, for a specific percentage markup rate, the probability of winning is higher as correlation among X_i ratios is increased. This gap between the probabilities of winning is considerably large from no correlated to highly correlated X_i values for each percentage markup.

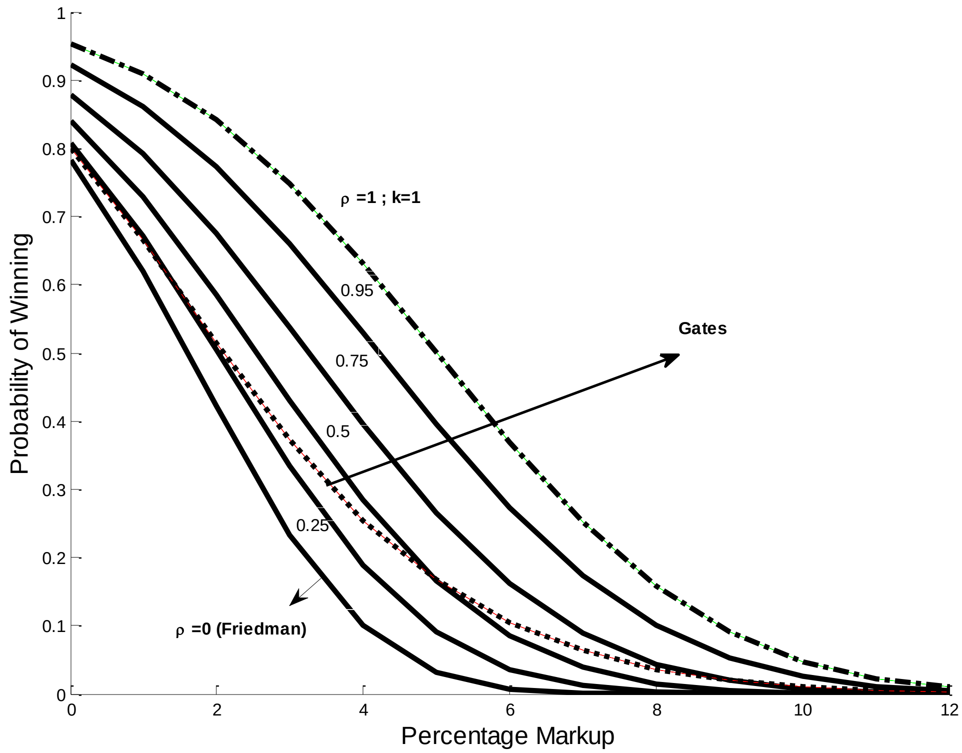

Fig.4.1: Probability of Winning vs. Percentage Markup

Fig.4.1 also demonstrates results of Gates and Friedman ($\rho = 0$) models. As can be recognized, for each percentage markup, Gates model has a higher probability of winning as compare with Friedman model and Gates values for probability of winning is close to the case with correlation coefficient of 0.5.

Fig.4.2 indicates the expected profit versus percentage markup for different values of correlation coefficient. This figure shows the fact that the maximum expected profit is around 2% markup in the case which there is no correlation among the ratios of X_i and 4.5% markup in the case which correlation coefficient is 0.95. This gives a sense of considerable consequences of existence of correlation among X_i ratios which is around two times of optimum percentage markup based on this range of correlation coefficient in order to reach maximum values of expected profit. An optimum markup for Gates model is 3% which is greater than optimum markup for Friedman model which is 2%.

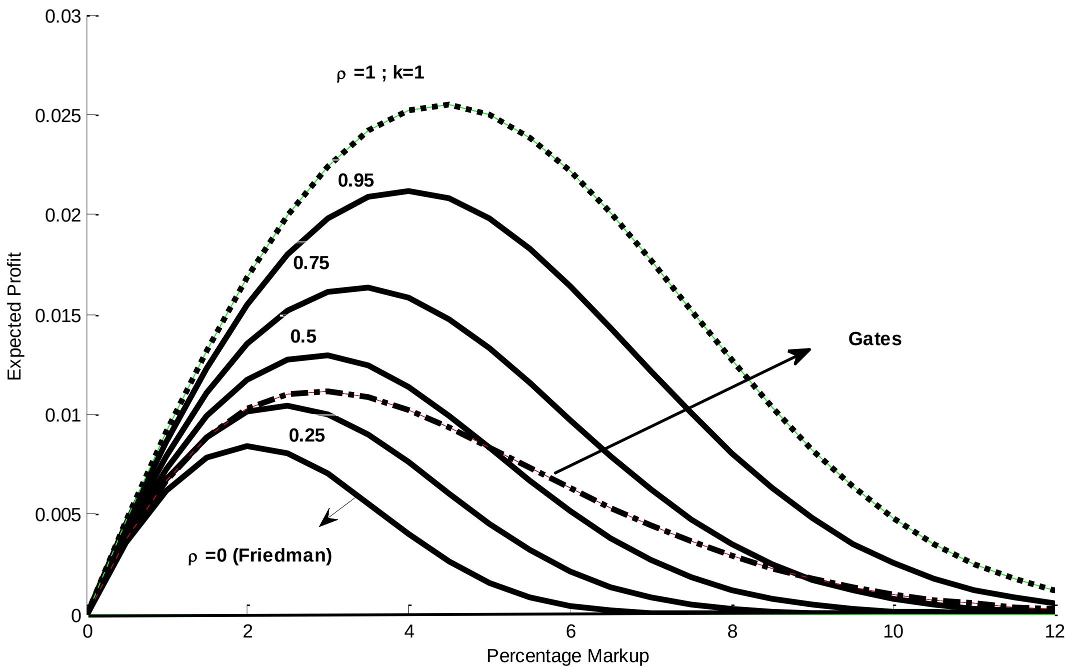

Fig.4.2: The Expected Profit vs. Percentage Markup

Fig.4.3 demonstrates that by increasing the correlation coefficient among X_i ratios, the value of optimum markup is also increased. The value of optimum markup is changed from 2% to 4.4% by changes of correlation coefficient from 0 to 1 which indicates importance of consideration of correlation coefficient among bid ratios. Friedman model has a value of optimum markup around 2% and Gates regardless of changes in correlation coefficient has an optimum markup around 3%.

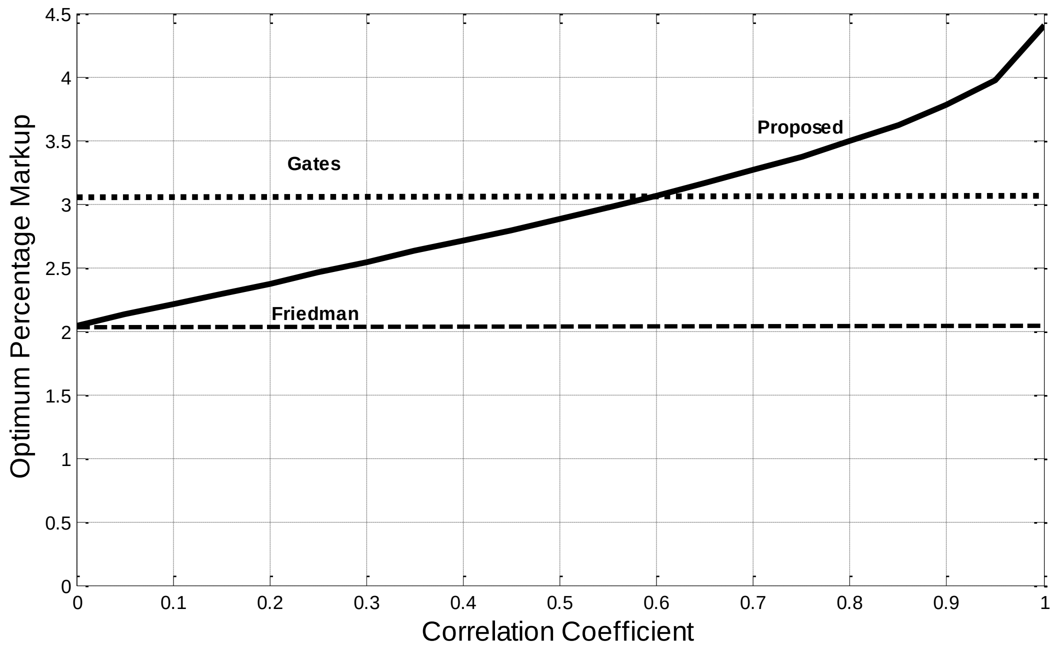

Fig.4.3: Optimum Markup Rates vs. Correlation Coefficient

Now consider the effects of correlation at different number of competitors (k=2,4,6,8). Again, assume the bid ratios have the same parameters $\mu = 0.05$ and $\sigma = 0.03$. In the following study, the correlation coefficient is fixed to 0.5.

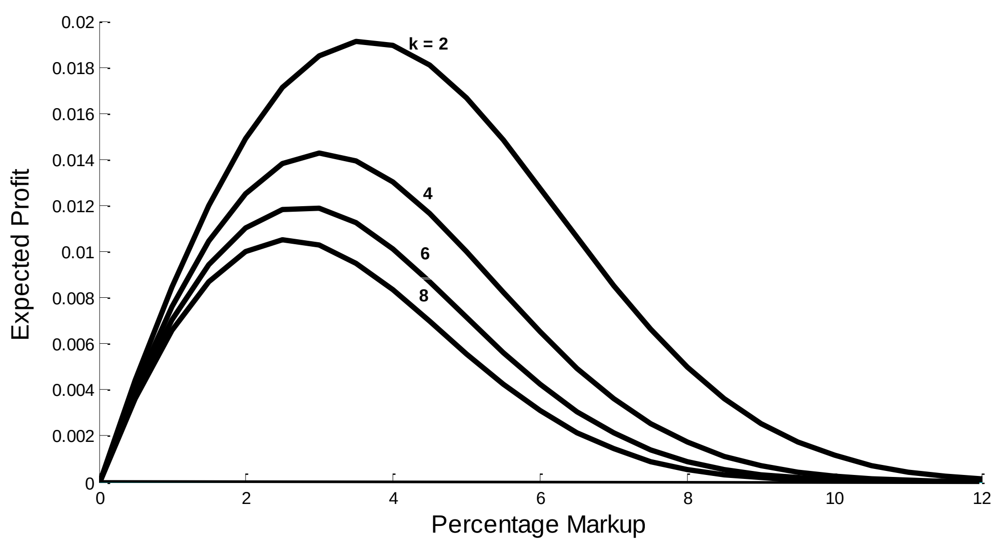

Fig.4.4: Expected Profit vs. Percentage Markup for Different Numbers of Competitors

As can be recognized from Fig.4.4, by increasing the number of competitors, an optimum markup rates is decreased. For the case when $k = 2,$ optimum markup is around 4%, however for a case when $k = 8$, the optimum markup rate is 3%.

Optimum markup rates for different number of competitors with the same parameters of bid ratios ($\mu = 0.05$ and $\sigma = 0.03$) for three different correlation coefficient values among bid ratios ($\rho = 0, 0.5\ and\ 1$) and Gates model, shown in Fig.4.5. As the number of competitors increased, an optimum markup rate is decreased; however, the rate of decreasing of optimum markup is higher when the correlation coefficient among bid ratios is lower. Fig.4.5 illustrates the fact that optimum markup rate for Gates model is very close to the case when the correlation coefficient among bid ratios is around 0.5 and it is greater than Friedman model.

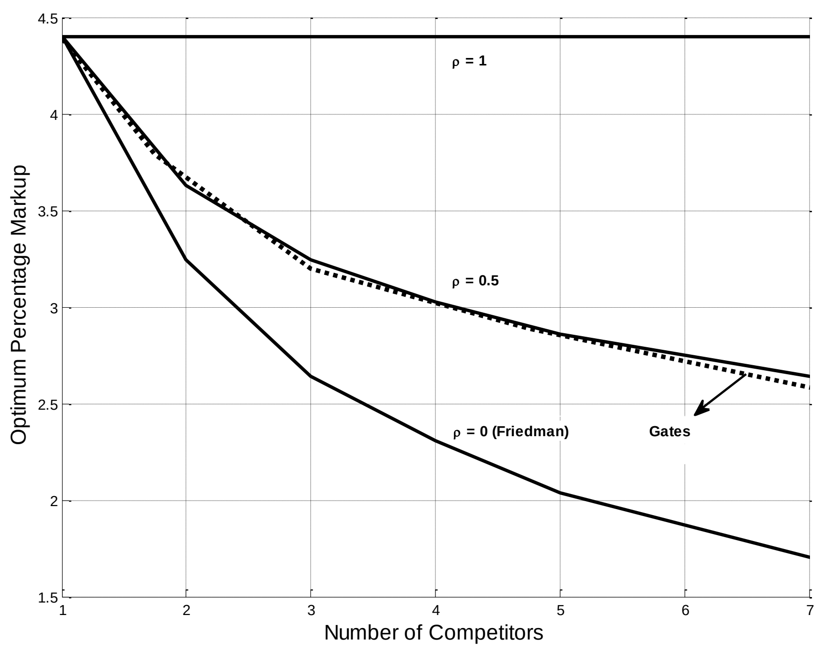

Fig.4.5: Optimum Markup Rates vs. Number of Competitors

In conclusion, the correlation coefficient among the bid ratios has a very significant effects on the probability of winning and hence the optimal markup which should be considered in bidding decision making model. In order to investigate the findings of sensitivity analysis, the case study based on a real bid data will be conducted in next chapter and the results of optimum markup of Friedman, Gates and proposed model with consideration of correlation coefficient is compared.

Chapter 5 Case Study

In order to illustrate the modeling methodology and statistical methods proposed in the previous two chapters, a case study of actual historical bidding data is presented in this chapter. The probability distributions that best fit the bid data are first selected by using the probability paper plots. Then the correlation coefficients of the bid ratios of any pairs of competitors are estimated by using both the maximum likelihood method and the Bayesian method proposed in Chapter 3. The optimal markup for the Contractor is calculated based on the proposed bidding model with consideration of correlation. Finally the results are compared with those from the Gates and Friedman models.

5.1 The Bidding Data

The data set from Skitmore and Pemberton (1994) is used for this study (Appendix A). The data set consists of 51 projects, for each of which the cost estimate of the Contractor (coded as Bidder No. 304 in the original data) is known. These bidding data have been collected by a construction company operating in London, England and covered all the company's building contract bidding activities during a twelve-month period in the early 1980s. Details of the project's type were not used in the analysis. An average number of bidders is seven in this data set and the total number of submitted bids is 352 for 51 projects. Three competitors, known as competitor number 1, 55 and 134, are selected to be investigated as they have the most available bid data. For ease of presentation the competitors 1, 55 and 134 are re-labelled as Competitor 1, Competitor 2 and Competitor 3, respectively.

The number of submitted bids and the average and standard deviation of the bid ratios of the three competitors are summarized and shown in Table 5.1.

Table 5.1: Basic Statistics of the Selected Competitors

	Competitor 1	**Competitor 2**	**Competitor 3**
Submitted Bids	34	20	12
Average (Bid Ratios)	0.951	1.018	1.005
Standard Deviation (Bid Ratios)	0.017	0.067	0.051
Average (log Bid Ratios)	-0.051	0.016	0.004
Standard Deviation (log Bid Ratios)	0.018	0.066	0.051

The scatter plots of the three pairs of bid ratios of the Competitors are shown through Figs.5.1 to 5.3. It can be seen from Figs.5.1 and 5.3 that correlation among bid ratios of Competitors 1-2 and also 1-3 are positive, however Fig. 5.2 shows that correlation coefficient among Competitors 2 and 3 is negative.

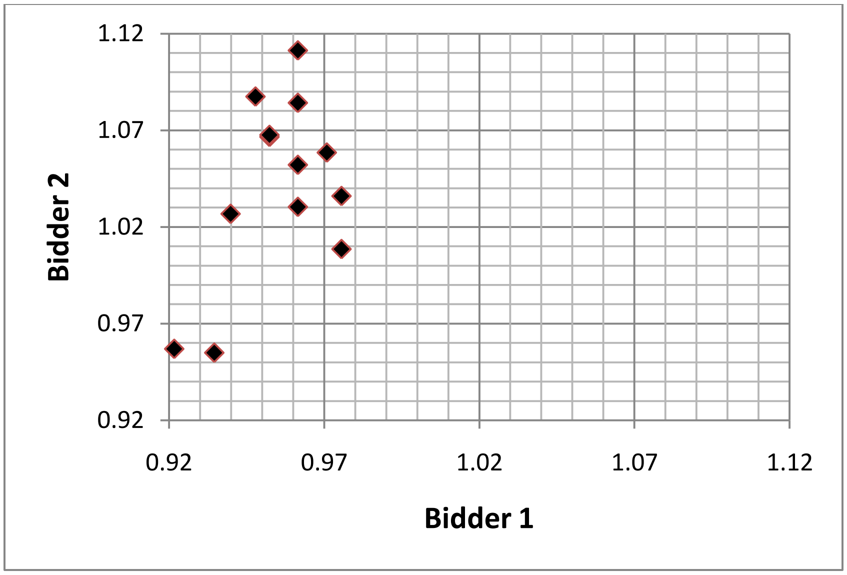

Fig.5.1: Scattered Plot of Bid Ratios (Bidder 1 and 2)

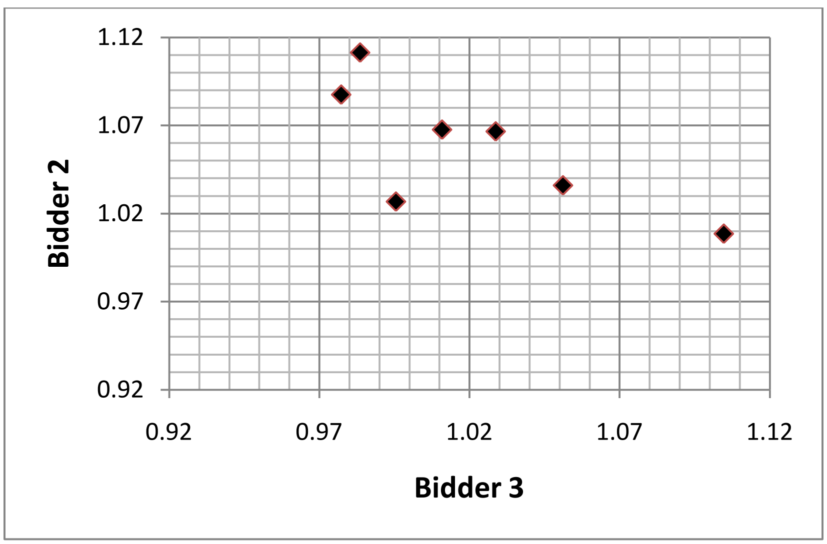

Fig.5.2: Scattered Plot of Bid Ratios (Bidder 2 and 3)

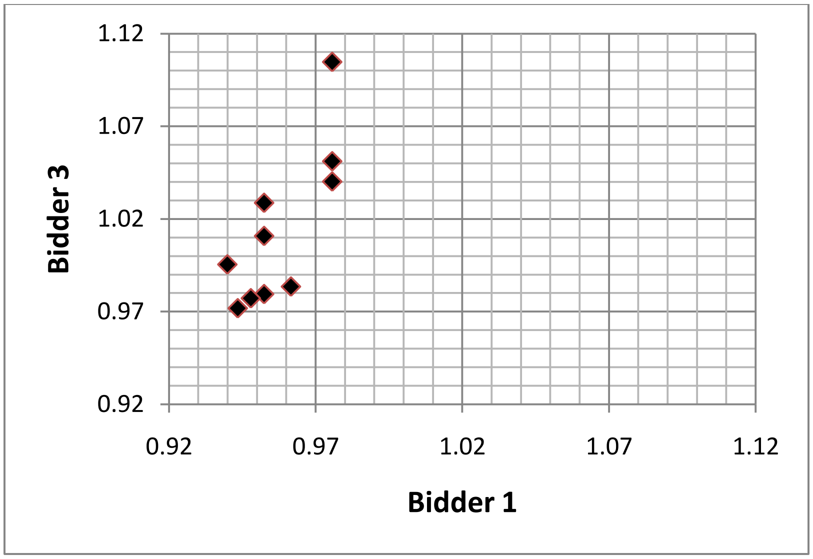

Fig.5.3: Scattered Plot of Bid Ratios (Bidder 1 and 3)

As shown in Table 5.1, Competitor 1 has the lowest mean value of bid ratios, μ_1=0.951. This means that this competitor has an estimated cost that is lower than the Contractor's estimated cost (C_0), which brings a fact of competition advantages in bidding environment as this competitor may utilized advanced equipment and technology in his estimated cost and can be considered as highly competitive competitor. Competitors 2 and 3 have eight pairs of bid ratios which is a small number to estimate a correlation coefficient. The negative correlation among Competitor 2 and 3 (Fig.5.2) can be a result of missing data and outliers, since by adding 2 or 3 data to the Fig.5.2, can be readily observed that correlation among Competitors 2 and 3 can be positive, therefore the proposed Bayesian approach with consideration of missing data used to show the possessiveness of this correlation coefficient.

5.2 Best-Fit Distributions of the Bidding Data

In this section probability paper plotting (PPP) technique is introduced for two reasons. First, to support the previous assumption of lognormal distribution for bid prices and estimated cost in Section 3.2 where the existence and positive sign of correlation coefficient explored, second, to find the best fit distribution for bid ratios in the case study.

Two distributions, normal and lognormal, selected as hypothesized distributions for the best fit analysis as described in following section.

5.2.1 The Probability Paper Plot Technique

In this section the probability plotting technique which helps to figure out the best fit distribution for available data of a random variable is introduced. The probability plot is a graphical technique to assess whether or not the variable is distributed for a hypothesized distribution. The steps of probability paper plotting method can be summarized as below (Montgomery, 2000)

1) Sample of size n from X variable, $x_1, x_2, \ldots, x_n$, is available.

2) Sort data set increasingly, $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$.

3) Rank the sorted data accordingly from 1 to n.

4) Calculate the cumulative frequency of sorted data, $(i - 0.5)/n$, known as edf.

5) Calculate the normal inverse of edf values, $Z = \Phi^{-1}(\text{edf})$.

6) Plot z_i scores against $[x]_{(i)}$ values for PPP of normal distribution.

7) Compare R^2 values and select the distribution with the higher R^2 value.

Probability paper plot (PPP) for estimated cost of Contractor, bid prices and bid ratios of Competitor 1 are shown in following sections. PPP studies of bid ratio and bid prices of Competitor 2 and Competitor 3 is presented in Appendix B.

5.2.2 Best-Fit Distribution of Estimated Cost

In this section the best fit distribution for estimated cost of Contractor is investigated and the following to figures are probability paper plots for estimated cost of Contractor.

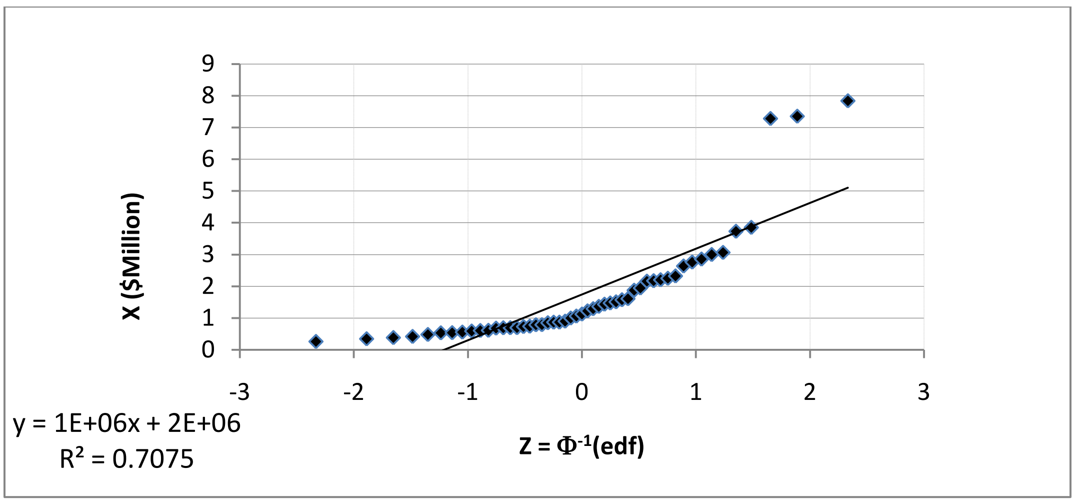

Fig.5.4: Normal PPP of the Contractor’s Cost Estimate

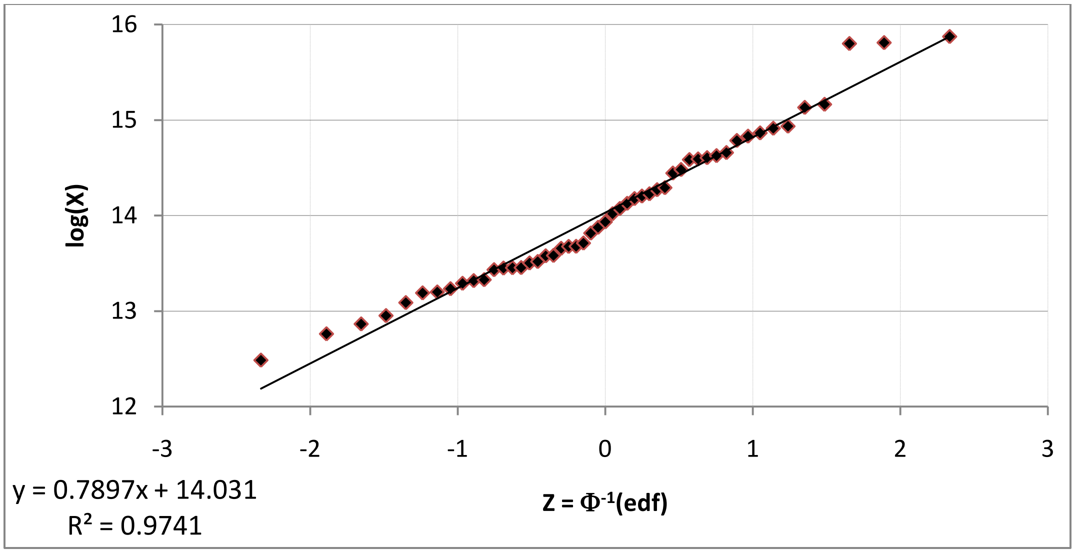

Fig.5.5: Lognormal PPP of the Contractor's Cost Estimate

Figs. 5.4 and 5.5 shows that lognormal distribution is better than normal distribution since plotted data are closer to the straight line and the value of R^2 is 0.97 which is closer to 1 in compare with normal distribution with R^2= 0.70. This would support the previous assumption of $C_0 \sim LN\ (\mu_0, {\sigma_0}^2)$ in chapter 3 in order to support the existence of correlation among bid ratios.

5.2.3 Best-Fit Distribution of Bid Prices

The best fit distribution of bid prices of Competitor 1 is investigated in this section. Figs.5.6 and 5.7 shows the probability paper plots of bid prices for normal and lognormal distributions.

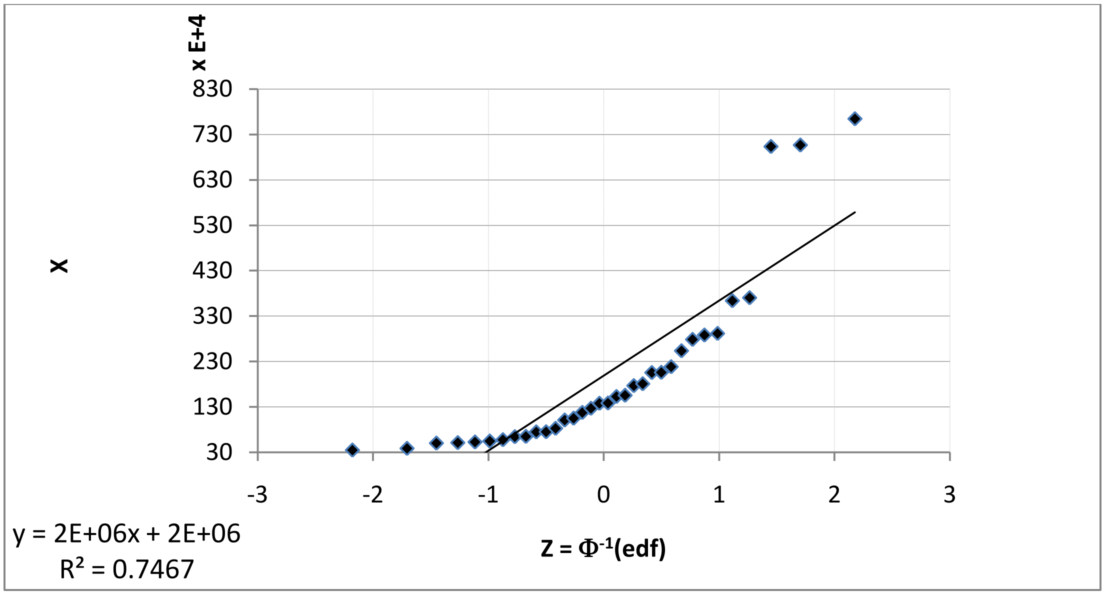

Fig.5.6: Probability Paper Plot for Bid Price (Normal)

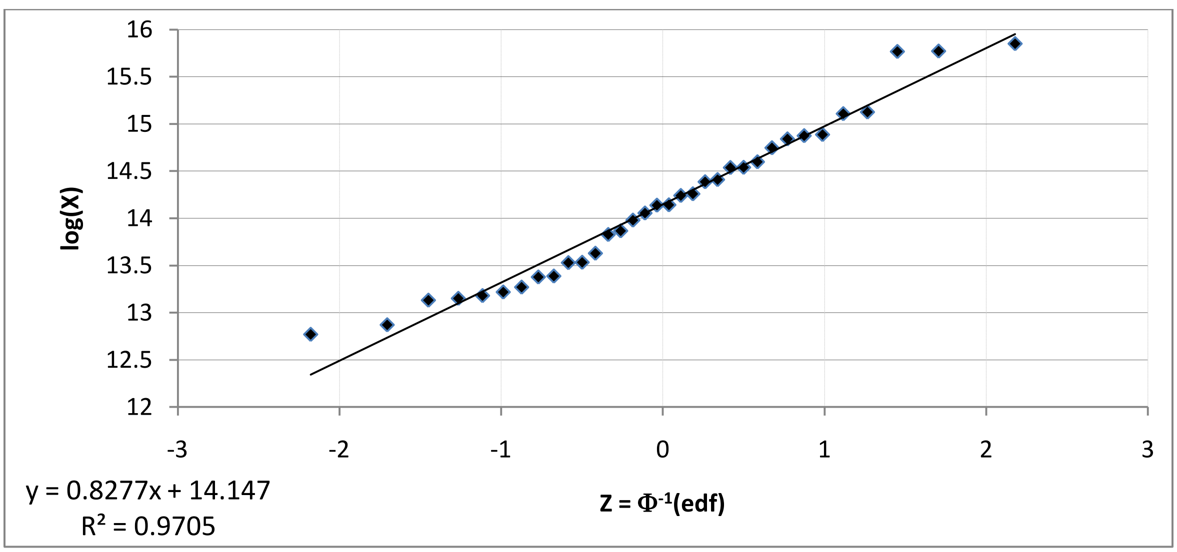

Fig.5.7: Probability Paper Plot for Bid Price (Lognormal)

The lognormal distribution is better than the normal distribution since it has a value of $R^2 = 0.9705$ which is closer to 1 in compare with normal distribution with a value of $R^2 = 0.7467$. This confirms the previous assumption in Section 3.2 that bid prices follow a lognormal distribution.

5.2.4 Best-Fit Distribution of Bid Ratios

Bid ratios best fit distribution is investigated for competitor 1 as an example based on normal and lognormal distributions as shown in Figs.5.8 and 5.9.

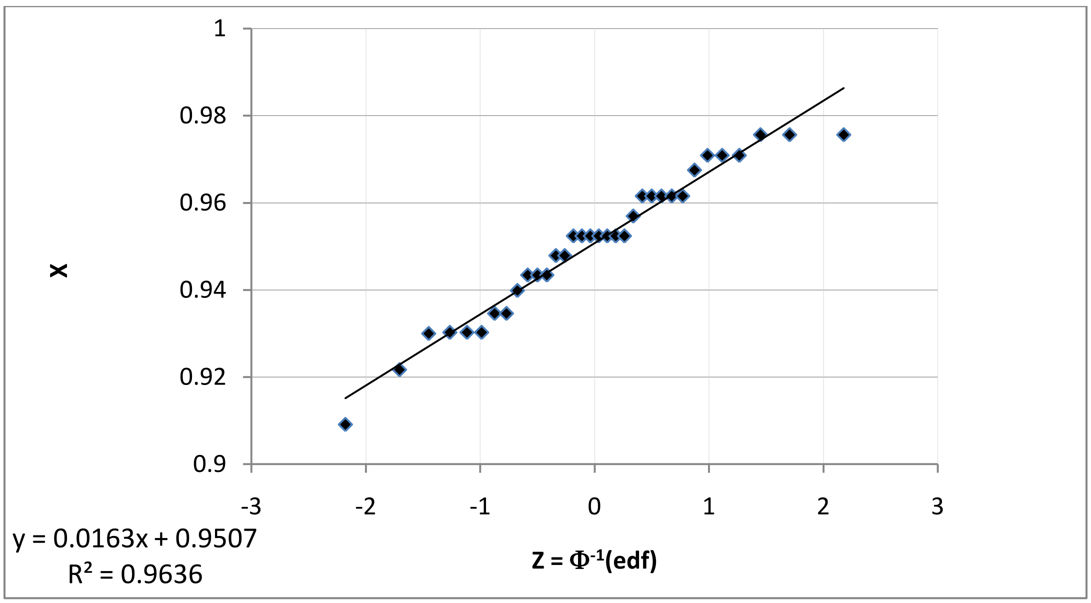

Fig.5.8: Probability Paper Plot for Bid Ratios (Normal)

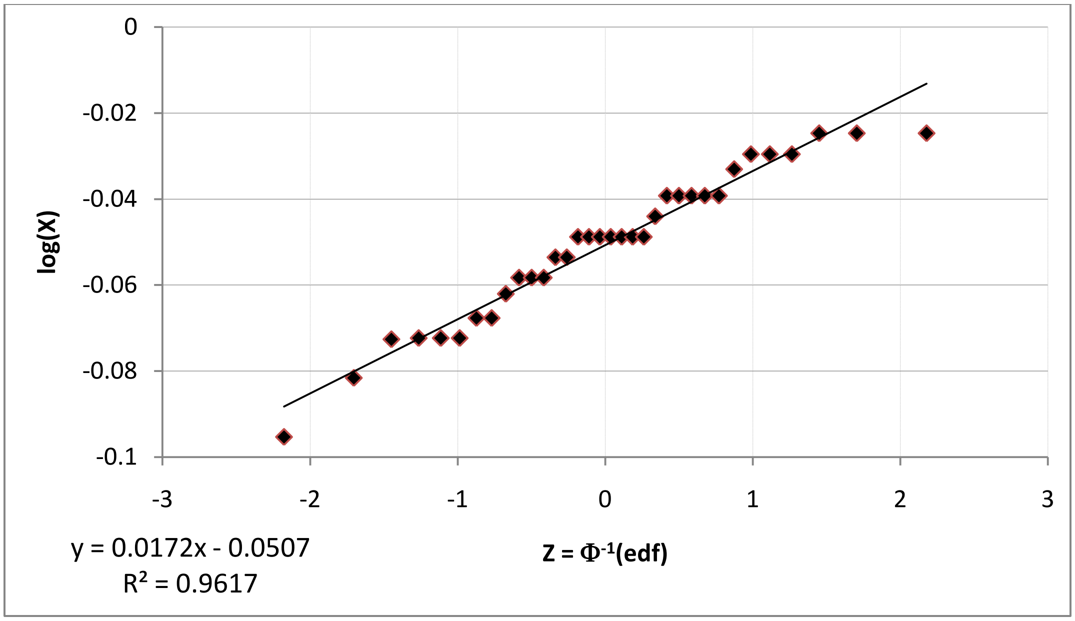

Fig.5.9: Probability Paper Plot for Bid Ratios (Lognormal)

As can be observed from Figs. 5.8 and 5.9 both normal and lognormal can describe the pattern of bid ratios very well since the values of R^2 for both distributions are close to 1 and data are close to the straight line for both distributions. In this case study, lognormal distribution selected in order to investigate the optimum markup decision making models.

5.3 Statistical Estimation of Correlation Coefficients

In this section correlation coefficient based on Maximum likelihood method and proposed Bayesian approach considering missing data for Competitors 1,2 and 3 has been estimated. First likelihood functions of correlation coefficient among three Competitors calculated. Second, Estimation of correlation coefficient with consideration of missing data obtained. Finally, correlation coefficient based on Bayesian approach is presented.

5.3.1 Maximum Likelihood Estimates

Since an estimated cost of Contractor for each project, C_0, is available, therefore bid ratios, X_i, of three competitors 1, 2 and 3 calculated for projects in which they have submitted their bids. Correlation coefficient estimation have been done based on Likelihood for complete data (considering missing data on one variable) in order to make the best use of available data set to estimate ρ.

Estimation of correlation coefficient, ρ, among competitor's X_i pairs as described in Section 3.3.1 can be calculated based on Equation (3.25).

The result of correlation coefficient based on complete data set for three Competitors is shown in Table 5.2.

Table 5.2: Maximum Likelihood Estimates of the Correlation Coefficients

Competitor	**1**	**2**	**3**
1	1	0.45	0.91
2	0.45	1	-0.43
3	0.91	-0.43	1

Likelihood estimation of correlation coefficient among bid ratios of competitors 2 and 3, ρ_{23}, has a negative value which has not been expected and resulted from effect of outliers. As explained in Section 3.3.2, proposed Bayesian approach which use prior information regarding a parameter of estimation can be used for bid ratios. As discussed in Chapter 3, likelihood function among Competitors calculated based on Equation (3.30) as this function is going to be used for Bayesian approach for estimation of correlation coefficient. Figs.5.10-5.12 represents corresponding likelihood functions.

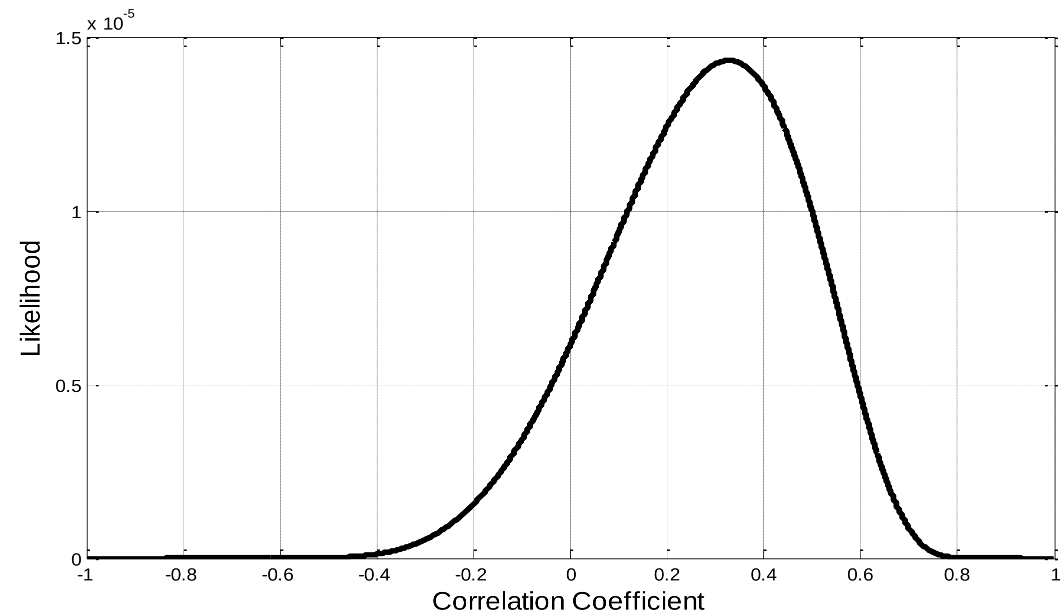

Fig.5.10: Likelihood function of ρ between Competitors 1 and 2

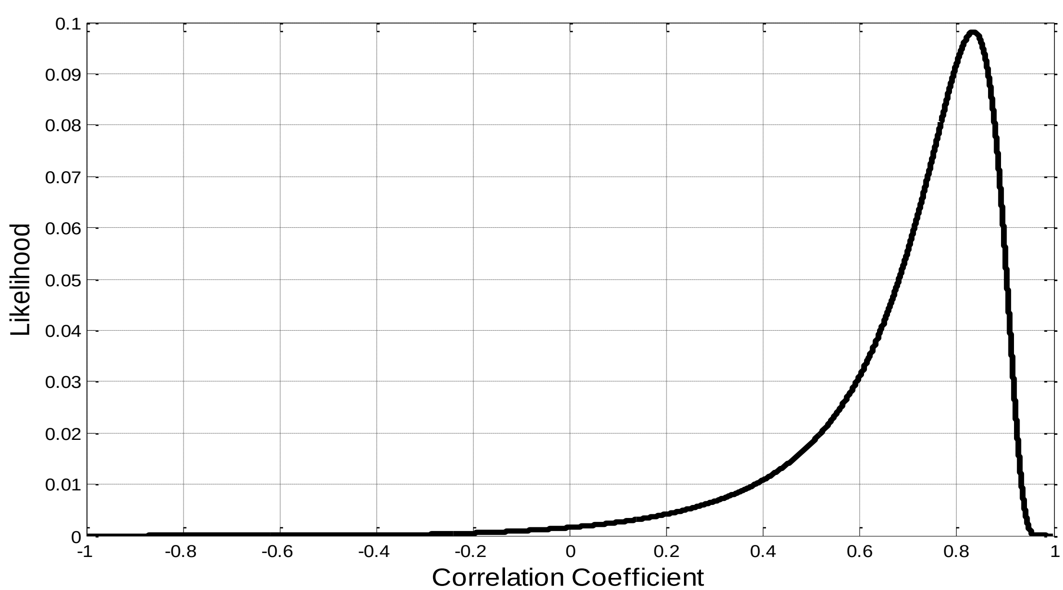

Fig.5.11: Likelihood function of ρ for competitors 1-3

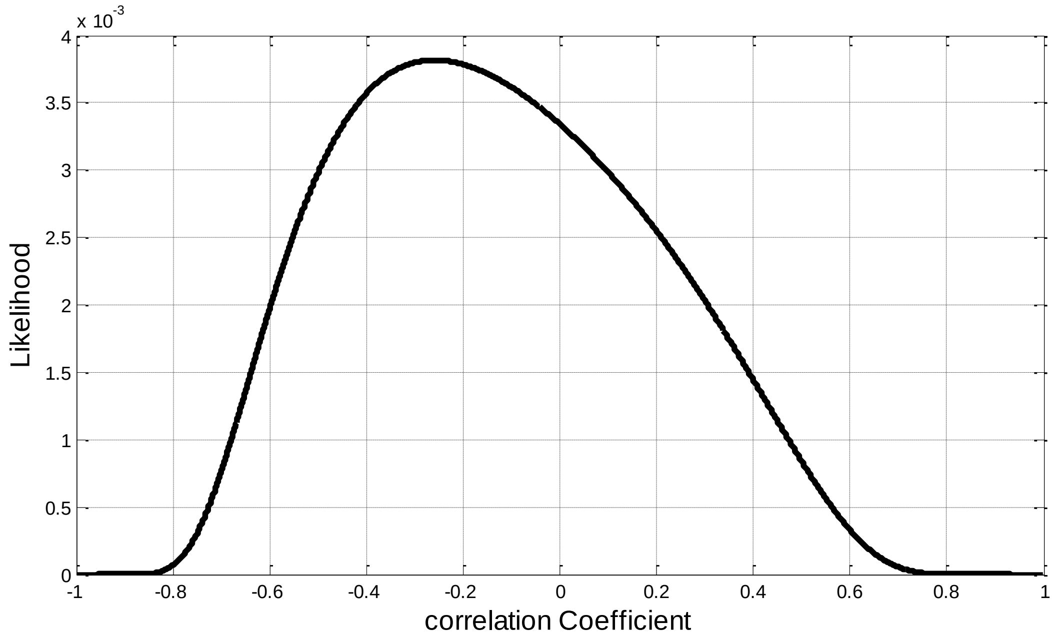

Fig.5.12: Likelihood function of ρ for competitors 2-3

Fig.5.12 illustrates that likelihood function of ρ for competitors 2 and 3 is skewed to the left, therefore the proposed Bayesian method will shift the negative mean of likelihood function by utilizing the prior information regarding correlation coefficient.

Next section estimation of correlation coefficient based on Bayesian method which can make use of advantages of prior information regarding parameter estimation presented.

5.3.2 Bayesian Inferences

As the proposed Bayesian approach in section 3.3.2 introduced, here correlation coefficient based on this method is estimated. First the posterior distributions of correlation coefficient without consideration of constent based on Equation (3.33) are presented and then estimation of correlation coefficient based on Equation (3.34) calculated. Figs.5.13 to 5.15 shows a pattern of posterior distribution of correlation coefficient among bid ratios of three Competitors as follows

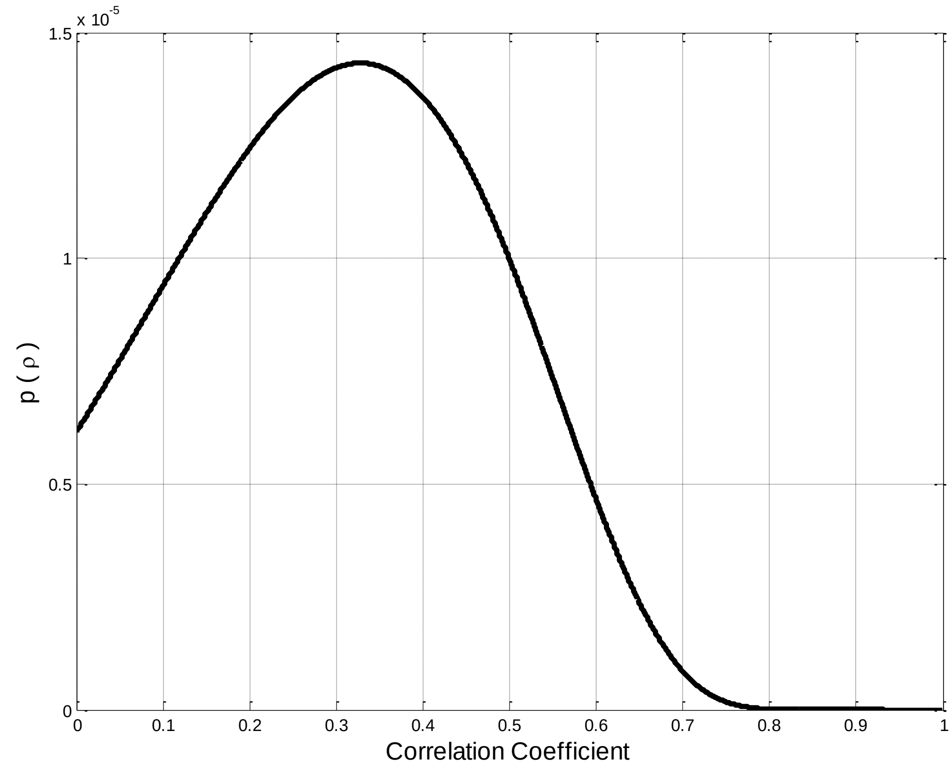

Fig.5.13: Posterior Distribution of ρ for Competitors 1 and 2

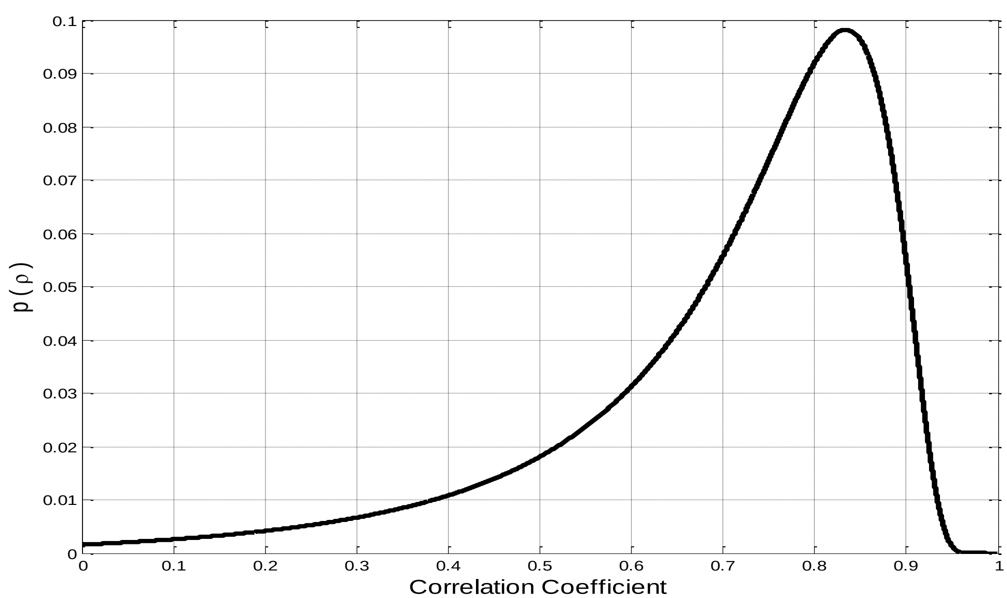

Fig.5.14: Posterior Distribution of ρ for Competitors 1 and 3

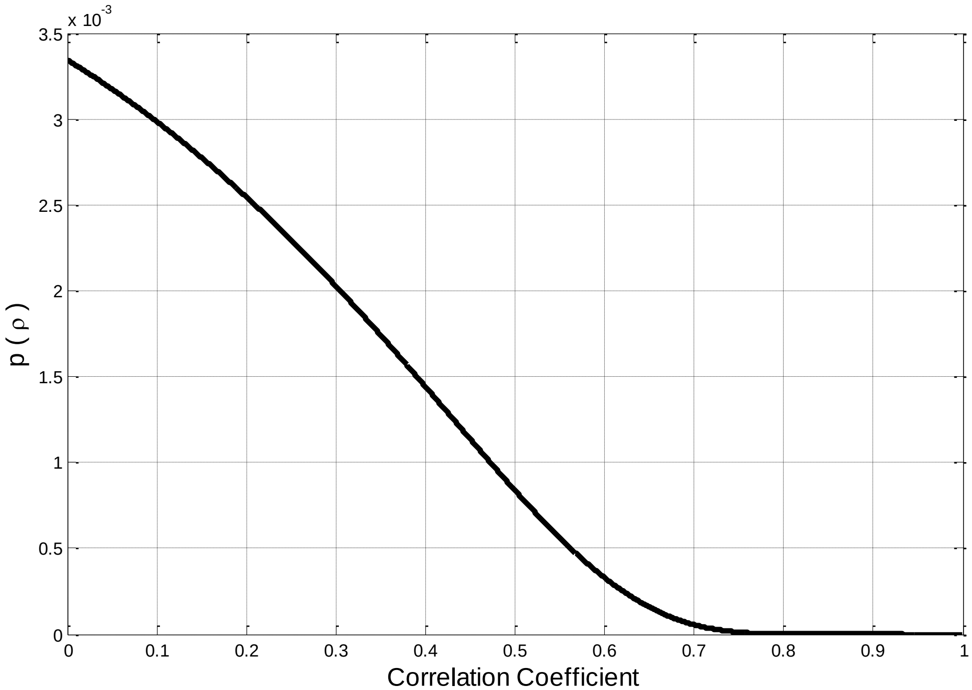

Fig.5.15: Posterior Distribution of ρ for Competitors 2 and 3

As can be recognized from Fig.5.15 the correlation coefficient among Competitors 2 and 3 which previously calculated as a negative value, is positive based on proposed Bayesian approach with consideration of missing data since this approach utilizes *a prior* information of correlation coefficient.

The results of Bayesian estimation of correlation coefficient among competitors 1, 2 and 3 can be derived since the likelihood functions of correlation, $L(\rho)$, among three competitors derived as shown in Section 5.3. As assumed that correlation coefficient follows a uniform distribution between [0, 1], therefore based on Equation (3.34) integrations should be calculated.

In order to determine $E_p[\rho]$, integrations calculated with $\Delta\rho = 10^{-4}$ among three Competitors. Results of correlation coefficient estimation based on proposed Bayesian approach is shown in table 5.3.

Table 5.3: Bayesian Results of the Correlation Coefficients

Competitor	**1**	**2**	**3**
1	1	0.31	0.71
2	0.31	1	0.22
3	0.71	0.22	1

Based on these results which are presented in Table 5.3, prior information regarding a sign of correlation coefficient has shifted a mean and results in a positive expected value for correlation coefficient of posterior distribution. In a case of estimation of ρ_{23} based on Likelihood approach for complete data set negative value of -0.43 calculated which has not been expected, however, in proposed Bayesian approach based on prior information this value has been calculated as positive 0.22. Next section presents results of optimum markup for proposed model and compares results with Friedman and Gates models.

5.4 The Optimum Markup

Section 4.1 introduced the proposed model by considering a correlation coefficient among bid ratios, therefore in this section probability of winning (P_w) of Contractor over three Competitors (Competitor 1,2 and3) has been calculated based on three models: Friedman, Gates and proposed model.

For proposed model, since correlation coefficient among Competitors calculated based on Bayesian approach as shown in Table 5.3, therefore the covariance matrix can be calculated by Equation 4.5. For the three Competitors 1,2 and 3, the result of the covariance matrix is

$$\mathbf{\Sigma} = \begin{bmatrix} 0.0003 & 0.00036 & 0.0006 \\ 0.00036 & 0.004 & 0.00074 \\ 0.0006 & 0.00074 & 0.003 \end{bmatrix}$$

and the mean vector is

$$\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} -0.051 \\ 0.016 \\ 0.004 \end{pmatrix}$$

Therefore the probability of winning for the proposed model can be calculated based on Equation (4.8) as explained in Section 4.1.

Friedman and Gates models introduced in chapter 2 are also investigated. For them, the probabilities of winning are calculated based on Equations (2.1) and (2.2), and results are compared with the probability of winning based on the proposed model, as shown in Fig.5.16. It is seen that Gates model over-estimates the probability of winning whereas Friedman model underestimates the probability. As a result, the expected profit is also overestimated by Gates model and underestimated by Friedman model, as illustrated in Fig. 5.17.

From Fig. 5.17, the optimum markup corresponding to the expected profit is around 3% based on the proposed model. Friedman model with an assumption of independency among bid ratios result in an optimum markup around 2% and the optimum markup for Gates model is around 4%. Table 5.4 summarized results of the three models.

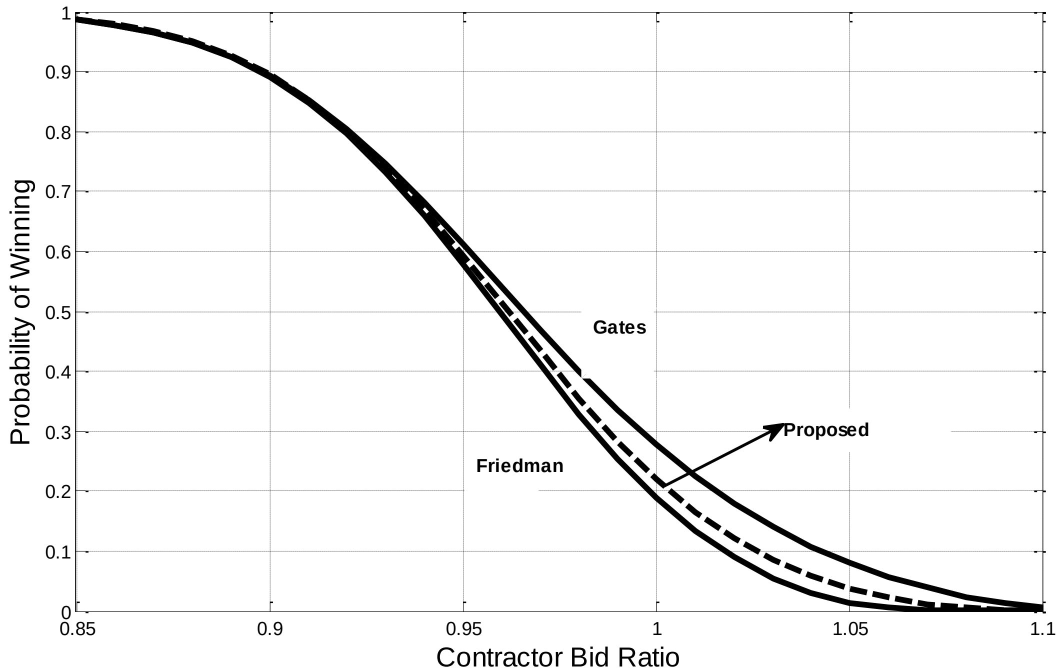

Fig.5. 16: Probability of Winning vs. Markup

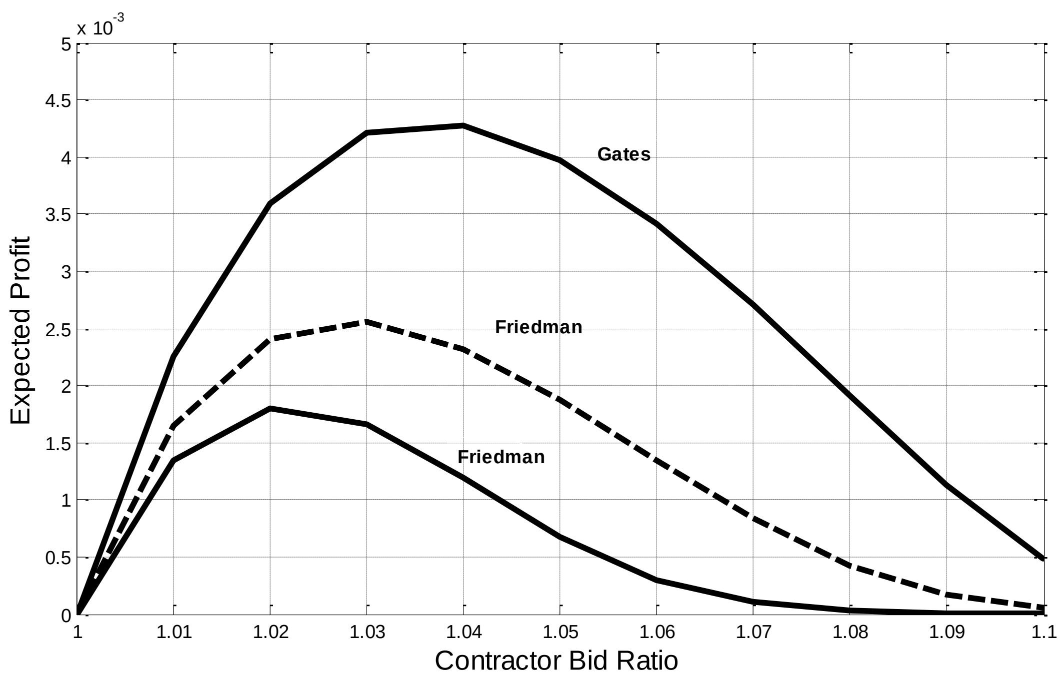

Fig.5.17: Expected Profit vs. Markup

Table 5.4: Optimum Markup for Contractor

Models	Proposed	Friedman	Gates
Optimum Markup	3%	2%	4%

The proposed model with consideration of correlation coefficient among bid ratios of three competitors suggest an optimum markup rate (3%) which is closer to the reality of the construction industry as previously in section 3.2 the rationale behind an existence of correlation proved and this case study support the existence of correlation among bid ratios which is not considered in Friedman and Gates models results. Following chapter presents the summary and conclusions of this project and recommendations for the future researches.

Chapter 6 Summary, Conclusions and Recommendations

6.1 Summary

The focus of the thesis is on the development of a mathematical model to support the decision making in competitive bidding, particularly for determining the markup size. Two major achievements of the study are the new competitive bidding model considering correlation of bids and the Bayesian method of estimating the correlation from historical bid data with missing values. Several major tasks have been accomplished in the study:

1) The Bidding procedure and uncertainties associated to the cost components of a project were reviewed before a relatively comprehensive literature review was conducted of the probabilistic bidding models.
2) The mechanistic reasoning of existence of correlation among bid ratios was presented and the significances of the correlation on optimum markup decision making were investigated.
3) A case study of the published bidding data was conducted to illustrate the effectiveness of the proposed modeling methodology and statistical method. Comparisons of results from the proposed model with those from Friedman and Gates were also performed.

To summarize, the proposed probabilistic bidding model aims to maximize the long-term profit of a general contractor (the decision maker). It is a decision-theoretic model that maximizes the objective function, by choosing an optimal markup:

$$\psi(m) = (mc_0 - c)P_w(m) \tag{6.1}$$

The detailed steps of for the implementation of the proposed model were summarized in Section 4.3 and are not repeated here.

6.2 Conclusions

A key conclusion that can be drawn from the study is that the correlation in bid data plays an important role in markup sizing. In comparison with the results from the models in which the correlation is ignored, a higher markup can be used to maximize the long term profit of the Contractor.

Based on the study some other important conclusions can be reached as follows:

1) Both mechanistic arguments and empirical evidences have shown that the correlation coefficient among bid ratios should be positive
2) The probability of winning increases with an increase in the correlation coefficient among bid ratios.
3) Optimum markup decreases as the number of competitors increases. However, with consideration of the correlation, the decreasing rate of the optimum markup is smaller than the decreasing rate when correlation is ignored.
4) Maximum likelihood estimate of correlation coefficient is sensitive to outliers; the estimated correlation can be negative when the data points are scarce and/or the missing rate is relative large. By utilizing *a prior* information of the correlation coefficient, the Bayesian approach provides a more robust estimation for the correlation coefficient.

6.3 Recommendations

The proposed model described in this thesis is still very much in its formative stage and clearly more work is required before it can be applied in real-world bidding situation. For future research to enhance the proposed model, a couple of recommendations can be made as follows:

First of all, a better statistical method is needed for estimating the correlation coefficient among bid ratios of competitors. This can be done probably in two ways: First, assemble more bid ratio pairs of data. This needs more investment of financial and time efforts since this information is often confidential. Second, more research is required for estimation of correlation coefficient in complex cases where there are missing bid data for in general pattern.

The second recommendation relates to the probability of winning for cases where some unidentified competitors are involved. In this case, since there is no historical bid record for those competitors, a statistical method needs to be developed to estimate the correlation coefficient for them.

References

AbouRizk, S. M., & Halpin, D. W. (1994). Fitting Beta Distributions Based on Sample Data. *Journal of Construction Engineering and Management, ASCE*, Vol. 120, No. 2, 288-305.

Ahmed, I., & Minkarah, I. (1988). Questionaire Survey on Bidding in Construction. *Journal of Management in Engineering*, Vol. 4, No. 3, 229-242.

Back, E. W., Boles, W. W., & Fry, G. T. (2000). Defining Triangular Probability Distributions from Historical Cost Data. *Journal of Construction Engineering and Management, ASCE*, Vol. 126, No. 1, 29-37.

Capen, C. E., Clapp, V. R., & Campbell, M. W. (1971). Competitive bidding in high risk situations. *Journal of Petroleum Technology*, Vol. 23, No. 6, 641-653.

Carr, R. (1982). General Bidding Model. *Journal of the Construction Devision, ASCE*, Vol. 108, No. 4, 639–50.

Casella, G., & Berger, R. L. (1990). *Statistical Inference.* Belmont, CA.: Duxbury Press.

Chua, D. K., & Li, D. (2000). Key Factors in Bid Reasoning Model. *Journal of Construction Engineering and Managemet, ASCE*, Vol. 126, No. 5, 349-57.

Clough, R. H., Sears, G. A., & Sears, K. S. (2005). *Construction Contracting : A Practical Guide to Company Management.* Hoboken, N.J.: John Wiley & Sons, Inc.

Fayek, A., Ghoshal, I., & AbouRizk, S. (1999). A Survey of Bidding Practices of Canadian Civil Engineering Construction Contractors. *Canadian Journal of Civil Engineering*, Vol. 26, No. 1, 1-13.

Friedman, L. (1956). A Competitive Bidding Strategy. *Journal of the Operational Research Society*, Vol. 1, No. 4, 104–12.

Fuerst, M. (1976). Bidding Models: Truths and Comments. *Journal of the Construction Division*, *ASCE*, Vol. 102, No. 1, 169-177.

Fuerst, M. (1977). Theory for Competitive Bidding. *Journal of the Construction Division, ASCE*, Vol. 103, No. 1, 139-152.

Gates, M. (1967). Bidding Strategies and Probabilities. *Journal of the Construction Devision, ASCE*, Vol. 93, No.1, 75–107.

Gross, A. L., & Quevedo, R. T. (1995). Estimating Correlations With Missing Data, A Bayesian Approach. *Psychometrika* , Vol. 60, No. 3, 341-354.

Halpin, D. W. (2006). *Construction Management.* Hoboken, NJ: John Wiley & Sons, Inc.

Hegazi, T., & Moselhi, O. (1995). Elements of Cost Estimation: A Survey in Canada and the United States. *Cost Engineering*, Vol. 37, No. 5, 27-33.

Holm, L., Schaufelberger, J., Griffin, D., Cole, T. (2005). *Construction Cost Estimating: Process and Practices.* N.J.: Prentice Hall.

Ioannou, P. G. (1988). Bidding models-Symmetry and State of Information. *Journal of the Construction Engineering and Management, ASCE*, Vol. 114, No. 2, 214-232.

King, M., & Mercer, A. (1991). Distributions in Competitive Bidding. *Journal of the Operational Research Society*, Vol. 42, No. 2, 151-155.

King, M., & Mercer, A. (1988). Recurrent competitive bidding. *European Journal of Operational Research*, vol. 33, No. 1, 2-16.

Little, R. J., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data.* Hoboken, NJ.: John Wiley & Sons, Inc.

Matlab Help Documentation:www.mathworks.com/help/toolbox/stats/mvncdf.html. Retrieved on Nov.3.2010.

Miller, J. B., & Garvin, M. J. (2000). Toward a New Paradigm: Simultaneous Use of Multiple Project Delivery Methods. *Journal of Management in Engineering*, Vol. 16, No. 3, 58-67.

Montgomery, D. C. (2000)., *Introduction to Statistical Quality Control.* New York: John Wiley & Sons Inc.

Montgomery, D. C., & Runger, G. C. (2003). *Applied Statisitics and Probability for Engineers.* New York: John Wiley & Sons, Inc.

Moselhi, O., & Hegazy, T. (1993). Markup Estimation Using Neural Network Methodology. *Journal of Computing Systems in Engineering*, Vol. 4, Nos. 2-3, 135-45.

Pandey, M. (1998). An effective approximation to evaluate multinormal integrals. *Structural Safety*, Vol. 20, No. 1, 51-67.

Park, W. B., & Chapin, W. (1992). *Construction Bidding: Strategic Pricing for Profit.* New York: John Wiley & Sons, Inc.

Rodgers, J. L., & Nicewander, A. (1988). Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician,*, Vol. 42, No. 1, 59-66.

Schexnayder, C., Knutson, K., & Fente, J. (2005). Describing a Beta Probability Distribution Function for Construction Simulation. *Journal of Construction engineering and Management, ASCE*, Vol. 131, No. 2, 221-229.

Shash, A. (1995). Competetive Bidding System. *Journal of Cost Engineering*, Vol. 37, No. 2, 19-20.

Shash, A. (1992). Factors Affecting a Contractor's Markup Size Decision in Saudi Arabia. *Construction Management and Economics*, Vol. 10, No. 5, 415-429.

Skitmore, M., & Pemberton, J. (1994). A Multivariate Approach to Construction Contract Bidding Mark-up Strategies. *Journal of Operational Research Society*, Vol. 45, No. 11, 1263-1272.

Stark, R. M., & Rothkopf, M. H. (1979). Competitive Bidding: A Comprehensive Bibliography. *Operations Research*, Vol. 27, No. 2, 364-390.

Wallwork, J. W. (1999). Is There a Right Price in Construction Bids? *Journal of Cost Engineering* ,Vol. 41, No. 2, 41-43.

Ward, S. C., & Chapman, C. B. (1988). Developing Competitive Bids: A Framework for Information Processing. *Journal of Operational Research Society*, Vol. 39, No. 2, 123-134.

Weverbergh, M., & Williams, T. M. (2002). A comment on 'Predicting the probability of winning sealed bid auctions: a comparison of models'. *Journal of the Operational Research Society*, Vol. 53, No. 10, 1156-1158.

Yuan, X. X., & Pandey, M. (2006). Analysis of Approximations for Multinormal Integration in System Reliability Computation. *Structural Safety*, Vol. 28, No. 4, 361-377.

Appendix A Data Set for Case Study

This Appendix consists of subset from bid data set which provided in journal paper by Martin Skitmore and John Pemberton in 1994. Data set consist of 51 projects in which competitors have been coded due to confidentiality of bidding information. An estimated cost for projects has shown by contractor 304 in original data set. Based on an objective of our case study, competitors 1, 55 and 134 have been chosen to be studied. We name these competitors in case study as competitor 1, 2 and 3 respectively.

Table A.1: Bid Prices and Estimated Cost

Project #	Cost Estimate	Competitor		
		1(1)	55 (2)	134 (3)
1	1475398	1386652	1514865	1468775
2	535608	505291		
3	1366863	1271146		
4	696743			
5	422297	389214	404110	
6	2161120	2058210		2116877
7	3065742	2919754	3269768	3153800
8	7351929	7035339		
9	902378			
10	1063337	1012702		
11	1947733	1811845		
12	1126816	1053099		
13	698005	652341	666545	
14	682802			
15	1511033		1717715	
16	870894			
17	348969			313203
18	483862		447021	

Table A.1: Bid Prices and Estimated Cost (Continued)

Project #	Cost Estimate	Competitor		
		1(1)	55 (2)	134 (3)
19	2999999	2884614	3333793	2950723
20	7837276	7646123	7904172	8657685
21	3854074	3705840	3971051	
22	615015	580203		597730
23	1610942	1558574		
24	1226589	1179413		
25	2762123		2685127	
26	540814	515061	486485	
27	1876612	1770389		
28	2175928	2062491		
29	608957		559596	619065
30	2639525	2538005	2861665	
31	732572			
32	559351	530190	608242	546641
33	853793		847621	
34	2325900			
35	871927	830407		
36	792474	754737		
37	7279854	7067819		
38	592096	550787		
39	1001254			
40	2205359			
41	1576905	1530976		
42	3732133	3641105	3866339	3922937
43	743578			
44	2252833	2187217	2384494	
45	1294986		1268733	
46	2857275	2787585		2972189
47	1436804	1381542	1511643	
48	789355	751767	842684	797926
49	264933			
50	386983	351803		
51	694297	645858		

Appendix B Probability Paper Plots for Contractors 2 and 3

In this Appendix probability paper plots of bid prices and bid ratios of two Competitors No.2 and No.3 for Lognormal and Normal distribution investigate.

Probability plots for bid ratios and bid prices of competitors No.2 and No.3 are presented bellow.

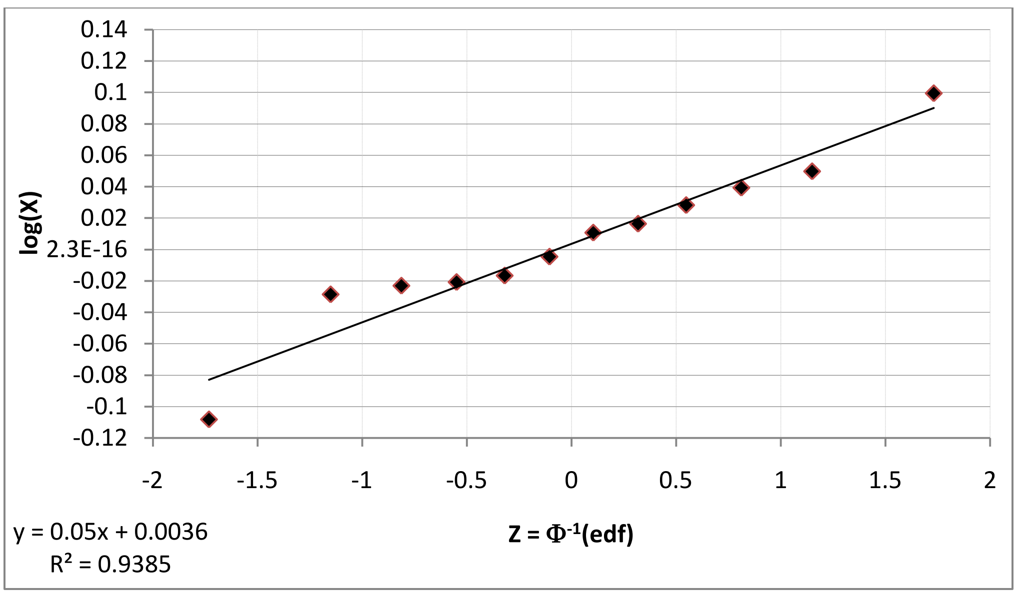

Fig.B.1: Log-normal PPP of the competitor No.2 (Bid Ratio)

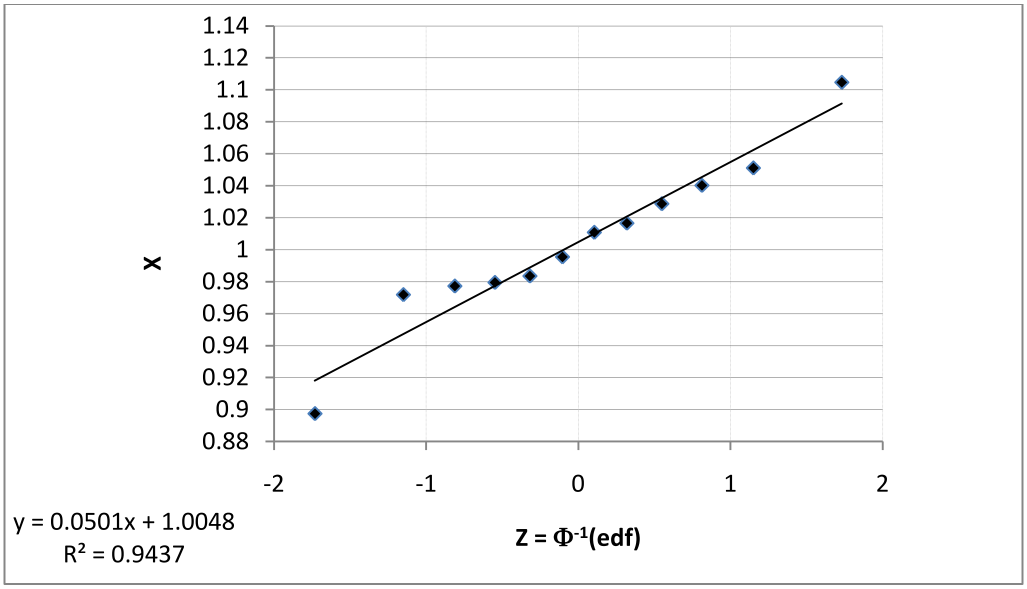

Fig.B.2: Normal PPP of the competitor No.2 (Bid Ratio)

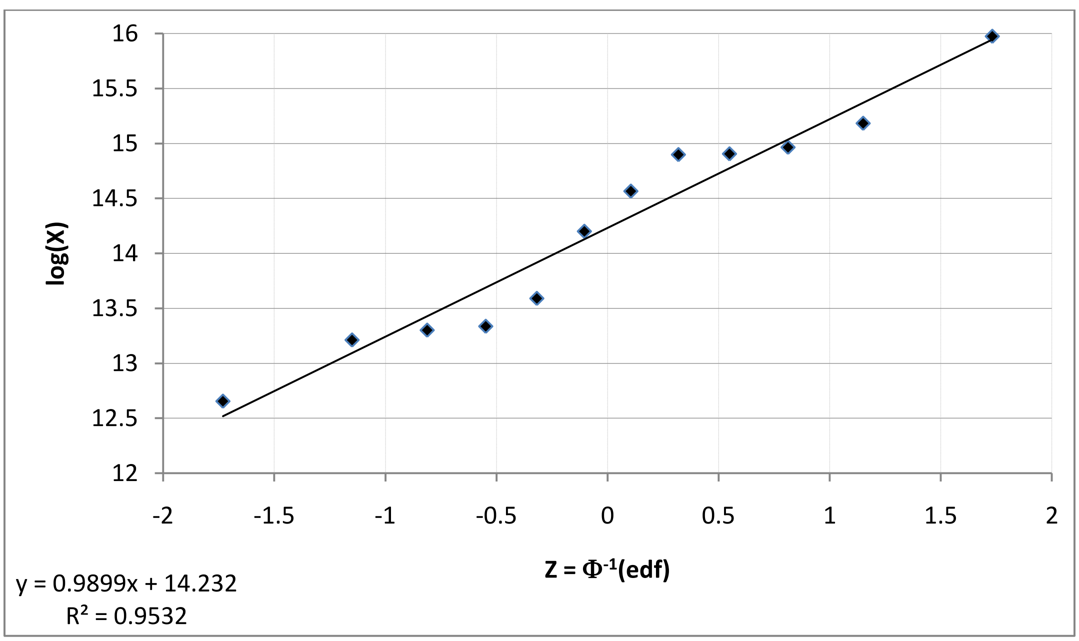

Fig.B.3: Log-normal PPP of the competitor No.2 (Bid Prices)

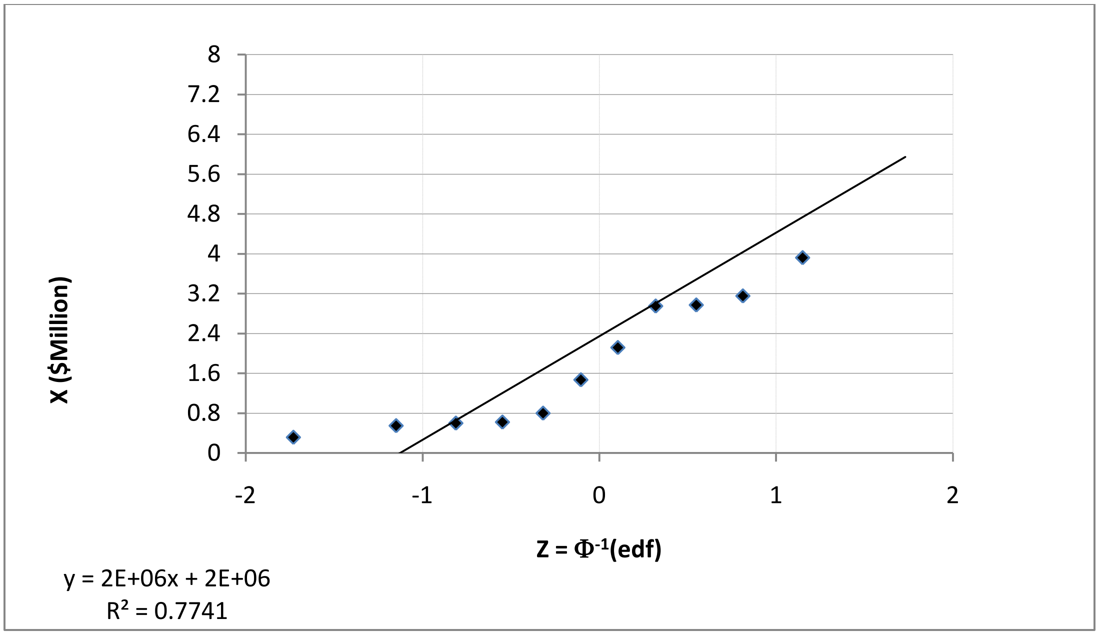

Fig.B.4: Normal PPP of the competitor No.2 (Bid Prices)

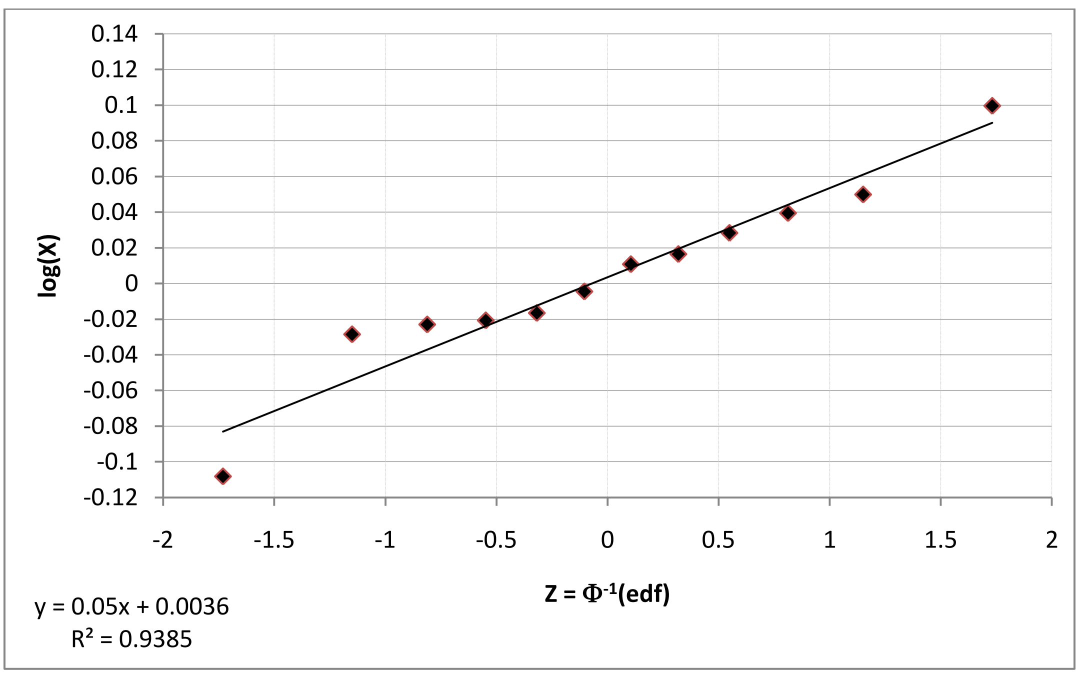

Fig.B.5: Log-normal PPP of the competitor No.3 (Bid Ratio)

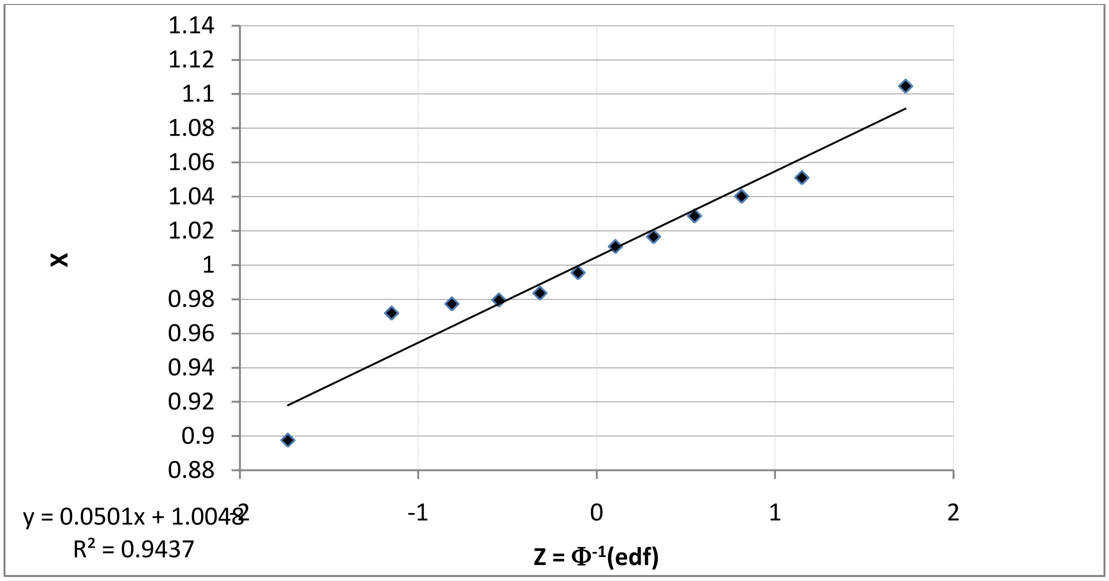

Fig.B.6: Normal PPP of the competitor No.3 (Bid Ratio)

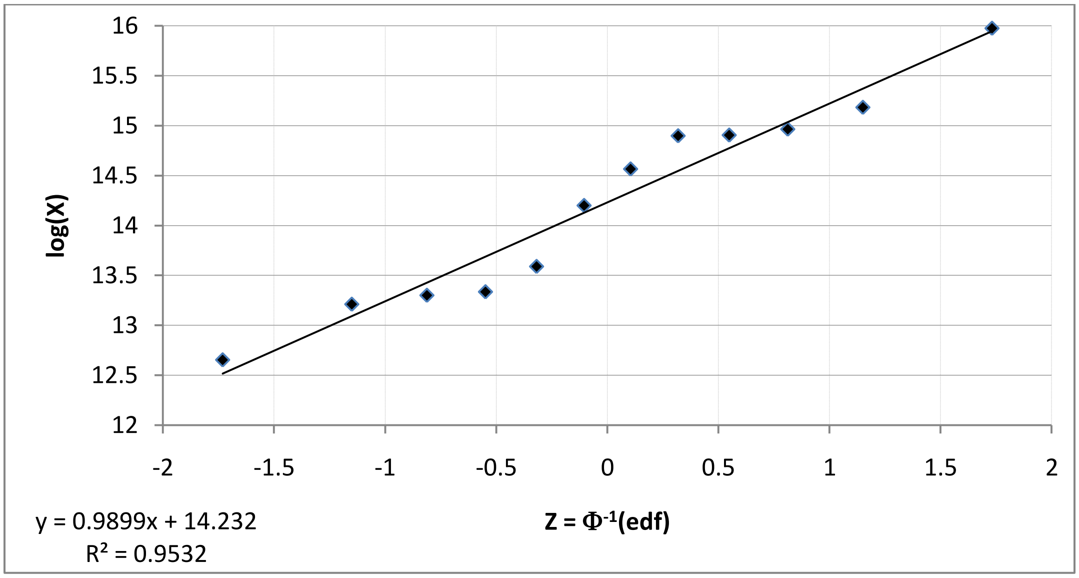

Fig.B.7: Log-normal PPP of the competitor No.3 (Bid Prices)

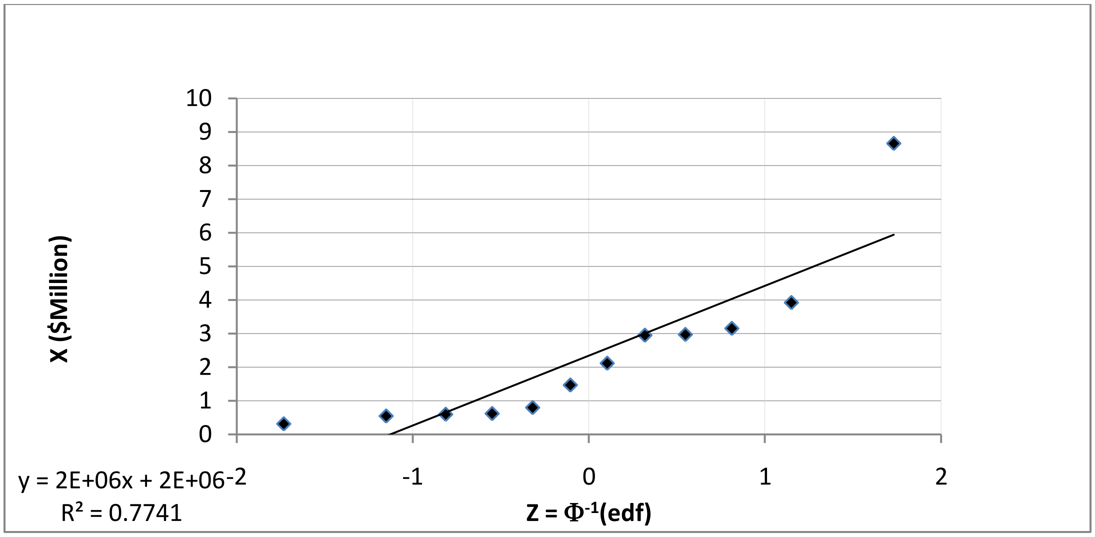

Fig.B.8: Normal PPP of the competitor No.3 (Bid Prices)

Appendix C Notation

Following is the list of notations which have been used in this study:

B_i	Bid Price of Competitor *i*
C_i	Estimated Cost of Competitor *i*
C_0	Contractor Estimated Cost
EP	Expected Profit
K	Number of Competitors
$l(\boldsymbol{\theta}\|\boldsymbol{z})$	Log-likelihood Function of Parameters
$L(\boldsymbol{\theta}\|\boldsymbol{z})$	Likelihood Function of Parameters
m_i	Competitors' Markup
P_w	Probability of Winning
$p(\rho)$	Posterior Distribution of Correlation Coefficient
r_0	Correlation among log-transformed Bid Prices and log-transformed Estimated Cost
r_1	Correlation among log-transformed Bid Prices
X_i	Competitors' Bid Ratio
ρ	Correlation Coefficient
$\pi(\rho)$	Prior Distribution of Correlation Coefficient