Prediction Model of Final Project Cost using Multivariate Probabilistic Analysis (MPA) and Bayes' Theorem

Wi Sung Yoo* · Fabian C. Hadipriono**

Abstract

This paper introduces a tool for predicting potential cost overrun during project execution and for quantifying the uncertainty on the expected project cost, which is occasionally changed by the unknown effects resulted from project's complications and unforeseen environments. The model proposed in this study is useful in diagnosing cost performance as a project progresses and in monitoring the changes of the uncertainty as indicators for a warning signal. This model is intended for the use by project managers who forecast the change of the uncertainty and its magnitude. The paper presents a mathematical approach for modifying the costs of incomplete work packages and project cost, and quantifying reduced uncertainties at a consistent confidence level as actual cost information of an ongoing project is obtained. Furthermore, this approach addresses the effects of actual informed data of completed work packages on the re-estimates of incomplete work packages and describes the impacts on the variation of the uncertainty for the expected project cost incorporating Multivariate Probabilistic Analysis (MPA) and Bayes' Theorem. For the illustration purpose, the introduced model has employed an example construction project. The results are analyzed to demonstrate the use of the model and illustrate its capabilities.

Keywords : Bayes' Theorem, change of uncertainty, cost overrun, multivariate probabilistic analysis (MPA)

1. INTRODUCTION

Today's construction project is often large, complex, and unique; and it encompasses various design and construction methods. Hence, such a project may experience cost overrun due to unpredictable project conditions and environments, as well as because of limited and deterministic information at the starts. Inaccurate prediction of dynamic uncertainties and changeable factors as a result from complications of project circumstances have caused hidden project value and decreased project competitiveness. Therefore, an effective tool is needed to diagnose the cost performance, specifically to monitor cost uncertainty and to quantify the risk level. Such a tool can also be used to manage the cost contingency under changeable environments.

For a successful project execution, it is an increasing need to predict a more precise project cost and contingency associated with the uncertainties that may appear within the project duration. Predicting project cost and contingency (defined as a buffer in the light of quantitative value) is useful if it is done at the early stage of the execution (Zwikael et al., 2000). This early stage is often defined as a phase in which at least 30% of project development is completed (CII, 1994). The prediction of project cost and contingency at the progressive stage and the quantification of the reduced uncertainties resulted

* 일반회원, Ph.D Candidate, Civil & Environmental Engineering & Geodetic Science, The Ohio State University, United States, wisungyoo@hotmail.com

** 일반회원, Professor, Civil & Environmental Engineering & Geodetic Science, The Ohio State University, United States, hadipriono@osu.edu

from any reported cost data are crucial to control and manage the costs of incomplete work packages and to provide project managers with better management actions and strategies. Hence, forecasting the dynamic uncertainties contributes to a flexible project control and management (Karlsen and Lereim, 2005). The latter can be exemplified as the return of excess contingencies to investors or owners, re-investment on other projects, resolution of emergencies, control of project schedules, and improvement of facilities (Ford, 2002).

In construction industry there are very few methods available to accurately predict and quantify changes of uncertainties on project cost and contingency. Thus, many projects have often overspent the planned cost, resulting in financial losses for the owners and investors. To overcome these losses project managers need tools for quantifying changes of uncertainties and re-estimating project costs and contingencies. Multivariate probabilistic analysis (MPA) and Bayes' Theorem are introduced in this paper as such tools for project managers to forecast expected project costs based on cost performance of completed work packages and to monitor the changed and reduced cost uncertainties when more information of an ongoing project is available. These tools provide project managers with a risk function (or probability distribution) to establish a risk level to avoid cost overrun. Such a thing can be done since these approaches update dynamic predictions that correspond to any reported cost data, and automatically diagnose project cost and contingency at an initially desired confidence level.

(1) Objectives of the study

One of the ultimate objectives in this study is to monitor the reduction of cost uncertainties of incomplete work packages by increasing cost information resulted from completed work packages considering the dependency between these packages. Also, others are to quantify its magnitude at an initially desired confidence level and to address the effects of informed cost data on the modifications of the costs of incomplete work packages and expected project cost in quantitative assessment. In brief, the reported cost data of an ongoing project is applied for revising them, considering that the costs of all work packages are not totally independent and that dependency provides any information to incomplete work packages. When the correlation among work packages exists, the actual costs of completed work packages affect the remaining project costs, and hence the expected cost and its uncertainty are re-computed and continually adjusted as a project progresses.

(2) Scope of the study

This study is concerned with the fact that the changes of the estimated costs are often caused by the dependencies of work packages and the unpredictable uncertainties under the complicated environments. If work packages are totally independent, then the changes of uncertainties resulted from the dependency are insignificant. Hence, the method introduced in this study focuses on projects whose work packages are somewhat or highly correlated among each other (which are usually the case); and so the variations of uncertainty will appear during a project execution. Furthermore, because there exist no perfectly satisfied correlation matrix in reflecting the dependency of cost variates, this study is concerned with two specific patterns of Correlation Function; Gaussian and Exponential Correlation Functions with the characteristics that are described in the next section.

2. BACKGROUND

There is no perfectly accurate method for predicting the future phases or potential outcomes of a construction project. In a construction project, the project manager plans to absorb the uncertainties for unexpected future outcomes affecting a project cost by estimating the contingency. One common way to calculate this contingency is to consider a percentage of expected project cost (e.g., 5% or 10%). Despite its simplistic approach, it does not quantify the degree of confidence that the contingency provides against cost overruns

(Touran, 2003). Pugh and Soden (1986) have addressed that to make a single-value deterministic cost estimate is unsafe because of the unpredictable changes of uncertainty on total cost caused by project's complications. The amount of actual contingency can also be assessed by either experts' subjective judgments or statistical analyses. Particularly, statistical methods, such as Monte Carlo simulations, regression models, and variance analyses are useful to analyze and range the amount of contingency when previous cost data was recorded systematically (Clark, 2001). These approaches are efficient to estimate an expected total cost and quantify a proper range of contingency. Nassar (2005) asserted that when they are used at the beginning and the variation of uncertainties during the development is unconsidered, it is difficult to apply new reported cost data in order to predict future outcomes.

Multivariate probabilistic approach (MPA) and Bayes' Theorem are provided to create a predictive probability distribution (or a risk function) for future project cost and contingency. This function is based upon a prior distribution of initial project cost and reported cost information from completed work packages. In statistical analyses, these methods have been known as a powerful tool to measure the uncertainty, which started from subjective or planned distribution. In particular, Bayes' Theorem is a logical tool to update the probability distribution of an unknown factor in estimating the uncertainties. Its main capabilities are described in the following:

1. To combine prior knowledge and actual reported data;
2. To automatically update initial estimates as soon as actual data is available; and
3. To reflect the reduced uncertainty as information is getting more available.

3. METHODOLOGY

(1) Correlation functions

Construction projects are often complicated and have various operations, and project managers may have difficulties to objectively and accurately assess the dependency among work packages that are defined as cost variates in this study. Even though there is available data from past similar projects, such assessments tend to depend on the experts' subjective judgment and experience. Correlation or covariance among specific cost variates is one element that represents the dependency. For instance, if two work packages that may affect each other are estimated, the costs may increase or decrease together. Correlation and covariance among work packages measure such a tendency. Statistically, it is important to satisfy the properties of correlation matrix. To apply this matrix to cost estimating, it is necessary that the matrix be square, symmetric, and positive definite. In determining whether the correlation matrix is satisfactory, the eigenvalues for all cost variates are computed; and if they are positive, the matrix is statistically logical to reflect the dependency among the operations.

In this section, the correlation functions (CF) are introduced to construct a legitimate correlation matrix. The CF represents the probability density of measuring the dependency of cost variates, and this creates the correlation matrix satisfying all the properties for applying it to the cost estimates. In this paper, two types of the Correlation Functions are employed to test the proposed model: Gaussian Correlation Function (GCF) and Exponential Correlation Function (ECF). The GCF shows the dependency from the preceding to the subsequent work packages that is initially decreasing slowly but subsequently decreasing rather quickly (Figure 1a). Conversely, the ECF shows a faster decrease in the dependency between subsequent work packages (Figure 1b). In building this matrix, a project manager determines which CF is more appropriate for a specific construction project.

$$GCF = Corr(x_i, x_j) = \exp[-\theta(i-j)^2]$$

$$ECF = Corr(x_i, x_j) = \exp[-\theta|i-j|]$$

Here, i and j represent the position of the correlation component in the matrix, and θ is a parameter to control the smoothness of the dependency change among the

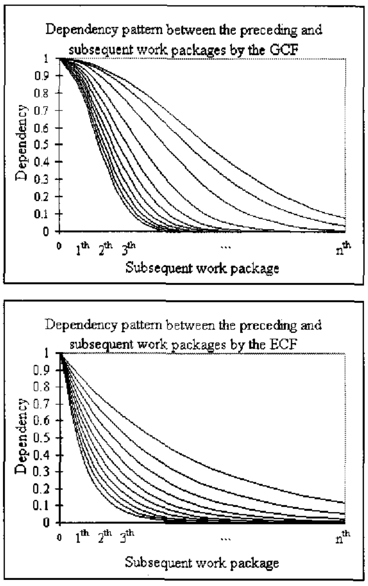

Figure 1 Gaussian (a) and Exponential (b) Correlation Function

successive cost variates. θ is controlled by a project manager; and it also governs the dependency on the next cost variates.

(2) Conditional probability theory and Bayes' Theorem

The knowledge of Bayes' Theorem statistics is one scientifically justifiable way to integrate informed expert judgments with reported information. Hence, Bayes' Theorem can be employed to modify the expectation on a project cost, and theoretically, based on a prior distribution of a project total cost, a posterior distribution is assessed through informed cost data (Johnson and Wichern, 2002). For instance, if work packages such as civil-work and foundation-work are performed in sequence, and civil-work is completed first with actual cost as X_1, then its cost will always be known as X_1 since the work was already completed. The conditional probability distribution on cost variate x_2 given that the cost variate x_1 having a specific value of X1 is $f_{2|1}(x_2|x_1)$. In this expression, X_1 is known before X_2. This can also mean that civil-work is completed before foundation-work is done. Hence, the joint probability density function on these cost variables is the product of one conditional distribution and one marginal distribution. In this case the joint probability distribution is a bivariate probability density function, where the marginal distribution is determined by integrating the joint distribution. The conditional probability density function is presented as follows

$$f_{2|1}(x_2|x_1) = \frac{f_{1,2}(x_1, x_2)}{f_1(x_1)}$$

It is obvious that if they are correlated any information about x_1 reduces the uncertainty in x_2. Hence, if work packages are dependent, cost information about the completed work can be used to generate better cost estimates of work packages that are yet to be completed. When a project manager has a consistent view of contingency, the project cost after the completion of civil-work will be set such that the probability of cost overrun can be kept with the same cost risk level, α. It indicates that whatever risk level (α) or confidence level ($1-\alpha$) is appropriate at the beginning of a project, they remain constant throughout the project life cycle. Thus, modified (B_{new}) total cost is computed by the conditional probability analysis and Bayes' rule, and the difference between a prior (B_0) and the posterior (B_{new}) is presented below, where k, the unit standard variate, also remains constant even though the uncertainty on project cost has changed.

$$B_0 - B_{new} = [\mu_1 + \mu_2 + k\sqrt{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2}] - [X_1 + \mu_2 + \left(\frac{\sigma_2}{\sigma_1}\right)(X_1 - \mu_1) + k\sigma_2\sqrt{1-\rho^2}]$$

$$B_0 - B_{new} = \overset{\mathrm{I}}{\boxed{[\mu_1 - X_1]}} + \overset{\mathrm{II}}{\boxed{[\rho\left(\frac{\sigma_2}{\sigma_1}\right)(\mu_1 - X_1)]}} + \overset{\mathrm{III}}{\boxed{k[\sqrt{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2} - \sigma_2\sqrt{1-\rho^2}]}}$$

In the above equations, μ_1 and μ_2 are prior expected costs, σ_1 and σ_2 are prior expected standard deviations, and ρ is the correlation coefficient between civil-work and foundation-work. The cost differential consists of three terms:

I. $\mu_1 - X_1$, which represents the direct savings if $\mu_1 > X_1$.

II. $\rho(\sigma_2/\sigma_1)(\mu_1 - X_1)$, which represents the expected savings on foundation-work, based on the reported cost data on civil-work and its correlation with foundation-work.

III. $k[\sqrt{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2} - \sigma_2\sqrt{1-\rho^2}]$, which represents a reduction in the contingency necessary to cover consistent cost risk level (α), corresponding to a reduced uncertainty caused by information of the completion of the previous work package. It is noted that this last term is independent from the actual cost, X_1.

The above term (cost differential) engenders two possible outcomes, i.e., cost reduction with the difference returned to the owners and uncertainty elimination with consistent confidence level (1− α) or risk level α. Of course, if the cost of completed work is unfavorable (μ〈X_1), or if the correlation is negative, then cost differential will become negative; in turn, this will cause the available contingency to go down. Consequently, this increases the risk of a cost overrun above the originally desired level α.

(3) MPA for cost update

Construction project consists of many work packages, and so general expressions in this paper are described by the joint multivariate probability analysis (MPA) in N-variates, where each variate represents a work package in a project.

- x=N×1 column vector or work package cost defined as one random variable in statistical analysis.
- μ=N×1 column vector of expected values of work package costs derived from experts' subjective judgment and historical information in past similar projects.
- V=N×N covariance matrix representing the relationship among random variables or work packages with a likely numerical value. This matrix often relies heavily on information-based judgment, but it has some restrictions such as "square," "symmetric," and "positive definite matrix" in multivariate probabilistic approach for estimating the project cost.
- |V| = Determinant of covariance matrix
- V^{-1}= Inverse of covariance matrix

From probability theory and notations above, the joint multivariate probability analysis for project cost estimate yields the following:

$$f_X(x) = \frac{1}{(2\pi)^{N/2}\sqrt{|V|}}\exp\left\{-\frac{(x-\mu)'V^{-1}(x-\mu)}{2}\right\}$$

The (x− μ)′ matrix is a transpose of (x− μ) and a N×1 column vector. It is needed to clarify a few restrictions on covariance matrix (V) for mathematically validating the application of joint multivariate probability analysis into a project cost estimate. From a prior condition of a project, it is possible to compute the required contingency to meet any significant confidence level. In this paper, 10% risk and 90% confidence level are defined as the likelihood that an actual total project cost exceeds any specific project cost. These levels are dependent upon the characteristics, conditions, and uncertainties of the project; and they often depend on experts' judgment.

The expected total cost for a project and its variance are computed with the Eqs.(1) and (2).

$$E[\text{Total Project Cost}] = \mu_1 + \mu_2 + \cdots + \mu_N = \sum^{N} \mu_i \quad \cdots\cdots(1)$$

$$Var[\text{Total Project Cost}] = \sum_{i=1}^{N}\sum_{j=1}^{N} V_{ij} \quad \cdots\cdots\cdots\cdots(2)$$

From the expectation and standard deviation for total project cost, a risk function (e.g. probability distribution) is created, and the degree and likelihood that the cost exceeds any specific value at the initial allowable risk level is computed. These actions are helpful to employ a better strategy for achieving the objectives within project duration. To understand the theoretical and mathematical model of multivariate probabilistic analysis (MPA), a construction project consisting of N^{th} work packages is considered, and for more convenient illustration N^{th} work package is first completed. Next, $(N-1)^{th}$ work package is also completed. The N^{th} work completion is known as specific information for revising the risk function of total project cost with N-1 remaining work packages. The revisions are performed repeatedly until the only remaining work package is x_1. The MPA represents mathematically partition information for the costs of the completed (x_N) and remaining work packages (x_1, x_2, ⋯, x_{N-1}) as follows:

$$\underset{N\times 1}{x} = \left[\begin{array}{c} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{N-1} \\ \hline x_N \end{array}\right] \quad x_1 \text{ ((N-1)x1 matrix)} \qquad \underset{N\times 1}{\mu} = \left[\begin{array}{c} \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_{N-1} \\ \hline \mu_N \end{array}\right] \quad \mu_1 \text{ ((N-1)x1 matrix)}$$

The partition for variance is created based upon a prior N×N covariance matrix and the reported data (x_N); this is presented below, where each component (v_{ij}) in this matrix is computed by prior standard deviations (s_i) and correlation coefficients (c_{ij}) between i and j work package ($1 \le i, j \le N$). For instance, v_{11} is calculated by c11×s1×s1.

V_{11} ((N-1)x(N-1) matrix) $\quad V_{1N}$ ((N-1)x1 matrix)

$$\underset{N\times N}{V} = \left[\begin{array}{cccc|c} v_{11}, & v_{12}, & v_{13}, & \cdots, & v_{1N} \\ v_{21}, & v_{22}, & v_{23}, & \cdots, & v_{2N} \\ v_{31}, & v_{32}, & v_{33}, & \cdots, & v_{3N} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \hline v_{N1}, & v_{N2}, & v_{N3}, & \cdots, & v_{NN} \end{array}\right]$$

V_{N1}(1x(N-1) matrix)

From the above mathematical definitions, (N−1)×1 of x variates is distributed as (N−1) multivariate normal density functions with (N−1)×1 matrix of μ and (N−1)×(N−1) covariance matrix. Here, (N−1) multivariate normal density functions are dependent upon the actual value of x_N variate. The conditional risk functions from $x_1, x_2, \cdots, x_{N-1}$ variates are re-created given the actual value of x_N variate. The following modified equations are presented for computing the conditional expectations and (N−1)×(N−1) covariance matrix for remaining work packages.

$$\underset{(N-1)\times 1|x_N=reported}{\mu} = \underset{(N-1)\times 1}{\mu} + V_{1N}V_{NN}^{-1}(x_N - \mu_N) \quad \cdots\cdots\cdots(3)$$

$$\underset{(N-1)\times(N-1)|x_N=reported}{V} = \underset{(N-1)\times(N-1)}{V} - V_{1N}V_{NN}^{-1}V_{N1} \quad \cdots\cdots\cdots(4)$$

Theoretically, the reported cost data of completed work package is utilized to create the modified conditional probability distribution of remaining work packages. Thus, the conditional (N−1) expectations and (N−1)×(N−1) covariance matrix are derived from Eqs. (3) and (4). As a result, project managers can calculate the required contingency to meet the initial cost risk level by re-calculating a posterior expected cost and variation with Eqs. (5) and (6).

$$E[\text{Total Project Cost} \mid x_N = \text{reported cost value}] = \mu_{1|x_N} + \mu_{2|x_N} + \cdots + \mu_{N-1|x_N} \quad \cdots\cdots\cdots(5)$$

$$Var[\text{Total Project Cost} \mid x_N = \text{reported cost value}] = Sum\ of\ V_{11|x_N} \quad \cdots\cdots(6)$$

Hence, the likelihood of cost overrun can be re-computed using the normal probability distribution table. From Eqs.(1) and (2), a prior total project cost (B_0) at the risk level is computed like the following.

$$B_0 = E[\text{Total Project Cost}] + k\sqrt{Var[\text{Total Project Cost}]}$$

As stated earlier, k is determined from the normal probability distribution table corresponding to the desired risk level (α). If the risk level is about 10%, then k is about 1.283, where k is defined as indication called the standard unit variate. The expected cost and standard deviation of total project cost conditional on the reported cost data are revised by Eqs.(5) and (6); the modified standard deviations ($\sigma_1, \sigma_2, \cdots, \sigma_{N-1}$) of incomplete work packages are obtained from a prior standard deviation (s_i) and correlation of each pair of work packages (c_{iN}) from Eq.(7).

$$\sigma_i = \sqrt{s_i^2(1 - c_{iN}^2)},\ 1 \le i \le N-1 \quad \cdots\cdots\cdots\textbf{Eq. (7)}$$

The revised correlation coefficients (ρ_{ij}) are also derived from c_{ij} and c_{iN} :

$$\rho_{ij} = \frac{(c_{ij} - c_{iN}c_{jN})}{\sqrt{(1-c_{iN}^2)(1-c_{jN}^2)}},\ 1 \le i \le N-1,\ 1 \le j \le N-1 \quad \cdots\cdots(8)$$

From Eqs.(5), (6), (7), and (8), the expectation and variance of project cost is re-estimated, such that the modified total project cost (B_{new}) is calculated like the below.

$$B_{new} = (\text{Modified } E[\text{Total Project Cost}]) + k\sqrt{(\text{Modified } Var[\text{Total Project Cost}])}$$

Like B_0, k is initially determined from the normal probability distribution table. The proposed methodology in this paper is tested by using data of a building project, and the results are analyzed in the next section.

4. TESTS AND VALIDATIONS OF MODEL

The required contingencies to cover the acceptable risk level of cost overrun is re-estimated as a project progresses and actual cost data is reported, compared to the initial cost estimates at the start. The MPA responds to informed cost data on an ongoing project. If total project cost re-estimated with the most recently reported cost data is higher than the initial cost, supposing no scope changes, more contingency is required and some corrective management is needed to develop the project within the planned cost. Otherwise, if the revised cost at completion is lower than the initial cost, the required contingency is decreased and the savings can be redistributed elsewhere. Hence, the costs of all incomplete work packages are re-estimated as the actual cost data of completed work package is obtained. As an illustration, a building project was considered to demonstrate the model. This project consisted of 14 main work packages; the best estimated cost for each work package is presented in Table 1.

Table 1 Initial best estimates of each work package

No.	Work package (wp)	Best Estimate of Cost
wp1	Site Work	$2,138,148
wp2	Civil	$2,762,211
wp3	Piling	$16,400,483
wp4	Structure and Concrete	$21,745,348
wp5	Masonry	$634,405
wp6	Moisture protection	$1,011,885
wp7	Finishes	$4,432,300
wp8	Carpentry	$2,198,509
wp9	Windows and glass	$2,174,364
wp10	Metals	$442,963
wp11	Furnishings	$2,173,146
wp12	Electrical	$1,606,907
wp13	Mechanical	$4,284,248
wp14	Landscape	$1,305,412
	Expected Total Cost	$63,310,331

The uncertainty on each work package cost was determined by the coefficient of variation, which was 30% for every individual work package. The coefficient of variation is the standard deviation divided by the expected value, and can be controlled by a project manager. Then, the expected total cost at completion is the sum ($=\mu_1+\mu_2+\mu_3+\cdots+\mu_{14}$) of the expectations of 14 work packages, which was $63,310,331. The standard deviation of total cost was computed by $16,796,517 ($=[\Sigma_i \Sigma_j v_{ij}]^{1/2}$, where $1 \leq i, j \leq 14$). This value indicates that the coefficient of variation of total project cost is about 26.5%, which is less than that for each work package taken separately. Figure 2 shows a prior distribution on total project cost, and Table 2 presents the correlation matrix used in this project, which was created by the GCF.

Table 2 Initial correlation matrix created by Gaussian Correlation Function (GCF)

	wp1	wp2	wp3	wp4	wp5	wp6	wp7	wp8	wp9	wp10	wp11	wp12	wp13	wp14
wp1	1	0.9	0.6561	0.3874	0.1853	0.0718	0.0225	0.0057	0.0012	0.0	0.0	0.0	0.0	0.0
wp2	0.9	1	0.85	0.5220	0.2316	0.0743	0.0172	0.0029	0.0	0.0	0.0	0.0	0.0	0.0
wp3	0.6561	0.85	1	0.8	0.4096	0.1342	0.0281	0.0038	0.0	0.0	0.0	0.0	0.0	0.0
wp4	0.3874	0.5220	0.8	1	0.78	0.3702	0.1069	0.0188	0.0020	0.0	0.0	0.0	0.0	0.0
wp5	0.1853	0.2316	0.4096	0.78	1	0.76	0.3336	0.0846	0.0124	0.0010	0.0	0.0	0.0	0.0
wp6	0.0718	0.0743	0.1342	0.3702	0.76	1	0.74	0.2999	0.0665	0.0081	0.0005	0.0	0.0	0.0
wp7	0.0225	0.0172	0.0281	0.1069	0.3336	0.74	1	0.72	0.2687	0.0520	0.0052	0.0	0.0	0.0
wp8	0.0057	0.0029	0.0038	0.0188	0.0846	0.2999	0.72	1	0.7	0.2401	0.0404	0.0033	0.0	0.0
wp9	0.0012	0.0	0.0	0.002	0.0124	0.0665	0.2687	0.7	1	0.68	0.2138	0.0311	0.0021	0.0
wp10	0.0	0.0	0.0	0.0	0.001	0.0081	0.0520	0.2401	0.68	1	0.66	0.1897	0.0238	0.0013
wp11	0.0	0.0	0.0	0.0	0.0	0.0005	0.0052	0.0404	0.2138	0.66	1	0.64	0.1678	0.0180
wp12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0033	0.0311	0.1897	0.64	1	0.62	0.1478
wp13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0021	0.0238	0.1678	0.62	1	0.6
wp14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0013	0.0180	0.1478	0.6	1

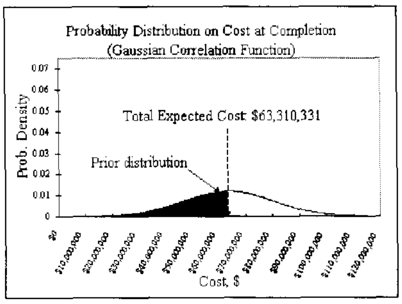

Figure 2 Planned risk function on the cost at completion

The correlation components in the matrix are created by the GCF. However, project managers can manage these values on the basis of their experiences and subjective judgments. When the contingency was set at the 90% confidence level, it represented that there was 10% chance that the cost overruns, and a 90% chance that total cost was not overrun. Standard unit variate for one-sided 90% confidence level is 1.283, so the required total cost at the 90% confidence level is $63,310,331 plus 1.283$16,796,517 = $21,549,931. When the site-work was completed, it cost $2,228,846, which means 4.24% higher than the expected value of its original estimate of $2,138,148. The re-estimates of remaining work packages are calculated by Eq. (3) in the below.

$$\mu_{(\text{Re-estimates of wp2, wp3, ..., wp14})\,|\,\text{wp1}=\$2,228,846} = \mu_{(\text{Prior estimates of wp2, wp3, ..., wp14})} + V_{13\times 1}V_{1\times 1}^{-1}(\$2,228,846 - \$2,138,148)$$

$$V_{\text{Re-estimated matrix among wp2, wp3, ..., wp14}\,|\,\text{wp1}=\$2,228,846} = V_{\text{Prior } 13\times 13 \text{ matrix among wp2, wp3, ..., wp14}} - V_{13\times 1} \times V_{1\times 1}^{-1} \times V_{1\times 13}$$

Table 3 Initial covariance matrix created by Gaussian Correlation Function (GCF)

V_{1x1} V_{1x13}

	wp1	wp2	wp3	wp4	wp5	wp6	wp7	wp8	wp9	wp10	wp11	wp12	wp13	wp14
wp1	7.31E+11	8.5E+11	4.54E+12	5.42E+12	1.42E+11	2.04E+11	8.06E+11	3.6E+11	3.2E+11	5.87E+10	2.59E+11	1.73E+11	4.14E+11	1.14E+11
wp2	8.5E+11	1.22E+12	6.16E+12	6.94E+12	1.72E+11	2.33E+11	8.69E+11	3.66E+11	3.06E+11	5.33E+10	2.22E+11	1.4E+11	3.17E+11	8.21E+10
wp3	4.54E+12	6.16E+12	4.3E+13	4.56E+13	1.07E+12	1.36E+12	4.76E+12	1.89E+12	1.5E+12	2.44E+11	9.57E+11	5.86E+11	1.21E+12	2.94E+11
wp4	5.42E+12	6.94E+12	4.56E+13	7.57E+13	1.72E+12	2.14E+12	7.32E+12	2.83E+12	2.18E+12	3.47E+11	1.33E+12	7.86E+11	1.59E+12	3.79E+11
wp5	1.42E+11	1.72E+11	1.07E+12	1.72E+12	6.44E+10	7.81E+10	2.6E+11	9.8E+10	7.36E+10	1.14E+10	4.25E+10	2.39E+10	4.84E+10	1.12E+10
wp6	2.04E+11	2.33E+11	1.36E+12	2.14E+12	7.81E+10	1.84E+11	5.31E+11	1.95E+11	1.43E+11	2.15E+10	7.81E+10	4.27E+10	8.43E+10	1.9E+10
wp7	8.06E+11	8.69E+11	4.76E+12	7.32E+12	2.6E+11	5.31E+11	3.14E+12	1.12E+12	7.99E+11	1.17E+11	4.14E+11	2.2E+11	4.23E+11	9.29E+10
wp8	3.6E+11	3.66E+11	1.89E+12	2.83E+12	9.8E+10	1.95E+11	1.12E+12	7.73E+11	5.35E+11	7.64E+10	2.62E+11	1.36E+11	2.53E+11	5.4E+10
wp9	3.2E+11	3.06E+11	1.5E+12	2.18E+12	7.36E+10	1.43E+11	7.99E+11	5.35E+11	7.94E+11	1.05E+11	3.5E+11	1.76E+11	3.19E+11	6.6E+10
wp10	5.87E+10	5.33E+10	2.44E+11	3.47E+11	1.14E+10	2.15E+10	1.17E+11	7.64E+10	1.05E+11	[illegible]	1.02E+11	4.96E+10	8.73E+10	1.76E+10
wp11	2.59E+11	2.22E+11	9.57E+11	1.33E+12	4.25E+10	7.81E+10	4.14E+11	2.62E+11	3.5E+11	1.02E+11	7.56E+11	3.58E+11	6.1E+11	1.19E+11
wp12	1.73E+11	1.4E+11	5.66E+11	7.66E+11	2.39E+10	4.27E+10	2.2E+11	1.36E+11	1.76E+11	4.96E+10	3.58E+11	4.41E+11	6.83E+11	1.29E+11
wp13	4.14E+11	3.17E+11	1.21E+12	1.59E+12	4.84E+10	8.43E+10	4.23E+11	2.53E+11	3.19E+11	8.73E+10	6.1E+11	6.83E+11	2.94E+12	5.37E+11
wp14	1.14E+11	8.21E+10	2.94E+11	3.79E+11	1.12E+10	1.9E+10	9.29E+10	5.4E+10	6.6E+10	1.76E+10	1.19E+11	1.29E+11	5.37E+11	2.73E+11

V_{13x1} V_{13x13}

Table 4 Modified correlation matrix after the completion of site-work

	wp1	wp2	wp3	wp4	wp5	wp6	wp7	wp8	wp9	wp10	wp11	wp12	wp13	wp14
wp1	Completed													
wp2		1	0.7889	0.4313	0.1514	0.0222	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
wp3		0.7889	1	0.7845	0.3884	0.1157	0.0177	0.0	0.0	0.0	0.0	0.0	0.0	0.0
wp4		0.4313	0.7845	1	0.7817	0.3723	0.1065	0.0180	0.0017	0.0001	0.0	0.0	0.0	0.0
wp5		0.1514	0.3884	0.7817	1	0.7817	0.3723	0.1065	0.0180	0.0017	0.0001	0.0	0.0	0.0
wp6		0.0222	0.1157	0.3723	0.7817	1	0.7405	0.3002	0.0666	0.0081	0.0005	0.0	0.0	0.0
wp7		0.0	0.0177	0.1065	0.3723	0.7405	1	0.7201	0.2688	0.0520	0.0052	0.0003	0.0	0.0
wp8		0.0	0.0	0.0180	0.1065	0.3002	0.7201	1	0.7000	0.2401	0.0404	0.0033	0.0001	0.0
wp9		0.0	0.0	0.0017	0.0180	0.0666	0.2688	0.70	1	0.6800	0.2138	0.0311	0.0021	0.0001
wp10		0.0	0.0	0.0001	0.0017	0.0081	0.0520	0.2401	0.68	1	0.66	0.1897	0.0238	0.0013
wp11		0.0	0.0	0.0	0.0001	0.0005	0.0052	0.0404	0.2138	0.66	1	0.64	0.1678	0.0180
wp12		0.0	0.0	0.0	0.0	0.0	0.0003	0.0033	0.0311	0.1897	0.64	1	0.62	0.1478
wp13		0.0	0.0	0.0	0.0	0.0	0.0	0.0001	0.0021	0.0238	0.1678	0.62	1	0.6
wp14		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0001	0.0013	0.0180	0.1478	0.6	1

Table 5 Modified covariance matrix after the completion of site-work

	wp1	wp2	wp3	wp4	wp5	wp6	wp7	wp8	wp9	wp10	wp11	wp12	wp13	wp14
wp1	Completed													
wp2		2.32E+11	8.77E+11	6.38E+11	6.63E+09	-4.22E+09	-6.78E+10	-5.18E+10	-6.42E+10	-1.49E+10	-7.89E+10	-5.08E+10	-1.64E+11	-4.99E+10
wp3		8.77E+11	1.48E+13	1.2E+13	1.81E+11	8.95E+10	-2.43E+11	-3.45E+11	-4.84E+11	-1.21E+11	-6.54E+11	-5.06E+11	-1.38E+12	-4.11E+11
wp4		6.38E+11	1.2E+13	3.55E+13	6.66E+11	6.26E+11	1.34E+12	1.64E+11	-1.9E+11	-9.82E+10	-5.94E+11	-5.13E+11	-1.48E+12	-4.63E+11
wp5		6.63E+09	1.81E+11	6.66E+11	3.67E+10	3.65E+10	9.26E+10	2.22E+10	5.37E+09	-1.27E+09	-1.41E+10	-1.43E+10	-4.44E+10	-1.46E+10
wp6		-4.22E+09	8.95E+10	6.26E+11	3.65E+10	1.07E+11	3.06E+11	9.44E+10	5.32E+10	5.1E+09	5.63E+09	-5.49E+09	-3.14E+10	-1.27E+10
wp7		-6.78E+10	-2.43E+11	1.34E+12	9.26E+10	3.06E+11	2.26E+12	7.26E+11	4.47E+11	5.26E+10	2.26E+11	3.06E+10	-3.26E+10	-3.22E+10
wp8		-5.18E+10	-3.45E+11	1.64E+11	2.22E+10	9.44E+10	7.26E+11	5.96E+11	3.78E+11	4.75E+10	1.35E+11	5.09E+10	4.97E+10	-1.8E+09
wp9		-6.42E+10	-4.84E+11	-1.9E+11	5.37E+09	5.32E+10	4.47E+11	3.78E+11	6.16E+11	7.91E+10	2.36E+11	1E+11	1.37E+11	1.63E+10
wp10		-1.49E+10	-1.21E+11	-9.82E+10	-1.27E+09	5.1E+09	5.26E+10	4.75E+10	7.91E+10	2.57E+10	6.08E+10	3.58E+10	5.41E+10	8.44E+09
wp11		-7.89E+10	-6.54E+11	-5.94E+11	-1.41E+10	5.63E+09	2.26E+11	1.35E+11	2.36E+11	6.08E+10	6.84E+11	2.98E+11	4.63E+11	7.88E+10
wp12		-5.08E+10	-5.06E+11	-5.13E+11	-1.43E+10	-5.49E+09	3.06E+10	5.09E+10	1E+11	3.58E+10	2.98E+11	3.72E+11	5.85E+11	1.02E+11
wp13		-1.64E+11	-1.38E+12	-1.48E+12	-4.44E+10	-3.14E+10	-3.26E+10	4.97E+10	1.37E+11	5.41E+10	4.63E+11	5.85E+11	2.7E+12	4.73E+11
wp14		-4.99E+10	-4.11E+11	-4.63E+11	-1.46E+10	-1.27E+10	-3.22E+10	-1.8E+09	1.63E+10	8.44E+09	7.88E+10	1.02E+11	4.73E+11	2.55E+11

$V_{13x13\,|\,wp1=\$2,228,846}$

Table 6 Modified best estimates of remaining work packages after the completion of site-work

No.	Work package (wp)	Best Estimate of Cost	Modified Estimate
wp1	Site Work	$2,138,148	Actual cost = $2,228,846
wp2	Civil	$2,762,211	$2,867,664
wp3	Piling	$16,400,483	$16,856,925
wp4	Structure and Concrete	$21,745,348	$22,102,710
wp5	Masonry	$634,405	$639,392
wp6	Moisture protection	$1,011,885	$1,014,966
wp7	Finishes	$4,432,300	$4,436,536
wp8	Carpentry	$2,198,509	$2,199,043
wp9	Windows and glass	$2,174,364	$2,174,473
wp10	Metals	$442,963	$442,967
wp11	Furnishings	$2,173,146	$2,173,149
wp12	Electrical	$1,606,907	$1,606,907
wp13	Mechanical	$4,284,248	$4,284,248
wp14	Landscape	$1,305,412	$1,305,412
Expected Total Cost		$63,310,331	$64,333,238

The modifications of expected project cost and standard deviation are computed based upon informed cost data of site-work by Eqs. (4), (5), and (6), and a posterior risk function is compared to the planned function as presented in Figure 3. In brief, given site-work cost information, it is possible to modify the best estimates for remaining work packages from civil-work to landscape using the multivariate probabilistic analysis (MPA), such that their revised values are given in Table 6. Based upon the actual cost of site-work ($2,228,846), the initial correlation and covariance matrix are re-created using Eqs. (6), (7), and (8) as shown in Tables 4 and 5.

The fact that site-work was completed for more than the prior best estimate of $2,138,148 indicates that there is some evidence that its initial cost estimate was under-estimated. Also, this underestimate cost may be due to the unpredictable complications that cannot be accounted for earlier. Consequently, there is a need to revise the estimates from civil to landscape work packages because of the correlations among them. Hence, the best estimate of total expected cost at completion is modified by $64,333,238, such that $1,022,907 is increased from the initial estimate. If project managers do not increase total project cost, so that the cost with contingency remains constant at the original value, the remaining contingency is only $17,644,263. This is positive, but still less than the required contingency to cover the remaining costs at the 90% confidence level. Site-work has actually used up some of project contingency, although no specific contingency is assigned to site-work. Thus, after the completion of site-work, the probability of cost overrun is about 11.4%, which is more than 10% risk level. This means that the total cost for remaining work packages corresponds to the 88.6% confidence level, a decrease from the 90% initial level. The revised probability distribution after the completion of site-work is plotted in Figure 3.

The width of the modified probability distribution is slightly narrower than the initial one. Next, the civil-work was completed for a cost of $2,949,061. This is slightly higher than the updated prediction, which is the best re-estimate for civil-work ($2,867,664) after site-work was completed.

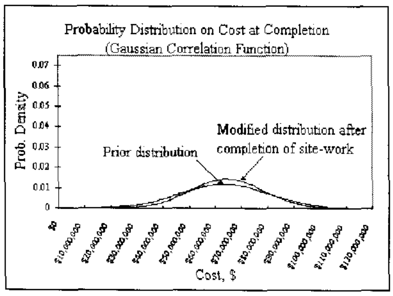

Figure 3 Modified risk function after the completion of site-work

However, $186,850 is higher than the original estimate. This additional evidence confirms that the initial project costs were under-estimated and running higher than the estimates. The revised values for the best estimates of incomplete work packages are calculated in the same previous way after obtaining cost data from the completions of site-work and civil-work. Then, the modified probability distribution of project cost is shown in Figure 4.

After each work package was completed, the variation in the revised total costs and in the expected costs at completion is shown in Figure 5. Whereas the variation resulted from the ECF is plotted in Figure 6. The objective of using the confidence limits in these figures is to keep

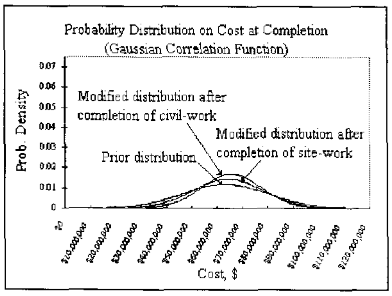

Figure 4 Modified risk function after the completions of site-work and civil-work

Table 7 Comparison between initial best estimates and actual costs

No.	Work package (wp)	Best Estimate of Cost	Actual Cost
wp1	Site Work	$2,138,148	Actual cost = $2,228,846
wp2	Civil	$2,762,211	Actual cost = $2,979,061
wp3	Piling	$16,400,483	Actual cost = $18,605,588
wp4	Structure and Concrete	$21,745,348	Actual cost= $22,084,531
wp5	Masonry	$634,405	Actual cost= $583,008
wp6	Moisture protection	$1,011,885	Actual cost= $868,864
wp7	Finishes	$4,432,300	Actual cost= $4,113,908
wp8	Carpentry	$2,198,509	Actual cost= $2,259,430
wp9	Windows and glass	$2,174,364	Actual cost= $2,624,082
wp10	Metals	$442,963	Actual cost= $628,353
wp11	Furnishings	$2,173,146	Actual cost= $2,290,456
wp12	Electrical	$1,606,907	Actual cost= $2,046,128
wp13	Mechanical	$4,284,248	Actual cost= $4,538,749
wp14	Landscape	$1,305,412	Actual cost= $1,282,895
Expected Total Cost		$63,310,331	$67,133,899

the confidence band positioned so that it envelops the actual cost at completion. No one can exactly predict the future with certainty, but this study tries to define a confidence limit where it is expected to find it with probability 90% in this case. The lower confidence limit is not shown with the fact that it is the expected cost at completion minus the contingency. However, the presented method at least has achieved the goal of keeping the 90% confidence level above the actual cost at completion ($67,133,899) for every completion in the process up to project completion. Although the costs increased over the original estimates and contingency was being used up for a few steps, this project was actually well behaved in which it never ran out of contingency.

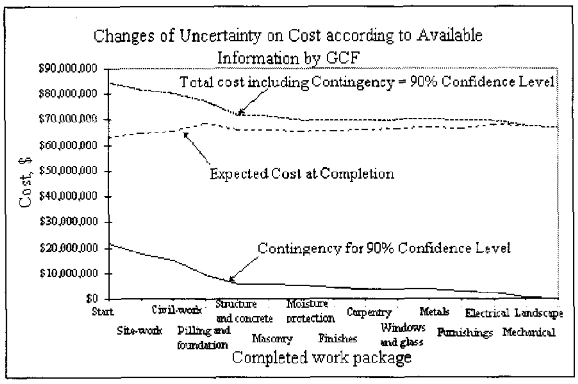

Figure 5 Variation in the revised total costs and expected costs based upon Gaussian Correlation Function

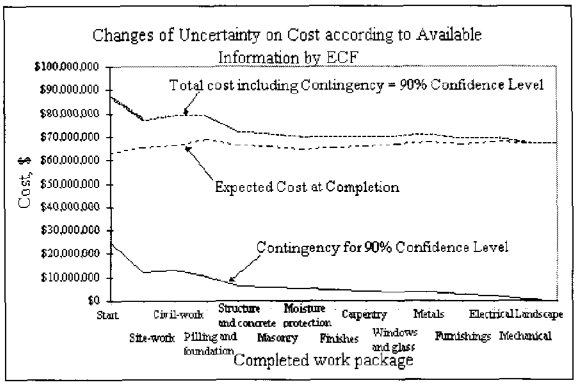

Figure 6 Variation in the revised total costs and expected costs based upon Exponential Correlation Function

As each subsequent work package was completed, the best estimate (or most likely value) for the cost at completion crept up or down somewhat, depending on the latest actual cost reports. In spite of this fluctuation, the uncertainty expressed as the width of the distribution always lessens for both of the GCF and ECF. Figures 5 and 6 appropriately illustrate how the project cost is changing and how well the degree of uncertainty is merging to zero on the target as actual cost data becomes more available.

5. CONCLUSIONS

This study introduced a cost re-estimating model using the Multivariate Probabilistic Analysis (MPA) and Bayes' Theorem to accurately predict construction project cost and contingency. The model has predicted the changes of the costs of incomplete work packages and expected project cost in quantitative assessment, resulted from actual cost data of completed work packages of ongoing project. Moreover, this has quantified the magnitude of the impacts of informed cost data on the re-estimates of the remaining project execution. If project managers foresee the behavior of expected project cost based on the actual costs of completed work packages and forecast it at the early stage, then they will obtain significant information to efficiently arrange the resources and to implement better strategies.

Commonly, a construction project has experienced cost overrun due to the use of limited and deterministic information and the inaccurate prediction of dynamic uncertainties and changeable factors as a result from complications of project circumstances at the starts. The proposed methodology in this paper contributes to modifying the costs of all work packages to be completed corresponding to reported cost data from completed work packages. Such revisions provide project managers with valuable information for embodying strategic management actions, and quantitatively analyzing project cost performance at the desired confidence level.

To illustrate the use of the model, data from a small construction project was employed and validated. The results show a consistent contingency strategy by keeping a specific risk level of cost overruns throughout the project executions. Further, these results confirm that the reduced uncertainties from increasing cost information had decreased the need for contingency. Consequently, project managers can examine the decrease amount of contingency, and thus, consider the potential to re-allocate project funding elsewhere. With these strategic management actions they can diminish the risks of cost overruns. In the further study, there are challenges to create a universal correlation function for all the construction projects and to furnish a way to practically re-allocate decreased risks resulted from the predicted variations.

REFERENCES

CII (1994) "Pre-Project Planning: Beginning a project the right way" Published by Construction Industry Institute, Texas.

Clark, D.E. (2001) "Monte Carlo Analysis: Ten Year of Experience" Cost Engineering, Vol. 43, No. 6, pp. 40-45.

Ford, D.N. (2002) "Achieving multiple project objectives through contingency management" Journal of Construction Engineering and Management, Vol. 128, No. 1, pp. 30-39.

Johnson, R.A. and Wichern, D.W. (2002) "Applied Multivariate Statistical Analysis" Published by Prentice Hall, New Jersey.

Karlsen, J.T. and Lereim, J. (2005) "Management of project contingency and allowance" Cost Engineering, Vol. 47, No. 9, pp. 24-29.

Nassar, K. (2005) "Cost contingency analysis for construction project using spreadsheets" Cost Engineering, Vol. 47, No. 9, pp. 24-29.

Pugh, L.A. and Soden, R.G. (1986) "Use of risk analysis techniques in assessing the confidence of project cost estimates and schedules" Project Management, Vol. 4, No. 3, August pp. 158-162.

Touran, A. (2003) "Probabilistic model for cost contingency" Journal of Construction Engineering and Management, Vo. 129, No. 3, pp. 280-284.

Zwikael, O., Globerson, S., and Raz, T. (2000) "Evaluation of models for forecasting the final cost of a project" Project Management Journal, March, pp. 53-57.

논문제출일: 2007.03.06
심사완료일: 2007.06.01