SIMULATION-FACILITATED FACTOR-BASED APPROACH FOR COST CORRELATION EVALUATION

Wei-Chih Wang

Associate Professor, Department of Civil Engineering
National Chiao Tung University, Hsin-Chu 300, Taiwan;
(886)3-5712121 ext. 54952; weichih@cc.nctu.edu.tw

ABSTRACT: Project cost becomes increasingly variable if many cost items for a construction project are correlated, and this can increase the uncertainty of completing a project within a target budget. This work presents a factor-based computer simulation model for evaluating project costs given correlations among cost items. Uncertainty in the total cost distribution of an item is transferred to several factor cost distributions according to qualitative estimates of the sensitivity of each cost item to each factor. Each cost distribution is then decomposed further into a family of distributions (children; costs given factor conditions), with each child corresponding to a factor condition. Correlations are retrieved by sampling from the child distributions with the same-condition for a given iteration of the simulation.

KEYWORDS: Computer Simulation; Cost Estimating; Risk Factor; Uncertainty

1. INTRODUCTION

Accurately estimating costs is an essential task in effectively managing construction projects. Each cost component, and thus project cost, is variable or probabilistic since future events are always uncertain [1]. Project cost becomes increasingly variable if several cost items are correlated, increasing the uncertainty of finishing a project to a target budget. Current research on correlated costs deals with theoretical issues concerning in the accuracy of correlations. For example, Touran and Wiser used a multivariate normal distribution to generate correlated cost variables for a precise simulation analysis, assuming that the correlation coefficients between variables are known [2]. The simulation model of Chau employed a percentile-based sampling procedure to influence the probability of sampling the same quantiles from two correlated probability density functions, according to whether the given correlation coefficient is positive or negative [3]. Finally, Ranasinghe highlighted some theoretical requirements, such as the conditions required to achieve a positive definite correlation matrix and the possibility of using an induced correlation to define the correlation between derived variables [4].

This paper presents a simulation-based cost model that considers correlations between cost items [5]. In contrast to existing cost related models in incorporating correlations, the proposed model is designed to meet the following three requirements which are considered practical in a cost management tool, namely: not requiring excessive input from management, introducing correlations indirectly (since this correlation information is not readily available) [2], and recognizing factor-based correlations when they occur in the field.

2. THE PROPOSED MODEL

2.1 Breakdown of uncertainty

The proposed model treats the cost of a bill item as a random variable. The cost variable is represented by a total cost distribution (that is, "grandparent" distribution) that combines a base cost with variations resulting from various factors. Variations owing to a particular factor are represented by a cost distribution, a "parent" distribution. The base cost is assumed to be deterministic, while the cost distribution for each factor is assumed to be a zero-mean random variable. Figure 1 schematically depicts this approach to break down the uncertainty. The base cost is taken to be the user's best estimate of an item's cost under expected factor conditions, and is the expected value of the total cost distribution for the item. Deviations from the expected value caused by various factors are introduced through the cost distributions.

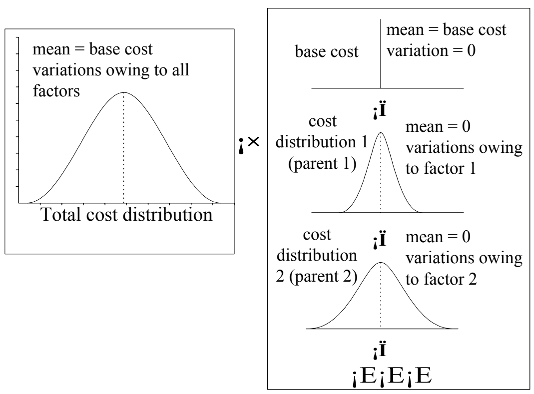

Figure 1. Breakdown of uncertainty

The model captures correlations by drawing cost samples from related portions of the cost distributions for cost items that are sensitive to a given factor. For example, the upper part of Fig. 2 classifies weather conditions into “better than expected,” “normally expected,” and “worse than expected.” Based on these three different weather conditions, the weather related cost distribution is disaggregated into three corresponding child distributions (illustrated in the lower half of Fig. 2), namely, cost given better than expected weather (that is, better than expected weather child), cost given normally expected weather (that is, normally expected weather child), and cost given worse than expected weather (that is, worse than expected weather child). Child distributions may also overlap, as presented in Fig. 2. Restated, the cost of an item may be the same under both better than expected and normally expected weather conditions; or the cost with normally expected weather conditions may be less than the cost with better than expected weather.

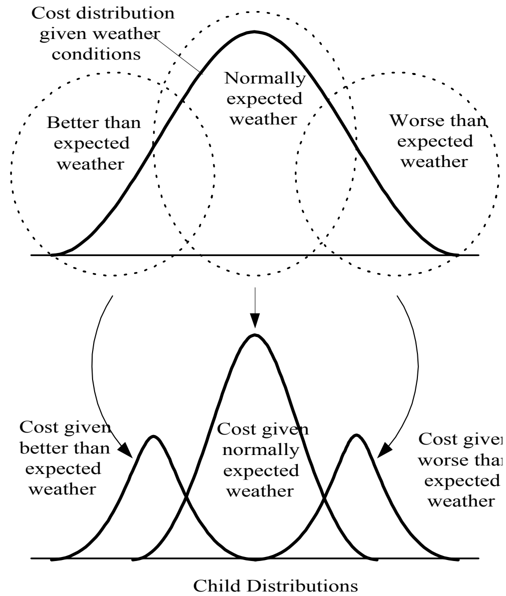

Figure 2. Decomposition of cost distribution into costs, given particular factor conditions

A cost model, in which the effect of uncertainty is broken down by factor, is derived from the unit cost perspective [5]. Following a series of derivation [5], C_i the cost of item i, may be expressed as

$$C_i = c_{i(0)} + \sum_{j=1}^{J} c_{i(j)} \tag{1}$$

where $c_{i(0)}$ is the estimated (or base) cost and the random variable $c_{i(j)}$, j = 1,..., J, is the cost (parent) distribution of cost item i due to factor j. Restated, Equation (1) displays the variations in the cost of an item, as a base cost and a series of cost distributions for various factors.

The model assumes that the costs of items are correlated only through the impact of shared factors. Different factors are assumed to cause independent effects. For example, assume that cost item 1 is sensitive to weather and labor, and cost item 2 is sensitive to weather and equipment. Only the weather-related cost distributions are correlated; the variations caused by labor and equipment are assumed to be independent. Then, regardless of the type of the marginal distribution of $c_{i(j)}$, the mean and variance of the cost of cost item i can be derived as [5]

$$M_i = m_{i(0)} + m_{i(1)} + m_{i(2)} + \ldots + m_{i(J)} = m_{i(0)} \tag{2}$$

$$\sigma_i^2 = SD_{i(0)}^2 + SD_{i(1)}^2 + SD_{i(2)}^2 + ... + SD_{i(J)}^2$$
$$= SD_{i(1)}^2 + SD_{i(2)}^2 + ... + SD_{i(J)}^2 \quad (3)$$

in which M_i and σ_i are the mean and standard deviation for C_i (the total cost distribution for item i), and $m_{i(j)}$ and $SD_{i(j)}$ are the mean and standard deviation for $c_{i(j)}$, with $SD_{i(0)} = 0$. The model finds M_i and σ_i for cost item i, and then determines $SD_{i(j)}$. In the example project presented herein, the three-point estimates of PERT are used to calculate M_i and σ_i.

In constructing a family of child distributions to represent changes in cost due to factor conditions, one goal is to preserve the mean and standard deviation of the cost distribution. In other words, the mean and standard deviation of the combination of the child distributions for a family should be the same as the mean and standard deviation of the cost distribution. Mathematically, this relationship can be represented [5]

$$m_{i(j)} = \sum_{h=1}^{H} p_{j(h)} \times o_{i[j(h)]} = 0 \quad (4)$$

$$SD_{i(j)}^2 = \sum_{h=1}^{H} p_{j(h)} \times (sd_{i[j(h)]}^2 + o_{i[j(h)]}^2) \quad (5)$$

in which H = number of child distributions; $p_{j(h)}$= probability of occurrence for child distribution h of factor j; and $o_{i[j(h)]}$ and $sd_{i[j(h)]}$ = mean and standard deviation, respectively, for child distribution h of factor j for cost item i. Equations (4) and (5) are valid for any type of statistical distribution. Steiner's theorem can be directly applied to justify (5) [6].

The mean of the child distribution for a given condition is the expected deviation from the mean of the cost distribution when the cost item is performed under the given condition. Means of child distributions are expressed through a variable x, the mean placement. The mean of each child distribution should be confined to a range that maintains the variance of the cost distribution. When x is equal to the limit, the child distributions will have zero standard deviations [5].

To construct a family of child distributions is to determine their means and standard deviations. Consider a cost distribution that is sensitive to factor j and has a variance of \$4 K. Assume that the user chooses the categories of better than expected, normally expected, and worse than expected conditions to describe the conditions of the factor. Then a family of three child distributions should be constructed. Assume that the probabilities of occurrence for the child distributions are equal; that is, $p_1 = p_2 = p_3 = 1/3$. Thus, based on (4) and (5), the mean and variance, respectively, of the combined child distributions are

$$\frac{1}{3} o_1 + \frac{1}{3} o_2 + \frac{1}{3} o_3 = 0 \quad (6)$$

$$\frac{1}{3}(sd_1^2 + o_1^2) + \frac{1}{3}(sd_2^2 + o_2^2) + \frac{1}{3}(sd_3^2 + o_3^2) = 4 \quad (7)$$

Assume $-o_1 = o_3 = x$ and $o_2 = 0$ so that (6) is satisfied, and let the child distributions have equal standard deviations, then (7) can be rewritten as

$$sd^2 + (2/3)x^2 = 4 \quad (8)$$

The limit of the value of x is found by requiring that the variance of the child distribution be non-negative. Namely,

$$sd^2 = 4 - (2/3)x^2 \geq 0 \quad (9)$$

Thus, the limit in this case is $x \leq \sqrt{6}$ = 2.45 (limit = 2.45). In other words, the values of 2.45 and -2.45 are the two extreme means for Child Distributions 1 and 3, respectively. The next step is to select the value of x between 0 and 2.45. Instead of specifying the exact value of x, the proposed model suggests that the value of x be selected according to the level of influence of the factor under consideration on the cost item under consideration. In this example, assume x is set to one-half of the limit. Then x is equal to 1.27. The properties of this family of three child distributions are thus Child 1 (p_1 = 1/3, o_1 = -1.27, sd_1 = 1.71), Child 2 (p_2 = 1/3, o_2 = 0, sd_2 = 1.71), and Child 3 (p_3 = 1/3, o_3 = 1.27, sd_3 = 1.71).

2.2 Qualitative estimates

Cost distributions are derived according to subjective information. Project planners are asked to estimate qualitatively the extent to which each factor influences the cost of each item. For example, a cost item would be considered to be highly sensitivity to weather if its cost varies greatly depending on the weather. This approach of qualitative estimates is practical because the impact of uncertainties is easily expressed linguistically. No inherent restriction is placed on the number of levels of influence used for each factor. The example included herein use four levels of influence, high, medium, low, and no influence.

2.3 Scale system

A scale system is used to transfer the uncertainty associated with total cost distribution to the cost distributions based on qualitative estimates of the uncertainty sensitivity of cost item i to factor j [5][7]. That is,

$$\begin{aligned}\sigma_i^2 &= SD_{i(1)}^2 + SD_{i(2)}^2 + \ldots + SD_{i(J)}^2 \\ &= (w_1[Q_{i(1)}] + w_2[Q_{i(2)}] + \ldots + w_J[Q_{i(J)}]) \times K_i \\ &= (\sum_{j=1}^{J} w_j[Q_{i(j)}]) \times K_i \end{aligned} \quad (10)$$

$$SD_{i(j)}^2 = w_j[Q_{i(j)}] \times K_i \quad (11)$$

where $Q_{i(j)}$ is the qualitative estimate of the sensitivity of cost item i to factor j, and $w_j[Q_{i(j)}]$ is a scale for each level of influence. For example, the values of the estimates of *high*, *medium*, *low*, and *no* sensitivity for factor j can be represented by w_j[High], w_j[Medium], w_j[Low], and w_j[No], respectively. K_i is an adjustment constant that ensures that σ_i^2 is preserved. Since $w_j[Q_{i(j)}]$ is fixed for a given factor j, K_i will be different for each cost item. The value of w_j[No] is always zero. The value of $w_j[Q_{i(j)}]$ is higher when $Q_{i(j)}$ represents a higher level of influence. Consequently, a larger portion of the variance is distributed to a cost distribution that has a higher sensitivity.

2.4 Sensitivity of project cost to uncertainty

When several cost items for a project are sensitive to particular factors, these factors are likely to dominate the cost performance of the project. Knowledge of factor-sensitivities gives management a better idea of what factors to control. For instance, management should focus on carefully scheduling weather-sensitive tasks and ensuring adequate equipment is available if weather and equipment performance exert the biggest influence on project cost. Controlling the factors that influence performance improves performance more than modifying or changing work methods. This study measures the uncertainty sensitivity of each cost item to a given factor based on its standard deviation divided by its mean. A project in which a certain factor has a high standard deviation is considered highly sensitive to that factor (since the mean of project cost is equal for each factor), and consequently project cost is more likely to be affected by a change in that factor.

3. COMPUTER IMPLEMENTATION

In the model, when cost distributions are sensitive to the same factor, a sample cost is independently drawn from a particular child distribution (given a specified probability of occurrence) for each cost distribution. For example, if better than expected, normally expected, and worse than expected weather are equally likely to occur, then one-third of a predefined number simulation iterations will have cost samples that are simultaneously and independently drawn from the better than expected weather child distributions; one-third will have normally expected weather child distributions; and one-third will have worse than expected weather child distributions. A simulation language, STROBOSCOPE [8], is used to execute the simulation-relevant procedure described in the model. This procedure was implemented on a 586 PC with 64 MB under a 32-bit Windows environment (namely, Windows 98). Making 1,000 analyses of twenty-four cost categories of the example project took approximately six minutes, which is acceptable for research.

4. EXAPME DEMONSTRATION

An example for a building project is used to compare the results obtained using the model with two analyses that do not consider correlations, namely: a standard PERT analysis (PERT) and a Monte-Carlo simulation, carried out using normally distributed costs with the same mean and standard deviation as the model's total cost distribution (W/O Correlation Normal). Meanwhile, three different scale systems (Scales 1, 2, and 3) are applied to investigate the effect of the scale system. This project comprises 20 direct-cost division items and 4 indirect-cost division items (that is, insurance, tax, profit, and contingency). The model requires two types of inputs, the three-point cost estimates for each division item and the qualitative estimates of the sensitivity of each division item to various factors. The analyses considered here involve 1,000 simulation iterations. The scales of Scale 1 are listed in Table 1.

Table 1. Scales of Scale 1

	Scales			
F1	$w_{F1}[H]=16$	$w_{F1}[A]=12$	$w_{F1}[L]=8$	$w_{F1}[No]=0$
F2	$w_{F2}[Yes]=12$			$w_{F2}[No]=0$
F3	$w_{F3}[H]=7$	$w_{F3}[A]=5$	$w_{F3}[L]=3$	$w_{F3}[No]=0$
F4	$w_{F4}[H]=4$	$w_{F4}[A]=3$	$w_{F4}[L]=2$	$w_{F4}[No]=0$
F5	$w_{F5}[H]=3$	$w_{F5}[A]=2$	$w_{F5}[L]=1$	$w_{F5}[No]=0$

where "H", "A", "L", and "No" represent high, average, low, and no sensitivity, respectively. "Yes" or "No" are used to describe the sensitivity of cost items to F2. F1 - F5 represent owner approval, weather, material delivery, labor, and equipment, respectively.

Meanwhile, the scales for Scales 2 and 3 (which exaggerate the differences between high, medium, and low sensitivities) are displayed in Table 2 and Table 3, respectively.

Table 2. Scales of Scale 2

	Scales			
F1	$w_{F1}[H]=8$	$w_{F1}[A]=5$	$w_{F1}[L]=1$	$w_{F1}[No]=0$
F2	$w_{F2}[Yes]=8$			$w_{F2}[No]=0$
F3	$w_{F3}[H]=8$	$w_{F3}[A]=5$	$w_{F3}[L]=1$	$w_{F3}[No]=0$
F4	$w_{F4}[H]=8$	$w_{F4}[A]=5$	$w_{F4}[L]=1$	$w_{F4}[No]=0$
F5	$w_{F5}[H]=8$	$w_{F5}[A]=5$	$w_{F5}[L]=1$	$w_{F5}[No]=0$

Table 3. Scales of Scale 3

	Scales			
F1	$w_{F1}[H]=100$	$w_{F1}[A]=10$	$w_{F1}[L]=1$	$w_{F1}[No]=0$
F2	$w_{F2}[Yes]=100$			$w_{F2}[No]=0$
F3	$w_{F3}[H]=100$	$w_{F3}[A]=10$	$w_{F3}[L]=1$	$w_{F3}[No]=0$
F4	$w_{F4}[H]=100$	$w_{F4}[A]=10$	$w_{F4}[L]=1$	$w_{F4}[No]=0$
F5	$w_{F5}[H]=100$	$w_{F5}[A]=10$	$w_{F5}[L]=1$	$w_{F5}[No]=0$

Results: project cost. The project costs obtained from various analyses (PERT, W/O Correlation Normal, With Correlation Scale 1, Scale 2, and Scale 3) are compared using several metrics, namely the mean, standard deviation, minimum and maximum project costs. Table 4 lists the analytical results, and yields the following observations:

- The mean and standard deviations for PERT and W/O Correlation Normal are approximately the same because of the effect of the Central Limit Theorem.
- The analytical results with and without correlation analyses reveal very little difference in mean project cost. Restated, the correlation affects the variance rather than the expected cost.
- Correlation produces a project cost that may be significantly lower than expectations (e.g., $117.96 K for Scale 1 versus $132K for W/O Correlation Normal) or significantly higher than expected (e.g., $184.25 K for Scale 1 versus $167.49K for W/O Correlation Normal). The correlation effect thus has the potential to create an unexpected cost overrun.
- The project standard deviations of the three With Correlation analyses are 153%, 137%, and 149% higher than for the W/O Correlation Normal analysis for Scales 1, 2, and 3, respectively. For this example project, the choice of scale systems does not markedly affect the analytical results, which fact applies even in the case of Scale 3 (highlighting the differences between sensitivities), because the correlation effect determined by Scale 3 is enhanced only when most activities have high sensitivities to the same factor or factors. It was found out that the correlation effect tends to be dominated by the lower-sensitivity factor cost distributions, rather than the higher-sensitivity ones.

Results: uncertainty sensitivity. Table 5 summarizes the results of uncertainty sensitivity to F1, F2, F3, F4, F5, and all factors of project cost for different scale systems. For Scales 1 and 2, the project cost is most sensitive to F4 (labor), followed by F1, F5, F3, and F2. This information tells management that controlling the quality and availability of labor deserves special attention. Meanwhile, in Scale

3, which increases the difference between high, medium, and low sensitivities, F1 becomes the most sensitive factor rather than F4. Notably, the PERT and W/O Correlation Normal models are unable to provide this type of sensitivity information.

Table 4. Comparisons of W/O correlation and the model analyses

Project Cost [a]	PERT	W/O Corr. Normal	Proposed model		
			Scale 1	Scale 2	Scale 3
Mean	150 [b]	149.97	150.06	149.69	150.33
Standard deviation	5.69	5.32	13.46	12.63	13.26
Min. cost	N/A	132	117.96	118.65	117.05
Maxi. cost	N/A	167.49	184.25	184.09	184.38

[a] The results are evaluated considering all factors.
[b] All data are expressed in thousands (K).

Table 5. Effect of scale systems

Factors	Standard deviation		
	Scale 1	Scale 2	Scale 3
1. Owner approval	7.9857[b][2][a]	6.8174 [2]	10.8664 [1]
2. Weather	1.5813[5]	1.9488 [5]	1.7313 [5]
3. Material delivery	2.7240[4]	1.9603 [4]	3.8811 [4]
4. Labor skills	8.3092[1]	8.9019 [1]	5.7696 [2]
5. Equipment breakdown	6.7841[3]	5.5836 [3]	4.0056 [3]
All factors	13.46	12.63	13.26

[a] [ ] indicates the rank of the sensitivity with respect to a given factor.
[b] All data are expressed in thousands (K).

5. CONCLUSIONS

This work has developed a simulation-facilitated factor-based model that allows correlation between cost items to be considered in cost analysis. The model is based upon the two-step breakdown of uncertainty. The correlation between cost distributions is caused by their sharing the same factor(s). Correlation is introduced by sampling from the child distribution representing a given factor condition. The use of qualitative estimates to describe the effect of factor-based uncertainty should make the user more comfortable in providing inputs than other approaches. Future research directions could include exploring ways to capture non-Normal cost distributions and total cost distributions; implementing time-dependent and non-time-correlated cost variables; and applying the proposed model to other practical projects.

6. ACKNOWLEDGEMENTS

The author thanks Professor Laura Demsetz from the College of San Mateo and Professor Hojjat Adeli from the Ohio State University for their valuable assistance. Dr. Julio Martinez is also commended for making STROBOSCOPE available.

7. REFERRENCES

[1] Adeli, H. and Wu, M., "Regularization Neural Network For Construction Cost Estimation", Journal of Construction Engineering and Management, ASCE, Vol. 124, No. 1, pp. 18-24, 1998.

[2] Touran, A. and Wiser, E.D., "Monte Carlo Technique With Correlated Random Variables", Journal of Construction Engineering and Management, ASCE, Vol. 118, No. 2, pp. 258-272, 1992.

[3] Chau, K.W., "Monte Carlo Simulation Of Construction Costs Using Subjective Data", Construction Management and Economics, Vol. 13, pp. 369-383, 1995.

[4] Ranasinghe, M., "Impact Of Correlation And Induced Correlation On The Estimation Of Project Cost Of Buildings", Construction Management and Economics, 18, pp. 395-406, 2000.

[5] Wang, W-C., "Simulation-Facilitated Model For Assessing Cost Correlations", Journal of Computer-Aided Civil and Infrastructure Engineering, Vol. 17:5, pp. 368-380, 2002.

[6] Levin, R.I. and Rubin, D.S., Statistics for Management, 5th Edition, Prentice Hall, Englewood Cliffs, New Jersey, 1991.

[7] Wang, W-C. and Demsetz L. A., "Model For Evaluating Networks Under Correlated Uncertainty – NETCOR", Journal of Construction Engineering and Management, ASCE, 126(6), pp. 458-466, 2000.

[8] Martinez, J. C., STROBOSCOPE: State and Resource Based Simulation of Construction Processes, Ph.D. Dissertation, University of Michigan, Ann Arbor, Michigan, 1996.