*Computer-Aided Civil and Infrastructure Engineering* **17** (2002) 368–380

# Simulation-Facilitated Model for Assessing Cost Correlations

Wei-Chih Wang*

*Department of Civil Engineering, National Chiao Tung University, Hsin-Chu 300, Taiwan*

**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 (COSTCOR) for evaluating project costs given correlations among cost items. Uncertainty in the total cost distribution of an item (grandparent) is transferred to several factor cost distributions (parents) 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. COSTCOR integrates the uncertainty effects caused by each factor at the project cost level, thus making it easier for management to determine what parts of the project need to be controlled.*

## 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 because future events are always uncertain. Many cost estimation models have been developed that account for the effects of uncertainties. These recent cost models involve neural networks (Adeli and Wu, 1998; Adeli and Karim, 2001), simulation (Touran and Wiser, 1992; Chau, 1995; Ranasinghe, 2000), experiential learning theory (Lowe and Skitmore, 1994), and other systematic approaches (Diekmann, 1983; Diekmann and Featherman, 1998; Oberlender and Trost, 2001; Wang, 2002). Only very few current cost models address correlations among cost variables in construction projects. Project cost becomes increasingly variable if several cost items are correlated, increasing the uncertainty of finishing a project to a target budget.

As widely known, however, correlations are often included in evaluations of the durations of construction projects. Related research on correlations of duration variables stresses that correlations result from uncertainties (such as weather, labor skills, site conditions, and management quality) that are shared between activity durations (Woolery and Crandall, 1983; Ahuja and Nandakumar, 1985; Levitt and Kunz, 1985; Padilla and Carr, 1991; Wang and Demsetz, 2000). Other construction activities are likely to be simultaneously influenced by, for example, the presence of poor weather while concrete is being laid. Poor weather increases the duration of every weather-sensitive activity, and vice versa. The variability of the project duration may be significantly increased by the correlations between numerous activities along critical or near critical paths.

Current research on correlated costs deals with theoretical issues concerning the accuracy of correlations. For example, Touran and Wiser (1992) used a multivariate normal distribution to generate correlated cost variables for a precise simulation analysis, assuming that the correlation coefficients between variables are known. The simulation model of Chau (1995) 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. Finally, Ranasinghe (2000) highlighted some theoretical requirements, such as the conditions required to achieve a positive definite correlation

*To whom correspondence should be addressed. E-mail: *weichih@cc.nctu.edu.tw*.

matrix and the possibility of using an induced correlation to define the correlation between derived variables.

This study proposes a simulation-based cost model, COSTCOR, that considers correlations between cost items. In contrast to existing cost-related models in incorporating correlations, COSTCOR is designed to meet the following three requirements that are considered practical in a cost management tool: (*1*) not requiring excessive input from management, (*2*) introducing correlations indirectly (since this correlation information is not readily available) (Touran and Wiser, 1992), and (*3*) recognizing factor-based correlations when they occur in the field. The following section presents the development of COSTCOR and is followed by an illustration of the simulation-facilitated computer implementation strategy. Finally, the correlation effect is assessed and COSTCOR is applied to an example for a building project.

## 2 THE COSTCOR MODEL

### 2.1 Hierarchical levels of project cost

The cost of a construction project is typically organized according to three levels of estimates (generally in a work breakdown structure based on the Construction Specification Institute masterformat) (Oberlender, 1993). The first is the project summary (or bid summary) level. This level summarizes various categories of costs (referred to herein as division items). Typical cost categories for a building project include direct costs, such as site work, concrete, equipment, and mechanical, and other indirect costs, such as tax, insurance, and overhead (or profits). The second is the cost item level (also called bill item or line item, and referred to herein simply as item). Each category of costs is subdivided into smaller cost items for particular construction processes. For example, the cost items for the first-level site-work cost category may be subdivided into clearing, excavation, compaction, and so on. The cost of an item equals a unit cost rate multiplied by its quantity. The third is the unit cost level. Unit costs are expressed as the cost required to complete a unit of work for a cost item, such as the cost for excavating a cubic yard of earth.

The term *cost item*, which is used in developing the COSTCOR model, can refer to either the cost category (division item) of the first level or the cost item of the second level, depending on the application purpose. The use of cost categories is more appropriate for applications subject to time constraints (for example, the tendering period of a bidding process is normally short). Meanwhile, the use of cost items is preferred for cost planning and cost control purposes, because although this approach is comparatively time consuming it yields more precise evaluation results (since the uncertainty can be estimated more precisely).

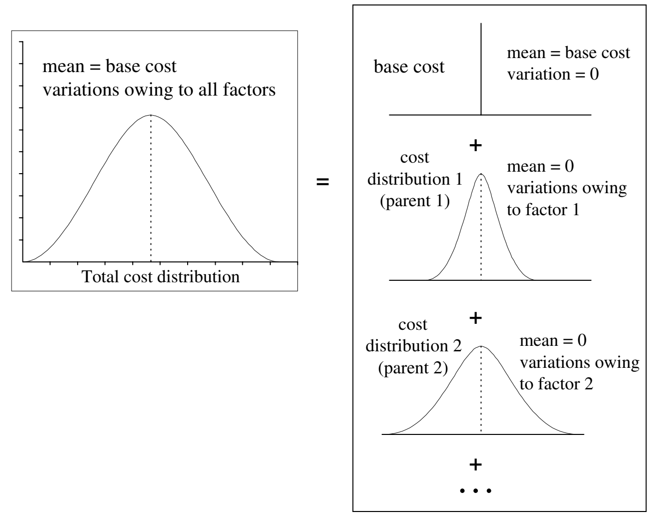


**Fig. 1.** Breakdown of uncertainty.

### 2.2 Breakdown of uncertainty

COSTCOR 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 it 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.

The COSTCOR 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 Figure 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 Figure 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).

In this example, the probability of observing any cost "C" of this cost distribution is the sum of the following three conditional probabilities: (*1*) Probability of "C" given better than expected weather times the probability of better than expected weather, (*2*) probability of "C" given

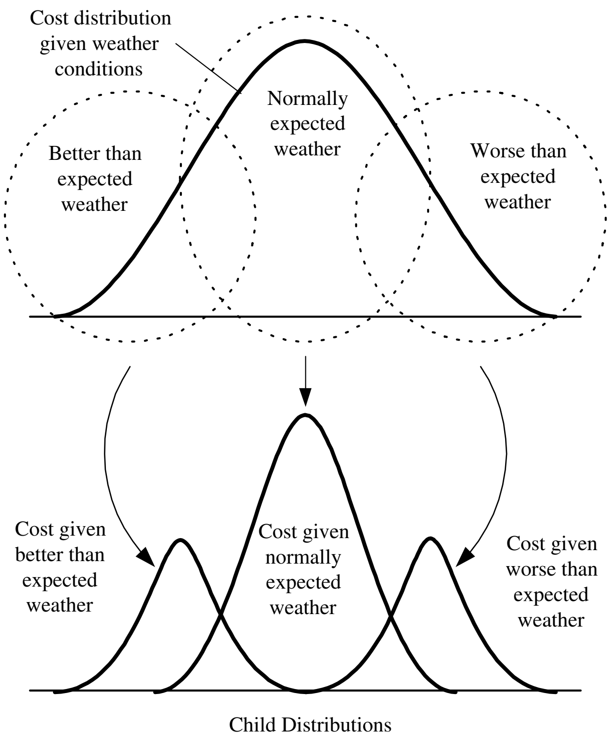


**Fig. 2.** Decomposition of cost distribution into costs, given particular factor conditions.

normally expected weather times the probability of normally expected weather, and (*3*) probability of "C" given worse than expected weather times the probability of worse than expected weather. Such conditional distributions and probabilities are such that the area under the cost distribution equals one. Notably, however, a computer simulation algorithm run with better than expected weather independently draws sample costs for weather-sensitive cost items from the better than expected weather child; likewise for normally and worse than expected weather conditions, and likewise for other factors.

Child distributions may also overlap, as presented in Figure 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. An extreme case is when child distributions overlap perfectly, in which case the cost samples will always be drawn from the same child distribution under any conditions and no correlation will exist. Meanwhile, when a cost item is highly sensitive to weather, the child distributions will be distinct.

*2.2.1 Cost modeling.* A cost model, in which the effect of uncertainty is broken down by factor, may be derived from the unit cost perspective. Unit cost is expressed as

$$\text{Unit Cost } (U) = \text{Cost/Quantity} \tag{1}$$

In a deterministic environment, the estimated unit cost for an item $i$, $U_{i(estimated)}$, can be represented thus:

$$U_{i(estimated)} = U_{i(0)} + \sum_{j=1}^{J} X_{i(j)} \tag{2}$$

where $U_{i(0)}$ is an estimated unit cost for item $i$ under expected outcomes of factors and $X_{i(j)}$ is an estimated constant representing the variation in unit cost of item $i$ with respect to factor $j$. The value of $X_{i(j)}$ can be negative, zero, or positive.

In a probabilistic environment, however, Equation (2) should be rewritten as

$$U'_{i(estimated)}$$

$$= U'_{i(0)} + \sum_{j=1}^{J} X'_{i(j)} \tag{3a}$$

$$= U'_{i(0)} + X'_{i(1)} + X'_{i(2)} + \cdots + X'_{i(j)} + \cdots + X'_{i(J)} \tag{3b}$$

where $U'_{i(estimated)}, X'_{i(1)}, X'_{i(2)}, \ldots,$ and $X'_{i(J)}$ are random variables. Each realization of $X'_{i(j)}$ represents the increase or decrease in the unit cost of item $i$ due to factor $j$.

Following (3), $C_i$, the cost of item $i$, may be expressed as

$$C_i = \text{Quantity} \times U'_{i(estimated)} \tag{4a}$$

$$= \text{Quantity} \times (U'_{i(0)} + X'_{i(1)} + X'_{i(2)} + \cdots + X'_{i(J)}) \tag{4b}$$

$$= c_{i(0)} + c_{i(1)} + c_{i(2)} + \cdots + c_{i(J)} = c_{i(0)} + \sum_{j=1}^{J} c_{i(j)} \tag{4c}$$

where $c_{i(0)}$ is the estimated (or base) cost and the random variable $c_{i(j)}, j = 1, \ldots, J$, is the cost (parent) distribution of cost item $i$ due to factor $j$. Restated, Equation (4) displays the variations in the cost of an item, as a base cost and a series of cost distributions for various factors.

*2.2.2 Factor-based correlation.* The COSTCOR 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.

COSTCOR applies the following assumptions to the cost model:

- $c_{i(0)}$ is a deterministic value. Restated, the value of $c_{i(0)}$ equals the cost that is estimated under the normally expected conditions of all factors.

- The expected values of the $c_{i(j)}$s are zero, that is, $m_{i(1)} = m_{i(2)} = \cdots = m_{i(J)} = 0$. Each sample of $c_{i(j)}$ thus represents a change from the expected cost.
- $c_{i(1)}, c_{i(2)}, \ldots,$ and $c_{i(J)}$ are independent of each other. Namely, for a given cost item, the impact of weather, the impact of labor skills, and the impact of other factors are assumed to be unrelated.

Then, regardless of the type of the marginal distribution of $c_{i(j)}$, the mean and variance of the cost of cost item $i$ are

$$M_i = m_{i(0)} + m_{i(1)} + m_{i(2)} + \cdots + m_{i(J)} \tag{5a}$$

$$= m_{i(0)} \tag{5b}$$

$$\sigma_i^2 = SD_{i(0)}^2 + SD_{i(1)}^2 + SD_{i(2)}^2 + \cdots + SD_{i(J)}^2 \tag{6a}$$

$$= SD_{i(1)}^2 + SD_{i(2)}^2 + \cdots + SD_{i(J)}^2 \tag{6b}$$

in which $M_i$ and $\sigma_i$ 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 $SD_{i(0)} = 0$. COSTCOR finds $M_i$ and $\sigma_i$ 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 $\sigma_i$ (Moder et al., 1983). However, the use of other methods is unrestricted as long as the values of $M_i$ and $\sigma_i$ can be determined.

*2.2.3 Breakdown of uncertainty by condition.* 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 as

$$m_{i(j)} = \sum_{h=1}^{H} p_{j(h)} \times o_{i[j(h)]} = 0 \tag{7}$$

$$SD_{i(j)}^2 = \sum_{h=1}^{H} p_{j(h)} \times \left(sd_{i[j(h)]}^2 + o_{i[j(h)]}^2\right) \tag{8}$$

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 (7) and (8) are valid for any type of statistical distribution. Steiner's theorem can be directly applied to justify (8) (Kreyszig, 1983). Note that the mean and standard deviation of the combination of a base cost and cost distributions have been preserved for the total cost distribution. (See Equations [5] and [6].)

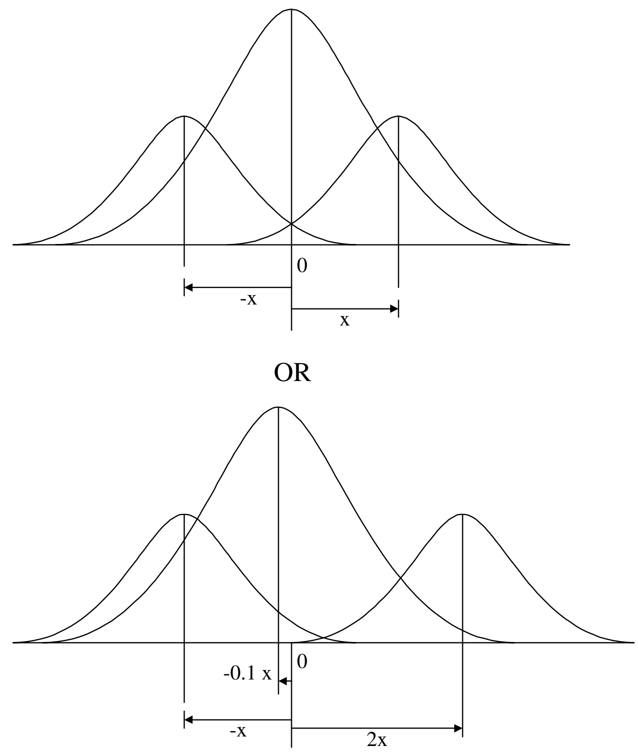


**Fig. 3.** Child distributions for different mean placements.

*2.2.4 Mean of child distributions.* 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. Figure 3 shows the means of three child distributions as represented by $(-x, 0, x)$ or $(-x, -0.1x, 2x)$. The mean of each child distribution should be confined to a range that maintains the variance of the cost distribution. Consider a family of three child distributions, as shown in Figure 4. As the mean placement $x$ approaches the limit, the standard deviations of child distributions must become smaller if the standard deviation of the parent is to be preserved. When $x$ is equal to the limit, the child distributions will have zero standard deviations.

*2.2.5 Constructing child distributions.* 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 (7) and (8), the mean and variance, respectively, of the combined

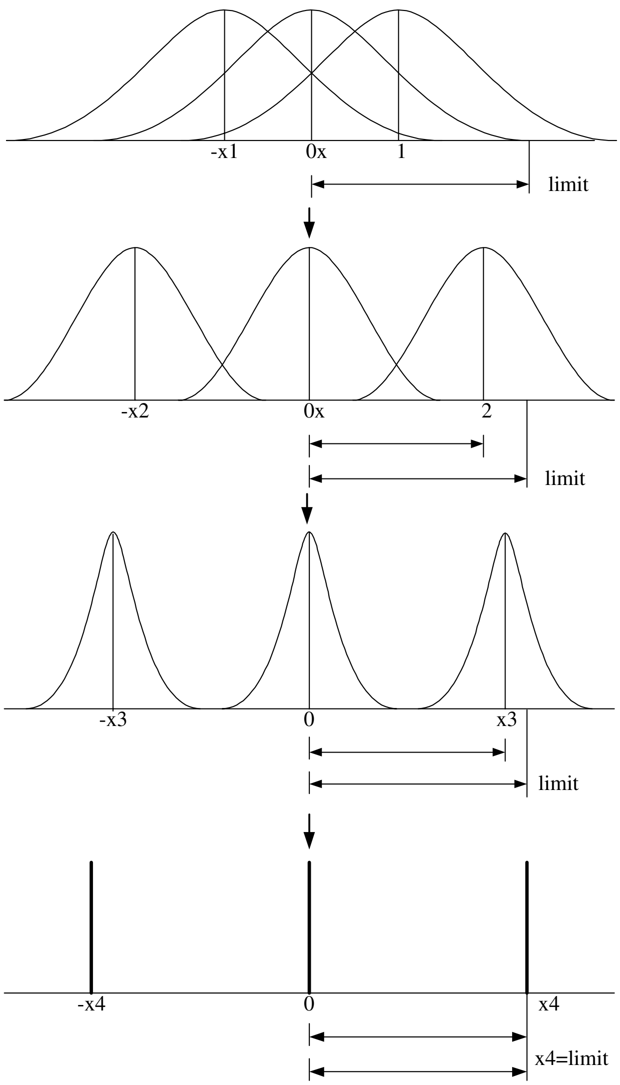


**Fig. 4.** Effect of mean placement.

child distributions are

$$(1/3)o_1 + (1/3)o_2 + (1/3)o_3 = 0 \quad (9)$$

$$(1/3)(sd_1^2 + o_1^2) + (1/3)(sd_2^2 + o_2^2) + (1/3)(sd_3^2 + o_3^2) = 4 \quad (10)$$

Assume $-o_1 = o_3 = x$ and $o_2 = 0$ so that (9) is satisfied, and let the child distributions have equal standard deviations, then (10) can be rewritten as

$$sd^2 + (2/3)x^2 = 4 \quad (11)$$

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

$$sd^2 = 4 - (2/3)x^2 \geq 0 \quad (12)$$

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 $x$, COSTCOR 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 ($p_3 = 1/3, o_3 = 1.27, sd_3 = 1.71$).

## 2.3 Qualitative estimates of uncertainty sensitivity

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 sensitive 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 (Chang, 1987). No inherent restriction is placed on the number of levels of influence used for each factor. The examples included herein use four levels of influence: high, medium, low, and no influence.

## 2.4 Scale system to break down uncertainty by factor

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 $j$ (Wang and Demsetz, 2000). That is,

$$\sigma_i^2 = \sum_{j=1}^{J} SD_{i(j)}^2 = SD_{i(1)}^2 + SD_{i(2)}^2 + \cdots + SD_{i(J)}^2 \quad (13a)$$

$$= (w_1[Q_{i(1)}] + w_2[Q_{i(2)}] + \cdots + w[Q_{i(J)}]) \times K_i \quad (13b)$$

$$= \left(\sum_{j=1}^{J} w_j[Q_{i(j)}]\right) \times K_i \quad (13c)$$

$$SD_{i(j)}^2 = w_j[Q_{i(j)}] \times K_i \quad (14)$$

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 $\sigma_i^2$ 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. The value of $w_j[Q_{i(j)}]$ is determined by the user according to the relative importance of factors. For example, if the user assumes that factor 1 causes more uncertainty than other factors, then values of $w_j[High]$, $w_j[Medium]$, and $w_j[Low]$ for factor 1 should be higher than those for other factors.

### 2.5 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 IMPLEMENTATION OF COSTCOR

In COSTCOR, 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 of 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.

Figure 5 displays the implementation strategy for the correlation modeling procedure. In each iteration, a condition is retrieved for each factor, a cost sample is drawn from a corresponding child distribution for each cost item, then the total cost distribution for each cost item is calculated by summing its base cost and any variations caused by factor conditions, and finally the calculated cost for each item and for the project are recorded. Following a predefined number of iterations, the recorded sample costs are used to compute various statistics, such as the mean, correlation coefficient, and standard deviation for the project cost.

A simulation language, STROBOSCOPE (Martinez, 1996), is used to execute the simulation-relevant procedure described in COSTCOR. 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 24 cost categories of the example project took approximately 6 min, which is acceptable for research.

## 4 ANALYSIS OF CORRELATION EFFECT

The cost of a project is the sum of the costs of all the items involved. Correlations occur between cost items that are sensitive to the same factor(s) and accumulate throughout the project. This section first examines the characteristics of input child distributions of COSTCOR. Scenarios with two identical cost items and scenarios with multiple identical cost items are analyzed; an example for a building project is presented, and finally a COSTCOR is compared with a theoretical model (considering correlations) and the significance of COSTCOR is elucidated. This section addresses how different families of child distributions, values of mean placement $x$, and the number of factors involved can influence cost correlations; how the performance of COSTCOR differs from current non-correlated models; and how COSTCOR can help to overcome traditional obstacles in correlation modeling. The analyses considered here involve 1,000 simulation iterations.

### 4.1 Input distributions

The COSTCOR user must provide data to calculate the mean and variance of the total cost distribution for each cost item. However, COSTCOR's simulation is run based on the derived child distributions. Given the scarcity of historic results concerning the cost distributions subject to a particular factor (and even a factor condition), the suitability of the input child distributions for a cost item may be evaluated by examining whether their integrated cost distribution and total cost distribution meet the following four criteria.

- The mean and variance (or standard deviation) of integrated cost distribution and total cost distribution are maintained. That is, the overall sensitivity of uncertainty (represented by the standard deviation divided by the mean) to a cost item should be equal to the sum of individual uncertainty sensitivities caused by each factor. The COSTCOR user's input of uncertainty sensitivities are accordingly preserved.
- The integrated cost distribution should best fit a statistical distribution. The Kolmogorov-Smirnov (K-S) test is performed to assess whether a significant difference between the simulated cost (or total cost) distribution and a theoretical probability distribution exists (Levin and Rubin, 1991), rather than evaluating higher moments

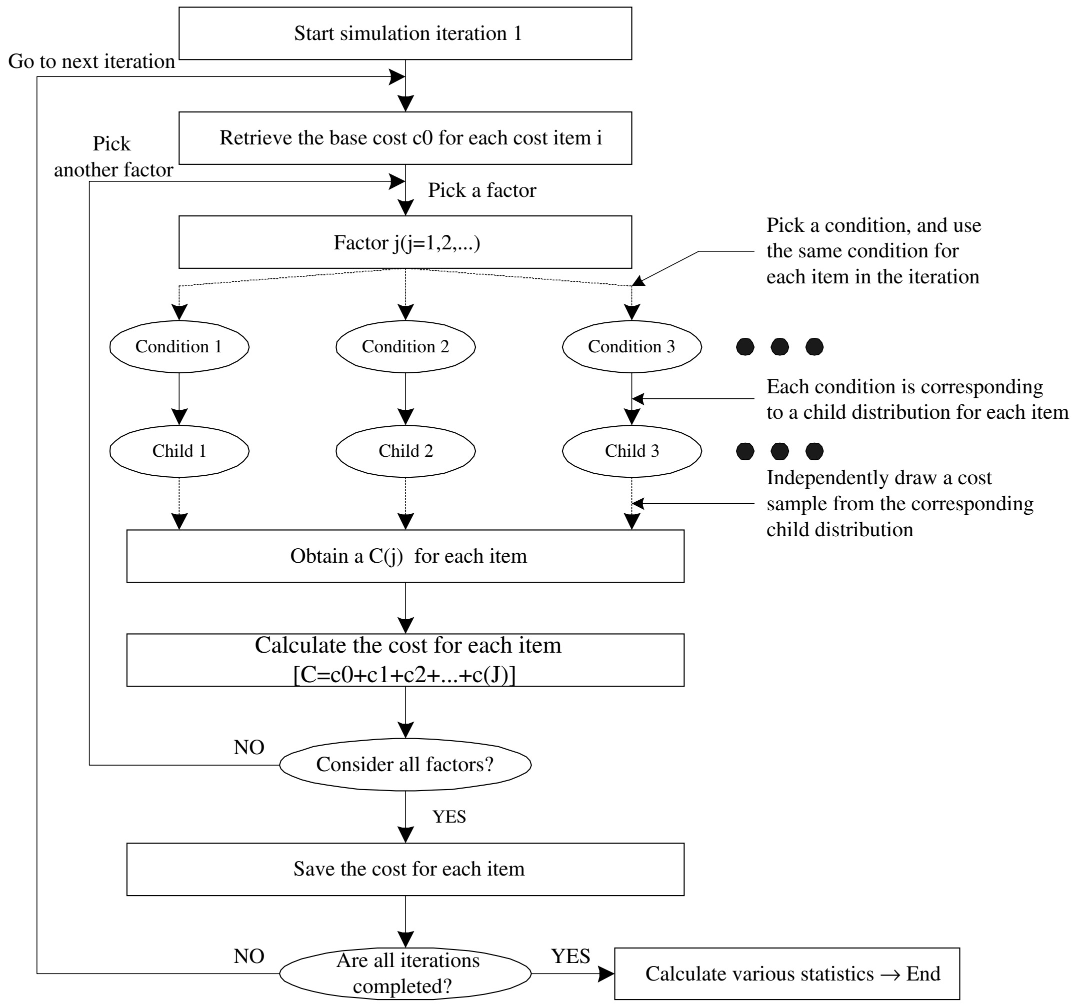


**Fig. 5.** Implementation strategy of COSTCOR.

(skewness and kurtosis) of the child distributions against cost distribution (or total cost distribution).

- Integrated total cost distribution should also best be described by a statistical distribution. Figure 6 gives an example of the simulated total cost distribution for a cost item based on a family of three Beta child distributions. In this example, the K-S test shows that a Normal distribution is best fit to the simulated total cost distribution, passing the test at a 20% significance level; the Beta distribution does not fit.
- The child distributions must be able to generate reasonable correlation coefficients. As a reference, in Touran and Wiser's (1992) database, the minimum and maximum positive correlation coefficients are 0.1 and 0.75, respectively.

Table 1 summarizes the evaluation results for 12 families of input child distributions following the aforementioned criteria. The means and variances of the integrated cost distributions and the total cost distribution are all maintained as each family of child distributions is derived from Equations (5) to (8). That is, all families meet the first criterion. If $x$ is set to be below the 0.9 limit, the families (2, 3, 5, 6, 8, and 9) constructed by Normal child distributions satisfy the second, third, and fourth criteria. This analysis does not recommend Beta child distributions for the current COSTCOR model. For example, the mean placement $x$ for family 12 established by Beta child distributions must be set at the 0.7 limit to yield a fitted Normal distribution, and only a very low correlation coefficient (that is, up to about 0.1) can be modeled. Thus, only Normal child distributions are applied in the following scenarios and example project (implying that their integrated cost distributions and total cost distributions follow a Normal distribution).

### 4.2 Scenarios with two identical cost items

Assume that the mean and standard deviation of the total cost distributions for two cost items (namely, items 1 and 2)

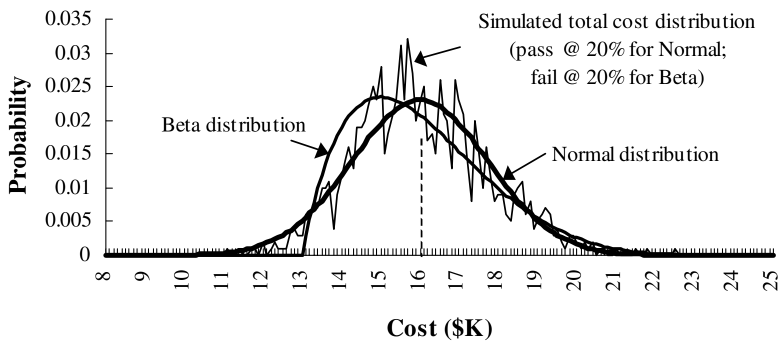


**Fig. 6.** Example of simulated total cost distribution, based on Beta child distributions.

are $M_1 = M_2 = \$10K$ (K = 1,000) and $\sigma_1 = \sigma_2 = \$1K$, respectively; and assume that both items are only sensitive to factor $j$. The corresponding standard deviations of their cost distributions are then $SD_1 = \sigma_1 = \$1K$ and $SD_2 = \sigma_2 = \$1K$, respectively.

*4.2.1 Inputs.* For a family of three child distributions (that is, child 1, 2, and 3), assume that the child distributions have the following properties:

- Probability of occurrence: $p_1 = p_2 = p_3 = 1/3$
- Standard deviation: $sd_1 = sd_2 = sd_3 = sd$
- Mean placement: $-o_1 = o_3 = x$, and $o_2 = 0$

The properties of child distributions with respect to different values of $x$ can then be calculated, as can the inputs of five and seven child distributions.

*4.2.2 Results.* Figure 7 presents the results of the correlation coefficient (CC) for different scenarios, with each scenario representing a combination of two identical mean placement values for two cost items. As expected, the value

**Table 1**
Evaluation of input child distributions

| *Family of input child distributions* | *Second criterion: Fit of cost distribution* | *Third criterion: Fit of total cost distribution* | *Fourth criterion: Ability to capture correlation value* |
|---|---|---|---|
| **Normal children with equal probabilities** | | | |
| Family 1: $x^a = 0.9$ limit | Does not pass[b] | Pass @ Normal[c] | Up to about 0.85 |
| Family 2: $x = 0.8$ limit | Pass @ Normal[c] | Pass @ Normal | Up to about 0.65 |
| Family 3: $x = 0.7$ limit | Pass @ Normal | Pass @ Normal | Up to about 0.50 |
| **Normal children with symmetric probabilities** | | | |
| Family 4: $x = 0.9$ limit | Does not pass | Pass @ Normal | Up to about 0.85 |
| Family 5: $x = 0.8$ limit | Pass @ Normal | Pass @ Normal | Up to about 0.65 |
| Family 6: $x = 0.7$ limit | Pass @ Normal | Pass @ Normal | Up to about 0.50 |
| **Normal children with non-symmetric probabilities** | | | |
| Family 7: $x = 0.9$ limit | Does not pass | Pass @ Normal | Up to about 0.85 |
| Family 8: $x = 0.8$ limit | Pass @ Normal | Pass @ Normal | Up to about 0.65 |
| Family 9: $x = 0.7$ limit | Pass @ Normal | Pass @ Normal | Up to about 0.50 |
| **Beta children with equal probabilities** | | | |
| Family 10: $x = 0.9$ limit | Does not pass | Pass @ Normal | Up to about 0.50 |
| Family 11: $x = 0.8$ limit | Does not pass | Pass @ Normal | Up to about 0.25 |
| Family 12: $x = 0.7$ limit | Pass @ Normal | Pass @ Normal | Up to about 0.10 |

Note: All families meet the first criterion (that is, maintenance of mean and variance) as they are derived from Equations (5)–(8).
[a] $x$: mean placement.
[b] Does not pass: The integrated cost distribution fails @ 20% to fit both a Normal and a Beta distribution.
[c] Pass @ Normal: The integrated distribution passes @ 20% to fit a Normal distribution, but fails @ 20% to fit a Beta distribution.

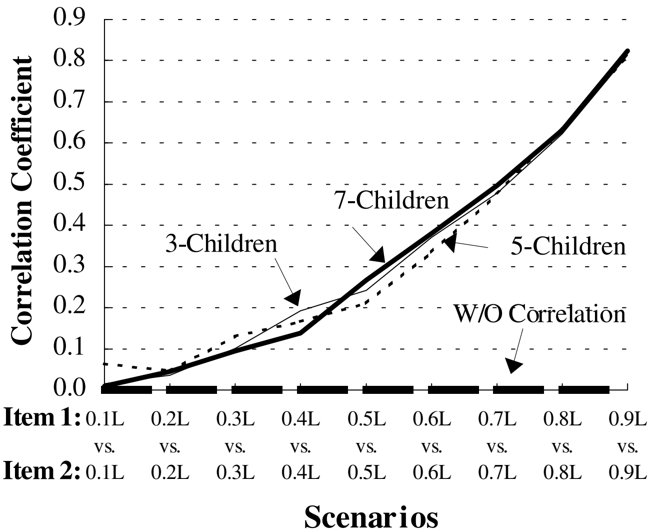


**Fig. 7.** Changes of correlation coefficients for two cost items.

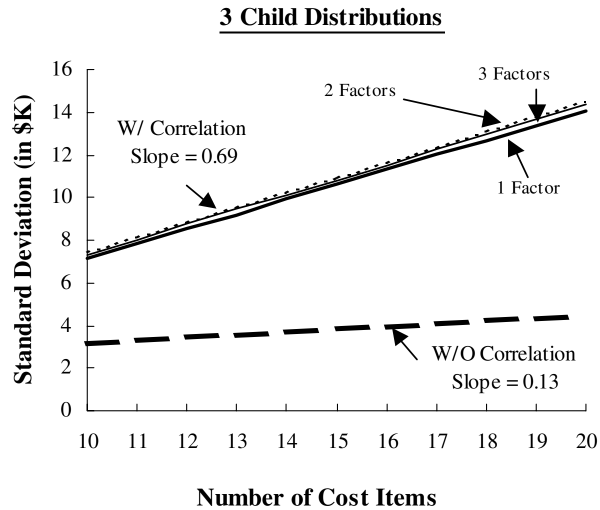


**Fig. 9.** Standard deviation for multiple items by different number of factors.

of CC increases with the value of $x$ for both cost items, that is, from 0.1L-0.1L ($x = 0.1$ limit for cost item 1 and $x = 0.1$ limit for cost item 2) to 0.9L-0.9L. Within the range investigated herein, the number of child distributions (that is, 3, 5, and 7 factor conditions) has little influence on the value of CC. As presented in Figure 8, the standard deviation of the project cost increases with the value of $x$. In quantitative terms, the project standard deviation when both cost items are highly sensitive to factor $j$ (take the 0.7L-0.7L scenario for example) is $24\%(= [\$1.75\text{K} - \$1.414\text{K}]/\$1.414\text{K})$ greater than when the correlation is neglected.

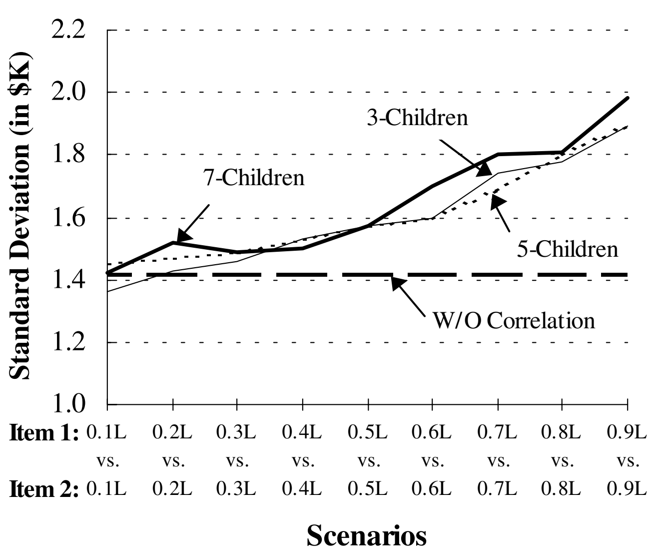


**Fig. 8.** Changes of standard deviation for two cost items.

### 4.3 Scenarios with multiple identical cost items

Scenarios involving between 10 and 20 identical cost items are analyzed to reveal the impact of correlation with an increasing number of cost items and factors. Although the mean and standard deviation of each cost item are maintained at \$10K and \$1K, respectively, the number of factors varies from one to three. A mean placement $x = 0.7$ limit is applied to represent each family of child distributions.

*4.3.1 Results.* Figure 9 illustrates the results of the standard deviation of project cost for various numbers of cost items. The project standard deviation increases as more cost items are involved, regardless of whether correlations are considered. Specifically, the slope with correlation analyses is about 0.69 (the increased standard deviation divided by the number of cost items involved), while without correlation analysis it is about 0.13. However, considering correlations significantly increases project standard deviation. For example, correlations produce an increase of 83.5% (from \$3.16K to \$5.8K) and 160% (from \$4.25K to \$11K) in standard deviation for the 10-item scenario and 20-item scenario, respectively. The above figure also reveals that the standard deviations of the projects are roughly the same, regardless of the number of factors involved, provided the total cost distribution is held.

### 4.4 Example project

An example for a building project is used to compare the results obtained using COSTCOR with two analyses that do not consider correlations: a standard PERT analysis (PERT) and a Monte Carlo simulation, carried out using normally distributed costs with the same mean and standard deviation

**Table 2**
Comparisons of W/O Correlation and COSTCOR analyses

| | | | *COSTCOR* | | |
|---|---|---|---|---|---|
| *Project cost*[a] | *PERT* | *W/O Correlation normal* | *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 |
| Minimum cost | N/A | 132 | 117.96 | 118.65 | 117.05 |
| Maximum 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).

as COSTCOR'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). COSTCOR 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 scale of scale 1 is as follows:

*Scale 1*

$$
\begin{array}{llll}
w_{F1}[\text{H}]=16 & w_{F1}[\text{A}]=12 & w_{F1}[\text{L}]=8 & w_{F1}[\text{No}]=0 \\
w_{F2}[\text{Yes}]=12 & & & w_{F2}[\text{No}]=0 \\
w_{F3}[\text{H}]=7 & w_{F3}[\text{A}]=5 & w_{F3}[\text{L}]=3 & w_{F3}[\text{No}]=0 \\
w_{F4}[\text{H}]=4 & w_{F4}[\text{A}]=3 & w_{F4}[\text{L}]=2 & w_{F4}[\text{No}]=0 \\
w_{F5}[\text{H}]=3 & w_{F5}[\text{A}]=2 & w_{F5}[\text{L}]=1 & w_{F5}[\text{No}]=0
\end{array}
$$

where "H", "A", "L", and "No" represent high, average, low, and no sensitivity, respectively. "Yes" and "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, scales 2 and 3 (which exaggerate the differences between high, medium, and low sensitivities) are as follows:

*Scale 2*

$$
\begin{array}{llll}
w_{F1}[\text{H}]=8 & w_{F1}[\text{A}]=5 & w_{F1}[\text{L}]=1 & w_{F1}[\text{No}]=0 \\
w_{F2}[\text{Yes}]=8 & & & w_{F2}[\text{No}]=0 \\
w_{F3}[\text{H}]=8 & w_{F3}[\text{A}]=5 & w_{F3}[\text{L}]=1 & w_{F3}[\text{No}]=0 \\
w_{F4}[\text{H}]=8 & w_{F4}[\text{A}]=5 & w_{F4}[\text{L}]=1 & w_{F4}[\text{No}]=0 \\
w_{F5}[\text{H}]=8 & w_{F5}[\text{A}]=5 & w_{F5}[\text{L}]=1 & w_{F5}[\text{No}]=0
\end{array}
$$

*Scale 3*

$$
\begin{array}{llll}
w_{F1}[\text{H}]=100 & w_{F1}[\text{A}]=10 & w_{F1}[\text{L}]=1 & w_{F1}[\text{No}]=0 \\
w_{F2}[\text{Yes}]=100 & & & w_{F2}[\text{No}]=0 \\
w_{F3}[\text{H}]=100 & w_{F3}[\text{A}]=10 & w_{F3}[\text{L}]=1 & w_{F3}[\text{No}]=0 \\
w_{F4}[\text{H}]=100 & w_{F4}[\text{A}]=10 & w_{F4}[\text{L}]=1 & w_{F4}[\text{No}]=0 \\
w_{F5}[\text{H}]=100 & w_{F5}[\text{A}]=10 & w_{F5}[\text{L}]=1 & w_{F5}[\text{No}]=0
\end{array}
$$

*4.4.1 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 2 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 (Moder et al., 1983).
- 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 (for example, \$117.96K for scale 1 versus \$132K for W/O Correlation Normal) or significantly higher than expected (for example, \$184.25K for scale 1 versus \$167.49K for W/O Correlation Normal). Restated, increased variability in project cost increases the uncertainty of completing the project within a specified budget. 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. If only a few activities are highly sensitive to the same factors (such as the case for this example project), the correlation effect on the total project cost may not be greatly highlighted by employing an exaggerated scale system (such as scale 3). The correlation effect tends to be dominated by the lower-sensitivity factor cost distributions, rather

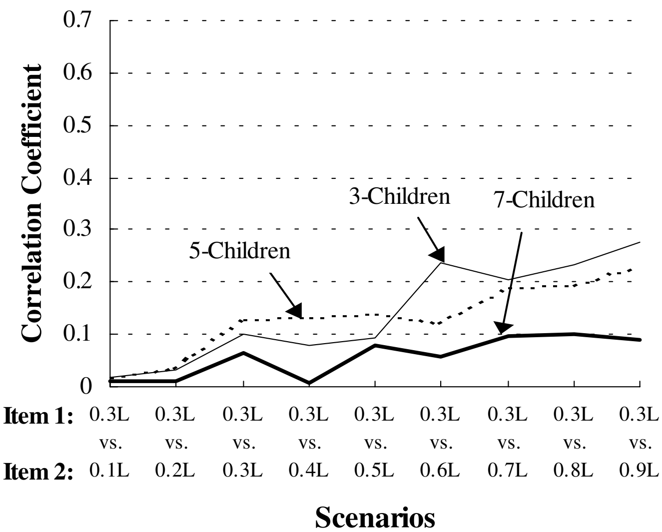


**Fig. 10.** Changes of correlation coefficients for two cost items, $x = 0.3$L for item 1.

than the higher-sensitivity ones. Figure 10 illustrates the correlation coefficients when cost item 1 is fixed at the $x = 0.3$ limit (low sensitivity) with varying $x$ (varying sensitivities) for cost item 2. In the figure, although a high value of $x$ leads to a relatively high correlation coefficient, the highest correlation coefficient is only around 0.28 (for 0.3L-0.9L scenario). Nevertheless, the exaggerated scale 3 system influences the relative priorities of the factors as more variance is distributed to the parent distributions due to higher sensitivities.

*4.4.2 Results: Uncertainty sensitivity.* Table 3 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.

### 4.5 Comparison of COSTCOR with a theoretical model

Comparing the accuracy of COSTCOR and existing simulation models may not be feasible due to the lack of sufficient field data (such as cost data and factor sensitivity information for a set of cost variables). However, COSTCOR's performance may be further distinguished by comparing the results of COSTCOR to those of a well-known theoretical model.

In the theoretical model, the standard deviation, STD DEV, of the project cost (sum of a number of cost items) can be defined as (Benjamin and Cornell, 1970)

$$\text{STD DEV}^2 = \sum_{i=1}^{I} \sigma_i^2 + \sum_{i=1}^{I} \sum_{h=i+1}^{I} 2\left(\rho_{i,h} \times \sigma_i^2 \times \sigma_h^2\right) \quad (15)$$

where $\sigma_i$ and $\sigma_h$ are the standard deviations of the total cost distribution for cost items $i$ and $h$, respectively. $\rho_{i,h}$ is the correlation coefficient for a pair of cost items, $i$ and $h$.

The example used in section 4.3 for scenarios involving 10 to 20 identical cost items is also considered here. The mean and standard deviation for each item are held at \$10K and \$1K, respectively. COSTCOR assumes only one factor is involved, and the $x = 0.7$ limit is applied to a family of three Normal child distributions for each item. Following Touran and Wiser (1992), the values of $\rho_{i,h}$ relating each pair of cost items are taken to be 0.1, 0.5, and 0.75 for the theoretical model. (0.5 is a randomly selected value.)

*4.5.1 Results.* Figure 11 compares the standard deviation of project cost for various numbers of cost items. For this example, COSTCOR's results (using the $x = 0.7$ limit to represent high sensitivity) are closer to those obtained by

**Table 3**
Uncertainty sensitivity of project cost by scale systems

| | *Standard deviation* | | |
|---|---|---|---|
| *Factors* | *Scale 1* | *Scale 2* | *Scale 3* |
| 1. Owner approval | 7.9857[a] [2][b] | 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] All data are expressed in thousands (K).
[b] [ ] indicates the rank of the sensitivity with respect to a given factor.

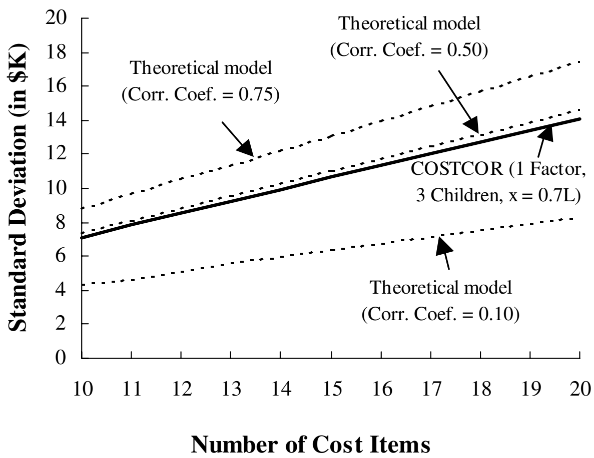


**Fig. 11.** Standard deviations from COSTCOR and the theoretical model.

the theoretical model using a correlation coefficient of 0.5. The ability to obtain correlation values, as displayed in Table 1, may suggest that the mean placement, $x$, used in COSTCOR should be set as high as the 0.85 limit to represent a high sensitivity, such that the value of 0.75 (maximum correlation coefficient in Touran and Wiser's [1992] database) can be captured.

### 4.6 Benefits of COSTCOR

The theoretical and current simulation models are limited by their need for correlation coefficient data, which are not easily available. Although the theoretical model may give much more precise solutions than do the simulation models, the latter provide the distribution of cost that is essential to probabilistic estimation and risk analysis (Touran and Wiser, 1992). COSTCOR, a simulation-based model, requires only reasonable and qualitative inputs (such as high, medium, and low uncertainty sensitivities) to evaluate indirectly the impact of correlation. A valuable improvement of COSTCOR over existing models is its capacity quantitatively to derive the factor sensitivities with an emphasis on the importance of controlling and understanding the factors that influence cost.

A simulation-based model, NETCOR (Wang and Demsetz, 2000), was recently developed by the author to evaluate schedule networks with correlated activity durations. The COSTCOR model may be considered to be an extension of NETCOR. The main difference is that COSTCOR addresses project cost, while NETCOR is concerned with project time. Project cost is the sum of individual cost items, whereas project time depends on the degree to which activities are concurrent.

## 5 CONCLUSIONS

This work has developed a simulation-facilitated factor-based model, COSTCOR, that allows correlation between cost items to be considered in cost analysis. The COSTCOR model is based upon the two-step breakdown of uncertainty. The first breakdown separates uncertainty on the basis of factor for each total cost distribution; that is, total cost distribution = base cost + cost distributions. The second breakdown separates uncertainty on the basis of condition for each cost distribution; that is, cost distribution = family of child distributions (that is, costs given factor conditions). 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.

The correlation between cost distributions is caused by their sharing the same factor(s). However, cost distributions sensitive to the same factor may not be correlated because the outcomes of any shared factor may not remain the same over a long period. Restated, correlation can be time dependent. Two outdoor cost items (for example, concrete and masonry) that are highly sensitive to weather may not be correlated if the construction tasks associated with these cost items are scheduled over different days or in different seasons. Nevertheless, if the second-level cost items are applied to COSTCOR and the factors involved are clearly defined, it is reasonable to attribute any correlation to the sharing of factors.

The COSTCOR model can be applied to either the cost category (division item) level or individual cost item level, depending on the application. Several examples have confirmed the effect of correlations on project cost. The example project also demonstrates how COSTCOR can provide management with information on the sensitivity of various factors to project cost. 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; collecting field data to justify child and cost distributions, the values of mean placement $x$, and the values of the correlation coefficient; and applying COSTCOR more widely.

## ACKNOWLEDGMENTS

The author thanks the reviewers of this paper for their careful evaluation and thoughtful comments. The author is indebted to Professor Laura Demsetz from the College of San Mateo (California, USA) for her valuable assistance. Dr. Julio Martinez is also commended for making STROBOSCOPE available and generously providing assistance in its use.

## REFERENCES


Adeli, H. & Karim, A. (2001), *Construction Scheduling, Cost Optimization, and Management—A New Model Based on Neurocomputing and Object Technologies*, Spon Press, London.

Adeli, H. & Wu, M. (1998), Regularization neural network for construction cost estimation, *Journal of Construction Engineering and Management*, ASCE, **124** (1), 18–24.

Ahuja, H. N. & Nandakumar, V. (1985), Simulation model to forecast project completion time, *Journal of Construction Engineering and Management*, ASCE, **111** (4), 325–42.

Benjamin, J. & Cornell, A. (1970), *Probability, Statistics and Decision for Civil Engineers*, McGraw-Hill, New York.

Chang, T. C. (1987), *Network Resource Allocation Using an Expert System with Fuzzy Logic Reasoning*, Ph.D. Dissertation, University of California, Berkeley, CA.

Chau, K. W. (1995), Monte Carlo simulation of construction costs using subjective data, *Construction Management and Economics*, **13**, 369–83.

Diekmann, J. E. (1983), Probabilistic estimating: mathematics and applications, *Journal of Construction Engineering and Management*, ASCE, **109** (3), 297–308.

Diekmann, J. E. & Featherman, W. D. (1998), Assessing cost uncertainty: lessons from environmental restoration projects, *Journal of Construction Engineering and Management*, ASCE, **124** (6), 445–51.

Kreyszig, E. (1983), *Advanced Engineering Mathematics*, 5th Edition, John Wiley & Sons, New York.

Levin, R. I. & Rubin, D. S. (1991), *Statistics for Management*, 5th Edition, Prentice Hall, Englewood Cliffs, NJ.

Levitt, R. E. & Kunz, J. C. (1985), Using knowledge of construction and project management for automated schedule updating, *Project Management Journal*, **XVI** (5), 57–76.

Lowe, D. & Skitmore, M. (1994), Experiential learning in cost estimating, *Construction Management and Economics*, **12**, 423–31.

Martinez, J. C. (1996), *STROBOSCOPE: State and Resource Based Simulation of Construction Processes*, Ph.D. Dissertation, University of Michigan, Ann Arbor, MI.

Moder, J. J., Philips, C. R. & Davis, E. W. (1983), *Project Management with CPM, PERT and Precedence Diagramming*, 3rd Edition, Van Nostrand Reinhold, New York.

Oberlender, G. D. (1993), *Project Management for Engineering and Management*, McGraw-Hill, New York.

Oberlender, G. D. & Trost, S. M. (2001), Predicting accuracy of early cost estimates based on estimate quality, *Journal of Construction Engineering and Management*, ASCE, **127** (3), 173–82.

Padilla, E. M. & Carr, R. I. (1991), Resource strategies for dynamic project management, *Journal of Construction Engineering and Management*, ASCE, **117** (2), 279–93.

Ranasinghe, M. (2000), Impact of correlation and induced correlation on the estimation of project cost of buildings, *Construction Management and Economics*, **18**, 395–406.

Touran, A. & Wiser, E. D. (1992), Monte Carlo technique with correlated random variables, *Journal of Construction Engineering and Management*, ASCE, **118** (2), 258–72.

Wang, W.-C. (2002), SIM-UTILITY: model for project ceiling price determination, *Journal of Construction Engineering and Management*, ASCE, in press.

Wang, W.-C. & Demsetz, L. A. (2000), Model for evaluating networks under correlated uncertainty—NETCOR, *Journal of Construction Engineering and Management*, ASCE, **126** (6), 458–66.

Woolery, J. C. & Crandall, K. C. (1983), Stochastic network model for planning scheduling, *Journal of Construction Engineering and Management*, ASCE, **109** (3), 342–54.