

Simulation-based estimation for correlated cost elements

I.-T. Yang

Department of Civil Engineering, Tamkang University, No. 151 Ying-chuan Road, Tamsui, Taipei 251, Taiwan

Received 12 August 2004; received in revised form 10 November 2004; accepted 1 December 2004

Abstract

In this paper, we present a general method to incorporate correlations between cost elements in the process of cost estimation. The proposed method first checks the feasibility of the correlation (Pearson or Spearman) matrix, adjusts it if necessary, then uses the correlations to generate correlated multivariate random vectors, which are employed to model possible outcomes of the cost elements. The method is applied to a full data set of 216 British office buildings to illustrate its practical use. The application result indicates that the impact of correlations is significant and may cause serious problems if neglected. The result is also used to validate that the proposed method can capture the correlations with relatively small deviations.

© 2005 Elsevier Ltd and IPMA. All rights reserved.

Keywords: Cost; Estimation; Risk; Simulation; Project management

1. Introduction

Cost estimation is a process of collecting, analyzing, and summarizing data in order to prepare an educated projection of the anticipated cost of a project [26]. This process begins in the early stages of the project and repeats frequently during the entire life cycle. The reliability of cost estimation is important to ensure the success of the project since it serves as the foundation for making critical decisions on financial investments and borrowings.

The prices of all the resources (material, equipment, and labor) are exposed to certain levels of uncertainty, particularly when the project life cycle is lengthy [25]. To manage the inherent uncertainty, Monte-Carlo simulation methods have been widely applied for various types of projects, such as [3,11,15,31]. In Monte-Carlo simulation, a mathematical model is constructed based on pre-specified probability distributions, which describes the possible outcomes of major cost elements (e.g., substructure, exterior walls, and electrical finishing) involved in a project, and run to see what the over-

all project cost will be for each simulation replication. After a certain number of replications, the collected samples are used to derive the output distribution of the overall project cost.

An enhancement of ordinary simulation methods has been directed to consider statistical correlations (dependencies) between cost elements. The correlation represents the co-movement of two cost elements; when one is more expensive, the other tends to cost more as well (or cost less for a negative correlation). Arguments and evidences for the existence of correlations and their profound impact on simulation results have been presented in the literature [10,23,33]. To treat the correlations, various approaches have been proposed, such as [29,30,34].

The goal of this paper is to present a simulation-based method to incorporate correlations between cost elements with more modeling capabilities. Specifically, the present method is unique in handling the following requirements:

1. To allow the distributions (i.e., marginal distributions) of individual cost elements to be of different types. Namely, some of them may only be expressed

E-mail address: ityang@mail.tku.edu.tw

with discrete and finite options whereas others can be expressed as continuous functions. In addition, those continuous distributions may come from different families (e.g., some are lognormal while some are beta).

2. To provide an automatic procedure to check the feasibility (a mathematical definition will be given later) of a correlation matrix and adjust it if infeasible.

This paper is organized as follows. In the following two sections, we introduce two different approaches (historical and subjective) to obtain necessary input for the analysis: the marginal distributions of individual cost elements and their correlations. The characteristics of the two approaches lead to the need for a more general method. In the section followed, we present the proposed method in two stages and provide computational guidelines. The proposed method is then applied to a modified British data set composed of 216 office buildings to illustrate its practical use. The results are used to validate the proposed method. Closing remarks are presented in the last section.

2. Required input: distributions and correlations

There are two sets of input data required to perform a simulation-based cost analysis considering correlations. The first set describes marginal distributions of individual cost elements and the second is a correlation matrix consisting of the correlation coefficients between pairs of cost elements. Both sets of data can be estimated in two ways: (1) by summary statistics on historical data, or (2) by subjective judgments. In what follows we give brief introduction to these two approaches and elucidate how their characteristics give rise to a more general simulation method.

2.1. Choice of distribution

When historical data is used to describe the marginal distributions, it involves an attempt to fit theoretical distributions to the data and verify goodness-of-fit statistically. While a complete review of goodness-of-fit tests and their theoretical backgrounds is given in [18], the fitting process can be done very efficiently by commercial software packages (such as Palisade BestFit, Arena, ProModel, and Crystal Ball). A typical result is a list of several “good” distributions and their associated parameters, based on which the estimator can select the most proper one. Previous studies suggested that the lognormal distribution fits historical cost data better than other well-known distributions, such as normal or beta [30,33].

Despite its theoretical maturity, using historical data to forecast possible outcomes has some pitfalls. First, actual values may lie outside the range of historical records due to new technology, equipment, and material. Second, historical data may not adequately represent the true underlying population because of sampling error [12]. Last, the prices of resources may not always be repeatable, thus, the historical approach may be fallacious [23].

In the absence of reliable historical data, the second best alternative is for a cost estimator to rely on his/her experiences to subjectively specify the marginal distributions. In the context of probabilistic estimation, it is usual to assume the underlying distribution is a beta distribution whose parameters are specified by three point estimates: the minimum, maximum, and most likely values [4]. Some recent controversy has been whether the beta distribution should be replaced with the triangular distribution since the former does not have clear-cutting bounds [16] and requires four parameters, which do not have a one-to-one correspondence with the three estimates [32].

Both approaches mentioned above (historical and subjective) may involve the following practical concerns. First, they may actually be mixed in practice. A cost estimator may not have historical data of *all* the cost elements if some works are usually outsourced or subcontracted. Thus, he/she can obtain summary statistics only on those data on hand and has to rely upon subjective estimation for the remaining cost elements. Another possibility occurs when the estimator has reasons to believe some of the cost elements in a new project have their own bounds (minimum and maximum) and thereby cannot be represented by historical data. For these particular elements, the subjective approach is more appropriate. Second, it has been frequently encountered that the price options of some cost elements are collections of discrete outcomes in lieu of continuous functions [2].

The practical concerns above give rise to the need for a more general simulation model, which should be able to treat all different types of distributions in one framework (some are discrete while some are continuous; some are lognormal while some are beta). This is the first goal of this paper.

2.2. Specification of correlations: product-moment and rank

Two different measures have been used to reflect the degree of relation between cost elements. The first one is an ordinary product-moment (Pearson) correlation coefficient and the second is a rank (Spearman) correlation coefficient. The product-moment correlation can be calculated as follows:

$$\rho_{xy} = \frac{n \sum_{i=1}^n X_i Y_i - \sum_{i=1}^n X_i \sum_{i=1}^n Y_i}{\sqrt{[n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2][n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2]}} \quad (1)$$

where X and Y denote two populations of variables, each of which contains n observations.

The fundamental assumptions behind the product-moment correlation are that the relationship between two variables is linear and both populations are normally distributed. When these two assumptions cannot be met, a more attractive alternative is to use the rank correlation.

The rank correlation is essentially the product-moment correlation coefficient between the “ranks” in each set of the two variables. To compute the rank correlation coefficient, one should sort each set of variables based on their values, replace the value by the rank, and compute the product-moment correlation coefficient to the rank. A useful equation is

$$r_{xy} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (2)$$

where d is the differences between ranks of corresponding X and Y .

Both coefficients can range between +1 and –1. A correlation coefficient of +1 signifies a perfect positive relationship, while –1 shows a perfect negative relationship. The smallest correlation is 0. Because correlation coefficients are defined between two cost elements, they are traditionally stored in a two-dimensional correlation matrix, which is always symmetric with unit diagonal.

Similar to the marginal distributions, the correlation matrix may be a result of statistic analysis on historical data or of subjective judgment. The historical approach requires a test on the significance of the correlation coefficients. A null hypothesis shall be tested against an alternative hypothesis to determine if a cost element is, not by coincidence, independent from others.

The historical approach has been used to find the product-moment correlations in an Australian data set drawn from 8 office-building projects [20], and in a British data set drawn from 216 office-building projects [33]. The statistical approach has also been applied to calculate the rank correlations in the samples of 131 American projects [29]. The empirical results of these studies suggest the existence of correlations.

On the other hand, subjective judgment also finds its application in specifying the correlations between cost elements qualitatively. To this respect, researchers can subjectively choose two groups of correlations to assess strong, moderate, and weak relations: {0.8,0.45,0.15} [28] and {0.85,0.55,0.25} [7].

When a correlation matrix is used in simulation, an important requirement is to ensure its feasibility, which restricts the matrix to be positive semidefinite regardless of its type (product-moment or rank) or the way it is estimated (historical or subjective) [19]. Being positive semidefinite means the eigenvalues of the correlation matrix must be non-negative. In a mathematical form, the requirement can be written as

$$\mathbf{E}^T \mathbf{M} \mathbf{E} \geq 0, \quad (3)$$

where \mathbf{M} is the correlation matrix, \mathbf{E} , the eigenvector, and \mathbf{E}^T , transpose of \mathbf{E} .

In the literature, it has frequently occurred that the correlation matrix is not positive definite as indicated by Ranasinghe [24]. This is particularly an issue when the number of dimensions increases because the possibility of having an infeasible correlation matrix will grow rapidly as the dimension increases [17].

Two procedures were proposed to adjust an infeasible correlation matrix. Touran’s approach was to reduce all the correlations slightly (say 0.01) and repeat until the correlation matrix becomes feasible [28]. This approach overlooks the possibility of increasing some correlations while reducing others. Ranasinghe [24] developed a computer program to iteratively calculate and list the bounds of each correlation to make the matrix positive semidefinite. The program then asks the estimator to change the original values and wait until the program re-checks the feasibility and new bounds. This process continues until reaching the feasibility. This approach, however, may be time consuming due to its iterative nature.

Our discussion leads to the second goal of this study: to have an automatic procedure to check the feasibility of the correlation matrix and adjust it if necessary. To attain efficiency, the adjustment should be performed only once and in a prior fashion.

3. Proposed method

The proposed method takes two sets of input: marginal distributions of the cost elements (measured in unit cost, for example £/m²) and a correlation matrix between these elements. The method is composed of two stages, which will be explained below.

3.1. Setup stage

The proposed method starts with a check on the feasibility of the original correlation matrix. If it is already positive semidefinite, one can immediately begin the simulation steps described in the following section; otherwise, we adopt the eigenvalue correction method from Ghosh and Henderson [13] to approximate the infeasible correlation matrix into a feasible one.

The setup stage consists of the following steps:

1. Decompose the correlation matrix \mathbf{M} into a diagonal vector \mathbf{D} of the eigenvalues and a full matrix \mathbf{V} whose columns are the corresponding eigenvectors so that

$$\mathbf{M}\mathbf{V} = \mathbf{V}\mathbf{D}. \quad (4)$$

2. Locate the negative eigenvalues and change them to a tiny positive number ε to yield a new diagonal vector $\bar{\mathbf{D}}$.

3. Adjust the correlation matrix \mathbf{M} by

$$\mathbf{M} = \mathbf{V}\bar{\mathbf{D}}\mathbf{V}^T. \quad (5)$$

4. Take the diagonal elements of \mathbf{M} and store their inverses as the diagonals in a full matrix \mathbf{E}

$$\mathbf{E}_{ij} = \begin{cases} 1/\sqrt{\mathbf{M}_{ij}} & \forall i = j, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

5. Normalize the diagonal elements to ensure unit diagonals (fundamental requirement for a correlation matrix)

$$\bar{\mathbf{M}} = \mathbf{E}\mathbf{M}\mathbf{E}, \quad (7)$$

where $\bar{\mathbf{M}}$ is the new (approximated) correlation matrix.

A preliminary test is performed to ensure the effectiveness of the setup stage. We use a documented 19×19 correlation matrix (Table 2 in [7]) as an example. This correlation matrix has been proved to be non-positive-semidefinite [24]. Our tasks here are to apply the proposed steps and to check if the differences between the approximated values and the original specifications are small enough. The differences are quantified in two metrics: L_{ave} (average) and L_{max} (maximum)

$$L_{ave} = \frac{\sum_{i>j} |\bar{\mathbf{M}} - \mathbf{M}|}{\frac{n(n-1)}{2}}, \quad (8)$$

where \mathbf{M} is the specified correlation matrix and $\bar{\mathbf{M}}$ is the approximated one; n is the number of cost elements

$$L_{max} = \max_{i>j} |\bar{\mathbf{M}} - \mathbf{M}|. \quad (9)$$

After performing the correction steps, L_{ave} is 4.8954×10^{-18} and L_{max} is 1.7208×10^{-15} . This shows empirically that the setup stage can, de facto, adjust the infeasible correlation matrix to a feasible one with ignorable changes.

The setup stage is to treat possible infeasibility, which may result from either erroneous input or inconsistent estimation. In other words, the setup stage would not be of any good if the correlation coefficients are incorrect or inconsistent. Thus, careful review is critical to ensure the correlation coefficients can reflect the true behavior of the correlation relationships.

3.2. Simulation stage

The fundamental concept of the simulation stage is to generate a vector of correlated normal variates, transform them into uniform variates by the aid of the cumulative normal probability function, and then map the variates into their individual marginal distributions by the inverse transform method [18]. The generated random variates are used to model the cost elements with the desired correlation structure. The procedure described here incorporates ideas from a new correlated multivariate generation technique (Normal To Anything, NORTA) [6]. In what follows, we enumerate all the steps and provide computational guidelines:

1. Apply the Cholesky decomposition to the correlation matrix, so that $\mathbf{M} = \mathbf{C}\mathbf{C}^T$, where \mathbf{C} represents the Cholesky triangular.
2. Generate an IID (independent and identically distributed) unit scaled uniform random vector, $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$, where n is the number of cost elements.
3. Translate \mathbf{Y} into a standard-normal random vector $\mathbf{P} = (P_1, P_2, \dots, P_n)$.
4. Transform \mathbf{P} into a correlated standard-normal random vector $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$.
5. Compute for $U_i = \Phi(Z_i)$ for $i = 1, 2, \dots, n$, where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function (CDF).
6. Compute for $X_i = F_i^{-1}(U_i)$ for $i = 1, 2, \dots, n$, where $F_i^{-1}(U_i)$ represents the inverse of the i th marginal CDF.
7. Return X_i as the estimate for cost element i .
8. Compute the total project unit cost by summing up all the cost elements.
9. Repeat Steps 2–8 for each simulation replication, $j = 1, 2, \dots, m$.
10. Return summary statistics on all simulation replications.

For Step 1, there exist several efficient algorithms to perform the Cholesky decomposition (see [14,22]). The generation of a uniform random vector in Step 2 is a standard feature supported by almost all the popular computer languages (such as C++, Java, Visual Basic, FORTRAN). The transformation in Step 3 can be approximated by the following equation [27]:

$$P_i = (Y_i^{0.135} - (1 - Y_i)^{0.135})/0.1975. \quad (10)$$

The transformation in Step 4 is

$$Z_i = \sum_{j=1}^i c_{ij} P_j \quad \text{for } c_{ij} \in \mathbf{C}. \quad (11)$$

Step 5 involves the following integral:

$$U_i = \int_{-\infty}^{Z_i} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx, \quad (12)$$

which can be approximated by [1]

$$U_i = \begin{cases} (1/\sqrt{2\pi} \times \exp(-Z_i^2/2)) \times (b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5) & \text{if } Z_i < 0, \\ 1 - (1/\sqrt{2\pi} \times \exp(-Z_i^2/2)) \times (b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5) & \text{if } Z_i \geq 0, \end{cases} \quad (13)$$

where:

$$\begin{aligned} t &= 1/(1 + 0.2316419 \times |Z_i|); \\ b_1 &= 0.319381530; \\ b_2 &= -0.356563782; \\ b_3 &= 1.781477937; \\ b_4 &= -1.821255978; \\ b_5 &= 1.330274429. \end{aligned}$$

The marginal CDF in Step 6 can be of any type of distribution as long as their inverses can be calculated either directly or via approximation. This is why the proposed method is able to treat different kinds of distributions simultaneously. It is much easier when the marginal distribution has a closed-form inverse (such as uniform or triangular). Otherwise, one has to rely on numerical approximation algorithms to find the inverses of the commonly-used distributions, such as beta [9], gamma [5], and normal distributions [21].

By using X_i as the estimate of the i th cost element, Step 8 is simply

$$E_j = \sum_{i=1}^n X_i, \quad (14)$$

where E_j is the j th observation of the unit cost of the project. To make the addition meaningful, the estimates of different cost elements should be converted into the same unit of measure, such as £/m².

4. Practical application

The proposed method is applied to the British data set described in [33] to demonstrate its practical use. The data set is drawn from 216 office buildings built between 1980 and 1994 and consists of 8 major cost elements. The data set has been standardized based on the times and locations the buildings were built.

All the cost elements and their marginal distributions are shown in Table 1. The value of each cost element is expressed as £/m². Here a cost element represents a relatively large work package, which may consist of several tasks. For example, “superstructure” involves formwork, steelwork, and concrete pouring. This level of granularity is suitable for higher level estimation. More-

over, the measure of £/m² can be changed to reflect the usual unit for progress measurement, if the proposed method is applied to other construction projects. For instance, a reasonable measure of cost elements for a high-

way project may be £/m while that for a residential community project may be £/house.

In the example, we consider three families of distributions, i.e., lognormal, beta, and discrete. The lognormal distributions are used because they fit the data better as argued by Wall [33]. The use of the other two is based on a pragmatic situation when a cost estimator prefers not using historical data but rather using a discrete distribution to describe possible outcomes of “fitting and furnishings”, and beta distributions (three points) to estimate the distributions of “services” and “external works”. These arrangements have been justified in previous sections.

Table 2 shows the rank correlation coefficients between the cost elements of the full data set. Before applying the proposed method, the rank correlation coefficients are reviewed and adjusted to verify (1) if they can reflect the actual behavior of the correlations and (2) if they, derived from past data, are suitable for the current project. This process is based on practical judgments and can complement pure mathematic analysis. In this example, the rank correlation coefficients between “external works” and other cost elements are adjusted to be zero.

A simulation experiment is designed to implement the proposed method and to evaluate the impact of correlations between cost elements. In the experiment, every simulation replication leads to a sample of the project

Table 1
Descriptive estimates for cost elements (distributions and parameters)

Cost elements	Descriptive estimate (£/m ²)
Substructure	Lognormal (47.2,30.9) ^a
Superstructure	Lognormal (263.6,82.4) ^a
Internal finishes	Lognormal (63.2,24.4) ^a
Fittings and furnishings	Discrete (7,0.2; 8,0.5; 9,0.2; 10,0.1) ^b
Services	Beta (150,180,220) ^c
External works	Beta (70,85,120) ^c
Preliminaries	Lognormal (76.4,47.3) ^a
Contingencies	Lognormal (21.2,13.2) ^a

^a Lognormal (mean, standard deviation); the lognormal distributions are estimated by the historical approach based on 216 buildings.

^b Discrete (outcome, probability); the discrete distribution is subjectively specified.

^c Beta (minimum, mode, maximum); the three parameters are subjectively specified.

Table 2
Rank correlation coefficients between cost elements^a

	Substructure	Superstructure	Internal finishes	Fittings and furnishings	Services	External works	Preliminaries	Contingencies
Substructure	1.00	–	–	–	–	–	–	–
Superstructure	0.33	1.00	–	–	–	–	–	–
Internal finishes	0.26	0.52	1.00	–	–	–	–	–
Fittings and furnishings	0.10	0.26	0.28	1.00	–	–	–	–
Services	0.28	0.57	0.64	0.33	1.00	–	–	–
External works	0.00 ^b	0.00 ^b	0.00 ^b	0.00 ^b	0.00 ^b	1.00	–	–
Preliminaries	0.35	0.37	0.44	0.18	0.39	0.00 ^b	1.00	–
Contingencies	0.23	0.28	0.34	0.21	0.29	0.00 ^b	0.36	1.00

^a Correlations above 0.10 significant at 95% confidence.

^b Subjective correlations.

Table 3
Statistics of two scenarios: including and excluding correlation (in £/m²)

Statistics	Excluding correlations	Including correlations
Mean	759.21	756.88
SD	108.92	149.55
Minimum	514.50	470.50
Q1 (25% percentile)	680.99	647.07
Q2 (Median)	759.21	756.88
Q3 (75% percentile)	823.71	843.35
Maximum	1147.20	1393.30
95% CI lower bound	590.30	522.00
95% CI upper bound	1024.00	1091.00
Estimate with 0.9 probability	903.52	958.50

cost by simply summing up cost elements drawn from individual distribution. The output statistics can then be used to assess the behavior of the true project cost. Before a full-scale simulation, validation runs are conducted to ensure the code is correct and the random variables have the specified distributions.

After 1000 simulation replications, Table 3 lists the descriptive statistics for the unit cost of the project. To assess the impact of correlations, we compare two scenarios: including and excluding correlations. Fig. 1 is a

box-and-whisker plot which is used to visually compare the distributions of the two scenarios. The first observation is that both distributions are skewed to the right because the mean (shown as the cross) is larger than the median.

The second observation is that the scenario of “including correlations” has a much longer tail to the right than that of “excluding correlation”. This indicates the former has a larger variability (uncertainty) than the latter. This conclusion is unsurprising because the former has a much greater standard deviation than the latter (149.55 versus 108.92, a 37% difference). Consequently, the 95% confidence interval of the former is much wider than that of the latter.

Fig. 2 plots the CDFs of both scenarios. A practical use of the chart is to estimate the unit cost of the project with a certain probability. Taking correlations into consideration, the unit cost with a 0.90 probability is 958.50 £/m², which would be profoundly underestimated as 903.52 £/m² if the correlations are neglected. The difference of 54.98 £/m² is greater than the cost of “substructure” (with a mean of 47.2 £/m² in Table 1). In other words, by neglecting the correlations, the error can be as serious as doing the substructure for free.

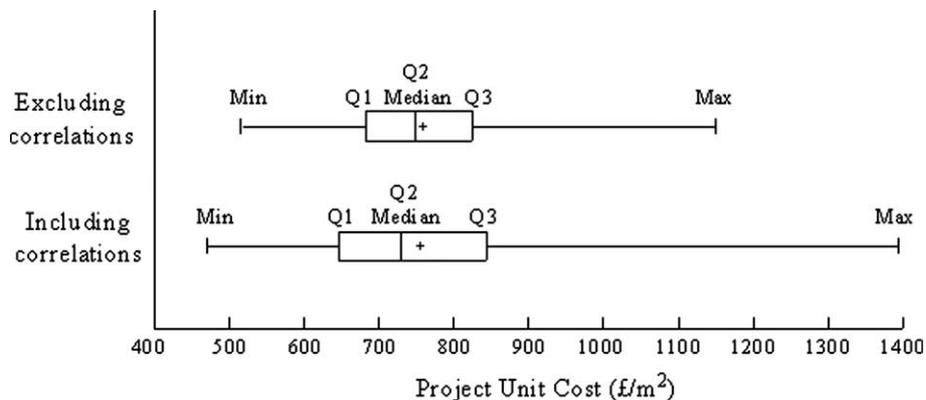


Fig. 1. Box-and-whisker plot for comparison between two scenarios: including and excluding correlations.

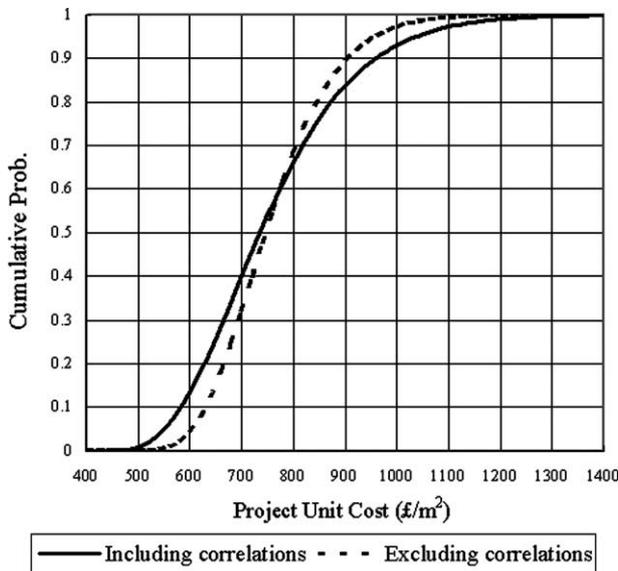


Fig. 2. Comparison on cumulative distribution functions of two scenarios: including and excluding correlations.

5. Validation

In this section, we validate whether the proposed method can capture the desired correlation matrix. Note, however, the proposed method is an approximation because of the following reasons. First, it is assumed that the correlation between X_i and X_j in Step 7 (denoted by \mathbf{M}_X) is close to the correlation between Z_i and Z_j in Step 4 (denoted by \mathbf{M}_Z). Theoretically, to find a proper \mathbf{M}_Z that leads to the desired \mathbf{M}_X requires solving $n(n-1)/2$ non-linear equations but the computation can be cumbersome [8]. Second, Steps 3, 5, and 6 require numerical approximation techniques.

Since the proposed method is an approximation, it is necessary to check the aggregated difference between the original specified correlation matrix and the generated one on the aforementioned metrics: L_{ave} and L_{max} defined by Eqs. (8) and (9). For this particular application, L_{ave} is 0.018 and L_{max} is 0.051. Moreover, the standard deviation of the differences is 0.015. Thus, the confidence limits for the mean of the difference is estimated to be 0.018 ± 0.0059 at a two-tailed significance level of 0.05 with 27 degrees of freedom. The empirical numbers give us confidence that the proposed method, despite being an approximation, can model the desired rank correlations with relatively small deviations and thereby can help assess the true impact of correlations on cost estimation.

6. Closing remarks

In this paper, we propose a simulation-based method to estimate project cost while considering correlations

between cost elements. The proposed method is more general than previous approaches because (1) it can treat different types of marginal distributions (discrete or continuous, different families distributions) for cost elements in one framework and (2) it can automatically adjust an infeasible correlation matrix into a close and feasible one very efficiently.

This paper introduces two sets of input data (marginal distributions of the cost elements and their correlations) and elaborates on how they are specified (historical and subjective approaches). The historical and subjective approaches possess their own characteristics and should both be allowed in the process of cost estimation. This leads to the development of the proposed method.

The modeling capabilities of the proposed method are empirically validated by an application to a modified British data set consisted of 216 office buildings. With the modeling capabilities, the proposed method helps cost estimators assess the true impact of correlations between cost elements on the project unit cost. The impact has been shown significant and should be considered with caution.

With the recognition that the proposed method is an approximation, users are suggested to perform the validation analysis as we demonstrated in this paper to ensure the deviation between the sampled and specified correlation matrices is within an acceptable range. It is also important to note that the proposed method relies heavily on the input of correlation data, which require not only mathematical computations but also practical judgments.

Acknowledgments

The author is thankful to Professor Shane Henderson and Dr. Soumyadip Ghosh at Cornell University and Professor Bruce Schmeiser at Purdue University for their helpful communications and discussions.

References

- [1] Abramowitz M, Stegun IA. Handbook of mathematical functions with formulas, graphs and mathematical tables 1993:932.
- [2] Attoh-Okine NO. Probabilistic analysis of factors affecting highway construction costs: a belief network approach. *Can J Civil Eng* 2002;29(3):369–74.
- [3] Badri MA, Mortagy A, Davis D, Davis D. Effective analysis and planning of R&D stages: a simulation approach. *Int J Proj Manag* 1997;15(6):351–8.
- [4] Benjamin JR, Cornell CA. Probability, statistics, and decision for civil engineers. New York: McGraw-Hill; 1970.
- [5] Best DJ, Roberts DE. The percentage points of the χ^2 distribution. *Appl Stat* 1975;24:385–8.
- [6] Cario MC, Nelson BL. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix.

- Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University; 1997.
- [7] Chau KW. Monte-Carlo simulation of construction costs using subjective data. *Construct Manag Econ* 1995;13:369–83.
- [8] Chen H. Initialization for NORTA: generation of random vectors with specified marginals and correlations. *INFORMS J Comput* 2001;13:312–31.
- [9] Cran GW, Martin KJ, Thomas GE. A remark on algorithm AS63: the incomplete beta integral, AS64: inverse of the incomplete beta function ratio. *Appl Stat* 1977;26:111–4.
- [10] Diekmann JE. Probabilistic estimating: mathematics and applications. *J Construct Eng Manag ASCE* 1983;109(3):297–308.
- [11] Elkjaer M. Stochastic budget estimation. *Int J Proj Manag* 2000;18(2):139–47.
- [12] Evans JR, Olson DL. Introduction to simulation and risk analysis. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall; 2002.
- [13] Ghosh S, Henderson SG. Behavior of the NORTA method for correlated random vector generation as the dimension increases. *ACM Transact Model Comp Simul* 2003;13(3):276–94.
- [14] Golub GH, Van Loan CF. Matrix computations. 3rd ed. Baltimore, MD: John Hopkins University Press; 1996.
- [15] Hayes RW, Perry JG, Thompson PA, Willmer G. Risk management in engineering construction, implication for project managers. Project management group. Thomas Telford Ltd; 1986.
- [16] Johnson D. The triangular distribution as a proxy for the beta distribution in risk analysis. *J Roy Stat Soc (Ser D): Statistician* 1997;46(3):387–98.
- [17] Kurowicka D, Cooke RM. Conditional, partial and rank correlation for the elliptical copula; dependence modeling in uncertainty analysis. Technical Report, Delft University of Technology, Netherlands; 2001.
- [18] Law A, Kelton WD. Simulation modeling and analysis. 3rd ed. New York: McGraw-Hill; 1999.
- [19] Lurie PM, Goldberg MS. An approximate method for sampling correlated random variables from partially specified distributions. *Manag Sci* 1998;44:203–18.
- [20] Newton S. Methods of analyzing risk exposure in the cost estimates of high quality offices. *Construct Manag Econ* 1992;10:431–49.
- [21] Odeh RE, Evans JO. The percentage points of the normal distribution. *Appl Stat* 1974;23:96–7.
- [22] Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes in C: the art of scientific computing. 2nd ed. Cambridge: Cambridge University Press; 1992.
- [23] Raftery J. Risk analysis in project management. E&FN Spon 1994.
- [24] Ranasinghe M. Impact of correlation and induced correlation on the estimation of project cost of buildings. *Construct Manag Econ* 2000;18:395–406.
- [25] Russell AD, Ranasinghe M. Analytical approach for economic risk quantification of large engineering projects. *Construct Manag Econ* 1992;10:277–301.
- [26] Schaufelberger JE, Holm L. Management of construction projects. Precedence Hall; 2002.
- [27] Schmeiser BW. Some myths and common errors in simulation experiments. In: Proceedings of the winter simulation conference, 2001. p. 39–46.
- [28] Touran A. Probabilistic cost estimation with subjective correlations. *J Construct Eng Manag ASCE* 1993;119(1):58–71.
- [29] Touran A, Suphot L. Rank correlation in simulating construction costs. *J Construct Eng Manag ASCE* 1997;123(3):297–301.
- [30] Touran A, Wiser EP. Monte-Carlo technique with correlated random variables. *J Construct Eng Manag ASCE* 1992;118(2):258–72.
- [31] Tummala VMR, Burchett JF. Applying a risk management process (RMP) to manage cost risk for an EHV transmission line project. *Int J Proj Manag* 1999;17(4):223–35.
- [32] Van Dorp JR, Kotz S. A novel extension of the triangular distribution and its parameter estimation. *J Roy Stat Soc (Ser D): Statistician* 2002;51(1):63–79.
- [33] Wall DM. Distributions and correlations in Monte-Carlo simulation. *Construct Manag Econ* 1997;15:241–58.
- [34] Wang WC. Simulation-facilitated model for assessing cost correlations. *J Comp Aided Civil Infrastruct Eng* 2002;17(5):368–80.