$\mathcal{I}^2\mathcal{SDS}$

The Institute for Integrating Statistics in Decision Sciences

Technical Report TR-2009-6

May 21, 2009

A Statistical Model for Estimating Combined Schedule and Cost Risks

Theodore S. Glickman
Department of Decision Sciences
The George Washington University

Junghee Hahn
Department of Decision Sciences
The George Washington University

A STATISTICAL MODEL FOR ESTIMATING COMBINED SCHEDULE AND COST RISKS

Theodore S. Glickman
Junghee Hahn

Department of Decision Sciences
School of Business
The George Washington University
Washington, DC 20052

ABSTRACT

In this integrated approach to project risk management, schedule risk is represented by the distribution of delay in project completion and budget risk is represented by the distribution of total project cost, which is conditional on the schedule risk. The derived expression for budget risk takes into account the dependence on schedule risk and is used to determine the budget reserve required to achieve a specific level of protection against schedule risk. The result extends previous research on delay-related cost contingency analysis.

I. INTRODUCTION

Schedule, budget, and performance—collectively referred to as the *project management triangle* [8] or the *triple constraint* [6]—are three of the most basic concerns in project management. The challenge for the project manager is to control the schedule and budget throughout the execution of the project while remaining mindful of project performance and customer expectations. This requires an integrated approach to risk management, in which schedule and budget risks are identified, assessed, and kept in balance. This need has been recognized by a number of authors [2], [4], [6].

Given the reliance on project networks in project management, it is not surprising that the assessment of schedule and budget risks is commonly approached by estimating the distributions of completion times and costs of the individual project activities and combining the distributions via analytical methods [1] or Monte Carlo simulation [7] . This *bottom-up* approach, however, has a number of limitations. It is data-intensive and time-consuming. Moreover, when schedule and budget risks are assessed separately, the interdependence between them is not evaluated, and when the assessment focuses exclusively on individual activities, it will not capture any schedule or budget risks that cut across activities.

In this paper, we discuss a *top-down* approach in which schedule risk is represented by the distribution of delay in project completion and budget risk is represented by the distribution of total project cost, which is conditional on the schedule risk. Based on the stipulated distributions, we develop an expression for budget risk that takes into account its dependence on schedule risk. We then show how this expression can be used to determine the budget reserve required to achieve a specific level of protection against schedule risk. Thus, the practical significance of this research is that it provides a basis for delay-related contingency analysis, which extends the previous research on cost contingencies exemplified by [5] and [9].

II. ANALYTICAL APPROACH

Project delays occur for a number of reasons, including failure to start all of the activities on time, overly optimistic estimates of the time required to complete some of the activities, and unforeseen developments that prolong the time it takes to perform certain activities. The greater the newness of the project or the complexity of the project, the more likely it is that a lack of experience will result in delay. When delays occur, there is an associated cost, which may have to be paid by the project executor, the customer, or both, and is likely to depend on the extent of the delay.

We begin by recognizing that the length of time a project is delayed and the associated cost of delay are random variables, where the value of the second is conditional on the value of the first. In the following development, we will begin by determining the joint distribution for delay time and delay-related cost from the assumed delay distribution and the conditional cost distribution. From that, we will then determine the marginal distribution of the cost of delay, irrespective of the length of delay. From the cumulative form of that distribution, we can then find the cost of delay at any given percentile level. Viewing the chosen percentile as the size of a cushion, the cumulative distribution is used determine the cost for any chosen size, or contingency level. For instance, at the 75th percentile, we would know the level of cost that would be exceeded only 25% of the time. If the budget for the project of concern were to be increased by this amount at the outset, we would in effect be incorporating a delay-related contingency cost.

Recognizing that risk is a combination of likelihood and consequence, the cumulative distribution of the cost of delay provides a means for relating the magnitude of that cost (the consequence) to the percentile of concern (the likelihood). This relationship will depend on the values of three parameters, the values of which will determine the severity of the risk. They are: (1) the rate of spending over the period of delay (i.e., the burn rate), (2) the maximum length of the delay period, and (3) the standard deviation of the cost of delay.

The cost of delay is assumed to be normally distributed and conditional on the length of delay, with a mean value that is proportional to the length of delay. For convenience, the standard deviation of the length of delay shall be assumed to be a third of the maximum reasonable cost, based on the fact that about 99.8% of the normal distribution is within three standard deviations of the mean. Selecting the normal distribution follows the guidance offered in [3], which states that it is the distribution that most often characterizes the underlying distribution function of a derived cost.

In contrast, the choice of distribution for the length is based on the intuitive notion that the probability of delay will not be an increasing function of the length of delay. Hence the length of delay is assumed to follow either a *uniform* distribution, where the delay is equally likely to have any value within some prescribed interval, or a *declining triangular* distribution, where the longer the delay, the less likely it will be. These two distributions are shown in Fig. 1.

III. MODEL FORMULATION

To represent the foregoing relationships algebraically, we proceed as follows. Let x = cost of delay, t = length of delay, b = burn rate, and x_{max} = maximum reasonable cost. The conditional distribution of x given t is then:

$$f_1(x|\,t) = N(\mu,\sigma^2) \tag{1}$$

where $\mu = bt$ and $\sigma = x_{max}/3$. If we let T be the maximum length of delay, then the distribution of t is either:

$$f_2(t) = 1/T \tag{2}$$

if t has a uniform distribution, or

$$f_2(t) = 2(1/T - t/T^2) \tag{3}$$

if t has a declining triangular distribution, where t ranges from 0 to T in both cases. The joint distribution of x and t is then:

$$f(x,t) = f_1(x|t)\,f_2(t)$$

from which we obtain the marginal distribution of x:

$$f(x) = \int_{-\infty}^{\infty} f(x,t)dt = \int_{-\infty}^{\infty} f_1(x\,|\,t) f_2(t)dt \tag{4}$$

where the limits of integration are 0 and T. The cumulative distribution of $f(x)$ is:

$$F(x_p) = \int_{-\infty}^{x_p} f(x)dx \tag{5}$$

where x_p = the $100p^{\text{th}}$ percentile cost of delay.

Performing the integration in (4) and (5) yields expressions for $f(x)$ and $F(x_p)$ in each of the two cases for $f_2(t)$, as follows.

Case 1. Uniform Delay Distribution

When the distribution f_2 (t) is uniform, substituting equations (1) and (2) into (4) yields

$$f(x) = \varphi(x) - \varphi(x - bT) \tag{6}$$

where

$$\varphi(x) = erf(\frac{x}{\sigma\sqrt{2}})/(2bT) \tag{7}$$

and

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}du \quad \text{(the error function).}$$

Fig. 2 shows how the shape of *f(x)* changes as the value of *T* increases and the other parameters are held constant. Since *b* and *T* only appear in the product *bT*, increasing *T* with *b* held constant has the same effect as increasing *b* with *T* held constant.

The cumulative distribution of x_p is determined by inserting (6) in (5), thereby producing this result

$$F(x_p) = \sigma\sqrt{2/\pi}\,[\omega(x) - \omega(x - bT)] + x\varphi(x\ ) - (x - bT)\varphi(x - bT)\Bigg|_{-\infty}^{x_p}$$

where

$$\omega(x) = e^{\frac{-x^2}{2\sigma^2}}/(2bT) \tag{8}$$

This culminates in

$$F(x_p) = \sigma\sqrt{2/\pi}\,[\omega(x_p) - \omega(x_p - bT)] + x_p\varphi(x_p) - (x_p - bT)\varphi(x_p - bT)\ + 0.5 \tag{9}$$

Case 2. Declining Triangular Delay Distribution

In the case where $h(t)$ is a declining triangular delay distribution, substituting (1) and (3) into (4) yields

$$f(x) = \psi(x - bT) - \psi(x) \tag{10}$$

where

$$\psi(z) = \frac{1}{b^2T^2}[\sigma\sqrt{2\pi}e^{-z^2/2\sigma^2} + (x - bT)erf(\frac{z}{\sigma\sqrt{2}})].$$

For this case, Fig. 3 shows how the shape of $f(x)$ changes as the value of T increases and the other parameters are held constant. Once again, b and T only appear in the product bT, hence the same observation holds as before.

The cumulative distribution in this case is determined by inserting (10) into (5), which results eventually in

$$F(x_p) = \frac{\sigma^2 + (x_p - bT)^2}{bT}\varphi(x_p - bT) - \frac{\sigma^2 + x_p(x_p - 2bT)}{bT}\varphi(x_p) + \frac{\sigma\sqrt{2/\pi}}{bT}[(x_p - bT)\omega(x_p - bT) - (x_p - 2bT)\omega(x_p)] + 0.5. \tag{12}$$

IV. COMPUTATION

This formula gives the probability that the cost of delay will not exceed a particular value x_p, but for practical purposes we need to know the inverse relationship. That is, to figure out how much contingency to add to a project budget in order to cover the $100p^{\text{th}}$ percentile value of the cost of lateness, we need to be able to find the value of x_p corresponding to a given value of $F(x_p)$. A closed-form expression for $x_p = F^{-1}(p)$ would serve this purpose, but since no such expression exists for the cumulative normal distribution—which is a far simpler distribution—it is unlikely to exist for $F(x_p)$ either. Hence we need to use numerical analysis instead.

To perform this analysis, we employ the Newton-Raphson method, which entails iterating with the following formula for determining the root x of the equation $\xi(x) = 0$:

$$x_i = x_{i-1} - \xi(x_{i-1})/\xi'(x_{i-1}),\ i = 1,2,\ldots$$

Substituting $F(x) - p$ for $\xi(x)$ and $f(x)$ for $\xi'(x)$ this becomes

$$x_i = x_{i-1} - (F(x_{i-1}) - p)/f(x_{i-1}),\ i = 1,2,\ldots \tag{13}$$

For the starting value x_o we can use the $100p^{\text{th}}$ percentile value of the normal distribution with mean μ_o equal to the expected value $E(x)$ of the cost of lateness and variance σ_o^2 equal to the variance $Var(x)$ of that cost. $E(x)$ and $Var(x)$ are determined by integration using the expressions for $f(x)$ in (6) and (10), yielding the following equations in each of the two cases:

Case 1. Uniform Delay Distribution

$$E(x) = \tfrac{1}{2}\, bT \tag{14}$$

$$Var(x) = b^2T^2/12 + s^2 \tag{15}$$

Case 2. Declining Triangular Delay Distribution

$$E(x) = \tfrac{1}{3}\, bT \tag{16}$$

$$Var(x) = b^2T^2/18 + s^2 \tag{17}$$

To illustrate the numerical procedure discussed above, suppose the delay distribution $f_2(t)$ is triangular and let $b = 10$, $T = 1$, and $\sigma = 3$, and that the level of concern is the 80th percentile of the cost overrun distribution, so that $p = 0.8$. Then from equations (16) and (17) we find that the starting values of μ_o and σ_o^2 are 3.33 and 25.67. By means of Microsoft Excel's NORMINV function, we then find that the 80th percentile of the normal distribution with those parameters is at $x_o = 7.58$. Then, using (13) in conjunction with equations (10) and (12), we find in just four iterations that $x_4 = 6.56$, with a deviation between the results of iterations 3 and 4 of only $x_4 - x_3 = 5.96 \times 10^{-7}$. The interpretation is as follows. If the delay follows a declining triangular distribution and is at most one month ($T = 1$), and the cost of delay is normally distributed with a mean of \$10,000 per month ($b = 10$) and a standard deviation of \$3,000 ($\sigma = 3$), then the contingency budget required to cover 80% of the possible cost of delay is \$6560 ($x_4 = 6.56$).

Thus the formulas for f(x) and F(x) in equations (6) and (9) for Case 1 and equations (9) and (12) for Case 2, coupled with the iterative formula in equation (13), provide a means of determining the reserve that will cover a given percentage of the cost of delay, over and above the baseline project budget, which presumes there will be no delay. The reserve may be expanded further to account for other uncertainties unrelated to delay (e.g., variability in the cost of inputs). The *F(x)* formulas can also be used to evaluate the benefits of potential risk management measures by changing the appropriate parameter values. For instance, a measure that would shorten the maximum duration of delay, *T*, would result in a reduction in *F(x)*. Such a measure would, therefore, have a beneficial impact on the risk and, if the cost of the measure were determined, a cost-benefit ratio could be calculated.

V. CONCLUSIONS

The results shown were based on the assumption that project delay follows either a uniform or declining triangular distribution and that the cost of delay, which is of course conditional on the delay itself, follows a normal distribution and is linked to the delay distribution by its mean, which is proportional to the length of the delay. In a sense, this approach is a first approximation to the solution of the problem of how to best describe the interdependence between schedule risk and cost risk, which may require different distributions and linkages than the ones assumed here. Nevertheless, it demonstrates how closed-form mathematical relationships can be derived and used to assess risk in lieu of simulation models, making it easier to estimate delay-related contingencies, analyze the sensitivity of risks to variations in key parameters, and evaluate the payoffs of risk reduction measures aimed at schedule risks, cost risks, or both. .

REFERENCES

[1] J. R. Birge and M. J. Maddox, "Bounds on expected project tardiness," *Operations Research*, vol. 43, no. 5, pp. 838-850, 1995.

[2] T. R. Browning and S. D. Eppinger, "Modeling impacts of process architecture on cost and schedule risk in product development," *IEEE Trans., Eng. Mgt*., vol. 49, no. 4, pp. 428- 442, Nov. 2002.

[3] P. R. Garvey, *Probability Methods for Cost Uncertainty Analysis: A Systems Engineering Perspective*, Boca Raton: CRC Press, 2000.

[4] D. T. Hulett and B. Campbell, "Integrated cost/schedule risk analysis," presented at the Fifth European Project Management Conference, PMI Europe, Cannes, France, 2002.

[5] J. T. Karlsen and J. Lereim*, "*Management of project contingency and allowance," *Cost Engineering*, vol. 47, no. 9, pp.24-29, Sept. 2005.

[6] U. Kuehn, *Integrated Cost and Schedule Control in Project Management*, Vienna, VA: Management Concepts, 2006.

[7] F. Raymond, "Quantify risk to manage cost and schedule," *Acquisition Review Quarterly*, pp. 147-155, Spring 1999.

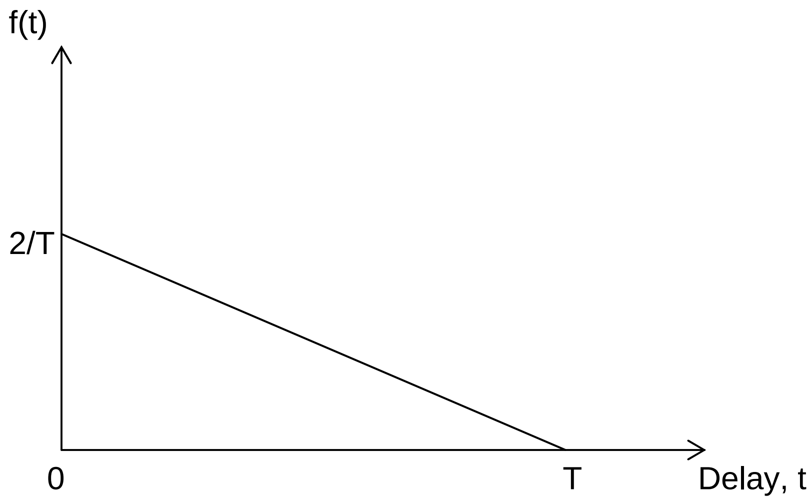
[8] W. J. Stevenson, *Operations Management,*10e, New York: McGraw-Hill/Irwin, 2009.

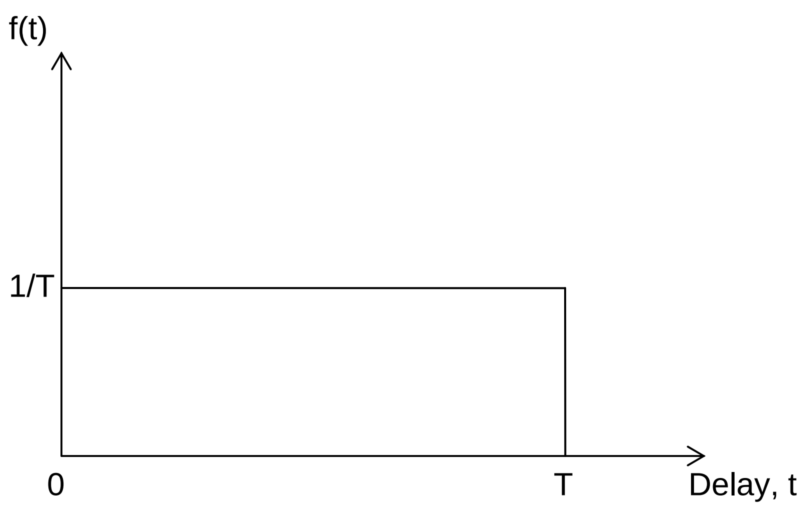
[9] A. Touran, "Probabilistic cost estimating with subjective correlations," *J. Construction Eng. and Mgt.*, vol. 119, no.1, pp. 58-71, March 1993.

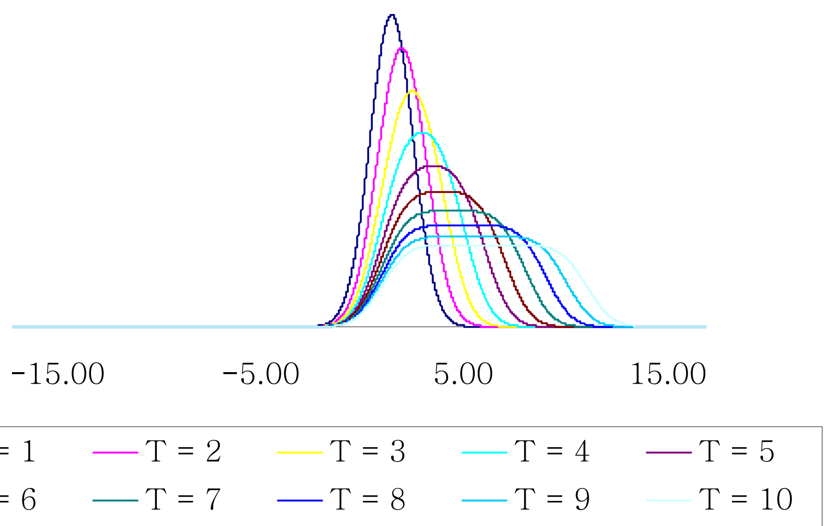
Figure 1 Uniform and triangular delay distributions, $f_2(t)$

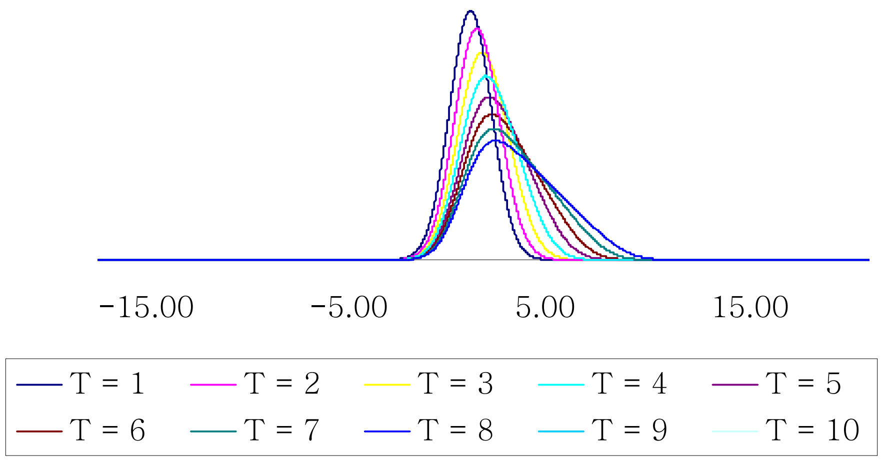
Figure 2. $f(x)$ when $f_2(t)$ is uniform and T varies from 1 to 10 ($b = \sigma = 1$)

Figure 3. $f(x)$ when $f_2(t)$ is triangular and T varies from 1 to 10 ($b = \sigma = 1$)


f(t)
2/T
0
T
Delay, t


f(t)
1/T
0
T
Delay, t


-15.00
-5.00
5.00
15.00
T = 1
T = 2
T = 3
T = 4
T = 5
T = 6
T = 7
T = 8
T = 9
T = 10


-15.00
-5.00
5.00
15.00
T = 1
T = 2
T = 3
T = 4
T = 5
T = 6
T = 7
T = 8
T = 9
T = 10