Quantifying the Relationship between Schedule and Cost

By Stephen A. Book, MCR

It is well known that a program's cost is related to its schedule, yet when a program's schedule estimate is updated, it is not often that the cost estimate is updated in parallel to reflect consistency with the new schedule estimate. One reason for this deficiency in estimating is probably the unfortunate wall of separation that seems to divide those analysts who do cost estimating from those analysts who do schedule estimating. Another appears to be a lack of understanding of the extent of the impact of schedule growth on cost growth.

While the first issue is a management problem that we cannot solve here, the second one is mathematical, and we can offer a framework for dealing with it. In technical papers [1,2] and a monograph [3], P.R. Garvey suggested the application of joint and conditional probability distribution theory to model the relationship between schedule and cost. Proceeding from Garvey's work, I will describe a practical method of understanding that relationship and applying it to estimating a project's cost based on its schedule and *vice versa*.

The Role of Risk Analysis

Because both schedule and cost estimating are versions of forecasting future events, there is considerable uncertainty in estimates of each. The solution to the problem is to treat the schedule and cost estimating process statistically, a technique referred to as schedule-risk analysis and cost-risk analysis, respectively. As it is now done, probability distributions are separately established to model the duration of each activity and the cost of each work-breakdown structure item. Then correlations among the activity durations and among the costs are estimated, and the schedule and cost distributions, separately, are summed statistically, typically by Monte Carlo sampling. The results are (1) a probability distribution of project schedule and (2) a probability distribution of project cost, from which one can obtain estimates of the median (50% confidence level), 70th percentile (70% confidence level), and other relevant estimates of interest. Commercial software is available to carry out the required statistical processes, including the correlation aspect: Risk+®, a third-party add-on to Microsoft Project, for schedule-risk analysis, and Crystal Ball®, a third-party add-on to Microsoft Excel, for cost-risk analysis. Each of these software products outputs, respectively, a probability distribution of project's schedule duration and a probability distribution of the project's cost.

Unfortunately, there doesn't seem to be any current commercial software that links the distribution of project schedule with the distribution of project cost; therefore, we are on our own in proceeding from the theory proposed by Mr. Garvey to the practicalities of assessing the impact of schedule on cost and cost on schedule.

The Mathematics

In the 1990s, studies at both The Aerospace Corporation and the MITRE Corporation found that the lognormal probability distribution almost always serves as a good model for both the schedule distribution and the cost distribution. The lognormal distribution is derived from the normal distribution in the following way: if X is a normally distributed random variable* having mean P and standard deviation Q, then $Y = e^X$ is said to have a lognormal distribution. The mean and standard deviation of that lognormal distribution can be calculated to be

$$\mu = e^{P+\frac{1}{2}Q^2}$$

and

$$\sigma = e^{P+\frac{1}{2}Q^2}\sqrt{e^{Q^2}-1}\ ,$$

respectively. (Derivations of all the formulas presented here can be found in Reference [3].) Because the

* *A random variable is a statistical quantity whose characteristics are described by a probability distribution, a mean, a standard deviation, and other statistical metrics.*

normal distribution is more familiar than the lognormal and is easier to work with in most situations, the lognormal distribution is often studied in terms of the normal distribution parameters P and Q, rather than the lognormal parameters μ and σ. The utility of this approach will become evident as we discuss the relationship between schedule and cost in more detail.

Suppose the distribution of project cost is represented by the lognormal random variable C that has mean μ_C and standard deviation σ_C, and the distribution of project schedule is represented by the lognormal random variable S that has mean μ_S and standard deviation σ_S. Suppose, also, that the correlation between cost and schedule, namely between C and S is λ. Let's translate all of these items into metrics of the underlying normal distributions. The algebraic expressions for the mean P and standard deviation Q of a normal distribution that is associated with a lognormal distribution having mean μ and standard deviation σ are, respectively,

$$P = \frac{1}{2}\ell n\left\{\frac{\mu^4}{(\mu^2+\sigma^2)}\right\}$$

and

$$Q = \sqrt{\ell n\left(1+\frac{\sigma^2}{\mu^2}\right)}$$

Furthermore, because the two lognormal distributions are correlated (with correlation value λ), so are the underlying normal distributions, and the correlation between the normal distributions, denoted by the Greek letter ρ is given by the expression

$$\rho = \frac{1}{Q_C Q_S}\ell n\left\{1+\lambda\sqrt{e^{Q_C^2}-1}\sqrt{e^{Q_S^2}-1}\right\}.$$

A picture of the relationship between a normal distribution with mean P and standard deviation Q and its derived lognormal distribution is displayed in Figure 1.

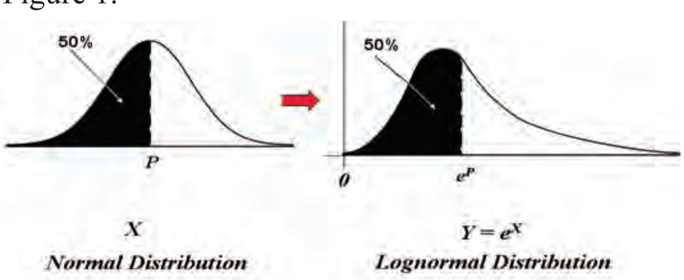

Figure 1. Transition from Normal to Lognormal.

The algebraic equation of the graph of the lognormal distribution on the right side of Figure 1 is called the "probability density function" and is, for cost

$$f_C(x) = \frac{1}{xQ_C\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\log x-P_C}{Q_C}\right)^2} \quad \text{for } 0<x<\infty$$

$$= 0 \quad \text{for} -\infty<x<0$$

and, for schedule,

$$f_S(y) = \frac{1}{yQ_S\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\log y-P_S}{Q_S}\right)^2} \quad \text{for } 0<y<\infty$$

$$= 0 \quad \text{for} -\infty<y<0$$

Using this information and the fact that the correlation between the two lognormal distributions is λ, we can calculate their "joint bivariate density," which is the key ingredient in being able to establish a relationship between a project's cost and its schedule:

$$h_J(x,y) = \frac{1}{2\pi xyQ_CQ_S\sqrt{1-\rho^2}}e^{-w(x,y)},$$

where

$$w(x,y) = \frac{1}{1-\rho^2}\left\{\left(\frac{\ell n(x)-P_C}{Q_C}\right)^2 - 2\left(\frac{\ell n(x)-P_C}{Q_C}\right)\left(\frac{\ell n(y)-P_S}{Q_S}\right)+\left(\frac{\ell n(y)-P_S}{Q_S}\right)^2\right\}.$$

The Practical Implementation

I'm sure all the mathematical theory described in the previous section is interesting to you but, to make practical use of it, we have to be able to actually make the computations that establish the relationship between project schedule and project cost. To do this task, we apply a technique from calculus known as Simpson's Rule. Simpson's Rule converts the process of calculating probabilities, which is a continuous process, into a discrete process whose steps can be sequentially programmed in Excel or a computer language such as C or C++.

The first computation we have to make is of the "conditional" probability of project cost, "given" project schedule. Using the notation of the previous section, we compute the probability that project cost is between *a* (million dollars) and *b* (million dollars) and schedule duration is *c* (months). In mathematical notation, this probability is expressed and calculated (using Simpson's Rule) as follows:

$$P\{a \le C \le b | S = c\} = \int_a^b \frac{h_J(x,c)}{f_S(c)}dx = \int_a^b g(x,c)dx$$

$$= \left(\frac{b-a}{3m}\right)\sum_{j=0}^{m-1}\left\{g\left(a+2j\left(\frac{b-a}{m}\right),c\right)+4g\left(a+(2j+1)\left(\frac{b-a}{m}\right),c\right)+g\left(a+2(j+1)\left(\frac{b-a}{m}\right),c\right)\right\}$$

Because we know the algebraic expressions for $h_J(x,y)$ and $f_s(c)$, we can easily set up the required algebraic expression for $g(x,c)$. We can perform the summation calculation in Excel by programming the successive terms into the appropriate cells or, even better, programming the entire summation process in Visual Basic. The larger the number m is, the more accurate the discrete representation of the continuous conditional probability is. For a start, $m = 100$ is good, $m = 1{,}000$ is better, and $m = 10{,}000$ is even better. The difference between them in amount of time required for the computer to complete the calculation is negligible.

The Results

Figure 2 is a portion of our Excel spreadsheet that implements the process of the calculating conditional probabilities of cost, given schedule.

The input cells in the upper left corner of the spreadsheet allow the analyst to enter the descriptive parameters of the cost (X) and schedule (Y) distributions. In the case shown (Fig. 2), the (lognormal) cost distribution has mean \$200M dollars and standard deviation \$30M, while the (lognormal) schedule distribution has mean 36 months and standard deviation 5 months. The correlation between cost and schedule is entered as 0.50.

Although the project schedule distribution has mean 36 months and standard deviation 5 months, the nature of probability distributions makes it possible for *any number of months at all* to be the actual duration of the project, albeit with differing probabilities. Generally, those durations closer to the mean have higher probabilities than those at the extreme ends of the distribution. In the example (Fig. 3), for each suggested value of project schedule duration, here 5, 25, 50, 80, 100, and 120 months, the spreadsheet computes the probability distribution of project cost and expressed that cost distribution as an "S-curve," or cumulative distribution curve. Of course, depending on the particular project involved,

Lognormal Distribution Parameters for Cost and Schedule															
	Variable	μ	σ	λ	P	Q	SQRT(eQ2-1)	ρ			Step 2: Starting with μ-5σ, enter numbers in cells R3 and				
Cost	*X*	200	30	0.50	5.287192062	0.14916638	0.150000	0.50259235			R4 until cells U3 and U4 have exponents E-15 or smaller				
Schedule	*Y*	36	5		3.573965725	0.138226	0.138889				(i.e., more negative). Then start with μ+5σ in cells V3 and				
1	Step 1: Enter lognormal parameters in the above yellow boxes.										V4 until cells Y3 and Y4 reach E-15 or lower.				
2	m =	200		0 =	0	x =	100	b-a =	100	(b-a)/m =	0.500				
3	n =	200		c =	0	d =	36	d-c =	36	(d-c)/n =	0.180		Difference between Successive Numbers in Th		
	P{a<X<b | Y = y} = (Constant) x Single Summation =				0.0000										
4			x	y											
5	j	k	a+j(b-a)/m	y	w1	w2	w	exp{-w/2}	DENOM h	h(x,y)	DENOM g(y)	EXPON g(y)	g(y)	h(x,y)/g(y)	Simpson
	0	1	0.000	36.000	-143.49941	0.06911	27564.93088	0.00000000	0.00000	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
6	1	1	0.300	36.000	-43.51627	0.06911	2537.71949	0.00000000	1.20960	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
7	2	1	0.600	36.000	-38.86947	0.06911	2025.07182	0.00000000	2.41920	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
8	3	1	0.900	36.000	-36.15126	0.06911	1751.97782	0.00000000	3.62879	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
9	4	1	1.200	36.000	-34.22266	0.06911	1570.20520	0.00000000	4.83839	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
10	5	1	1.500	36.000	-32.72672	0.06911	1436.06567	0.00000000	6.04799	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
11	6	1	1.800	36.000	-31.50445	0.06911	1330.91094	0.00000000	7.25759	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
12	7	1	2.100	36.000	-30.47104	0.06911	1245.12277	0.00000000	8.46719	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
13	8	1	2.400	36.000	-29.57586	0.06911	1173.11962	0.00000000	9.67679	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
14	9	1	2.700	36.000	-28.78625	0.06911	1111.38827	0.00000000	10.88638	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
15	10	1	3.000	36.000	-28.07992	0.06911	1057.58143	0.00000000	12.09598	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
16	11	1	3.300	36.000	-27.44097	0.06911	1010.05730	0.00000000	13.30558	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
17	12	1	3.600	36.000	-26.85765	0.06911	967.62511	0.00000000	14.51518	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
18	13	1	3.900	36.000	-26.32105	0.06911	929.39532	0.00000000	15.72478	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
19	14	1	4.200	36.000	-25.82423	0.06911	894.68700	0.00000000	16.93438	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
20	15	1	4.500	36.000	-25.36171	0.06911	862.96801	0.00000000	18.14397	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
21	16	1	4.800	36.000	-24.92905	0.06911	833.81508	0.00000000	19.35357	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
22	17	1	5.100	36.000	-24.52263	0.06911	806.88639	0.00000000	20.56317	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
23	18	1	5.400	36.000	-24.13944	0.06911	781.90218	0.00000000	21.77277	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
24	19	1	5.700	36.000	-23.77698	0.06911	758.63075	0.00000000	22.98237	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
25	20	1	6.000	36.000	-23.43311	0.06911	736.87824	0.00000000	24.19196	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
26	21	1	6.300	36.000	-23.10603	0.06911	716.48088	0.00000000	25.40156	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
27	22	1	6.600	36.000	-22.79416	0.06911	697.29920	0.00000000	26.61116	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
28	23	1	6.900	36.000	-22.49616	0.06911	679.21350	0.00000000	27.82076	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000
29	24	1	7.200	36.000	-22.21084	0.06911	662.12031	0.00000000	29.03036	0.00000000	12.473323	0.002388	0.079979851	0.00000000	0.00000000

Figure 2. Portion of Excel Spreadsheet for Calculating the Conditional Distribution of Project Cost, Given Various Values of Project Schedule.

some of these durations may not be feasible and others may be of interest instead. The point is that, for any feasible project schedule, the distribution of project cost can be estimated. A portion of the output of the spreadsheet's computations is seen in Figure 3.

The graph in Figure 4, based on the output information in Figure 3, illustrates the relationship between project schedule and project cost that is modeled by the computations in the spreadsheet. The graph displays the S-curves of the cost probability distributions, expressed as the conditional probability distribution of program cost, given a specific program schedule. As the schedule lengthens, the cost distribution moves to the right (more dollars) and widens (more uncertainty). This chart, as noted earlier, assumes a correlation of 0.5 (representing a 25% joint impact) between cost and schedule, as recommended by P.R. Garvey in one of this technical papers. Several research studies are in progress to determine the appropriate correlation between schedule and cost.

What information about the project can be drawn from the S-curves of Figure 4? Well, if there is schedule slip to 50 months (from the mean of 36 months), we can expect the median to be about \$220M and the 80th percentile cost to be about \$250M. If the schedule slips further to 80 months, we move to the 80-month S-curve and find that median cost has grown to \$300M and the 80th percentile to about \$340M.

The spreadsheet can be "tweaked" to also provide S-curves of schedule duration, given specific possible cost values. To do this, we simply have to replace the cost numbers in the

$P\{X \leq x \mid Y = y\}$

x Y=	5	25	50	80	100	120
60.00	0.162	0.000	0.000	0.000	0.000	0.000
70.00	0.582	0.000	0.000	0.000	0.000	0.000
80.00	0.893	0.000	0.000	0.000	0.000	0.000
90.00	0.984	0.000	0.000	0.000	0.000	0.000
100.00	0.998	0.000	0.000	0.000	0.000	0.000
110.00	0.999	0.001	0.000	0.000	0.000	0.000
120.00	0.999	0.009	0.000	0.000	0.000	0.000
130.00	0.999	0.039	0.000	0.000	0.000	0.000
140.00	0.999	0.118	0.000	0.000	0.000	0.000
150.00	0.999	0.257	0.000	0.000	0.000	0.000
160.00	0.999	0.440	0.001	0.000	0.000	0.000
170.00	0.999	0.625	0.005	0.000	0.000	0.000
180.00	0.999	0.777	0.016	0.000	0.000	0.000
190.00	0.999	0.881	0.042	0.000	0.000	0.000
200.00	0.999	0.943	0.091	0.000	0.000	0.000
210.00	0.999	0.975	0.169	0.002	0.000	0.000
220.00	0.999	0.990	0.275	0.005	0.000	0.000
230.00	0.999	0.996	0.401	0.013	0.001	0.000
240.00	0.999	0.999	0.531	0.029	0.002	0.000
250.00	0.999	1.000	0.654	0.057	0.006	0.001
260.00	0.999	1.000	0.758	0.101	0.013	0.001
270.00	0.999	1.000	0.839	0.162	0.027	0.004
280.00	0.999	1.000	0.898	0.241	0.050	0.008
290.00	0.999	1.000	0.939	0.333	0.085	0.016
300.00	0.999	1.000	0.965	0.433	0.134	0.031
310.00	0.999	1.000	0.980	0.534	0.197	0.053
320.00	0.999	1.000	0.989	0.630	0.272	0.085
330.00	0.999	1.000	0.994	0.716	0.357	0.128
340.00	0.999	1.000	0.997	0.789	0.446	0.183
350.00	0.999	1.000	0.999	0.848	0.535	0.249
360.00	0.999	1.000	0.999	0.893	0.621	0.323
370.00	0.999	1.000	1.000	0.927	0.698	0.402
380.00	0.999	1.000	1.000	0.952	0.766	0.484
390.00	0.999	1.000	1.000	0.969	0.823	0.564
400.00	0.999	1.000	1.000	0.980	0.870	0.640
410.00	0.999	1.000	1.000	0.988	0.906	0.708
420.00	0.999	1.000	1.000	0.993	0.933	0.769
430.00	0.999	1.000	1.000	0.996	0.954	0.821
440.00	0.999	1.000	1.000	0.997	0.969	0.864
450.00	0.999	1.000	1.000	0.998	0.979	0.898
460.00	0.999	1.000	1.000	0.999	0.986	0.925
470.00	0.999	1.000	1.000	0.999	0.991	0.946
480.00	0.999	1.000	1.000	1.000	0.994	0.962
490.00	0.999	1.000	1.000	1.000	0.996	0.973
500.00	0.999	1.000	1.000	1.000	0.998	0.982
510.00	0.999	1.000	1.000	1.000	0.999	0.987
520.00	0.999	1.000	1.000	1.000	0.999	0.992
530.00	0.999	1.000	1.000	1.000	0.999	0.994
540.00	0.999	1.000	1.000	1.000	1.000	0.996
550.00	0.999	1.000	1.000	1.000	1.000	0.998
560.00	0.999	1.000	1.000	1.000	1.000	0.998
570.00	0.999	1.000	1.000	1.000	1.000	0.999
580.00	0.999	1.000	1.000	1.000	1.000	0.999
590.00	0.999	1.000	1.000	1.000	1.000	1.000
600.00	0.999	1.000	1.000	1.000	1.000	1.000

Figure 3. A Portion of the Output of the Spreadsheet's Computations – Cost Probability Distributions Corresponding to Various Possible Values of the Schedule Duration.

upper left corner of the spreadsheet with the schedule numbers and the schedule numbers with the cost numbers in their respective cells.

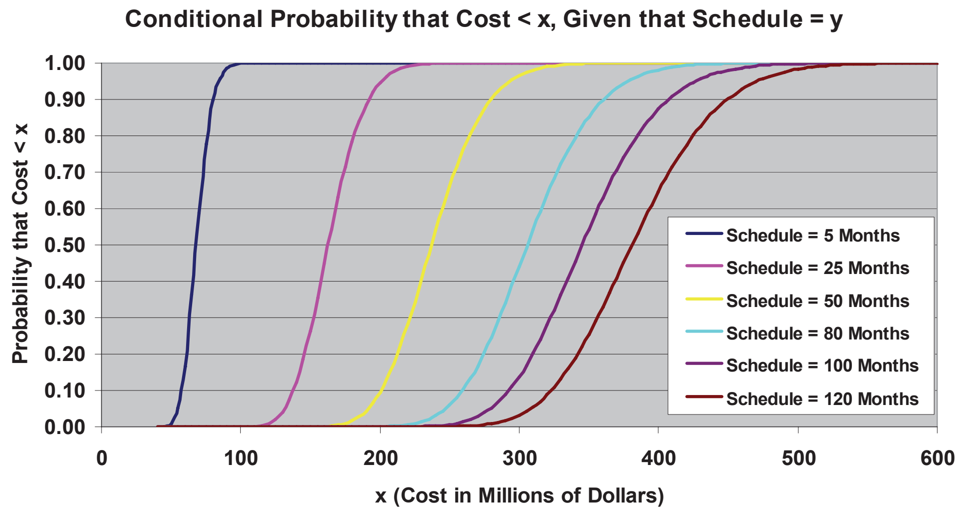

Figure 4. Cost S-Curves Corresponding to Different Project Schedule Durations.

References

1. Garvey, P.R., "Modeling Cost and Schedule Uncertainties – A Work Breakdown Structure Perspective," *Military Operations Research*, Vol. 2, No. 1, Spring 1996, pp 37–43.
2. Garvey, P.R. and Taub, A.E., "A Joint Probability Model for Cost and Schedule Uncertainties," *The Journal of Cost Analysis*, Spring 1997, pp 29–38.
3. Garvey, P.R., *Probability Methods for Cost Uncertainty Analysis: A Systems Engineering Perspective*, New York: Marcel Dekker, ©2000, "Modeling Cost and Schedule Uncertainties", pp 308–335.

About the Author

Dr. Book is Chief Technical Officer of MCR. He can be reached at 310.640.0005 x244 or sbook@mcri.com.