An Overview of Correlation

Key Points

- Recognize that correlations exist not only among cost elements in a project but among the cost elements in different projects in a program.
- As correlation grows, the probabilistic cost distribution curve broadens (higher standard distribution), requiring higher budgets at a given confidence level.
- As the number (n) of correlated activities in a project or projects in a program grows, so too does the variance in total project costs (proportional to n^2 at higher n). Correlation effects increase with the number of cost elements in a project or projects in a program.
- In the absence of any correlation, the probabilistic cost distribution narrows as the number of activities or projects increases (proportional to $1/\sqrt{n}$).[1]
- Correlation does not change the expected costs of individual cost elements, but instead only changes the "portfolio"[2] standard deviation.
- The behaviors in schedule risk analysis depend on whether tasks/projects are executed serially, rolled up, or executed in parallel.
- Some sources of correlation are identified and recommendations provided.

Introduction

One of the most under-considered elements of cost and schedule risk is the correlation that exists within various WBS (work breakdown structure) elements of a project or across projects comprising a program. Failure to adequately consider correlation between various activities and projects compounds the impact of other factors present in large complex projects. These include:

- MAIMS – "Money Allocated Is Money Spent"
- Parkinson's Law – work expands to fill the time allotted
- Overconfidence in assessing uncertainties
- Complexity with hidden coupling – risk events are likely to affect multiple cost elements with the potential for cascading impacts
- State of technology
- Common management, staff, and work processes
- Optimism bias
- Overly simplistic probabilistic cost analysis (PCA)

1

This Executive Insight looks at correlation in project and program risk assessments and some of the impacts of a failure to adequately consider such correlation in project risk assessments related to both cost and schedule.

Correlation and Its Impacts in Projects and Programs

Before delving into correlation more fully, it is important to highlight the effects of correlation and some of the erroneous behaviors failing to consider it may drive. In simple terms, as correlation grows, the probabilistic cost distribution curve broadens (higher standard distribution) requiring higher budgets at a given confidence level (P65; P80 etc.).

This can be seen in Figure 1.[3]

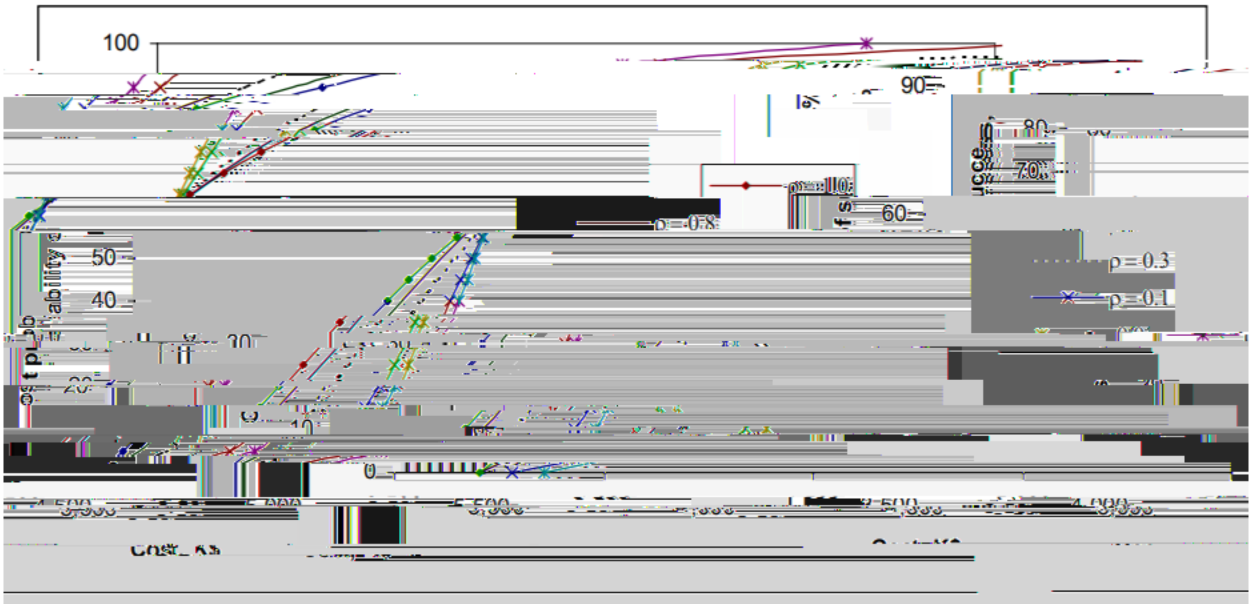

Figure 1

As the number of correlated activities in a project or projects in a program grows, so too does the variance in total project costs (proportional to n^2 at higher n). Correlation effects increase with the number of cost elements in a project or projects in a program.[4]

[3] Kujawski, Edouard & Alvaro, Mariana & Edwards, William. (2004); Figure 1b
[4] Book, 1999 and 2000/2001

Figure 2[5]

Conversely, in the absence of any correlation (independent random variables), the probabilistic cost distribution narrows as the number of activities or projects increases (proportional to $1/\sqrt{n}$). This often leads to a management decision to unacceptably decompose cost elements, which leads to a normally distributed cost distribution. This conflicts with the reality that large complex projects are characterized by distributions skewed to the right (in Figure 2) and much broader than what decomposition might suggest.

Positively correlated elements have the effect of spreading the distribution of total cost.

It must also be recognized that correlations exist not only among cost elements in a project, but among the cost elements in different projects in a program. The former are driven by characteristics such as complexity, common staff, and processes, while the latter arise from organizational and programmatic factors common across projects.

---

[5] See Book 2000/2001. This graph illustrates the importance of working with the numeric correlations between WBS items. Assuming these correlations to be zero causes a detrimental effect on the estimation of total-cost uncertainty. Shown is the percentage by which the sigma value (standard deviation) of the total-cost distribution is underestimated, assuming WBS inter-element correlations to be zero instead of the actual value (usually represented by ρ, the Greek letter rho). The horizontal axis tracks ρ, and the vertical axis tracks the percentage by which the total-cost sigma value is, for each nonzero correlation value, underestimated if the correlations are instead assumed to be zero. Each curve is keyed to a unique value of n, the number of elements in a roll-up. As the four curves show, the percent by which sigma is underestimated also depends on the number of WBS items for which the pairwise correlations are incorrectly assumed to be zero. For example, if $n = 30$ WBS items, and all correlations between WBS items (ρ) are 0.2, but the estimator assumes they are all zero, the total-cost sigma

So, What is Correlation?

Correlation measures linear dependence between two or more random variables. As such, it provides only a partial picture of their dependence. It does not indicate causality. Even a correlation coefficient of 1.0 does not indicate causality, only perfect dependence. Correlation typically refers to Pearson's product moment coefficient.[6] When data are nonlinear, non-parametric correlation may be more robust.

Correlation does not tell the whole story as was demonstrated by Anscombe's quartet, where four data sets with nearly identical correlation and other significant statistical properties look very different when graphed.[7] (Figure 3)

Figure 3: ! ýãÒ2ϒōŜ0ă Ijdz·ŴúŜű

The correlation coefficient is dependent on the variance[8] of the data and degrades with volatility of the data. Correlation does not change the expected costs of individual cost elements, but instead only changes the "portfolio"[9] standard deviation. This in turn changes the shape of the traditional S-curve, increasing budget requirements for confidence levels greater than the expected value (P50) (where

---

[6] Other definitions for correlation include rank correlation and Kendall's Tau. Both are non-parametric measures.

[7]

Sources of Correlation

Having discussed the implication of correlation on cost and schedule probabilistic assessments, it is worthwhile to identify some sources of correlation often found in large complex programs. These are summarized in Table 1. Other sources of correlation exist. The table is meant to be suggestive.

Table 1

S#7q14Td[seW*nB0gdh!2

Other Factors Impacting Project and Program Risk Assessment

Several factors beyond correlation can impact the output and, more importantly, the validity and conclusions from a probabilistic risk assessment. Some of these are outlined in the introduction and succinctly discussed here.

MAIMS – "Money Allocated is Money Spent"

MAIMS is the financial analog of Parkinson's Law and is a major contributor to cost overruns or higher than necessary expenditures in the delivery of a program. One telltale sign that this effect is in full play is in multi-project programs where the final cost performance index is at 1.00 for a large number of the individual projects. This is not the result of "perfect" management, but rather the willful consumption of any underrun that may have existed. The MAIMS principle effectively makes any potential savings from underruns unavailable to cover overruns elsewhere in the program.

Typical project cost analysis assumes an "ideal" project or program, where savings on one element are made available to other elements. The presence of MAIMS in program or project contexts drives to an alternative strategy on establishing budgets and dynamically managing contingency and risk pools.

MAIMS acts to increase a probability distribution function's mean and to reduce its standard deviation. (Effectively, values less than the allocated amount are assumed to be equal to the allocated budgets in the statistical assessment of total project or program costs.)

Optimism Bias

People have a systematic bias towards overconfidence. Thus, many cost distribution approaches that rely on expert judgement to set several values (minimum, most likely, and maximum, for example) lead

Recognize other human and organizational influences on project cost analysis and consider them together rather than separately.

Recognize systems thinking is essential when addressing correlation in large complex projects and programs.

References

M. Alpert and H. Raiffa, "A progress report on the training of probability assessors," in Judgment under Uncertainty: Heuristics and Biases, D. Kahneman, P. Slovic, and A. Tversky (Editors), Cambridge University Press, pp. 294-305, 1982.

P. Bakhshi, A. Touran, A Method for Calculating Cost Correlation among Construction Projects in a Portfolio; *International Journal of Architecture, Engineering and Construction* Vol 1, No. 3, September 2012.

S.A. Book, Why Correlations Matter in Cost Estimating, 32nd Annual DOD Cost Analysis Symposium, Williamsburg, 1999.