Risks and Opportunities
September 28, 2021

NAC Executive Insights

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] 1 divided by the square root of n

[2] All WBS cost elements in a project or projects in a program.

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]

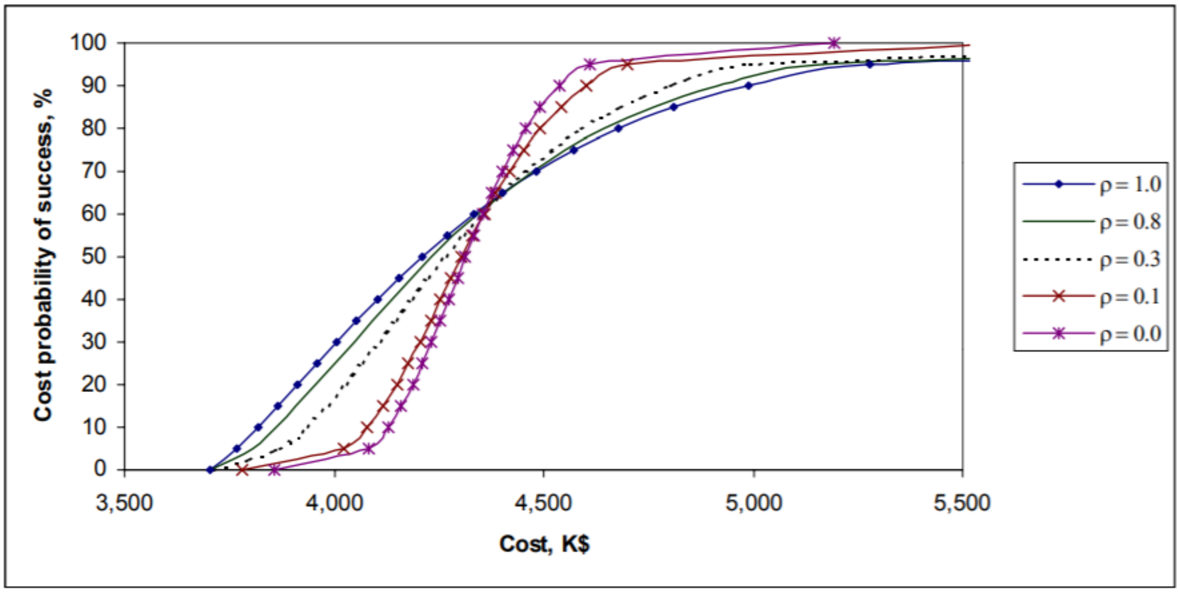


Figure 1

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

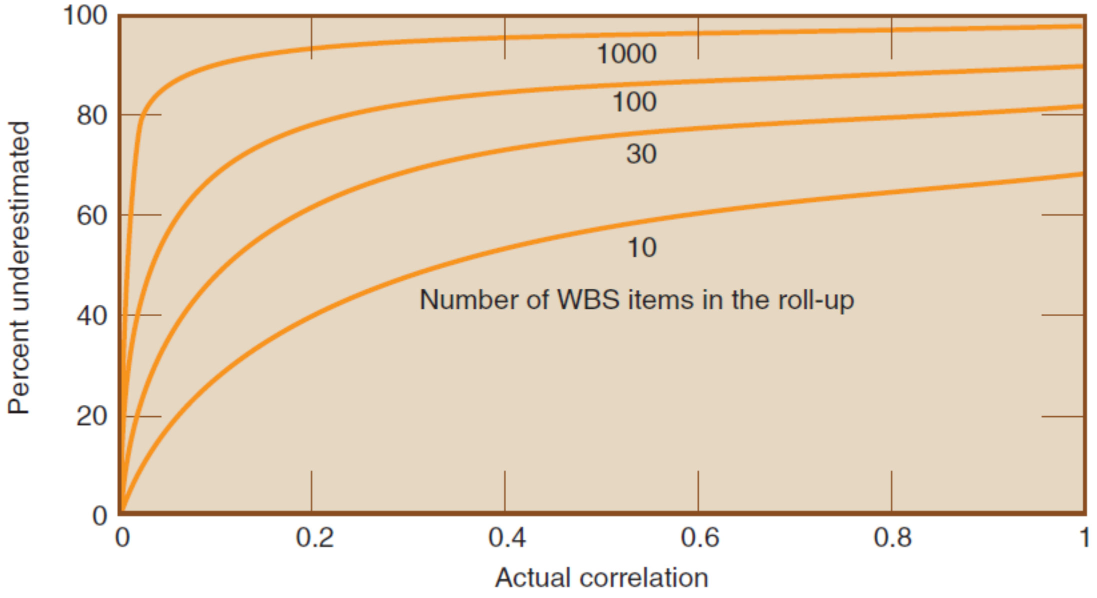


Figure 2[5]

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/√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 values would be underestimated by about 60 percent. (This is meant to be a generic illustration and therefore is only approximately true in any specific case. It is assumed that the sigma values for the WBS items are the same throughout the entire structure.)

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)

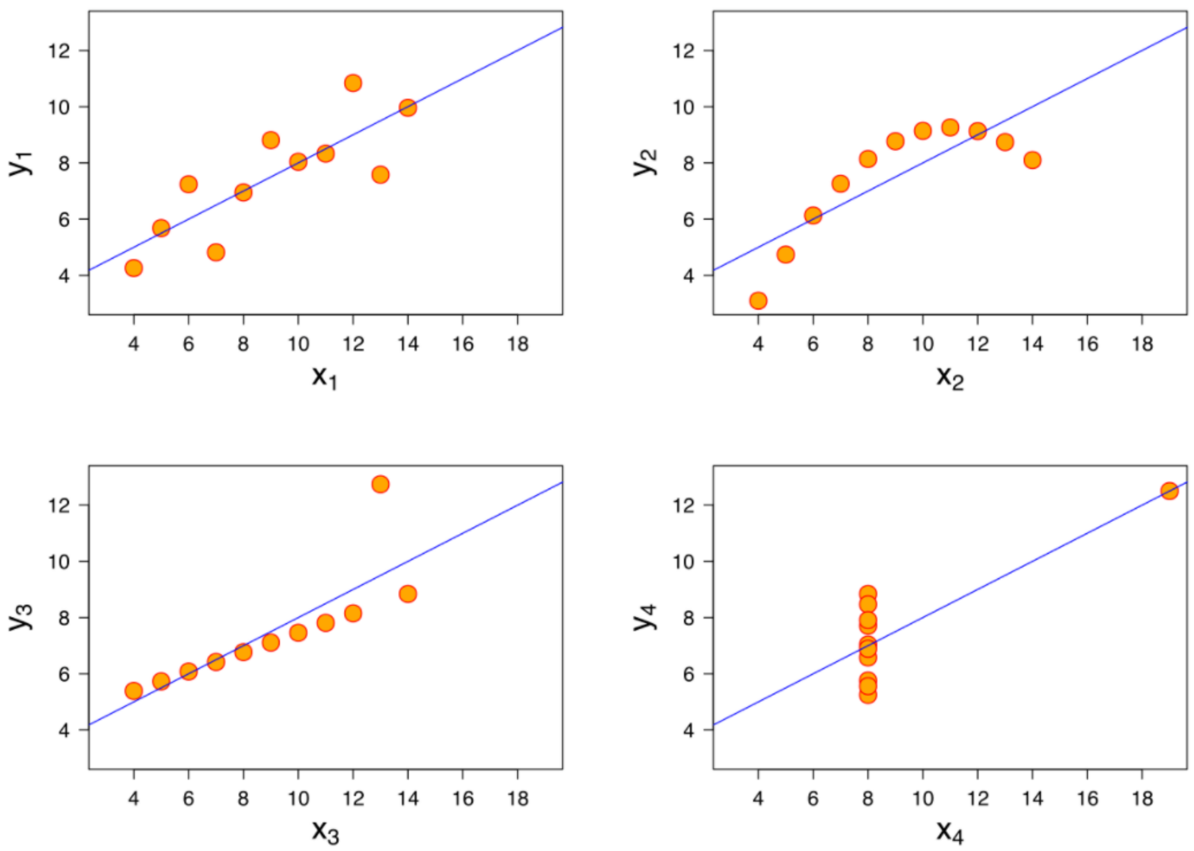


Figure 3. Anscombe's quartet

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.

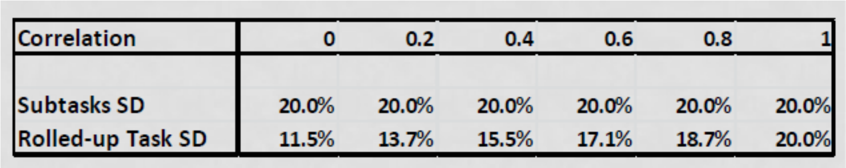
[7] https://en.wikipedia.org/wiki/Anscombe%27s_quartet

[8] Square of the standard deviation

[9] All WBS cost elements in a project or projects in a program.

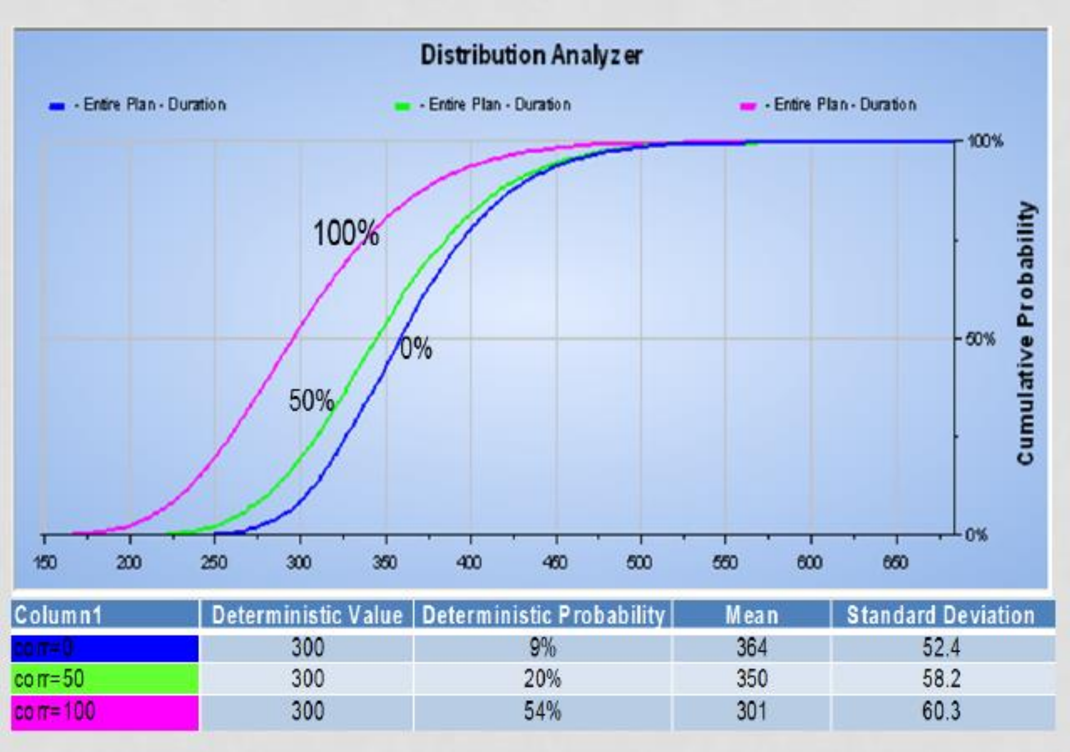
point estimates carry less confidence). Conversely, below expected values we are actually more confident in point estimates than without correlation. These relationships can all be seen in Figure 1.

The behaviors in schedule risk analysis depend on whether tasks/projects are executed serially, rolled up, or executed in parallel. Serial execution has a high correlation coefficient, which tends to tilt the S-curve. The variance of rolled-up tasks is dependent upon the variances of the individual subtasks and the degree of correlation. This can be seen in the following table for a specific example.



Correlation	0	0.2	0.4	0.6	0.8	1
Subtasks SD	20.0%	20.0%	20.0%	20.0%	20.0%	20.0%
Rolled-up Task SD	11.5%	13.7%	15.5%	17.1%	18.7%	20.0%

In the situation where various tasks or projects are executed in parallel, increasing correlation in parallel tasks can be seen. This reduces mean duration (more synchronized to dominant task/project), but increases the variance as seen in Figure 4.



Column1	Deterministic Value	Deterministic Probability	Mean	Standard Deviation
corr=0	300	9%	364	52.4
corr=50	300	20%	350	58.2
corr=100	300	54%	301	60.3

Figure 4

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
Sources of Correlation

Cost Correlation	**Event Correlation**
Project competence	Changed requirements
Project organization	Stakeholder influence/action
Project management and processes	Market conditions
Decision and approval processes	Economic trends
Estimation and risk processes	F/X rates
Wages, benefits, payroll taxes	Trade actions
Productivity	Regulatory changes/actions
Raw material costs	Low-frequency, high-impact events of scale
Design development	Archaeology finds
Means & methods	
State of technology — common new technology/materials	
Uncertainty factors/known unknowns	
Budgeting and contingency management strategy and approach	
Packaging and contracting strategy	
Schedule precedencies	
Shared/common assumptions	
Failures/delays at interfaces	
MAIMS – “Money Allocated Is Money Spent”	
Parkinson’s Law – work expands to fill the time allotted	
Optimism and other biases consistently applied	
Location factors	

Correlation between pairs of projects in a program can be calculated as the sum of the products of the standard deviation for each common risk divided by the product of the standard deviation of all risks in each project. Selection of appropriate distributions for each risk is important. As the number of projects in the program grows, calculation of overall program correlation becomes more challenging and more important. Ignoring dependency among cost components results in underestimation of total cost variance.

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 to distributions that are too tight and even weaker assessments on extreme values. Methods exist to reduce bias in assessing uncertain quantities, but are not embraced in the engineering and construction industry.[10]

Overly Simplistic Probabilistic Cost Analysis (PCA)

Numerous over-simplifications are often present in engineering and construction risk analysis. These include:

- Selection of distributions for various cost elements, in particular the use of triangular distributions, which is often blamed for unrealistically low and high estimates. The triangular distribution leads to underestimates and potentially significant underestimates because it has an upper bound.
- Omission of interrelationships among cost elements (correlation, discussed above).
- Process by which budgets are established and allocated (is MAIMS likely to be present?).
- Approach to management of contingencies (levels and process for release).
- Confidence levels in underlying assumptions (available equals good; tough to ascertain is ignored).
- How low-frequency, high-impact events are considered.

---

[10] Direct Fractile Assessment Method; Alpert and Raiffa, 1982

- Absence of adequate sensitivity analysis.

Summary

Ignoring correlation results in an underestimation of total cost variance, which can grow more significantly when dealing with a multi-project program or portfolio. Similarly, excluding correlation between variables in schedule estimation is significant. While there may be inadequate historical data to calculate correlation coefficients, ignoring a correlation’s presence is not an acceptable strategy. Sensitivity analyses for different levels of correlation may be performed. Guidance is provided in the recommendations that follow.

Correlation is especially important when projects are concurrent. Selection of the concurrency period, however, should consider overall program or contract durations (especially in task order type contracts) or sequential contracts with added coupling hidden in complexity.

Recommendations

- Treat correlation among cost elements and projects in a program realistically. Seldom are projects or programs “ideal,” with no correlation between cost elements.
- Correlations in large complex engineering and construction programs of 0.3 to 0.6 are not atypical.

> ***Rule of thumb**: 0.4 is now almost the de facto correlation coefficient for cost estimates, 0.5 for schedule.*

Consider sensitivity to variation in assumed correlation levels. Rule of thumb: 0.4 is now almost the de facto correlation coefficient for cost estimates, 0.5 for schedule.

- Consider use of either the open-ended three-parameter Lognormal or Weibull distribution for cost analysis.[11] There are other acceptable distributions, but emphasis should be placed on the values of input data (to minimize bias) and an open-ended distribution (triangular is not open-ended).
- Recognize that all statistical analysis represents a macroscopic view and that a complete assessment of risks and risk response also requires a microscopic view using other risk assessment methods such as decision trees. This microscopic view for low-probability, high-impact events helps address overconfidence and optimism biases. Ensure good quality information at the tails.
- Recognize the standard probable cost assessment assumes an “ideal” project and provides management with a false sense of confidence. Standard probable cost assessments are a source of major cost overruns, even when high contingencies are deemed to have been provided.
- Use an appropriate number of elements in the work breakdown structure in a project and the number of projects in a program. Subdividing costs into too many smaller pieces leads to a false sense of security and an erroneous outcome.
- In large programs, specifically incorporate the MAIMS principle in cost modeling and budget management practices.

---

[11] Kujawski, Edouard & Alvaro, Mariana & Edwards, William. (2004)

- 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.

S.A. Book, Estimating Probable System Cost, Crosslink 2 (1) (Winter 2000/2001), pp. 12-21.

EPA/630/R-97/001, Guiding Principles for Monte Carlo Analysis; March 1997.

Ince, P. and Buongiono, J., “Multivariate stochastic simulation with subjective multivariate normal distribution.” Symposium on System Analysis in Forest Resources, Charleston, SC, United States, 1991.

F. Kuo, Everything You Want to Know About Correlation But Were Afraid to Ask; NASA, available online.

Kujawski, Edouard & Alvaro, Mariana & Edwards, William. Incorporating Psychological Influences in Probabilistic Cost Analysis. *Systems Engineering*. 7. 195 - 216. 10.1002/sys., 20004.

About the Author

Bob Prieto was elected to the National Academy of Construction in 2011. He is a senior executive who is effective in shaping and executing business strategy and a recognized leader within the infrastructure, engineering, and construction industries.