



THE RELATIONSHIP BETWEEN COST GROWTH AND SCHEDULE GROWTH

*Richard L. Coleman, Jessica R. Summerville,
and Megan E. Dameron*

Considerable speculation has been put forward regarding schedule changes and cost growth. Surprising new findings from Selected Acquisition Reports will show the connection between cost and schedule as well as cost growth and schedule changes. The probability distribution function of the schedule size growth data will be presented. The paper will demonstrate that knowledge of program cost does not allow for prediction of program duration, neither is the reverse possible. A longer version of this paper has been received enthusiastically at the Department of Defense Cost Analysis Symposium, the Society for Cost Estimating and Analysis, the Program Management Institute Symposium, the Air Force Institute of Technology, and the Aeronautical Systems Center/Industry Cost and Schedule workshop.

Considerable speculation has been put forward regarding the effect of schedule changes and cost growth, as well as the relationship between program length and cost. This article will discuss a study that puts forward actual data, obtained from Selected Acquisition Reports, to show the connection between cost and schedule as well as cost growth and schedule changes. The study discusses both the effects of schedule compression and extension. It also corrects for the effects of program size and presents the probability distribution function of the schedule growth data.

In addition, this article will show that historical data indicate that knowledge of program cost does not allow for prediction of program duration, neither is the reverse possible. On the other hand, knowledge of changes in baseline schedule durations will allow for improved estimates of cost growth.

This study is important to arm Program Managers with a reasonable expectation for program length and schedule growth, and to help Acquisition officials know whether there is a tendency for programs to be late, early, or on time. We hope that a reasonable expectation

will aid all parties in setting reasonable expectations. Knowing the results show most are late, we hope less stigma is attached to finishing late, as it is the rule rather than the exception.

THE DATA

We analyzed data from the RAND Selected Acquisition Report Cost Growth Database¹ (1993 release) with both the following characteristics: 1) Programs with an Engineering and Manufacturing Development phase (E&MD) only, because growth is different² for those with and without Program Development and Risk Reduction (PDRR) phases, and 2) programs with schedule data in the requisite fields. There were 59 points satisfying both criteria, which were the points used.

THE ANALYSIS

The basic descriptive statistics, derived from the data points used, are shown in Table 1.

The data are clearly skewed positively, i.e., the majority of programs stretch (this is seen, e.g., from the fact that the median exceeds the mean.) The average program is thus 29 percent late (as seen by noting that the mean is 1.29)! Almost two-thirds (64.4 percent) of all programs are late! This was to be expected, knowing the conventional wisdom that stretches are not at all uncommon, but surprising in the sheer scale of it.

It was unusual, however, that so many programs finished *precisely* on time. From a probabilistic point of view, this is unlikely, because if the distribution is continuous, the probability of finishing exactly on time is zero — so perhaps one or two programs out of 59 might have an exact on-time finish, rather than twelve. This presents an analytic problem for modeling. We solve this problem by asserting that finishing on time is a discrete probability, and other outcomes are continuous. This odd hybrid distribution would be unlikely in events with no special emphasis on being on time,

Table 1. Schedule Growth Statistics

Mean	1.29
Standard Deviation	0.54
CV	42%
75th %-ile	1.46
61st %-ile	1.29
50th %-ile	1.11
25th %-ile	1.00
Programs that finished early	9/59 = 15.3%
Programs that finished on time	12/59 = 20.3%
Programs that finished late	38/59 = 64.4%

and suggests that finishing on time may be so important that programs are tempted to declare completion and rely upon follow-on work to fix errors in the “shipped product.”

The authors would like to make the analytically unsupported but anecdotally reasonable hypothesis that an undue emphasis is being placed on finishing on time to the detriment of quality. As an alternative for program managers, a modest slip might be preferable to a cosmetic on-time finish.

Whatever the case for finishing on time, we felt it appropriate to pursue what the actual model for lateness might be (a descriptive model) rather than what might happen if “cosmetic on-timeliness” were not to occur (a prescriptive model). It is important to note that this latter prescriptive model is purely speculative, but we intended to see if there was a persuasive model embedded in the data, and

we did find one, but the result is outside the scope of this article. To find more about this embedded model, we refer the reader to our longer presentation³ and for many other points of interest. The descriptive model is shown in Figure 1.

We found that an Extreme Value (Gumbel⁴) distribution fit best (using a Kolmogov-Smirnov-type test⁵ as the selection criterion, and a significance level of 0.95), and we felt that this was sensible, since in effect a schedule is constrained by the last of a number of events.⁶ The parameters of the fitted extreme value distribution were $\mu = 1.12$ and $\beta = 0.28$. We injected a discrete probability of finishing on time, 0.203, since no continuous model could ever predict so many on-time finishes.

The next issue addressed was whether there was any correlation between cost and schedule length, and between cost

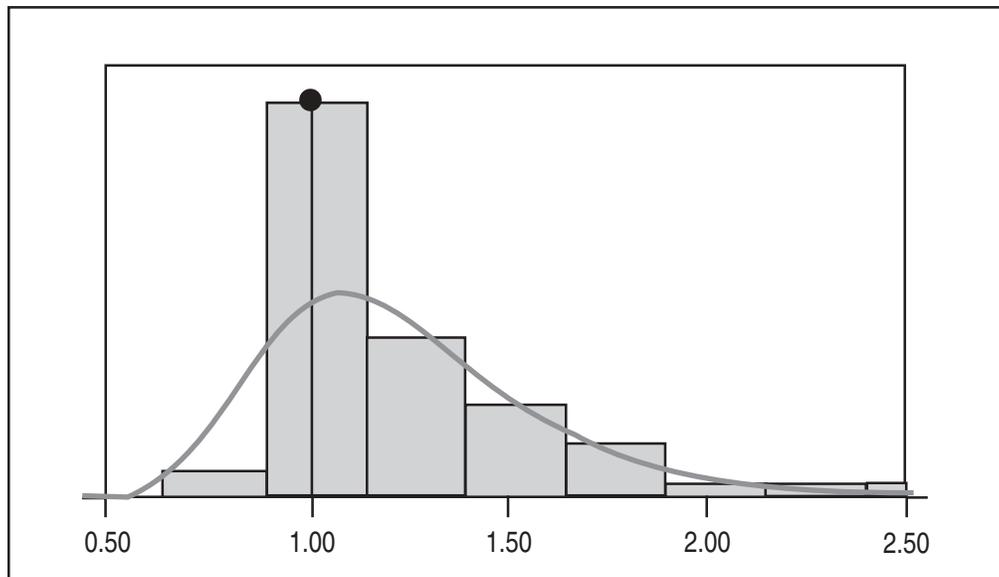


Figure 1. The Probability Distribution of Program Completion

growth and schedule growth. We tested cost versus schedule length for linear correlation using the Pearson parametric test. We were unable to reject the null hypothesis at a significance level of 0.95, and thus could not find correlation. This conclusion tells us that knowing the length of the program is of no added value in predicting the cost, and knowing the cost is of no added value in predicting the length. This should not surprise the reader, since no one would like to hazard a guess at the cost of a program from its length, nor vice versa. In any event, surprising or not, it is nonetheless true for this data.

For cost growth and schedule growth, we tested using a parametric test (the Pearson parametric test for correlation), and for general correlation using a non-parametric test⁷ (the Cox Stewart test for trend). We were again unable to find correlation. We conclude then, from these two tests, that there is no tendency for longer programs to cost more, nor is there a tendency for programs that stretched more to have more cost growth than programs that stretched less.

We could go on and on about the number of tests we conducted to challenge this finding, since we knew it would be astonishing to some, but the general reader is no doubt relieved we will not do so. Suffice it to say that we did our best to undo this conclusion fair-and-square, and we had no stake in either finding, but it would not come undone. We hope that it suffices to say that these results were somewhat surprising, since conventional wisdom holds that there is

a direct linear connection between schedule growth and cost growth. This conventional wisdom was not reflected in the data, and so we challenge other analysts to do studies in this area, or review our findings in depth.

We were unable to find a functional connection between amount of schedule growth and of cost growth, but we were interested to see if there was a difference if we simply grouped by regime (programs that compress, those that finish on time, and those that stretch). We observed large differences in the cost growth of these three regimes, as listed in the descriptive statistics above. We thus concluded that, although there is a great deal of difference among the regimes, within the regimes there is no discoverable difference with more or less growth, as evidenced by the tests for correlation, especially by the non-parametric test.

CONCLUSION

We found that program schedules do tend to grow, as we expected. However, there was no correlation found between cost and schedule length. In addition, we did not find the anticipated connection between cost growth and schedule growth, except for a general tendency for a difference by regimes (programs that compress, those that finish on time, and those that stretch). We challenge other analysts to take up this problem, so that conventional wisdom, or these findings, may be corrected as necessary.

The Relationship Between Cost Growth and Schedule Growth



Richard L. Coleman graduated from the U.S. Naval Academy and has an M.S. in operations research from the Naval Postgraduate School. He retired as a Captain, U.S. Navy, and works as a cost and risk analyst at TASC, supporting customers such as the DD(X) Design Agent Team, Littoral Combat Ship, Naval Air Systems Command, and the intelligence community. Coleman is National vice president of the Society of Cost Estimating and Analysis.

(E-mail address: rcoleman@northropgrumman.com)



Jessica R. Summerville has a B.A. in mathematics from the University of Michigan and an M.S. in operations research from the College of William and Mary. She works as a cost and risk analyst at TASC, supporting customers such as the Missile Defense Agency, DD(X) Design Agent Team, Littoral Combat Ship, Naval Air Systems Command, and the intelligence community.

(E-mail address: jsummerville@northropgrumman.com)



Megan E. Dameron graduated summa cum laude from the College of William and Mary with a B.S. in mathematics and economics. She now works as a cost and risk analyst at TASC, supporting customers such as the DD(X) Design Agent Team, Naval Air Systems Command, National Imagery and Mapping Agency (NIMA), National Reconnaissance Office Cost Group (NCG), the intelligence community, and the Society of Cost Estimating and Analysis (SCEA).

(E-mail address: mdameron@northropgrumman.com)

ENDNOTES

1. This database is available from RAND Publications or from the authors.
2. *NAVAIR Cost Growth Study*, International Society of Parametric Analysts (ISPA)/Society of Cost Estimating and Analysis (SCEA) 2001, 34th DoD Cost Analysis Symposium (DoDCAS) and ISPA/SCEA 2001, R. L. Coleman, M. E. Dameron, C. L. Pullen, J. R. Summerville, and D. M. Snead. (This paper is available from the authors.)
3. *The Relationship between Cost Growth and Schedule Growth*, 35th DoDCAS, SCEA 2002, Integrated Program Management Conference (IPMC) 2002, R. L. Coleman, and J. R. Summerville. (This paper is available from the authors.)
4. The Gumbel distribution is often found in “extreme events” such as the largest earthquake, the heaviest rainfall, the latest finisher in a race, etc.
5. Tests for distribution are either parametric, requiring assumptions, or they are non-parametric. Non-parametric tests are usually either Chi-Squared Goodness of Fit tests, or Komolgorov-Smirnov-type (K-S) tests. In general, the K-S tests are less vulnerable to being compromised (rigged) by the choice of bins and are less powerful. The reader can pursue this further in most standard statistics books, but the K-S test is less often found, and must be uncovered in rarer non-parametric textbooks. In any event, the K-S test is available in the usual statistical packages and spreadsheet add-ins.
6. Given that finishing a program is akin to waiting for the last event to finish, it is appealing that the Gumbel, or extreme value distribution, is the best fit to the data. This is an argument of “attractiveness” or reasonableness only; the case is actually made by the statistical test.
7. The parametric test is more often used, but the parametric test requires an assumption of normality, which is often challenged. Given that schedule finish data tested as being distributed Gumbel, we felt it inconsistent to assume normality for convenience. The non-parametric test requires fewer distributional assumptions, and is thus less compromised. We ran both since we felt the more common test, Pearson, would be expected, and its absence would raise eyebrows. In the event, the conclusion was the same, which makes the case all the stronger.

The Relationship Between Cost Growth and Schedule Growth
