# Cost Risk Analysis Based on Perception of the Engineering Process

Edwin B. Dean and Darrell A. Wood
NASA Langley Research Center
Mail Stop 444
Hampton VA 23665
(804) 865 4894

Arlene A. Moore and Edward H. Bogart
PRC Kentron International Inc.
3221 W. Armistead Ave.
Hampton VA 23665

Introduction:

In most cost estimating applications at the NASA Langley Research Center (LaRC), it is desirable to present predicted cost as a range of possible costs rather than a single predicted cost. A cost risk analysis generates a range of cost for a project and assigns a probability level to each cost value in the range. Constructing a cost risk curve requires a good estimate of the expected cost of a project . It must also include a good estimate of expected variance of the cost.

Many cost risk analyses are based upon an expert's knowledge of the cost of similar projects in the past. In a common scenario, a manager or engineer, asked to estimate the cost of a project in his area of expertise, will gather historical cost data from a similar completed project. The cost of the completed project is adjusted using the perceived technical and economic differences between the two projects. This allows errors from at least three sources. The historical cost data may be in error by some unknown amount. The managers' evaluation of the new project and its similarity to the old project may be in error. The factors used to adjust the cost of the old project may not correctly reflect the differences.

Some risk analyses are based on untested hypotheses about the form of the statistical distribution that underlies the distribution of possible cost. The usual problem is not just to come up with an estimate of the cost of a project, but to predict the range of values into which the cost may fall and with what level of confidence the prediction is made. Risk analysis techniques that assume the shape of the underlying cost distribution and derive the risk curve from a single estimate plus and minus some amount usually fail to take into account the actual magnitude of the uncertainty in cost due to technical factors in the project itself.

This paper addresses a cost risk method that is based on parametric estimates of the technical factors involved in the project being costed. The engineering process parameters are elicited from the engineer/expert on the project and are based on that expert's technical knowledge. These are converted by a parametric cost model into a cost estimate. The method discussed makes no assumptions about the distribution underlying the distribution of possible costs, and is not tied to the analysis of previous projects, except through the expert calibrations performed by the parametric cost analyst.

The  Cost  Risk  Methodology:

A detailed approximation of the probability distribution underlying the cost of a project can be obtained by viewing the engineering process as a tree structure with each node in the tree being an engineering decision which adjusts the final project cost.  Each possible limb of the tree culminates in a final project cost and can be described by a specific parameter vector.  A Monte Carlo process over this complete parameter space will provide and excellent approximation of the cost risk distribution.  However this requires a PRICE run or parametric cost estimate for each parameter vector.  Few organizations including LaRC can afford this precision.  Thus at LaRC an alternate approach was chosen.

A possible project cost of the above cost tree can be generated by adding one of either the low, the perceived, or the high values  from each item on the WBS.  If the cost for each item is selected  randomly, the sum will lie between the minimum and maximum possible project cost. Repeating this process a number of times produces an approximation of the distribution of possible  project  costs.

The critical step in this method is generating a low,  a perceived, and a high cost for each element on the Work Breakdown Structure (WBS)  for the project.  Each WBS element  cost estimate is generated by querying the persons most familiar with that WBS  item to obtain qualified estimates of the best-case, worst-case and perceived value for each of the engineering process input parameters for that item.  In the examples presented in this paper, the RCA PRICE parametric cost model is used.  Any other parametric cost model can be used in exactly the same way under this methodology.

For each element of the WBS, low, high and perceived values are elicited for each box of the appropriate PRICE Input Data Worksheet (IDW).  For a hardware part or assembly for instance, a PRICE H Basic Modes sheet would be used.  For example, the expert would be asked to give a best engineering estimate of the weight of the item to be produced based on item design, materials, manufacturing methods, and so on.  Estimates of the lowest likely weight and highest likely weight based on the same engineering factors would be elicited.  Three values for each factor on the IDW would be elicited in the same way.  These values are used to generate a low IDW, high IDW, and a most likely IDW, each containing either the low, high or perceived value from each box.  When these input sheets are run through PRICE-H, they produce three parametric estimates for the cost of that item. The low and high estimates define an unbiased estimate of the cost range for the item.   The third cost is a perceived estimate of the item cost based on the expert's knowledge of the engineering factors involved in the production of the item.

A cumulative distribution plot of the cost sums represents the risk curve for project costs with the median at the 50% point.  Low and high limits of cost are derived respectively by selecting the sum of the low WBS item costs and the sum of the high WBS item costs.

Example   1.

For the purposes of example, a data file was created by selecting ten data items from several real data files.  Each data row contains a low,  a perceived, and a high estimate for a single WBS element of a project.  The data, shown in Figure 1 becomes the input file for the program "Cost_Risk" developed at LaRC.

| Low | Perceived | High |
|-----|-----------|------|
| 3 7 | 6 6 | 6 6 |
| 4 6 | 5 0 | 5 6 |
| 4 0 | 4 4 | 6 9 |
| 1 1 | 1 3 | 1 7 |
| 5 3 | 6 5 | 8 0 |
| 5 4 | 6 6 | 7 3 |
| 4 0 | 6 8 | 9 7 |
| 2 2 | 2 7 | 3 3 |
| 1 5 | 1 7 | 1 9 |
| 2 9 | 4 2 | 6 0 |

Figure 1.  Data Input Tableau.

When the program "Cost_Risk" is run one data item is selected at random from the three data items on each line.  The data items are summed over the WBS set, generating a possible project cost.  This process is repeated n times.  The interval between the smallest and the largest cost sum is divided into k intervals or bins and each of the n sums is tallied in the appropriate bin. The distribution of sums within bins is plotted in Figure 2.  This plot can be interpreted as the cost density distribution for the project.
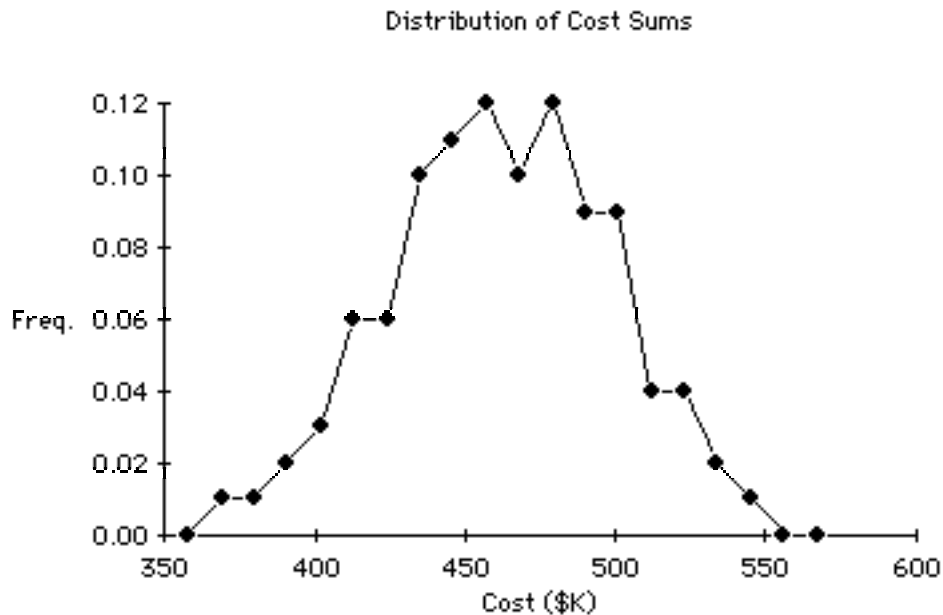


Figure 2.  Project Cost Probability or Risk Density Function.

A cumulative plot of the cost sums across cost bins produces the Cost Risk curve shown in Figure 3. The values on the Y axis, here marked "Risk" is the probability that the project cost will be at or below the coresponding cost on the X axis. The cost corresponding to the .5 risk point is an estimate of risk balance point for which there is a 50/50 chance that the project can be completed for that cost. Another value of importance is the estimate of the distribution mean which gives the expected delivered cost. If the distribution is normal the median will equal the mean. This has not been found to true based upon estimates generated at LaRC. In fact, the expected delivered cost has generally been found to be considerably greater than the mean. This would indicate that the choice of the 50/50 cost would generally result in a cost overrun. Further, the ratio between the two compares favorably with NASA cost overrun experience.
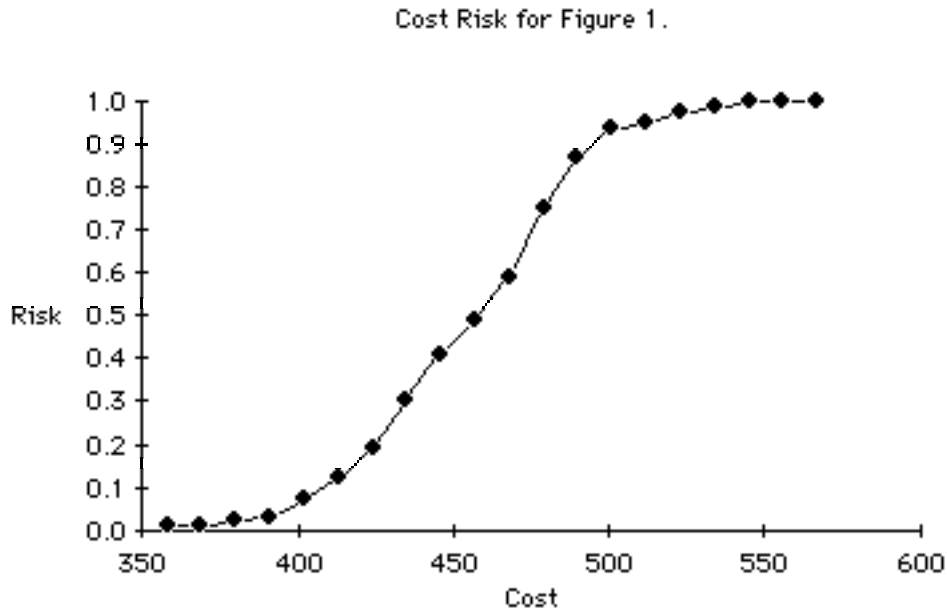
Cost Risk for Figure 1.



Figure 3. Project Cost Probability or Risk Distribution Function.

Example 2.

The second example demonstrates the reduction of the range of uncertainty in the Cost Risk curve when the final cost of some items in a system are known. The cost estimates shown in Figure 4. are for the same hypothetical project shown in Figure 1. but in this case, the delivered cost of three of the items (lines 1, 4, and 6) are known. This is typical of the later stages of a project where some some parts of the project are complete and the actual cost is known. It also is typical of proposals where the purchase cost for some items are known, such as fixed price quotes, or projects for which GFE is supplied.

| Low | Perceived | High |
|-----|-----------|------|
| 5 4 | 5 4 | 5 4 |
| 4 6 | 5 0 | 5 6 |
| 4 0 | 4 4 | 6 9 |
| 1 5 | 1 5 | 1 5 |
| 5 3 | 6 5 | 8 0 |
| 6 6 | 6 6 | 6 6 |
| 4 0 | 6 8 | 9 7 |
| 2 2 | 2 7 | 3 3 |
| 1 5 | 1 7 | 1 9 |
| 2 9 | 4 2 | 6 0 |

Figure 4.  Reduced Variance Data Input Tableau

When this data is run through the Cost_Risk program, using the same run parameters as in Example 1, the cost risk density and cost  risk distribution curves shown in Figures  5 and 6 is produced.  The shape of these curves remain about the same but the range of costs on the X Axis is reduced.  This is due to the reduction of uncertainty in the project cost.
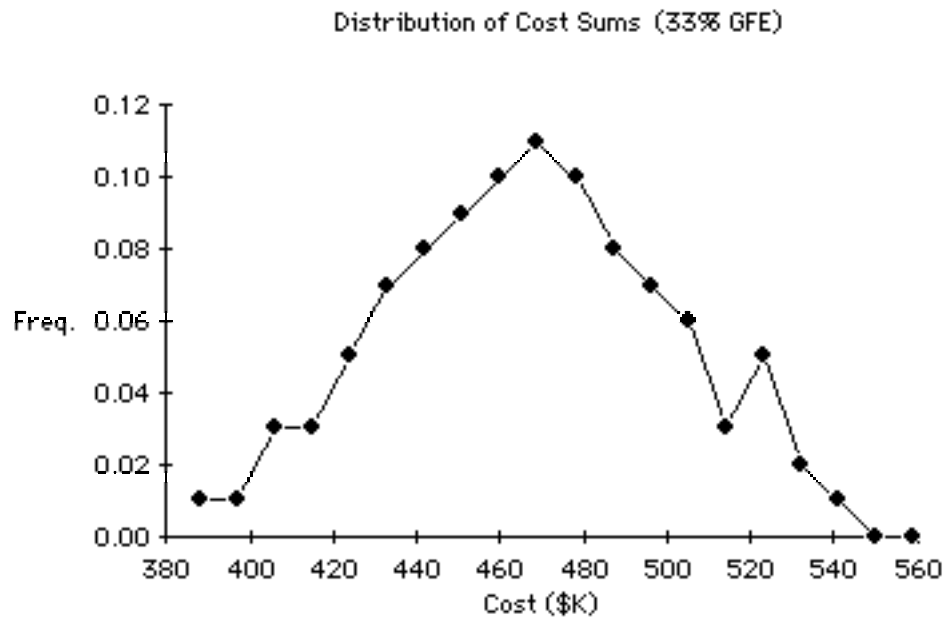


Distribution of Cost Sums (33% GFE)

Figure 5.  Reduced Variance Project Cost Probability or Risk Density Function.
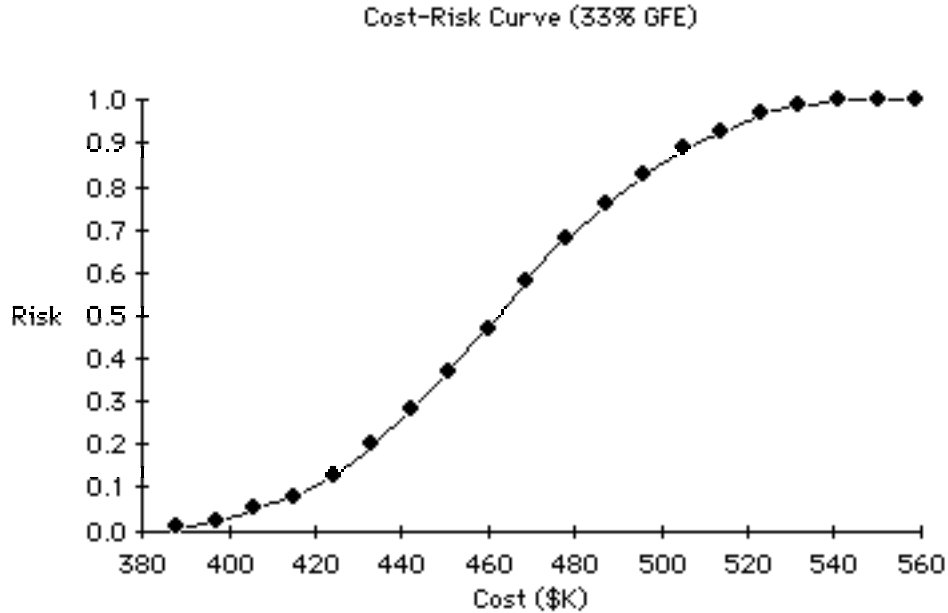
Cost-Risk Curve (33% GFE)



Figure 6.  Reduced Variance Project Cost Probability or Risk Distribution Function.

## Conclusion

The initial challenge which led to the development of this technique was to devise a method which used as much engineering definition and as few statistical assumptions as possible in order to increase the credibility of cost estimates in the engineering and management environment at LaRC.  The second challenge was to be able to afford the method.  The method discussed has succeeded in  both challenges.

In practice it has been easy to obtain the additional sets of high and low parameters.  Since these parameters are derived from the expertise of the engineers and managers, they feel the input to the parametric models is credible.  When they disagree with the resulting costs, we review  the engineering process parameters with them and change the input parameter set only if refined engineering definition justifies a change.  Expectedly, we often disagree with the managers and engineers on cost, but we do agree with them that we have described the engineering process expectations as best we can.  This provides a substantial degree of credibility.

In practice the method takes only a small additional amount of time and two additional parametric model runs.  The additional credibility is well worth the cost.

The resulting cost analysis benefits greatly from the additional perspective of variance.  Cost estimates at LaRC are performed at the very early stages of a project where one would expect the estimating variance to be greatest.  Variances between the high and low cost estimates experienced range from a factor of 2  to a  factor of 10.  These factors seem to be quite consistent with the degree of engineering definition available.  The variances also are considerably greater than the variances shown,  for example, in the output of the PRICE program.  PRICE variances correspond to the variance of cost based upon the data to which the PRICE equations are fit, not to the variance of the engineering process for a particular project.  The relatively large ratio between the variances obtained with this method and the PRICE variances is consistent with expectations.  As explained to a manager recently, this indicates that the estimating precision of the PRICE model is far better than our ability to provide engineering definition at early stages.  That adds another measure of credibility to the estimate.

The only actual data point to date resulted in an expected delivered cost which was considerably greater than the 50/50 cost projection,  but in the end was approximately ten percent less than actual.  Incidentally, this project estimate had a very wide variance because of new processes being used.  Both engineering process definition and cost estimating calibration uncertainties were included in input data.