Scientific Letter

MEOSAR Cost Escalation Risk

Background

On 10 February 2014, Senior Cost Analyst, Peter Iburg, CMA, of Directorate of Costing Services, (D Cost S), requested DRDC CORA's assistance with determining the cost escalation risk for the Medium Earth Orbit Search and Rescue (MEOSAR) project. Following a project meeting on 12 February 2014, Mr. Iburg provided us with subject matter expert(s) (SME) cost escalation data on 21 February 2014 in the form of three point estimates (low-expected-high) with confidence levels on 62 MEOSAR budget line items. Mr. Iburg asked that we determine the median and 85th percentile on the total cost of the project, the cost of the definition phase, and the cost of the implementation phase, using SME inputs.

Statement of Results

The SME dataset contains two estimate types – base and conservative – reflecting the SME's sensitivity analysis of their cost estimates. Senior project management's input provides the majority opinion in the base estimates while DCostS's revisions generate the conservative estimates.

Our estimates yield:

- **Base assumptions:**
  - The median total project cost is 180 million dollars, and the 85% quantile gives a total project cost of at most 231 million dollars.
  - The median cost of the definition phase is 39 million dollars, and the 85% quantile gives a definition phase cost of at most 41 million dollars.
  - The median cost of the implementation phase is 144 million dollars, and the 85% quantile gives an implementation phase cost of at most 190 million dollars.

- **Conservative assumptions:**
  - The median total project cost is 192 million dollars, and the 85% quantile gives a total project cost of at most 248 million dollars.
  - The median cost of the definition phase is estimated at 40 million dollars, and the 85% quantile gives a definition phase cost of at most 43 million dollars.
  - The median cost of the implementation phase is estimated at 155 million dollars, and the 85% quantile gives an implementation phase cost of at most 207 million dollars.

We display the distribution functions for each set of assumptions and project phase in figures B.1 – B.6

Methods

Mathematically, the MEOSAR cost problem reduces to finding the distribution function for the sum of the given 62 random cost variables. (In probability and statistics the distribution of the sum of random variables is called a convolution. See the annex for more details.) The SME provide two datasets comprising of base and conservative assumptions respectively, each composed of a definition and implementation phase. Each dataset contains budget line item estimates in three forms:

- three distinct dollar value estimates representing low-expected-high costs with an attendant confidence range, stated as a probability;

- two distinct dollar value estimates and a confidence range, stated as a probability, that one of the two outcomes will materialize in a binary sense. (For example, the budget line item representing the Prototype Contract has two data points, the first of which represents the contract's cost if the contract proceeds as anticipated, and a second point which represents the cost in the unlikely event of a contract renegotiation.); and

- one value stating the exact dollar cost of the budget line item (funds already committed or otherwise expended).

In the first form, we approximate the probability distribution with a lognormal random variable, and in the second form, we use a discrete random variable with the two outcomes and associated probabilities given by the SME dataset. We treat all random variables as independently, but not identically, distributed.[1]

We treat each budget line item as an independent random variable except for the cost of the known line items. Instead of relying on Monte Carlo methods, we directly convolve the underlying probability distributions.

Since we have no prior information on the probability distributions from the SME, we use the lognormal distribution as an approximation in the three-distinct-point SME estimate class. In general, price processes are often well approximated by a geometric random walk [1], which implies a lognormal distribution for price fluctuations. The lognormal distribution is characterized by two pieces of information which we can extract from the three point SME estimates. The convolution of independent lognormal distributions is not known in closed form and thus we use the Fenton-Wilkinson approximation [2] for convolving the lognormal approximated budget line items. We compute the exact convolution for the discrete distributions provided by the SME.

Direct convolution for the independent budget line items has advantages over Monte Carlo techniques, especially in the presence of heavy tailed distributions (such as the lognormal). In budget escalation risk analysis, decision-makers desire an understanding of upper quantiles – or the tail risk of the project's budget. In such circumstances Monte Carlo methods require large sample sizes to ensure proper convergence in the tails and the problem compounds as the number of independent heavy tailed distributions increases. Analytic and semi-analytic methods not only provide greater insight than exclusive reliance on Monte Carlo techniques but also remove the need for heavy computational (and time consuming) performance.

Discussion of Results

We encourage caution by decision-makers in directly applying our results. Our analysis stands on the quality of SME estimates of the probability and severity of unfavourable events. In any SME opinion extraction process – especially those which focus on probability estimates – we must use great care to account for human factors in reporting. To have full confidence in SME opinion, we require carefully calibrated questionnaires, independent sessions to limit group-think contamination, and input from social scientists trained in locating reporting bias. To further improve the analysis, we would need historical data on projects of similar complexity. Such data would not only help us understand possible correlations among budget line items, but it would also help anchor SME opinion. Given these caveats, we urge caution in interpreting the results of our analysis as a final statement on project cost escalation risk.

Conclusion

We find that under the base and conservative assumptions for the MEOSAR project, the median total project estimated cost is 180 million dollars and 192 million dollars respectively. The respective 85% quantiles for the total project costs lie at 231 million dollars and 248 million dollars. The base and conservative assumptions imply a median cost of the respective definition and implementation phase of 39 million dollars and 144 million dollars, and 40 million dollars and 155 million dollars. The 85% quantiles respectively sit at 41 million dollars and 190 million dollars, and 43 million dollars and 207 million dollars.

---

[1] We do not have any data that would allow us to construct a correlation matrix; we therefore assume independence throughout. Correlation effects can weaken or strength our estimates, depending on the direction.

Canada

We generate all results using exact convolution for discrete random variables and semi-analytic approximations for continuous random variables. Finally, we use all SME data as reported, making no correction for potential bias or other human factor issues in the SME dataset, and we assume independence among all the budget line items.

References

[1] Cochrane, J. (2005), Asset Pricing, Princeton University Press.

[2] Fenton, L. F. (1960), The sum of lognormal probability distributions in scatter transmission systems, *IRE Transactions on Communication Systems*, CS-8, 57–67.

[3] Mehta, N. B., Wu, J., Molisch, A. F., and Zhang, J. (2007), Approximating a sum of random variables with a lognormal, *IEEE Transactions on Wireless Communications*, 6(7).

Prepared by: **David W. Maybury.**

Attachments

Annex A: Convolutions of Probability Distribution Functions and Lognormal Fitting

Annex B: Cost Distribution Figures

This page intentionally left blank.

Annex A: Convolutions of Probability Distribution Functions and Lognormal Fitting

To determine the MEOSAR project's budget probability distribution function, we require the distribution function of the sum of all the random variables provided by the SME, namely,

$$\mathbb{P}(X_1 + X_2 + \ldots + X_N \leq x), \tag{A.1}$$

where each X_i denotes an independent, though not necessarily identical, random variable. Consider two independent discrete random variables, X and Y, and their sum, $Z = X + Y$. The probability that $Z = x + y$ is given by

$$
\begin{aligned}
\mathbb{P}(Z = z \equiv x + y) &= \cup \{\text{probability } (X = x) \text{ and probability } (Y = z - x)\} \\
&= \sum_i \mathbb{P}(X = x_i)\, \mathbb{P}(Y = z - x_i)
\end{aligned} \tag{A.2}
$$

where the sum is taken over all possible values of the random variable X. In continuous form (for positive random variables), the density function of $Z = X + Y$, $h(z)$, links the density function for X, $f(x)$, with the density function for Y, $g(y)$, with an analogous expression,

$$h(z) = \int_0^z f(x) g(z - y)\, dx. \tag{A.3}$$

While in principle we can use the convolution formula iteratively to generate the density function for an arbitrary sum of independent random variables, the lognormal distribution, which we use in this letter, presents a challenge in that eq.(A.3) does not have a closed form representation. Numerical integration techniques with the lognormal distribution suffers from the heavy tail behaviour of the distribution, which either requires a large number of function evaluation points or the careful application of asymptotic methods. The Fenton-Wilkinson method gives a simple approximate solution to the convolution of independent lognormal random variables by exactly matching the first two moments of the sum to another lognormal distribution. That is, the Fenton-Wilkinson method approximates the convolution of independent lognormal random variables with another lognormal random variable. Moment matching ensures that the approximating solution accurately models the right tail of the distribution. For more details on the Fenton-Wilkinson method and alternative approximation schemes (including cases of correlation) see [3].

Two parameters characterize the lognormal distribution, $LN(\mu, \sigma)$. To determine μ and σ for each budget line item with three distinct points, (a, c, b), given by the SME, we assume that the central value (c) represents the SME estimate of the expectation and that the range $([a, b])$ represents the SME's confidence level, (L). Thus, we must solve the system of equations for each relevant budget line item,

$$
\begin{aligned}
c &= \exp\left(\mu + \frac{\sigma^2}{2}\right) \\
L &= \frac{1}{2}\operatorname{erf}\left(\frac{\ln(b) - \mu}{\sqrt{2}\sigma}\right) - \frac{1}{2}\operatorname{erf}\left(\frac{\ln(a) - \mu}{\sqrt{2}\sigma}\right),
\end{aligned} \tag{A.4}
$$

where $\operatorname{erf}(\cdot)$ denotes the error function.

This page intentionally left blank.

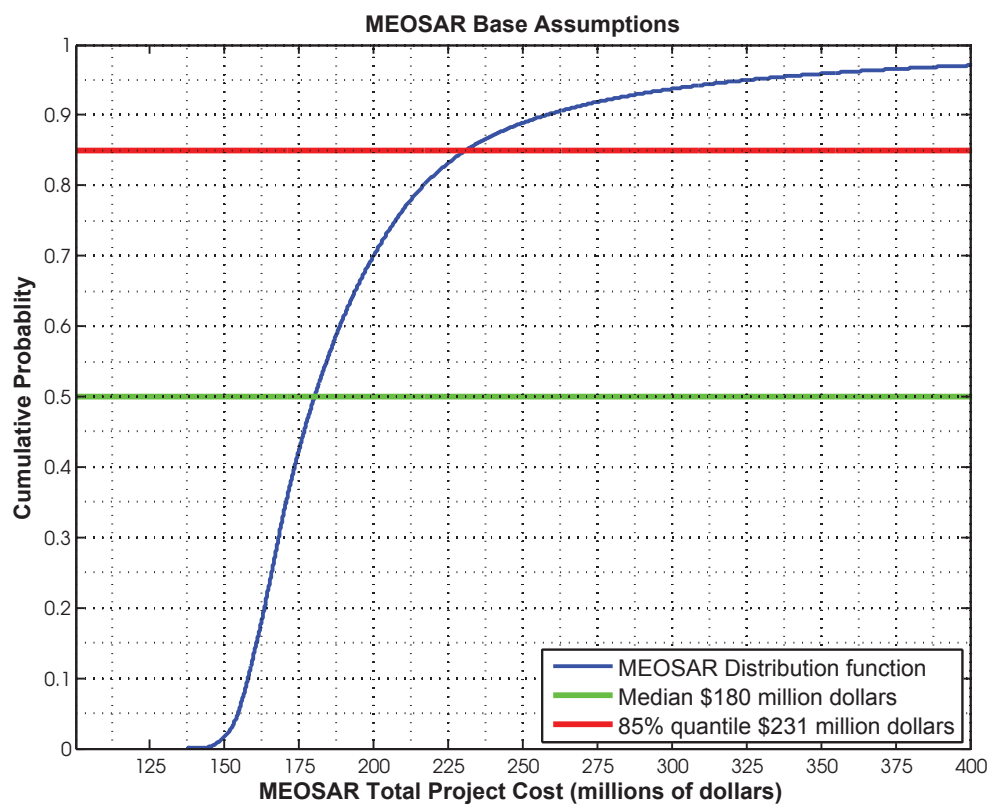# Annex B: Cost Distribution Figures

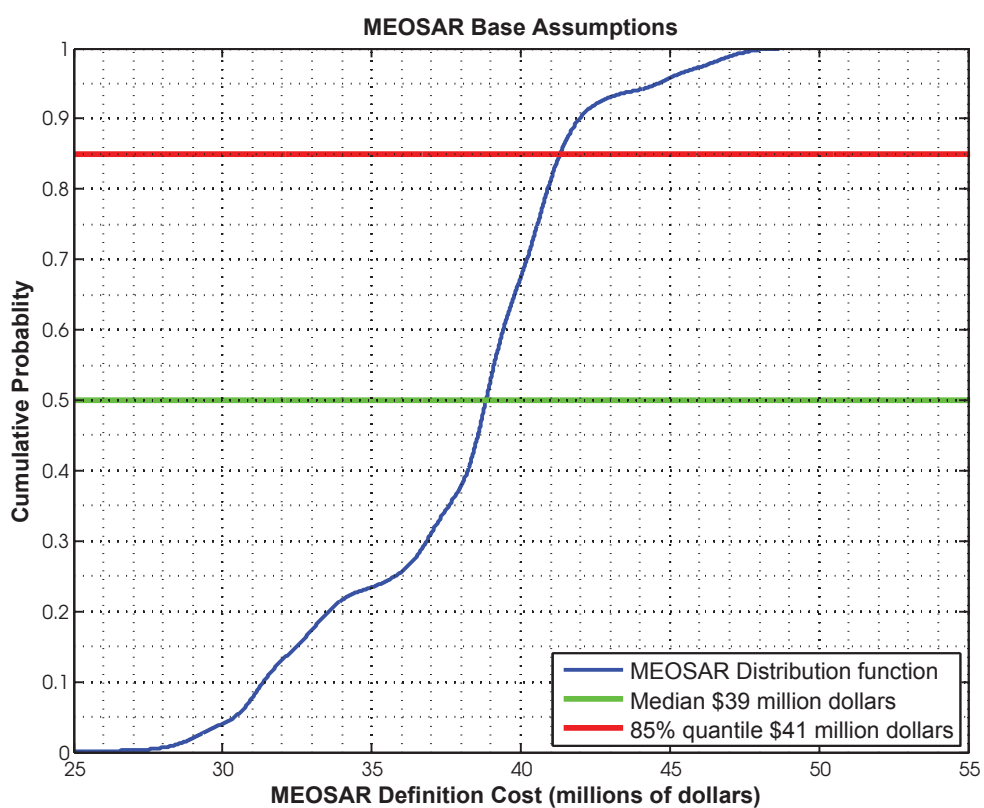**Figure B.1:** *MEOSAR total project cost distribution function under base assumptions.*

Figure B.2: MEOSAR cost distribution function for definition phase under base assumptions.
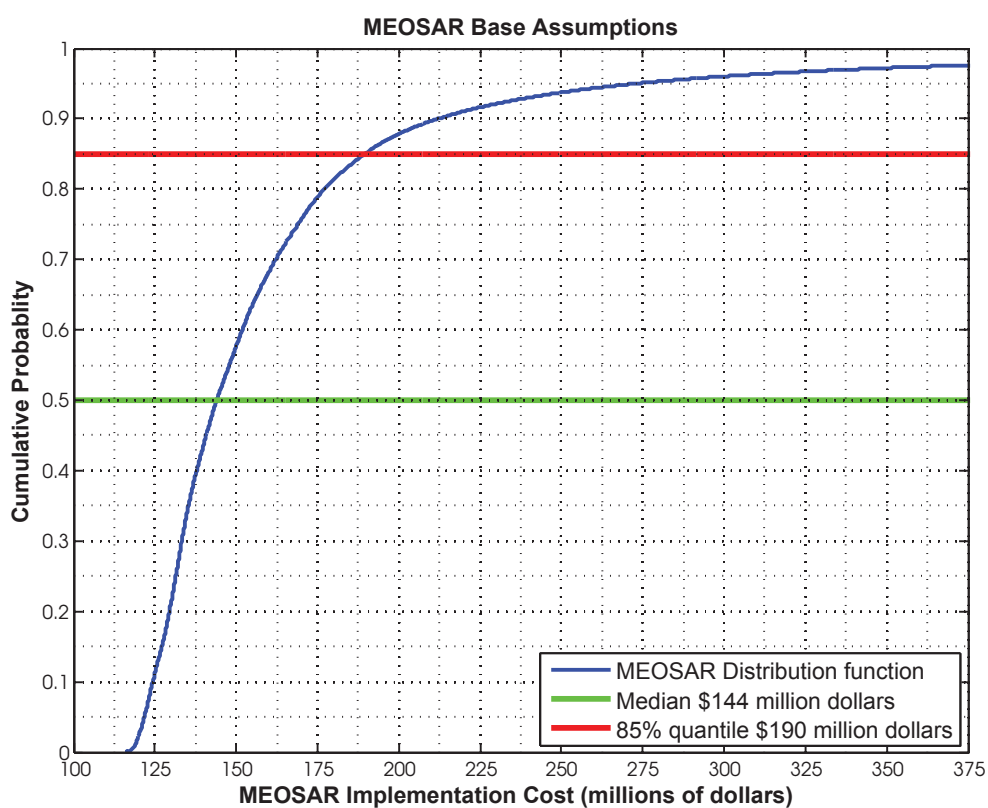
Figure B.3: *MEOSAR cost distribution function for implementation phase under base assumptions.*
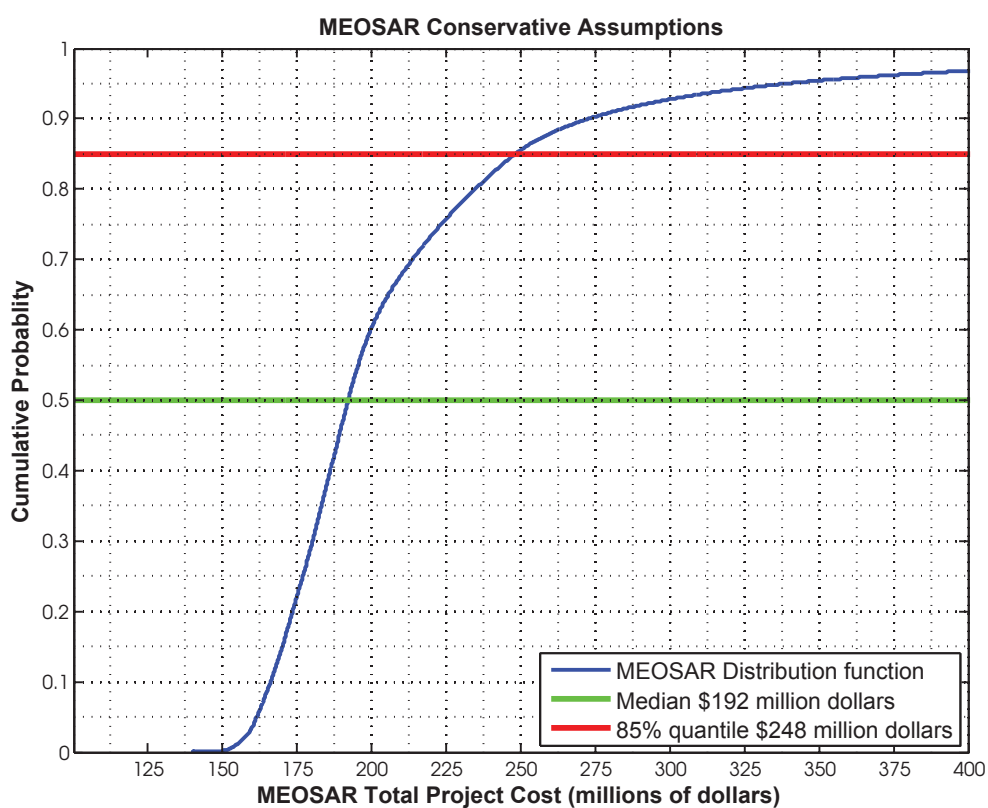
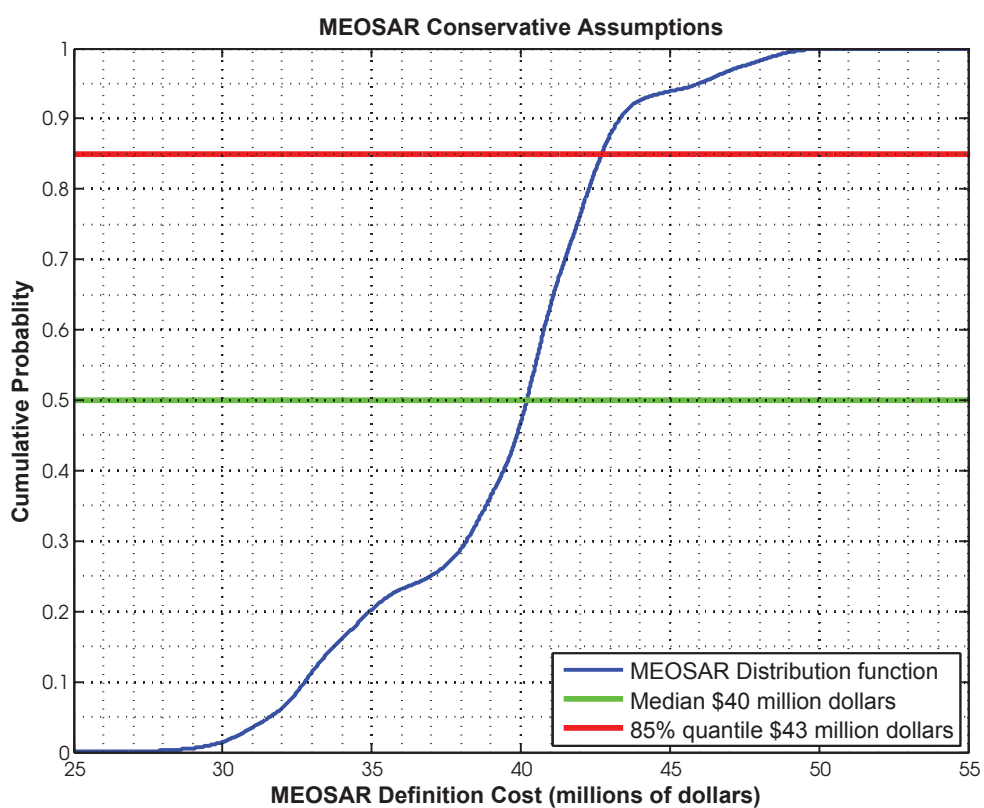**Figure B.4:** *MEOSAR total project cost distribution function under conservative assumptions.*

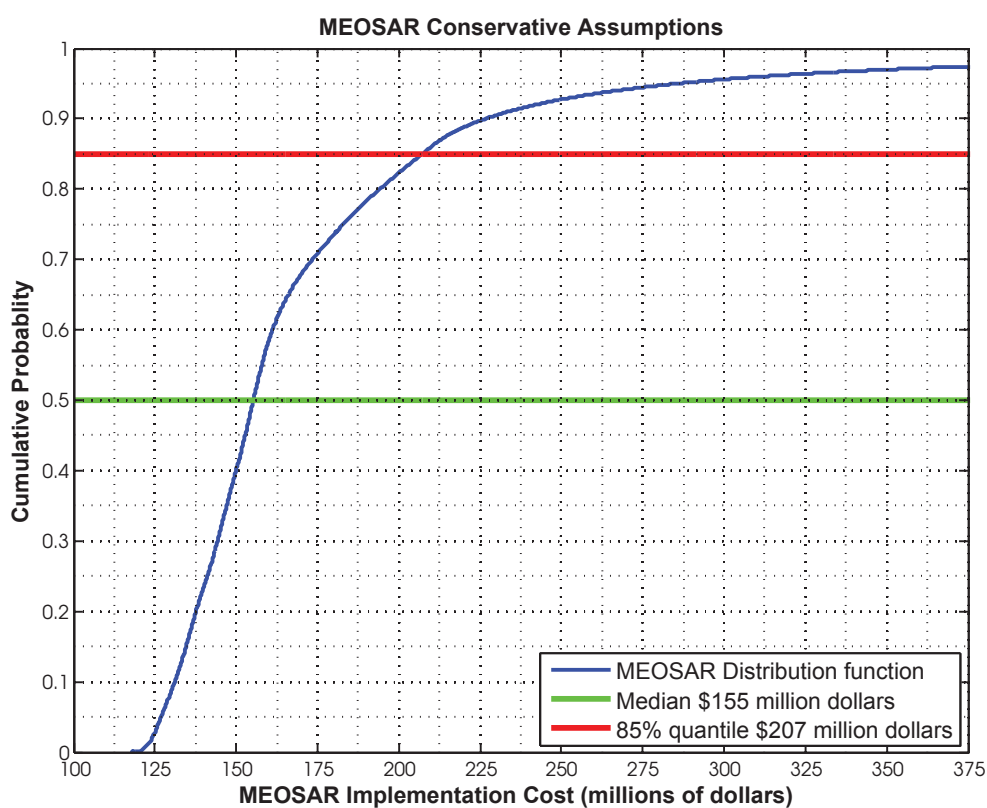**Figure B.5:** *MEOSAR cost distribution function for definition phase under conservative assumptions.*

MEOSAR Conservative Assumptions

Figure B.6: MEOSAR cost distribution function for implementation phase under conservative assumptions.