

AD-A161 920

①

Cost Uncertainty Assessment Methodology:
A Critical Overview

by

K. T. Wallenius
Clemson University

Presented at the
19th Annual Department of Defense
Cost Analysis Symposium
Xerox Training Center
Leesburg, Virginia
September 17 - 20 1985

DTIC
ELECTE
S DEC 4 1985 D
B

The Cost Analysis Symposium was sponsored
by: OSD (PA&E) Resource Analysis, Cost
Analysis Division, Pentagon, Washington,
DC 20301

DTIC FILE COPY

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

11 21-85 041

COST UNCERTAINTY ASSESSMENT METHODOLOGY:
A CRITICAL OVERVIEW

K. T. Wallenius
Clemson University

OVERVIEW

An appraisal of the quality of methodology and software for performing cost[@] uncertainty analysis will be given. It will be argued that

- the standard practice of developing a "most likely cost" and then generating an uncertainty distribution associated with that figure is a consequence of historical precedent and places the cart before the horse. Decision theoretic considerations require that uncertainty assessment precede total cost estimation. This fact underscores the primacy of cost uncertainty analysis as a tool in the overall cost analysis process.
- methodology for encoding subjective probability distributions are out of date, being, for the most part, modifications of the PERT techniques of the sixties. These techniques, ignoring the implications of the vast behavioral science literature on uncertainty elicitation and calibration, sacrifice human cognitive capability on the altar of mathematical convenience. This may explain, at least in part, the general reluctance of cost analysts to undertake a serious assessment of cost uncertainty.

[@] The term "cost" is used in a generic way this paper. Its meaning will depend on the purpose of the analysis. (e.g. first unit cost, nonrecurring cost, LCC, etc.)



Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By PER FORM 50	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

- methodology for processing elemental uncertainty assessments into uncertainty distributions over higher order structures (e.g. subsystem CER's or total system cost) often ignores important sources of variability. This may explain why uncertainty distributions over total cost are usually too tight. CER compendia which omit covariance data (see reference [4], for example) are of little value in quantifying uncertainty.
- extant computer software designed to implement cost uncertainty analysis methodology is often poorly written and, in some cases, may lead to conclusions which are inconsistent with the inputs. While "garbage in, garbage out" is axiomatic, users of computer software who expend the time and effort to generate reliable inputs have the right to expect output fidelity.

THE PRIMACY OF COST UNCERTAINTY ANALYSIS

Providing an answer to the question "How much will it cost?" is job one for the cost analyst and has a long historical precedence over the (more!) important question "How sure are you of that figure?". This is probably due to the facts that it's easier to come up with a single number than a probability distribution, and, budgets of the past were based on most likely costs. Many of today's directives require a quantification of cost uncertainty and some budgeting processes allow component-program cost uncertainties to enter into the total budget calculations (e.g. TRACE). The form of these directives is revealing: usually, they refer to the "uncertainty associated with a most likely cost" implying the cost estimate comes first. The modern approach [18] to decision making dictates that uncertainty considerations come first. It is one of the two prerequisite inputs needed for developing an optimal estimate of cost (the other being a consideration of the relative seriousness of overestimates and underestimates.)

The problems of generating an uncertainty

distribution over total cost are discussed below but, for now, suppose design specifications, engineering information, contractor performance requirements, historical data, subjective assessments, etc. have been correctly processed resulting in such an uncertainty distribution. As such, this distribution summarizes the totality of available information. Besides answering important probabilistic questions, this distribution is the cornerstone for generating estimates of total cost. Standard texts on decision making, for example [18], develop optimal estimates which account for the relative seriousness of the difference between the estimate and the actual cost. If $\hat{C}$ and $\tilde{C}$ denote the estimate and the cost, respectively, and $e = \tilde{C} - \hat{C}$ is the error ($e > 0$ is an underestimate or a cost overrun), it is customary to code the error seriousness in terms of a regret function $R(\hat{C}, \tilde{C})$ having the properties that R is nonnegative, is zero when $e = 0$, and is nondecreasing as the error moves away from zero in either direction. Some typical examples are shown in figure 1.

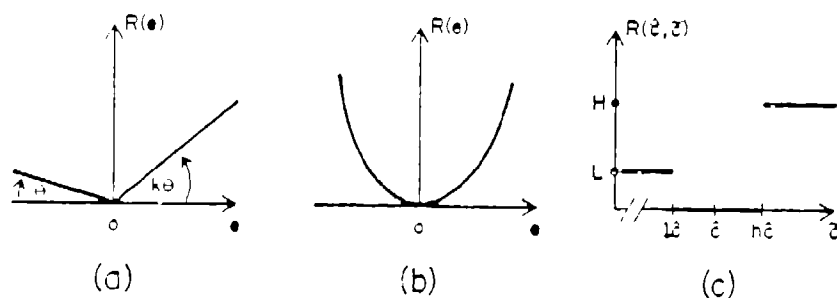


FIGURE 1: Some Regret Functions

In figure 1a, the seriousness of an error is proportional to the absolute value of the error with underestimates k times as serious as overestimates. In this case, the optimal² cost

² An optimal estimate for the purposes of this paper is one that minimizes expected regret. There are other definitions of optimal. The cost uncertainty distribution and the regret function are needed for most reasonable optimality criteria

estimate is the $k/(k+1)$ th percentile of the uncertainty distribution. If $k = 3$ for example, the 75th percentile would be optimal. In figure 1b, the seriousness of an error is proportional to the square of the error. In this case, the optimal estimate is the mean of the uncertainty distribution. In figure 1c, the regret is zero if the estimate is within tolerance, and very large otherwise. A regret function similar to this might be appropriate when errors outside a given range are not tolerable (e.g. $h = 1.15$ might be used to protect against a Nunn Amendment cancellation). In this case, optimal estimates depend on the values of l , L , h , and H but are near the mode of typical unimodal uncertainty distributions.

Most cost uncertainty directives and technical papers use the term "most likely cost" referring, presumably, to some sort of modal value. This implicitly says something universal about the relative seriousness of estimation errors which, in fact, will differ between organizations (perhaps even between different levels within the same organization). The point here is that the optimal estimate logically depends on its intended use. The cost uncertainty distribution is of fundamental importance in generating the optimal estimate and is the job of the cost analyst. Generating the cost estimate is the prerogative of the cognizant command and should reflect the relative seriousness of errors. This functional separation allows each organizational unit to focus on tasks best suited to their expertise and responsibility.

Thus, the cost uncertainty distribution should not be an afterthought, tagging along behind the point estimate. It is the cornerstone upon which that estimate is based. So, what is the state-of-the-art in cost uncertainty assessment methodology? What tools (training

programs, books, computer software, documentation, etc.) are available to assist cost analysts to perform this pivotal task? It is not feasible to survey the entire literature in a paper of this type nor is the entire literature available for such a survey (viz. proprietary documents and software). Comments will be organized around two component tasks: (1) quantifying uncertainty at the most minute level of disaggregation, (2) processing of these "particle" assessments to produce an uncertainty distribution over total cost.

1. QUANTIFYING UNCERTAINTY

Disaggregation of complex systems into cost components allows the analyst to focus on subproblems with relatively few variables thus simplifying individual assessments. (Of course, this requires aggregation of these component assessments to generate a total cost uncertainty distribution but that's more a mathematical problem than a cognitive one.) Suppose the quantity of interest is dollars-per-athe of debugged software for a midcourse guidance computer. What guidelines and decision aids are available to help the cost analyst quantify uncertainty about this component cost? If he has read the 1984 state-of-the-art survey in the Journal of Cost Analysis [15], he might try to fit a 4-parameter beta distribution by specifying 3 quantities: a mode (m), a least possible cost (l), and a highest possible cost (h). A fourth specification pertaining to the spread (s) of the distribution is also required to mathematically nail down a unique member of the beta family. The analyst may alleviate the cognitive pain of specifying s by using some ad hoc nominal value (e.g. the PERT formula for s) but, by so doing, a decision about s is nevertheless made. Three of the four specifications, (l , h , and s) are quite difficult for me to think about and m is also a little unnatural. It's hard to think about a number h such that there is positive probability that the cost can get arbitrarily close to h and,

at the same time, feel secure in betting your life against a jellybean that the cost will not exceed h!

Since cost element uncertainties must be aggregated to generate a total cost uncertainty distribution, it is necessary to express them mathematically. This usually takes the form of some parametric family of probability density functions such as the beta, gamma, or Weibull families. Over 90% of the cost uncertainty literature advocates use of the beta family for this purpose. See, for example references [1] [2], [6], [9], [16], [19], and [23]. The state-of-the-art survey paper [15] focuses exclusively on fitting beta distributions. The beta family is shape-rich and fairly adequate in representing unimodal uncertainty situations. Its popularity in cost analysis is due, in part, to the early 60's success of PERT analysis (based on beta representations of completion time uncertainty). Many of the 112 references cited in the bibliography of [15] deal with PERT. In all the beta fitting techniques cited, the user is required to specify l and h plus either m or the mean (μ). There is an obvious mathematical convenience in this since l and h are two of the four parameters needed. m (or μ) determines a linear relation between the two shape parameters so that just about any fourth condition makes solving for the shape parameters relatively easy. In looking at the papers in which this methodology was developed, mathematical convenience was the dominate, if not the only, consideration in designing the subjective input requirements.

But people have to do it. There is an enormous literature describing how people make judgments under uncertainty, how biases and heuristics influence judgment, and about the degree to which people (ordinary folks and experts) are calibrated in their judgments about uncertain events. Reference [8] is a collection of thirty papers with over 650 references. Several state-of-the-art papers are "must" reading for cost uncertainty methodology developers and users. Among them is "Calibration of

Probabilities: The State-of-the-Art to 1980" [11] and "Encoding Subjective Probabilities: A Psychological and Psychometric Review" [22]. A particularly relevant finding is that people are not well calibrated relative to rare events: in most studies there is strong evidence that decision makers systematically underestimate tail probabilities. Thus it may happen that actual costs exceed corresponding elicited 95th percentiles one time in five rather than one time in twenty. It is not too surprising that calibration is less precise in the tails of a distribution than in the fleshier parts since we haven't had as much experience with rare events as with nominal events. Alternately, even if the decision maker is viewed as an ideal data processor, estimates of large and small percentiles from data have larger variances than estimates of more central percentiles in general. If this empirical observation holds true with cost analysts, then requiring them to specify h and l of a beta distribution (the most extreme percentiles) would tend to result in a systematic understatement of uncertainty. It would be unfair to criticize the developers of PERT or the early cost researchers who modified PERT methodology to deal with cost uncertainty. Requiring (l, m, h) assessments was a matter of mathematical expedience and the behavioral research did not exist then. The fact that the 112 references in the cost uncertainty state-of-the-art paper [15] and the 650 references in [8] have no common entry can be viewed as an exciting opportunity to improve cost uncertainty methodology by exploiting a relevant body of related research.

There are no theoretical reasons why the beta family ought to enjoy such a position of prominence in the cost uncertainty liturgy. Nor is there any particular reason for avoiding its use. Because of its flexibility in shape and history of application, it will probably continue to be used in cost uncertainty analysis. That being the case, it may be worthwhile to dispel some common misconceptions:

- a) It takes 4 conditions to specify a beta density.
- b) The "asymmetry quotient" $AQ = (h-m)/(m-l)$ is good measure of skewness.
- c) Since cost is bounded above and below, the beta family with its finite range is better suited to represent cost uncertainty than are distributions with infinite range.

Assertion a) is correct, of course, but only in a strict sense. Suppose, instead of (l,m,h) plus a 4th condition, some less extreme assessments are made, say the three quartiles are elicited (25th, 50th, and 75th percentiles). These are more natural quantities to think about [6]: for example, they break the range of costs into 4 equally likely regions. Standard elicitation techniques require relatively simple scaling tasks with even, two-to-one, or three-to-one odds. These three inputs have the effect of tying the cumulative distribution function (CDF) down at three points called "knots." While there are infinitely many beta distributions whose CDF's pass through the knots, the functional form of the beta family has the effect of making these different distributions practically indistinguishable. Who, but God, is so finely calibrated in his cognitive abilities as to be able to distinguish between the four beta CDF's (shown dotted) and the gamma CDF (shown solid) in figure 2? Corresponding densities are superimposed.

It is hard to think of a practical situation in which the total cost uncertainty distribution would be effected by variation in component distributions as negligible as those shown. Thus, the four parameter beta distribution is, for most practical purposes, a three parameter family - as long as one chooses the correct three parameters, here percentiles, to assess. Similar results hold for other choices of knots. By contrast, fixing (l,m,h) and varying s over unimodal beta distributions produces shapes ranging from something nearly flat (uniform) to a virtual spike at m in the limit. The limiting case for the beta with three fixed knots is a well behaved gamma.

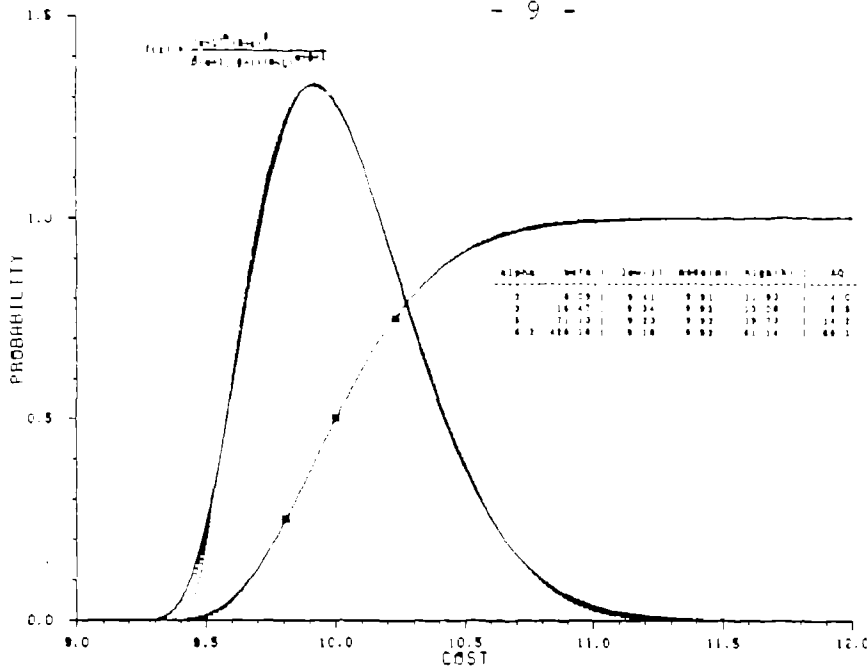


FIGURE 2: Knotted Beta and Gamma CDF'S:
Distributions with common quartiles

Figure 2 also sheds light on misconceptions b) and c). The four nearly indistinguishable beta distributions have asymmetry quotients $AQ = (h-m)/(m-l)$ ranging from 4 to 69. There is no relation between this measure of "asymmetry" [15] and any statistical definition or cognitive concept of asymmetry. It appears that the asymmetry quotient, like most of the beta-fitting methodology, was invented for mathematical convenience. Its cognitive value in helping cost analysts quantify uncertainty is nil. Concerning c), none of the distributions shown in figure 2 have any meaningful amount of probability to the right of 12 (less than .00001). Four of the five distributions (the betas) have finite upper bounds ranging from about twelve to over sixty. The fifth (gamma) has an infinite right tail. For all practical purposes, they are identical. The upper truncation point h of the beta density is just a parameter which obviously has very little to do with where the probability mass is located

(another argument for not eliciting h). So why should the cost analyst concern himself over inconsequential issues as confusing as finite versus infinite range?

Modifying the uncertainty encoding techniques of PERT to fit the cost analysis scenario was a reasonable first step in 1965 [19]. The current state-of-the-art in elemental uncertainty quantification as described in [15] suggests that progress in the last 19 years has been scant. With the advent of inexpensive and powerful personal computers, computational complexity is no longer a limitation in designing interactive software for probability encoding. Cognitive considerations rather than mathematical expedience should dominate the search for new methodology. The challenges are exciting. Some new initiatives are discussed in [21].

2. PROCESSING OF COMPONENT UNCERTAINTIES

Developing an uncertainty distribution over total cost or over some intermediate level of disaggregation such as WBS cost elements is a matter of applying the probability calculus to elemental uncertainty assessments - a purely mathematical task. As with subjective probability elicitation methodology and software, there is a large gap between what is available to the cost analyst and what could be available. A parallel thrust to simultaneously upgrade both elicitation and processing capabilities is needed: state-of-the-art processing tools won't improve misspecified uncertainty distributions (garbage in, garbage out) nor are finely honed subjective assessments worth the trouble if they are improperly processed. Two of the more important processing tasks are a) uncertainty associated with parametric costing, and b) uncertainty associated with a convolution (the summing of uncertain quantities such as WBS cost elements).

Parametric cost estimation is one of the most widely used techniques for developing point estimates of cost. General overviews of the topic can be found in [10], [13], or [17]. Empirical

model building techniques (usually regression analysis) are applied to historical cost data on systems judged more or less analogous to the object system resulting in a cost estimating relation (CER). Point estimates and prediction intervals for the cost of the object system can be obtained using standard regression techniques. But these quantities do not characterize the uncertainty distribution for the object system's cost. For concreteness, consider the the first unit cost CER for launch operations and orbital support (LOOS) for apogee kick motor equipped satellites, one of 37 CER's listed in [4].

$$fy79K\$ = 27.44 + .2992x \text{ (wet wt lbs)}$$

Stats: $n=12$, $r\text{-squared}=.80$, $P=38.81$

std error (SE) = 146.33 $485.9 \leq \text{lbs} \leq 3694.3$

The CER estimate of LOOS cost for a 1000 pound unit would be \$326,640. Assuming the model is appropriate, there are three sources of uncertainty to recon with: 1) statistical uncertainty: randomness associated with the estimates of the model parameters, 2) predictive uncertainty: random deviation of the cost of the object system from its predicted value, and 3) input uncertainty: uncertainty about the ultimate value of wet weight which is likely to differ from its expected or nominal value (1000 lbs). This last source of uncertainty is sometimes called "strategic uncertainty" [3]. Failure to include any of the three may result in substantial understatement of LOOS cost uncertainty. Only input uncertainty is treated in [13], whereas only input and predictive uncertainties are included in [24]. A schematic diagram for a formal cost uncertainty analysis is shown in figure 3. All of these sources of uncertainty are recognized in [4] but not dealt with explicitly. Ad hoc

rules replace the type of formal analysis shown in figure 3. The output is in the form of (l, m, h)

$$\text{MODEL : COST} = a + b \cdot (\text{LBS}) + \epsilon, \epsilon \sim N(0, \sigma^2)$$

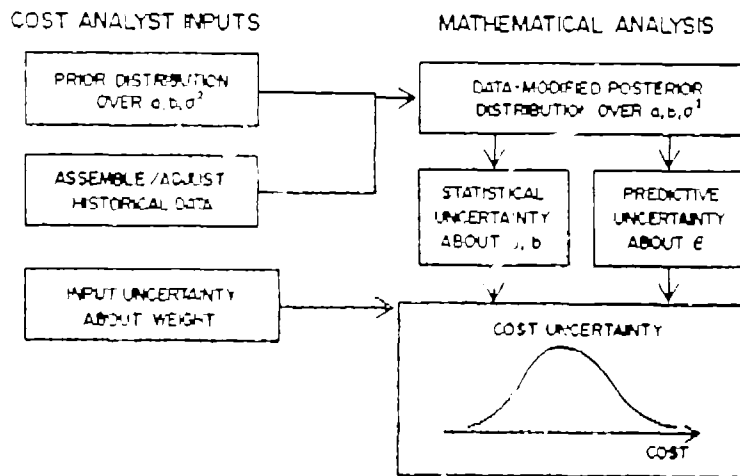


FIGURE 3: CER Uncertainty Processing

estimates, not an uncertainty distribution. In addition, since l and h are computed as $m \pm \text{SE}$ (i.e. not dependent on the value of the input variable), their meaning is not clear. If one wished to attempt a formal uncertainty analysis, the form in which the data has been summarized (see above stats) is incomplete since the sufficient statistics (sample means and covariances) are not given.

Direct assessment of a prior joint probability distribution over the model parameters is cognitively quite difficult. There are new indirect methods [7] supported by interactive computer codes which may prove useful if adapted to the cost analysis scenario. This approach is discussed in [21].

The final step in generating an uncertainty distribution over total cost consists of convolving the uncertainty distributions of the

cost elements into which total cost was disaggregated (probably the WBS, although it may be advantageous to use some other disaggregation scheme). It is tempting to take the bookkeeper approach and simply sum the individual estimates to generate a "best" estimate of total cost. This approach, while much criticized and devoid of mathematical justification [12], is nevertheless widely practiced. The CER compendium [4] is a case in point. In fact, most likely subsystem costs are irrelevant to the convolution process. What counts is their uncertainty distributions. Furthermore, summing most likely costs will result in systematically underestimating total cost if, as argued by some [14], only right-skewed distributions make sense in quantifying cost uncertainty. While there are undoubtedly many reasons why cost estimates tend to be low, this one is easy to avoid by simply using correct methodology.

The two most widely used methods of convolving component uncertainty distributions are Monte Carlo simulation (used in [2], [6], [9], and [16]) and a cumulant summation technique called the method of moments (used in [14], [19], [23], [24], and [25]). It is assumed that the reader is familiar with the ideas behind these two techniques. The output of the Monte Carlo method is sample of observations from the total cost uncertainty distribution. The comforting feature of this approach is that the empirical distribution function of the sample data converges to the "true" total cost uncertainty distribution as the number of replications becomes large. By "true" we mean the distribution that is the exact convolution of the component uncertainty distributions. With today's inexpensive and powerful microcomputers, the expense of large samples is no longer a limitation of this approach.

The output of the cumulant summation method is the first four moments of the "true" total cost uncertainty distribution from which that distribution must be estimated. Estimating a distribution from four moments requires an

assumption about the family to which the "true" distribution belongs - another potential source of error. But unless one chooses such a family, about all that can be said about the total cost uncertainty distribution is to provide probability bounds [26] based on these moments. These bounds are usually fairly broad and hence of limited practical value. The degree to which four, (in this case [5], even five!) moments fail to nail down a distribution is shown in figure 4.

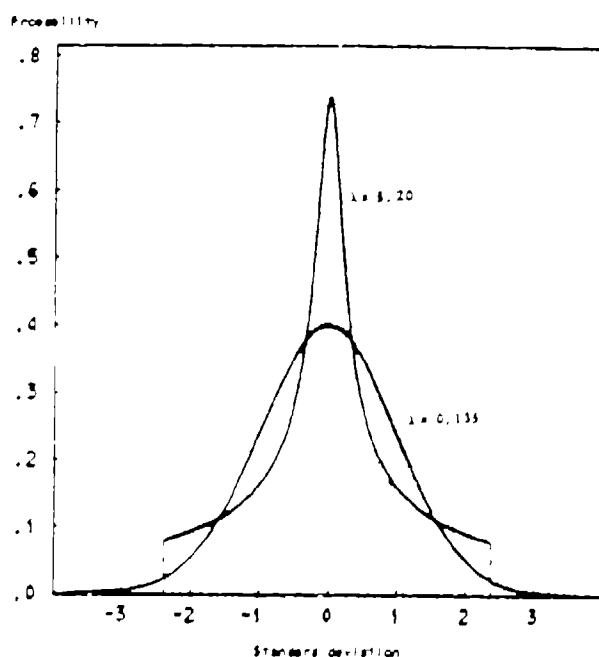


FIGURE 4: Tukey Lambda Densities. Moments up to and including order five are identical.

While cumulant summation methods can be expected to yield reasonable results in most nominal cost analyses, it is difficult to tell how well they will work in any specific application. In addition, moment methods are often nonrobust, i.e. sensitive to small changes in the data. Other comparisons are given in [12]. For these reasons, as methodologies, the Monte Carlo method is better

suited to cost uncertainty analysis.

No matter how good or bad a methodology is, poor implementation (algorithms, computer codes, ease of use, etc) will reduce its utility. Two of the better known techniques used in DOD, one employing a Monte Carlo approach [6] and one using the cumulant summation method [14] are discussed in [12] and [20]. Experience with actual data is reported. Neither approach was wholly satisfactory. For example, variance calculations in [6] resulted in negative numbers. Using the formulas given in [14], it is possible to fit a Weibull distribution to its own moments and come out with a different Weibull distribution as shown in figure 5.

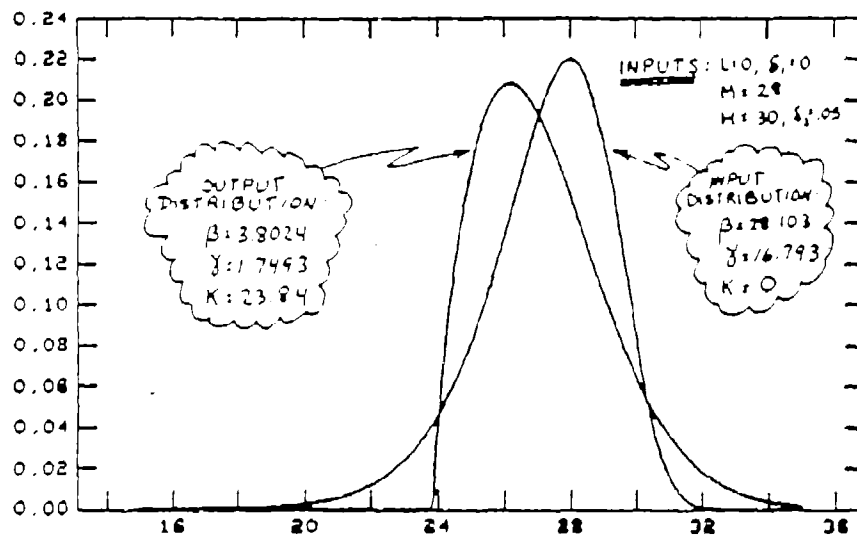


FIGURE 5: Weibull Densities. The output density resulted from applying the formulas of [14] to the moments of the input density.

CONCLUSIONS

A cost uncertainty analysis should logically precede the choice of a "best" estimate of total cost. Historical evolution of the cost estimation process has resulted in a reversal of these two tasks. Likewise, the PERT approach of the sixties seems to have had an inhibiting influence on creativity in developing cost uncertainty assessment methodology. Recent advances in psychometric research offer hope of providing the cost analyst with applicable tools which take his capabilities into consideration. Development of user-friendly software for today's inexpensive but powerful microcomputers will facilitate the probability encoding tasks. There are exciting opportunities for research and development in the important area of cost uncertainty analysis.

BIBLIOGRAPHY

- [1] Black, R. (1979) Fitting a beta distribution from moments, Grumman Aerospace Corporation, PDM-OP-79-115
- [2] Dienenmann, P. F. (1966) Estimating cost uncertainty using Monte Carlo techniques, The Rand Corporation, RM-4854PR
- [3] Fisher, G. H. (1971) Cost Considerations in Systems Analysis, American Elsevier Publishing Company, New York
- [4] Fong, P., Heydinger, G., Koscielski, M., Schmidt, J., Barnum, R., Nordli, R., Tomlinson, R., and Wilcox, L. (1981) Space Division Unmanned Spacecraft Cost Model Fifth Edition, Space Division/ACC, Los Angeles, Calif.
- [5] Joiner, B. L. and Rosenblatt, J. R. (1971) Some properties of the range in samples from Tukey's symmetric lambda distribution, Journal of the American Statistical Association, 66, 394
- [6] Jordan, H. R. and Klein, M. R. (1975) An application of subjective probabilities to the problem of uncertainty in cost analysis, Resource Analysis Group, CNO, U.S. Navy

- [7] Kadane, J. B., Dickey, J. M., Winkler, R. L., Smith, W. S. and Peters, S. C. (1980) Interactive elicitation of opinion for a normal linear model, Journal of the American Statistical Association, 75, pp. 845-855
- [8] Kahneman, D., Slovic, P. and Tversky, A. (1982) Judgement Under Uncertainty: Heuristics and Biases, Cambridge University Press, New York
- [9] Klein, M. R., (1973) Treating uncertainty in cost analysis: the beta distribution, Resource Analysis Group, CNO, U. S. Navy
- [10] Korkotsides, A. S. and Wallenius, K. T. (1981) Parametric cost methodology, IEEE Proceedings of Reliability and Maintainability Symposium
- [11] Lichtenstein, S., Fischhoff, B. and Phillips, L. D. (1982) Calibration of probabilities: The state-of-the-art to 1980, a paper in ref. [8]
- [12] Ling, R. F. and Wallenius, K. T. (1982) Cost Realism, T.R. 400, Dept. of Mathematical Sciences, Clemson University
- [13] McNichols, G. R. (1976) On the treatment of uncertainty in parametric costing, Report T-330, Inst. of Management Science and Engineering, The George Washington Univ., Washington, D. C.
- [14] McNichols, G. R. (1979) Treatment of uncertainty in life cycle costing, Proceedings of IEEE Symposium on Reliability and Maintainability
- [15] McNichols, G. R. (1984) The state-of-the-art of cost uncertainty analysis, The Journal of Cost Analysis, 1, pp. 149-174
- [16] Malik, D. and Stallworth, H. (1982) Cost uncertainty /management reserve analysis, Cost Management and Technology Div., Directorate of Cost Analysis, Eglin AFB, Florida
- [17] Miller, B. M. and Soverign, M. G. (1973) Parametric cost estimation with applications to sonar technology, Tech. Report NPS 55Z073091a, USNPS, Monterey, Calif.
- [18] Pratt, J. W., Raiffa, H. and Schlaiffer, R. (1965) Introduction to Statistical Decision Theory, McGraw-Hill, Inc. New York

- [19] Sobel, S. (1965) A computerized technique to express uncertainty in advanced system cost estimates, Tech. Report ESD-TR-65-79, The Mitre Corporation, Bedford, Mass.
- [20] Wallenius, K. T. (1983) Some observation on a subjective probability estimation technique (SPET) TR 437, Dept. of Mathematical Sciences, Clemson University.
- [21] Wallenius, K. T. (1985) Cost uncertainty analysis methodology: New initiatives, Proceedings, 19th Annual DOD Cost Analysis Symposium
- [22] Walsten, T. S. and Budescu, D. V. (1983) Encoding subjective probabilities: A psychological and psychometric review, Management Science, 29
- [23] Wilder, J. J. (1977) An analytical method for cost risk analysis, Grumman Aerospace Corp., PDR-OP-T79-12
- [24] Wilder, J. J. and Black (1981) Using moments in cost risk analysis, Proceedings of Management of Risk and Uncertainty Symposium, U. S. Air Force Academy, Colorado Springs, Colorado
- [25] Worm, G. H. (1980) Application of risk analysis in the acquisition of major weapon systems, AFBRMC Wright-Patterson AFB, Ohio
- [26] Zellen, M. (1954) Bounds on the distribution function which are functions of moments to order four, RP 2556, Journal of Research of the National Bureau of Standards, 53

END

FILMED

1-86

DTIC