LIFE CYCLE COST (LCC)
RISK ANALYSIS

DETERMINATION OF COST PAIRS CORRELATION

Abstract

Life Cycle Cost (LCC) decisions require the probable value range that the LCC may take, about the point estimate. This paper describes an algorithm for the determination of a valid and defensible LCC-Range, given the risks associated with the estimates for each component. ***Key Words****: Life Cycle Costing; LCC; cost risk ;risk correlation.*

Aim

1. Investigate automated determination of correlation between pairs of cost equations Yi and Yj, where there are N equations (cost components) and i<j.
2. Transfer the resultant matrix of correlations to the cost risk algorithm to permit calculation of actual cost spreads for lowest level cost components, intermediate cost aggregations and, finally for the LCC aggregation.

Definitions

3. Correlation between a pair of cost elements ((Rho (i,j)) is the effect that a variation in the value of one has on the value of the other, due to each being a function of common variables and, thus, are dependent to some degree.
4. Correlation (Rho) is expressed as an index; Rho = {1 or -1} for perfect correlation (total interdependence), for zero correlation (total independence) and somewhere between -1 and 1 for partial correlation.

Objective

7. The objective is to determine the most likely spread of the Life Cycle Cost (LCC) about the nominal value, given that:

- LCC is an aggregation:

  ⇒ directly, of cost components at the lowest indenture level, or

  ⇒ indirectly, of lower level sets of Aggregated Costs;
- an Aggregated Cost is the correlated sum of subordinate, lowest level cost components (at the lowest indenture level).
- each lowest level cost component may vary about its nominal value according to its individual risk spread; and
- that the risk spread of an Aggregated Cost will be a function of the correlation between all lowest-level subordinate cost pairs.

Why Needed?

7. In comparing two LCCs, eg for two or more tenders, it is not enough to compare the nominal values determined for each LCC.

8. Because there is a risk spread associated with each LCC estimate, these spreads can overlap, in which case there is a 'risk' that the nominally smaller LCC may not in fact be the smallest (or the biggest is not necessarily the biggest).

9. The size of the overlap is a measure of risk which, if greater than a given figure, means that the overlap is significant, in which case the two nominal LCC should be considered statistically equal for the purposes of comparison, ie there is a reasonable chance that the true nominal LCCs for the two distributions are in fact the same.

10. Given an overlap, say between LCC1 and LCC2 (between (C1) and C2)), and the assumed distributions of risk (spread about the nominal (mean), the risk would be calculated as follows:

- determine probability of LCC1 being between C1 and C2 [P(1)]
- determine probability of LCC2 being between C1 and C2 [P(2)];
- multiply P(1) and P(2) to obtain the joint probability (P(1)*P(2)) of the two LCCs overlapping; and
- compare the joint probability with the maximum figure set by management to determine significance or otherwise.
- However, the point of statistical significance to be set by management can be rather subjective. What is significant? 0.1, 0.2, 0.5, 0.10, 0.20 or higher probability of overlap?

Discussion

7. Theoretically, proper determination of Rho (i,j) is possible, given complete and exact formulae for every cost element in the set that has to be summed. However, this requires a very complex and time-consuming process because of the large number of combinations of cost pairs, the combinations of variables used across the cost equations and the effect of coefficients to each variable as used in each cost equation.

8. In practice, complete and exact formulae are rarely able to be determined due to lack of data and the sheer complexity of variables at play. Even if data was known about the past, its application to the future has inherent uncertainty. Cost estimates may be made without formulae, being effectively a constant, yet be related to other costs elements, of which the relationship would not or never be known.

9. The combinations of variables used across the cost equations and their respective coefficients, in any number of terms to the equation, means that it is difficult to determine an exact value of Rho (i,j). A suitable compromise by way of approximation is called for in the case of partial independence which can vary between -1 and 1. Full independence and full dependence can be readily tested but partial dependence is a very different problem.

10. Even estimations of Rho (i,j) based on expert opinion would require considerable effort and spreadsheet manipulation. At present, the cost risk algorithm provides for only 20 cost elements in the set to be summed. However, even this small number would require 190 (20*19/2) values of Rho (i,j) to be estimated subjectively, based only on the experience of the analyst as both cost analyst and logistician. Because it would be largely a futile exercise because of lack of data, no attempt has been yet made to go back to the basics of establishing exact formulae for each of the cost elements, in terms of a range of variables, and comparing these for dependencies.

11. In an earlier cost risk algorithm developed by the author, for the sake of simplicity, values of Rho(i,,j) were set as follows:

- set to 0, if the analyst believed there was no significant correlation between a pair;
- set to 1 or -1, if the analyst believed there was total correlation (+ or -) between a pair; and
- otherwise, set to one of a few intermediate values according to simple criteria.

In LCC practice, one can expect most cost components to be positively correlated, if dependent. Reliability (MTBF) is one important exception.

7. It would also help immensely if the algorithm could be automated.

8. Consequently, to be tractable and practicable, a risk evaluation algorithm should:

- limit the number of elements in a set to be summed;

- provide for a simple yet reliable method to determine values for Rho (i,j); and
- be automated.

9. The number of cost elements is currently limited to 20 but this figure can be readily altered to suit the LCC case, if warranted.

10. Any method to determine values for Rho (i,j) short of establishing and using exact cost equations would be a comprise, but a necessary and pragmatic compromise. Fortunately, the approximation of Rho() described in the new algorithm below is considered to give acceptable results, ie accurate enough to estimate the true spread of probable LCC about its nominal (mean) value.

11. It should be note here that, as will be shown subsequently by specific cases, the maximum and minimum spreads about a nominal LCC can be readily established. It is honing in on what the true spread is between the extremes that is difficult. While knowing both extremes is both necessary and important, also knowing the true spread may prevent serious errors, ie not recognising two LCCs as statistically equal.

New Methodology

Aim

12. To determine correlation between equations Yi and Yj, let:

Y1 = a1*v1+a2*v2+……+at*vt
Y2 = a1*v1+a2*v2+……+at*vt
Y3 = a1*v1+a2*v2+……+at*vt
Y4 = a1*v1+a2*v2+……+at*vt
Y5 = a1*v1+a2*v2+……+at*vt
…
Y20
Where:
{Y1, Y2, …Yt} is the set of LLC to be aggregated to an LCC;
{v1, v2, …vt} is a defined set of variables of which all Yi are considered be functions; and
{a1, a2,…at} is the corresponding set of coefficients of {v} for each Yi.

Procedure

13. Define the Cost Breakdown Structure (CBS). Standardised CBSs should be developed for defined system categories.

14. Define all variables {V1, …Vt} of which lowest level cost components {LLC} will be a function. These can and should be a standardised set, but may be varied for the case.

15. Identify (flag) the Lowest Level Costs {LLC} (lowest indenture level cost components).

16. Identify {LLC} for at least the following cost aggregations (if applicable) to a given LCC:

- Research and Development Cost (if applicable);
- Acquisition Cost (may include O&S costs covered under acquisition contract);
- Operating & Support Cost (that of the through-life support contract); and
- Disposal/Redeployment Cost aggregation (if applicable).

17. For each Lowest Level Cost (LLC):

- insert its cost equation in terms of defined variables, if known (virtually never known), otherwise leave blank;
- under each defined variable, enter a value of 1, 0 or -1, according to whether the LLC is a function of that variable and its direction:

⇒ 1 for directly proportional,

⇒ -1 for inversely proportional, and

⇒ 0 for not a function of the variable.

General Method for Rho(ij) [see bibliography references]

7. The following process is according to well established statistical theory and practice, for the determination of Rho(ij) (if data was available):

- LCC = SUM (aggregated costs) = SUM (CA1, CA2, …CAz); Nominal and spread values.
- Variance (Aggregation z) = SUM (Variance (i) + 2* Covariance (ij))
- Covariance (ij) = SUM SUM (Sigma(i) * Sigma(j) * Rho(ij)), i<j; j = 1,...N; i=1,...(N-1)
- Sigma(i) = STDEV(i) = SQRT(VAR(i))
- Sigma(j) = STDEV(j) = SQRT(VAR(j))
- Rho(ij) = [values as may be estimated]

Recommended Method to Approximate Rho(ij) [1]

7. For each pair of equations (Yi,Yj):

- Determine the number of 'relevant variables' (Tij); [2]
- Determine the number of common variables (matches) (Mij) [3]
- Determine the number of negative matches (NMij)
- Determine the pseudo-correlation between Yi and Yj.

⇒ T(i,j) = Count IF((ABS(Yi) + ABS(Yj)) <> 0)

⇒ M(i,j) = Count IF((ABS(Yi) * ABS(Yj)) > 0)

⇒ NM(i,j) = Count IF(Yi * Yj) < 0)

⇒ If T(ij) = 0, Then Rho(ij) = 0,

⇒ Else: Rho(ij) = Sum(Matches)/T(ij) = (M(ij)-2*NM(ij))/T(ij) [4]

Testing the Algorithm

7. Given the statistical basis of the algorithm, it must produce certain results for certain input data.

8. A few test cases suffice and Cases 2, 3 and 4 from Bibliography Reference 1 are used (Case 1 is used to solve the specific case).

Case 2

9. Case 2 is a special case giving the ***maximum*** extreme (does not occur in LCC practice, given that all ***component distributions are assumed identical*** and ***assumed perfectly correlated***, but is useful in understanding the process and testing the algorithm). In this case, the test should produce the following results:

- Mode (z) = n * Mode (i)
- Sigma (z) = n * √ Variance (i) = n * Sigma (i),

---

[1] This section is the intellectual property of ALSC and may not be used without written permission of ALSC.

[2] Relevant variables are those that comprise the pool of variables covering all cost equations under consideration. However, few if any cost equations would be a function of all relevant variables.

[3] A 'match' is when a pair of LLC are functions of the same variable, eg if a pair of LLC equations have three common variables for which coefficients are not zero, there would be three matches.

[4] "-2*NM(ij)" is necessary to correctly calculate the sum of positive correlations less the sum of negative correlations, given that Mij comprises both positive and negative matches.

- Range (z) = LCC (Nominal) + or - 2 * Sigma (z).
- where:
  - ⇒ i = 1 to n
  - ⇒ n = number of cost component distributions to be summed.

10. While Case 2 is simplified (all *component distributions are assumed identical),* it does indicate how the maximum spread is established for a nominal LCC, ie when *all pairs are assumed to be interdependent and perfectly correlated.* Sigma (z) = n * Sigma (i).

Case 3

11. Case 3 is a special case giving the ***minimum*** extreme (does not occur in LCC practice, given that all ***component distributions are assumed identical*** but all ***assumed to be independent and of zero correlation***). In this case, the test should produce the following results:

- Mode (z) = n * Mode (i)
- Variance (z) = n * Variance (i)
- Sigma (z) = √ n * Sigma (i)
- Range (z) = LCC (Nominal)+ or - 2 * Sigma (z).

12. While Case 3 is simplified (all *component distributions are assumed identical),* it does indicate how the minimum spread is established for a nominal LCC, ie when *all pairs are assumed to be independent with zero correlation.* Sigma (z) = n * Sigma (i).

13. Note the relationship between the maximum and minimum spreads for Cases 2 and 3, being a ration of n/√ n = √ n. Therefore, for a LCC comprised of 20 LLC, the theoretical maximum spread could be 4.47 time the minimum theoretical spread. Neither case is likely, but Case 2 has some usefulness in that it is conservative. It is also the most commonly used assumption in adding standard deviations of cost components. The assumption of complete independence between cost components (as in Case 3) is not used in financial practise simply because there is no simple way to determine Sigma (z), when all of the Sigma (i) are different. If it could be determined, it would give the minimum value required for contingency, ie to provide for cost risk.

Case 4

14. Case 4 is a special case between the extremes (does not occur in LCC practice either in that it assumes all ***component distributions to be identical)*** which assumes all pairs to be correlated to some extent. In this case, the test should produce the following results:

- Mode (z) = n * Mode (i)
- Variance (z) = n * Variance (i) + 2 * Variance (i) * ΣΣ[Rho (i,j)], i<j,
- Sigma (z) = Variance (z)
- Range (z) = LCC (Nominal) + or - 2 * Sigma (z), for a Confidence Level of 95 per cent

Test Results

7. Sample test data and expected results are shown in **Table 1** through **Table 4**. The summary sheet of the actual CRM run result for Case 4 are shown at Table 4, which can be seen to agree with the manually calculated figures in Table 3.

Table 1
LCC Risk Algorithm – Test Data – Cost Distributions

CBS	Cost Component	Cost-Nominal	- Sigma i	+ Sigma I
1	LCC	450	to be determined	to be determined
1.1	LLC1	90	10	20
1.2	LLC2	90	10	20
1.3	LLC3	90	10	20
1.4	LLC4	90	10	20
1.5	LLC5	90	10	20
		Count	**5**	**5**

Table 2
LCC Risk Algorithm – Test Data – Functions (for Case 4 only [1])

CBS	Cost	V1	V2	V3	V4	V5
1.1	LLC1	1	-1	1	1	1
1.2	LLC2		1		1	
1.3	LLC3	1	1	-1	1	1
1.4	LLC4	1		1		-1
1.5	LLC5	1	1	1	-1	1
Note 1: Test data sets only Rho(1,1) =1. All other pairs [i,j] are set to zero						

Table 3
LCC Risk Algorithm – Expected Test Results

Case	Parameter	Formulae	- Result	+ Result
2	Mode (z)	= n * Mode (i)	450	450
2	Rho (i,j)	= 1, for all pairs [i,j]	1	1
2	Variance (i)	= Sigma (i) ^ 2	100	400
2	Variance (z)	= n^2 * Variance (i)	2 500	10 000
2	Sigma (z)	= n * Sigma (i)	50	100
2	Range (z)	= LCC (Nominal) + or – 2 * Sigma (z)	**450-100 = 350**	**450+200 = 650**
			Correct results	Correct results
3	Mode (z)	= n * Mode (i)	450	450
3	Rho (i,j)	= 0, for all pairs [i,j]	0	0
3	Variance (i)	= Sigma (i) ^ 2	100	400
3	Variance (z)	= n * Variance (i)	500	2 000
3	Sigma (z)	= √ n * Sigma (i)	22.36	44.72
3	Range (z)	= LCC (Nominal) + or - 2 * Sigma (z)	**450-44.72 = 445.28**	**450+89.44 = 539.44**
			Correct results	Correct results
4	Mode (z)	= n * Mode (i)	450	450
4	Rho (i,j)	= {set by analyst doing test}	see test data	see test data
4	Variance (i)	= Sigma (i) ^ 2	100	400
4	Variance (z)	= n * Variance (i) + 2 * Variance (i) * ΣΣ[Rho (i,j)], i<j, [1]	500 + 200 * ΣΣ [Rho (i,j)], i<j,	2 000 + 800 * ΣΣ [Rho (i,j)], i<j,
4	Variance (z)	ditto	740	2,959
4	Sigma (z)	= √ Variance (z)	27.2	54.4
4	Range (z)	= LCC (Nominal) + or - 2 * Sigma (z)	**450 - 54.4 = 395.6**	**450 + 108.8 = 558.8**
			Correct results	Correct results

Table 4
LCC Risk Algorithm – Actual Test Results (Case 4)

EVALUATION OF COST RISK Sheet 1	SUMMARY	COST RISK MODEL Version 2.1 (CRM 2.1)

PROJECT :	Demonstration
TENDERER :	
OPTION :	
AGGREGATION :	

Use 1,2 or 3 sigma only.

AGGREGATION COST	Range	%
High	558.80	24.18
Nominal	450	100.00
Low	395.60	-12.09

% Range	95%
+ - N(Sigma)	2
LCC (Plus) =	108.80
LCC (Minus) =	54.40

Conclusion

7. Current limitation on the size of the risk matrix in the Cost Risk Model is 20x20. It may be increased by changing the Dimension statements and initial values for variables in the Visual Basic code. Ultimately, the code will be amended to provide for the matrix size as either a simple input by the user or a program determined value by counting the number of LLC in the CBS.

8. If subordinate LLCs are all assumed to be fully dependent, respective means and variances can be simply added to give the nominal LCC and its maximum and minimum values. This is the most commonly used assumption in practice when adding standard deviations of cost components. The assumption provides the maximum spread (+/- x sigma about the mean) and is conservative.

9. At the other extreme, where all subordinate LLCs are assumed to be fully independent of each other, respective means can still be simply added to give the nominal LCC but accurate estimation of maximum and minimum values about the mean LCC requires a complicated algorithm. The assumption would provide for the minimum spread (+/- x sigma about the mean) and, if calculated, would give the minimum contingency value. However, this figure would be somewhat optimistic. A more realistic contingency provision would be a spread determined by a minimum and maximum which are averages of the minimums and maximums of the two extreme cases.

10. In the new algorithm described and recommended herein, 'Correlation' is in fact a pseudo-correlation in that, while similar to multiple-correlation treated in statistics texts, is not exactly the same. In fact, this pseudo-correlation methodology had to be developed because application of multiple-correlation, as treated in statistics texts and adapted as Excel functions, could not produce logical results for the LCC aggregation problem.[5]

11. The core assumption of this method is that the pseudo-correlation is a satisfactory approximation of the true correlation, if it were ever to be known.

M.R. Flint
Director
Acquisition and Logistics Study Centre
29Dec21

Bibliography

1. Life Cycle Cost (LCC) Risk Analysis, by M.R. Flint, Canberra, 2002.
2. Statistical Analysis for Administrative Decisions, Clark and Schkade, Second Edition, South Western, 1974 (pp 527,544).
3. Operations Research (p341), Frederick S. Hillier and Gerald J. Lieberman, Second Edition, Holden-Day 1974 (p341).
4. Introductory Statistics for Business and Economics, Wonnacott and Wonnacott, John Wiley and Sons, 1972 (pp 103-107).

[5] Standard formulae and Excel functions do not work properly because they assume that any pair of variables being compared for correlation both have values for the same number of common points. In practice, cost components may be functions of different sets and numbers of variables, but with common variables. In this case, determination of correlation needs to recognise the full set of variables used in at least one cost function and not that every cost component is a function of the same set of variables.