Analytical and approximate variance of total project cost

Skitmore, Martin; Ng, S. Thomas

Published in:
Journal of Construction Engineering and Management

DOI:
10.1061/(ASCE)0733-9364(2002)128:5(456)

Analytical and approximate variance of total project cost

Prof. Martin Skitmore, School of Construction Management and Property, Queensland University of Technology, Gardens Point, Brisbane Q4001, Australia.

Dr. S Thomas Ng, Department of Civil Engineering, University of Hong Kong, Pokfulam, Hong Kong.

for

Journal of Construction Engineering and Management

Corresponding Author
Professor Martin Skitmore
School of Construction Management of Property
Queensland University of Technology
Gardens Point
Brisbane Q4001
Australia

rm.skitmore@qut.edu.au
phone: +61 (07) 3864 2234; fax +61 (07) 3864 1170

(Vers 4)
17 September 2001

Word count

Main text	2200
Equations	600
Refs	144
Table 1	630
Table 2	315

Total	**3889**
	====

Analytical and approximate variance of total project cost

Prof. Martin Skitmore, School of Construction Management and Property, Queensland University of Technology, Gardens Point, Brisbane Q4001, Australia.

Dr. S Thomas Ng, Department of Civil Engineering, University of Hong Kong, Pokfulam, Hong Kong.

Abstract

The statistical variance of total project cost is usually estimated by means of Monte Carlo simulation on the assumption that exact analytic approaches are too difficult. This paper tests that assumption and shows that, contrary to expectations, the analytic solution is relatively straightforward. It is also shown that the coefficient of variation is unaffected by the size (floor area) of the project when using standardized component costs. A case study is provided in which actual component costs are analyzed to obtain the required total cost variance. The results confirm previous work in showing that the approximation of the second moment (variance) under the assumption of independence considerably underestimates the exact value. The analysis then continues to examine the effects of professional judgement and, with the simulated data used, the approximation is shown to be reasonably accurate – the professional judgement absorbing most of the intercorrelations involved. An example is also given in which the component unit quantities are priced by their average unit costs and which again shows that the approximation to be close to the true value. Finally, this is extended to show how the exact total project cost variances may be obtained for each project.
Keywords: Cost, components, distribution, variance, covariance, coefficient of variation, independence assumption, correlation, Monte Carlo simulation, professional judgement.

Introduction

Touran and Wiser (1992) and Touran (1993) have proposed estimating the statistical variance of total project cost through Monte Carlo simulation "… because direct analytical approaches tend to be difficult and are sometimes infeasible" (Touran, 1993:58). No details of these difficulties are provided, but it is presumed that these are caused by two independent problems. One of these problems is that the component costs, being modeled as random variables, may take on a variety of distributional forms (e.g. normal, uniform, beta) resulting in an impossibly complicated distributional form of the sum of these variables. The other problem is that the component costs are likely to be intercorrelated, making estimates of the total project cost second moment (variance) under the usual assumption of intervariable independence overly inaccurate. Using Touran and Wiser's data, Moselhi and Dimitrov (1993) have shown how improved estimates may be made analytically, but no exact solution has yet been proposed.

This paper derives the exact value of the second moment of the total project cost distribution (the value of the first moment being, trivially, the sum of the means of the component costs) by using the mean estimated-actual differences of standardized (per unit floor area) component costs of previous projects to calculate the coefficient of variation of an ‘average’ project. This is applied to a set of 29 Hong Kong building projects and their standardized component costs, and the result is contrasted with the approximation obtained under the assumption of intervariable independence, confirming Touran and Wiser’s (1992) assertion, i.e., the approximation under the independence assumption significantly underestimates the variance of the total project cost.

The paper then continues to examine the more practical situation where the mean component costs have been adjusted by professional judgement, and it is shown that, with the assumed estimated costs used, the accuracy of the approximate method for an ‘average’ project is considerably improved, the professional judgement having absorbed most of the correlation effects. This is also shown to be the case when the component unit quantities are priced by their mean unit costs. Finally, it is shown how the exact variance may be calculated for an individual project, and this is contrasted with the approximation under the independence assumption.

Exact derivation of total project cost variance

Let c_{ip} and e_{ip} denote the respective actual and estimated standardized values of cost component i = 1, 2, ... , n for project p = 1, 2, ... , m; ie., $c_{ip} = c'_{ip}/a_p$ and $e_{ip} = e'_{ip}/a_p$, where c' and e' are the original (dollar) values of the respective actual and estimated component costs and a is the project gross floor area. The total standardized actual and estimated costs are therefore given by $\sum c_{ip}$ and $\sum e_{ip}$ respectively. The standardized estimated-actual component cost difference is $d_{ip} = e_{ip} - c_{ip}$ so that the total standardized estimated-actual cost difference is $t_p = \sum d_{ip}$. Now, if D_i and D_j are random variables from which d_{ip} and d_{jp} are values, the variance of $\sum D_i$, and therefore t_p, is the well known

$$\mathrm{var}\left[\sum_{i=1}^{n} D_i\right] = \sum_{i=1}^{n} \mathrm{var}[D_i] + 2\sum_{i<j}\sum \mathrm{cov}[D_i, D_j] \qquad (1)$$

which, in terms of the standardized estimated total project cost, gives a coefficient of variation of

$$cv_p = \frac{100\sqrt{\mathrm{var}\left[\sum_i D_i\right]}}{\sum_i e_{ip}} \tag{2}$$

Note: the cv_p value is the same for both standardized and unstandardized data as

$$cv_p = \frac{100\sqrt{\mathrm{var}\left[\sum_i a_p D_i\right]}}{\sum_i a_p e_{ip}} \tag{3}$$

and the a_p values cancel.

Now we are interested in evaluating the approximation assuming independence, ie.,

$$\mathrm{var}\left[\sum_i D_i\right]^* = \sum_i \mathrm{var}[D_i] \tag{4}$$

and use the ratio

$$v = cv_p^* / cv_p \tag{5}$$

where

$$cv_p^* = \frac{100\sqrt{\mathrm{var}\left[\sum_i D_i\right]^*}}{\sum_i e_{ip}} \tag{6}$$

so that v<1 indicates the true cv is underestimated and v>1 that the true cv is overestimated.

Case study

Estimation by mean component costs

Cost analyses of 29 school projects have been collected from the Architectural Services Department of Hong Kong. The projects were tendered between 1986-1998, and the costs in the cost analyses were updated to 3rd quarter of 1998 costs. Table 1 summarizes the updated standardized component costs[1] (c_{ip}) for the 29 projects. The columns contain the c_{ip} values for the components **G**ross **F**loor **A**rea, **PREL**iminaries, **SUB**structure, **SUP**erstructure, **M**echanical **& E**lectrical services, **E**xternal **W**orks, **DRAIN**age, **SITE D**evelopment, **FURN**iture and equipment and **CONT**ingencies. The last column gives the standardized total project costs ($\sum c_{ip}$). The last three rows give the means, standard deviations and coefficients of variation respectively for each c_{ip} column, with the final (total) column showing the mean and standard deviation of the standardized total project costs to be HK$7987.17[2] and HK$1671.68 per m[2] respectively, representing a coefficient of variation of 22.06%.

In the absence of any other information, we assume an 'average' future project, *p*. Here, $e_{ip} = \bar{c}_i$ where

$$\bar{c}_i = \frac{1}{m}\sum_{p=1}^{m} c_{ip} \qquad (7)$$

which means that

$$d_{ip} = \bar{c}_i - c_{ip}$$

Therefore, from Table 1

$$\sqrt{\operatorname{var}\left[\sum D_i\right]} = \sqrt{\left[\sum C_i\right]} = s = sd = 1761.68 \qquad (8)$$

where C_i is a random variable to which $\bar{c}_i$ belongs

and $\sum e_{ip} = \sum ci_p = 7987.17$

and so, from (2)

cv_p=100(1761.68)/7987.17 = 22.06%

also

[1] The values of the cost components are standardised by division by the gross floor area of the building.
[2] US$1=HK$7.32 (16 Sep 2001)

$$\sqrt{\text{var}[D_i]} = \sqrt{\text{var}[C_i]} = s = sd = 286.37, 157.34, etc$$

which means

$\text{var}[D_i] = \text{var}[C_i] = s^2 = 82007,\ 24756$, etc

and so, from (6)

$cv_p^* = 100\sqrt{(82007+247564+\text{etc})}/7987.17 = 14.59\%$

and therefore, from (5)

$v = 14.59/22.06 = 0.661$

Note 1: The identical answer to (8) is obtained by (1) using the correlation matrix.
Note 2: The cv_p and cv_p^* values are identical for both standardized and unstandardized data.

With a v value of 0.661, the approximate project variance underestimates the true value by 32.9%. The correlation matrix shows why this is the case, with many of the coefficients being significantly positively correlated.

Estimation by subjectively derived component costs

The above analysis is appropriate for project cost forecasts where the forecast is obtained by simply summing the means of each c_i for the projects in the database. In the above case, therefore, the forecast for a new, non-database, project would be, in the absence of any project details, HK$7987.17 per m2 floor area with a coefficient of variation of 22.06%. However, the usual practice is to estimate the component costs by professional judgement based on the mean value of the component costs in the database and taking the details (project type, size, specification, etc) of the project into account. In this case, it is appropriate to consider the efficacy of previous such judgements in order to quantify the uncertainties involved. So now, instead of modeling the difference between the c_{ip} values and their unadjusted means, as above, this time we model the difference between the c_{ip} values and their means adjusted by professional judgement. Again assuming an 'average' project, the estimated component costs become $e'_{ip} = \bar{c}_i + u_{ip}$, where u_{ip} is the subjective adjustment and thus $d_{ip} = e'_{ip} - c_{ip}$. As u_{ip} data were not available, reasonable values were assumed. The mean and standard deviation of the estimates of project costs this time are HK$–687.11 and HK$1583.09 (cv = 21.69%), indicating a bias towards underestimates..

Now, from above

$$\sqrt{\operatorname{var}\left[\sum D_i\right]} = s = sd = 1583.09 \quad (9)$$

and $\sum ei_p = \sum c_{ip} + \sum d_{ip}$ = 7987.17 (from Table 1) - 687.11 = 7300.06

and so, from (2)

cv_p=100(1583.09)/7300.06 = 21.69%

also

$$\sqrt{\operatorname{var}[D_i]} = s = sd = 196.98, 94.93, etc$$

which means

var[D_i] = s^2 = 38801,9012,etc

and so, from (3)

cv_p^* = 100√(38801+9012+etc)/7300.06 = 19.74%

and therefore, from (5)

v = 19.74/21.69 = 0.910

Note 1: The identical answer to (9) is obtained by (1) using the correlation matrix of d_{ip} values (not shown).
Note 2: The cv_p and cv_p^* values are identical for both standardized and unstandardized data.

Thus, in this hypothetical situation, the introduction of professional judgement has slightly decreased the likely distribution of forecast project costs from a coefficient of variation of 22.06% to 21.69%[3]. Interestingly, however, the coefficient of variation, assuming independence, in this case is 19.74% which, with a v of 0.910, is 91% of the true value. The reason that the coefficient of variation approximation assuming independence is so much improved is because the estimates themselves now contain much of the correlation in the data - the effect being to substantially cancel out the correlations involved.

[3] The correction for bias is straightforward.

Estimation by mean component unit costs

In practice, e_{ip} values are obtained by pricing component unit quantities, such as wall area, number of doors, etc, by component unit costs (eg., dollars per m^2 wall area, dollars per door, etc). Here, we use the standardized component unit costs, which is the component unit cost divided by the gross floor area. So now $c_{ip} = q_{ip}r_{ip}$ and $e_{ip} = q_{ip}\bar{r}_i$ where

$$\bar{r}_i = \frac{1}{m}\sum_{p=1}^{m} r_{ip} \tag{10}$$

Multiplying the mean standardized component unit cost for each component by its associated project component unit quantity and subtracting the actual project standardized component cost provides a set of estimated-actual standardized component cost differences. For example, the estimated Preliminary mean standardized component cost for project 1 is 11.38x120=1365.6, which gives, when the actual standardized component cost of 1440.54 is deducted, a difference of –75.36. The standardized component costs for each project with assumed q_{ip} values have coefficients of variation ranging from 22.41 to 105.48, which are quite typical of those found in practice (Beeston, 1974). The mean of the total differences for each project is 830.05 with a standard deviation of 1490.17 .

Now, for an 'average' project

$$\sqrt{\operatorname{var}\left[\sum D_i\right]} = s = sd = 1490.17 \tag{11}$$

and $\sum e_{ip} = \sum c_{ip} + \sum d_{ip}$ = 7987.17 (from Table 1) + 830.15 = 8817.32

and so, from (2)

cv_p=100(1490.17)/8817.32 = 16.90%

also

$$\sqrt{\operatorname{var}[D_i]} = s = sd = 170.61{,}114.10{,}etc$$

which means

var[D_i] = s^2 = 29107,13019,etc

and so, from (3)

cv_p^* = 100√(29107,13019+etc)/8817.32 = 17.82%

and therefore, from (5)

v = 17.82/16.90 = 1.054

Note 1: The identical answer to (11) is obtained by (1) using the correlation matrix .
Note 2: The cv_p and cv_p^* values are identical for both standardized and unstandardized data.

As can be seen, the coefficient of variation for the forecast total project cost is now 16.90%, again a typical figure found in practice for this kind of estimate (eg., Ashworth and Skitmore, 1983). The approximate coefficient of variation, assuming independence is 17.82% which, with a v of 1.054 is overestimated by 5.4%

This method can be extended to provide variances for individual projects for, as $d_{ip} = q_{ip}$ ($\bar{r}_i - r_{ip}$) and var [D_i] = var [R_i] so

$$cv_p = \frac{100\sqrt{\text{var}\left[\sum_{i=1}^{n} q_{ip} R_i\right]}}{\sum e_{ip}} \tag{12}$$

From which it can be shown that, for project p

$$\text{var}\left[\sum_{i=1}^{n} q_{ip} R_i\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} q_{ip} q_{jp} \,\text{cov}[R_i, R_j] \tag{13}$$

The approximate cv is given by

$$cv_p^* = \frac{100\sqrt{\text{var}\left[\sum q_{ip} R_i\right]^*}}{\sum e_{ip}} \tag{14}$$

setting all the off-diagonal elements of R_i and R_j to zero.

Table 2 gives the estimated standardized total project cost, standard deviation and coefficient of variation for each project plus the approximate coefficient of variation (**Appcv**) under the independence assumption. This method can clearly be used to estimate the coefficient of variation for any future project providing the component unit quantities are known. As can be seen from the v values in Table 2, the approximate method is quite accurate in most cases being generally within 10% of the true variance.

Conclusions

The paper has described a method for calculating the variance of total project cost based on standardized component costs for a set of database projects. For the sample analyzed, the correct variance for an 'average' project, taking in account intercomponent variability, was found to be much greater (cv=22.06) than the approximation under the assumption of independence (cv=14.59), confirming previous similar studies in this field. It was also found that the coefficient of variation is invariant of the gross floor area and so the method can be used when the gross floor area is not known. However, considering the difference between actual component costs and typical component cost estimates by professional judgement suggests that the independence assumption provides a reasonable approximation of the variance of the difference between actual total project costs and the associated estimates of total project costs, the professional judgement absorbing most of the intercomponent correlations involved. Similarly, where component costs are estimated via component unit quantities, the cost variance for an 'average' project is also shown to be reasonably approximated by the independence assumption. Following this, a method is proposed by which the coefficient of variation of individual projects can be derived from unit cost estimates and this also is shown to be estimated reasonably well by the approximate method.

The major limitation of the research described in this paper is in the simulation of component costs and quantities. Future research should undertake an empirical analysis of actual component unit cost and quantity estimates as a check on the validity of the simulations and hence these results.

Acknowledgements

Many thanks to Prof. Graeme Newell, of the University of Western Sydney, for help in finding eqns (1) and (13).

References

Ashworth, A., and Skitmore, R.M. (1983). "Accuracy in estimating." The Chartered Institute of Building *Occasional Paper* No 27 (ISBN 0 906600 57 X).

Beeston, D.T. (1974). "One statistician's view of estimating." Cost Study number 3, Building Cost Information Service, Royal Institution of Chartered Surveyors, 12 Great George Street, London W1.

Mood, A.M. (1974). *Introduction to the theory of statistics*. McGraw-Hill (ISBN 0 07 042864 5).

Moselhi, O., and Dimitrov, B. (1993). "Monte Carlo Technique with Correlated Random Variables." Discussion, *Journal of Construction Engineering and Management*, ASCE **119**(3) 658-61.

Spooner, J.E. (1974). "Probabilistic estimating." *Journal of the Construction Division*, ASCE, **100**(1) 65-77.

Touran, A. (1993). "Probabilistic cost estimating with subjective correlations." *Journal of Construction Engineering and Management* **119**(1) 58-71.

Touran, A., and Wiser, E.P. (1992). "Monte Carlo Technique with Correlated Random Variables." *Journal of Construction Engineering and Management*, ASCE **118**(2) 258-72.

Project	Type	GFA (m^2)	PREL	SUB	SUP	M&E	EW	DRAIN	SITED	FURN	CONTG	TOTAL
1	1	6530	1440.54	317.87	5818.36	1294.80	1131.59	336.82	1961.34	25.31	813.66	13140.29
2	1	7484	1250.53	364.04	3536.75	952.68	1099.10	401.56	354.80	7.47	522.62	8489.55
3	1	7484	589.86	443.42	3324.04	1043.31	1151.41	267.80	1424.42	9.22	516.12	8769.60
4	1	7484	816.09	396.75	3887.24	1218.25	673.77	166.73	810.67	8.75	490.10	8468.35
5	2	8150	535.73	306.61	4378.83	1083.05	893.69	228.45	448.80	16.25	738.28	8629.69
6	2	8150	716.69	345.70	4236.18	996.67	765.75	200.41	340.96	16.25	738.28	8356.89
7	2	8150	144.87	792.02	4240.65	1057.08	864.08	198.86	1722.40	32.69	1188.17	10240.82
8	1	4713	400.67	353.27	2997.83	350.93	790.63	242.03	974.96	3.55	310.17	6424.04
9	1	4713	236.16	237.47	2803.39	282.68	502.01	162.86	582.47	38.55	332.30	5177.89
10	1	4713	237.56	190.64	2751.29	281.74	430.28	159.65	522.02	42.98	332.30	4948.46
11	1	4713	233.32	234.99	2710.45	281.79	561.42	157.24	809.19	38.55	332.30	5359.25
12	1	6238	537.53	440.50	3644.81	907.49	1174.17	232.77	871.28	14.58	757.49	8580.62
13	1	6290	378.51	130.84	4206.50	783.01	842.10	243.04	3453.34	34.68	972.86	11044.88
14	2	6964	115.19	126.56	3257.21	399.58	689.79	126.33	530.54	67.91	397.23	5710.34
15	2	6964	478.81	265.65	4160.06	463.91	740.92	135.45	572.14	675.24	383.11	7875.29
16	2	6964	475.98	306.07	4160.06	463.91	740.48	135.45	495.81	675.24	383.11	7836.11
17	2	6964	467.76	217.76	3798.54	435.71	863.44	210.85	549.12	634.20	359.82	7537.20
18	2	6964	466.68	228.54	3798.54	435.71	922.31	207.84	465.84	634.20	359.82	7519.48
19	2	8393	416.42	174.69	3065.08	856.94	554.13	196.45	554.94	4.17	521.76	6344.58
20	2	8393	778.59	220.36	3336.51	973.27	589.10	151.52	423.95	884.55	776.13	8133.98
21	2	8393	387.11	161.58	3757.71	1040.29	650.18	179.33	21.77	823.93	684.60	7706.50
22	2	8393	530.92	188.61	3607.18	917.84	792.95	232.30	533.73	28.88	806.61	7639.02
23	2	6060	543.32	260.24	4502.96	2174.02	596.01	204.05	739.39	0	487.17	9507.16
24	2	8150	516.26	780.06	3736.16	1148.24	840.25	168.55	2168.83	18.62	964.83	10341.80
25	2	13518	432.01	244.34	3842.54	448.91	516.73	215.51	356.02	638.00	370.75	7064.81
26	2	16300	452.79	330.38	3946.49	864.95	559.59	153.96	1269.96	18.62	1074.85	8671.59
27	2	13518	297.76	132.42	3573.32	430.08	544.63	145.62	590.86	240.87	359.74	6315.30
28	1	6238	500.84	225.91	3767.80	560.70	831.80	198.20	1291.54	3.14	535.49	7915.42
29	1	6238	163.80	341.96	4027.10	532.34	849.03	192.19	1221.45	15.76	535.49	7879.12
	mean		*501.46*	*302.04*	*3754.26*	*782.06*	*764.18*	*201.79*	*898.71*	*194.90*	*587.76*	*7987.17*
	sd		*286.37*	*157.34*	*619.30*	*409.17*	*199.61*	*58.56*	*696.49*	*296.89*	*245.01*	*1761.68*
	cv		*57.11*	*52.09*	*16.50*	*52.32*	*26.12*	*29.02*	*77.50*	*152.33*	*41.69*	*22.06*

Note: Type 1 = Primary School; Type 2 = Secondary School

Table 1: Hong Kong public school standardized component costs and unit quantities

Proj	TOTAL	SD	CV	AppCV	*v*
1	14904.25	3433.24	23.04	23.60	1.02
2	10004.68	2468.99	24.68	24.84	1.01
3	10652.02	1875.57	17.61	18.80	1.07
4	9163.44	2160.73	23.58	24.22	1.03
5	11039.72	3148.92	28.52	28.86	1.01
6	10564.85	3060.87	28.97	29.01	1.00
7	12225.21	2868.92	23.47	23.80	1.01
8	7082.88	1745.91	24.65	25.63	1.04
9	6097.48	1508.71	24.74	24.19	0.98
10	7322.38	2203.33	30.09	29.26	0.97
11	7146.15	1813.62	25.38	25.43	1.00
12	12618.14	2837.79	22.49	22.90	1.02
13	11644.64	2792.49	23.98	24.97	1.04
14	7487.55	2149.99	28.71	28.65	1.00
15	10458.53	3301.79	31.57	30.32	0.96
16	10671.03	3469.24	32.51	31.29	0.96
17	7969.73	1953.83	24.52	23.47	0.96
18	7172.70	1633.39	22.77	21.52	0.95
19	5178.65	882.51	17.04	17.78	1.04
20	7204.10	1434.06	19.91	18.25	0.92
21	8610.98	2343.20	27.21	25.80	0.95
22	7669.22	1558.92	20.33	20.58	1.01
23	10796.31	2866.98	26.56	26.96	1.02
24	8716.63	1229.64	14.11	15.90	1.13
25	4756.75	952.62	20.03	18.47	0.92
26	7391.85	1184.58	16.03	17.61	1.10
27	5643.42	1432.40	25.38	25.95	1.02
28	7235.57	1210.11	16.72	17.74	1.06
29	8270.65	1805.47	21.83	22.46	1.03

Table 2: Individual project cvs