Methods of analysing risk exposure in the cost estimates of high quality offices

Sidney Newton & V.K. Smith

Methods of analysing risk exposure in the cost estimates of high quality offices

SIDNEY NEWTON

Department of Architectural and Design Science, University of Sydney, New South Wales 2006, Australia

The paper considers two non-deterministic methods of analysing the risk exposure in a cost estimate. Both methods are applied to the same data sample of eight high quality office buildings. An assessment of each method is made in the context of this practical application. The common practice of allowing for risk through an all-embracing contingency sum or percentage addition is challenged. Rather than excluding conventional, deterministic methods, they are here presented as possibly the only effective foundation on which to base risk management in cost estimating.

Keywords: Risk analysis, risk exposure, cost estimating, simulation.

Introduction

Any building cost estimate will involve the client in a degree of risk exposure. Since building projects tend to involve long term, and substantial investments, the associated risk exposure is likely to be significant. This paper considers two methods of analysing the risk exposure in a cost estimate. The first method (referred to as the 'conventional statistical' method) analyses cost data directly, to describe a probability distribution for total costs. The second method (referred to as the 'Monte Carlo simulation' method) interprets cost data indirectly, to generate a probability distribution for total costs from the descriptions of elemental cost distributions. (Note: these are specific applications of the two methods, and do not exhaust the full potential or range of possible applications for these techniques.)

Each method is applied to the same data set: a sample of costs for eight high quality office buildings, constructed in Sydney, Australia between June 1985 and April 1990. Each method is assessed within the context of this practical application.

Risk exposure is defined as the probability that a given cost estimate will be less than the actual costs incurred by the client. It takes no account of risk attitude (i.e. whether the client is inclined to take risks or to be more conservative), risk response (i.e. what might be done to adjust the level of risk exposure), or the more general considerations of risk management. The purpose of this paper is to demonstrate how the risk exposure associated with any particular estimated might be calculated. The analysis of risk exposure is not an end in itself, but should be viewed within the context of risk management more generally. For a broader picture see Perry (1986), Fong (1987), and Flanagan (1990).

The paper is divided into three parts. Part One provides a background to the paper. It examines the notions of risk and uncertainty, and explores how current practice deals with assessing risk exposure. It also establishes a basic premise of this paper, that conventional estimating practice could and should form an intrinsic part of risk management. The critical

need for non-deterministic methods of estimating does not exclude conventional practice as an effective platform on which to establish full risk management practices.

Part Two looks in detail at two methods of analysing risk exposure. The first uses a conventional statistical method to describe the extent of the risk exposure, directly from the cost data. The second uses Monte Carlo simulation to generate a description of the risk exposure, that uses the cost data only indirectly. The intention is to provide immediate access to two basic implementations of two simple techniques for the analysis of risk exposure. Both techniques will be described in practical terms. Descriptions come complete with a 'starter-pack' of the formulae and computer coding needed to apply the techniques in practice. (Note: the computer coding is provided in Appendix A: Algorithms to compute the standard deviation, and to select values from a triangular probability distribution in a Monte Carlo simulation.)

Part Three applies the two methods described in the previous section. Data on the cost of high quality offices in Sydney, Australia is analysed to determine risk exposure using both methods, and the results are assessed.

Part one: Background

Cost and price

When we speak of cost in estimating we most often actually mean price. A 'cost' relates directly to the goods and services consumed in the production of a commodity, a 'price' reflects the amount we are prepared to pay in exchange for that commodity. Price therefore involves the additional notions of utility and value, and appeals (in a market economy) to the forces of demand and supply in determining an appropriate pricing level. The title of this paper should more correctly refer to the 'price' estimates of high quality offices, but the terms are used here synonymously.

The cost estimate

A cost estimate (as considered in this paper) is intended to predict the total anticipated expenditure required by the client, to complete a particular building project. Of course, the proposal is likely to undergo various modifications and changes during the design process, and be subject to variations and price rises during construction. The cost estimate will therefore include an allowance, commonly in the form of a contingency sum or percentage addition, for any subsequent modifications or changes to the proposal.

The data to be used in this paper comes from a set of eight final account costs, broken into principal cost elements. The risk exposure is a measure of the probability that a given cost estimate will be less than the actual costs incurred by the client. The higher the probability, the larger the risk exposure.

Risk and uncertainty

There is some disagreement in the literature regarding the distinction between risk and uncertainty. There are those who maintain that risk and uncertainty are necessary distinctions (American Association of Cost Engineers, 1983; Black, 1984; Hillebrandt, 1985).

In this case, risk is used to refer to measurable quantities, and uncertainty is associated with the bits we cannot measure objectively. The distinction places limits on the application of analytical technqiues, by recognizing that certain events cannot be predicted quantitatively within useful limits. The counter argument to this (O'Donnell and Rhodes, 1983; Perry and Hayes, 1985) recognizes that risk and uncertainty are inevitably defined in terms of one another, and that to distinguish them might even be unhelpful. For example, 'risk is a measurable uncertainty while uncertainty is an unmeasurable risk' would usefully summarize the distinction, but tells us little of use.

To add to this confusion, an entire field of study has grown around the combined notions of risk and uncertainty, commonly referred to as 'risk management' (Perry, 1986; Fong, 1987; Flanagan, 1990). Risk management is a methodology developed for the identification, evaluation and control (as far as it is possible) of the risk and uncertainty in selected, non-deterministic processes. The focus of this paper is on a single aspect of the evaluation part only, of risk management. In an attempt to highlight this narrower focus, the term 'risk exposure' is used.

Risk exposure in cost estimation

There should be little doubt that differences can and do arise between cost estimates and final costs. (Morrison and Stevens, 1981; Ashworth and Skitmore, 1983). In statistics, these differences are referred to as 'error'. This is an inappropriate term when applied to building, where the cost estimates and final costs can describe very different building projects. A building proposal inevitably will develop during the course of design and construction, requiring corrections, clarification, and involving particular items not originally included in the drawings or not thought of. In preference to the term 'error', this paper will therefore use the term 'mismatch'. Like most things, cost estimating involves some risk of mismatch.

In certain situations, deterministic methods still prove to be a quite satisfactory means of dealing with risk. Those differences that do arise might either be small enough to be ignored, or be sufficiently constant to a be accounted for in some form of percentage addition. For certain kinds of work, such as repetitive housing, estimators are able to predict costs relatively accurately. Even where estimates are prone to a more significant mismatch, for many years the degree of mismatch has been fairly constant and predictable. Studies in the UK have indicated a mismatch somewhere between 15–20%, improving perhaps to around a 13–18% coefficient of variation at detailed design stage, immediately prior to receiving tenders (Ogunlana and Thorpe, 1987). In consequence, it is common practice to include a percentage addition around 20%, to cover for all manner of potential variations in an estimate.

Of late however, many large and complex projects have varied considerably in cost. Variations of the order of 50%, 200% and significantly higher are no longer unheard of. Variations of that order simply cannot be handled with a general percentage addition. The required cost additions across the board would be too prohibitive in most cases. Rather, each percentage addition must be determined on its own merits, and that demands an analysis of the risk.

In addition, the scale of percentage additions being applied has already grown. It is not unusual to find additions of 25–30% in Australia, and clients are beginning to demand substantiation of such additions. The high percentages may in fact be justifiable, but any justification first requires an analysis of the risk and uncertainty.

Current practice, which uses a percentage addition to cover subsequent design or construction changes, is based on deterministic methods. Such methods have served well under stable conditions, but as the scale and range of variations increase, the utility of this approach is lost. With large-scale and wide-ranging variations, the need is for an explicit assessment of the risk exposure, and deterministic methods are simply unable to provide this. The situation demands non-deterministic, stochastic methods.

Non-deterministic methods are often suggested as an alternative, or replacement to conventional practice (Mathur, 1982; Bowen *et al.*, 1987). But non-deterministic methods of estimating do not, of necessity, deny conventional practice. On the contrary, for certain non-deterministic methods, conventional estimating practice may provide the only effective source of data. The wealth of established expertise in conventional practice make this a particularly attractive possibility. A basic premise of this paper is that conventional estimating practice could and should form an intrinsic part of analysing risk exposure. In Part Three of this paper conventional practice is taken as the platform from which to establish the full range of risk management practices.

Part two: methods

The conventional statistical method

The conventional statistical method of analysis should be reasonably familiar to any graduate in construction management and economics. None-the-less, it is sometimes difficult to extract the requisite formulae for a particular application, and this paper seeks to do just that. Only relatively simple methods are considered here, having well-defined limitations. It is generally possible to avoid the limitations in more sophisticated applications (see, for example, Beeston, 1983).

Statistics do at least two important things for estimators. Firstly, from the confines of a small data sample, they provide a window on to the wider world of the population at large. This is particularly useful in cost estimating, since there is no possibility of collecting every piece of cost data. All possible cost data (the 'population') has to be represented within a small selection (the 'sample'). Secondly, statistics describe how likely specific values are to occur within the range of possible values. The population is represented, and the likelihoods of specific values are described, in the form of a frequency distribution.

Figure 1 illustrates the frequency distribution for a given data set. A frequency distribution can be represented either as a series of discrete frequencies (the boxes) or as a continuous line (also termed a probability distribution where the scale of the axis is adjusted so that the positive area under the line sums to 1). Where the frequencies are added to each other across the range of possible values, we have cumulative frequency curve. The cumulative frequency curve enables probabilities to be determined more easily.

To determine risk exposure, trace the vertical line from the value under consideration to its intersection with the cumulative frequency line, and then horizontally across to the risk exposure axis. For example, in Fig. 1, say, the value 5 has a risk exposure of 0.65.

As an alternative to carrying around a complete data set, statistics provides a short-hand description of the frequency distribution. Many enhancements are possible, but the two most usual descriptions of a frequency distribution are the mean and the standard deviation.
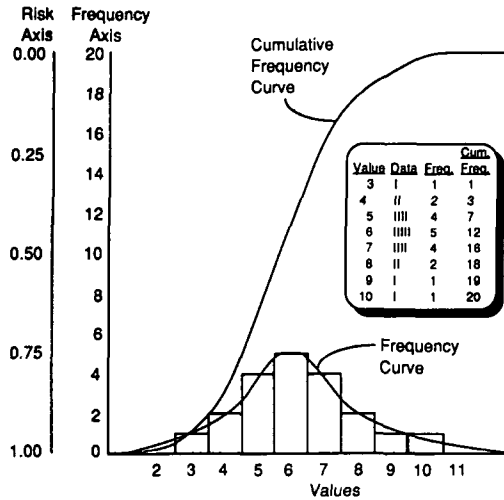
Fig. 1. An example frequency distribution and cumulative frequency curve for the data given. Figures are for illustration only.

The mean of a frequency distribution is more commonly called the arithmetic average. It is calculated by adding all values together and dividing by the number of values.

$$\text{mean of a data set} = \frac{\text{(the sum of all values in the data set)}}{\text{(the number of values in the data set)}} \tag{1}$$

The standard deviation gives a measure of the spread of a distribution. Its formal statement is rather a mouthful. The simplest formulation sums the data, then sums the squares of the data. From the sum of the squares is subtracted the average of the square of the data sum (*the sum multiplied by itself and divided by the number of values*). All of this is then divided by the number of values. The square root of the result is the standard deviation (see also Appendix A)

$$\text{SD} = \sqrt{\frac{((\text{sum of squares} - (\text{square of sum/number of values}))}{(\text{number of values})}} \tag{2}$$

However, this only provides the standard deviation for the sample data set. To approximate the standard deviation for the data population, the sample standard deviation can be adjusted using Bessel's correction. Without the Bessel's correction, the standard deviation tends to be too small (making it over-optimistic). The correction is a function of the sample size.

$$\text{Bessel's correction} = \text{SD} \times \sqrt{(\text{number of values/(number of values} - 1))} \tag{3}$$

Unfortunately the significance of the standard deviation value is related to the mean of its distribution. A standard deviation of 5 would be far more significant (i.e. represent a relatively greater spread) in a distribution where the mean was 10 than in one where the mean

value was 100. To enable some comparison of spreads between distributions with varying mean values, we must calculate the coefficient of variation. The coefficient of variation is the standard deviation expressed as a percentage of its mean.

$$\text{Coefficient of variation} = (\text{SD}/\text{mean}) \times 100 \qquad (4)$$

In summary, these three statistical measures provide a general description of the uncertainty in a sample set of cost data. The mean locates the distribution of uncertainty on a range of possible values. The standard deviation indicates the extent of the spread of uncertainty away from the mean. The coefficient of variation enables a comparison of standard deviations across distributions with different mean values.

In practice these three measures can be used in a variety of ways, as demonstrated in Part Three of this paper. The measures themselves describe the distribution of costs. For those estimators familiar with the characteristic values for equivalent cost data, the distribution can provide a useful indicator of risk. It is a simple step from these measures to generate a potential cost range. Further, a comparison of the various coefficients of variation for different elements will highlight those elements in a cost estimate build-up responsible for the major centres of risk exposure. The most significant elements are those with the largest standard deviations. Finally, the frequency distribution for total costs can be used to provide a measure of the risk exposure.

The Monte Carlo simulation method

The important constraint placed on construction cost data is the small size of the available data samples. Statistical methods tend to work better the more representative the sample is of the population. One way of getting a more representative sample is by collecting more and more data, and creating larger data sets. If, as with construction cost, the data is not available in sufficiently large numbers (buildings tend to be 'one-offs'), then the only real alternative is to 'create' a data set by generating the data. That is, forget about a data set that consists only of 'actual' costs (the cost data from buildings that actually exist). Rather, the data set is a collection of hypothetical data, about non-existing buildings. Importantly however, the hypothetical data (in a statistical sense) would come from the same family of data as would the actual costs. The statistical description of the hypothetical data set would be the same as the statistical description of the actual cost data (except the hypothetical data set would be much larger). This generative process can be done quite readily, based on the probability distributions of component, elemental costs. The most popular of the techniques used in this way is Monte Carlo simulation.

A further distinction between the conventional statistical method and the Monte Carlo simulation method comes from how they each determine the final distribution of risk exposure. In the conventional statistical method, uncertainty is inherent in the data set and is carried through to the final distribution in a single analysis. In the Monte Carlo simulation method, uncertainty is also inherent in the data set, but it is transferred to the final distribution through a large number of analyses. It is this large number of analyses that provide the additional, 'hypothetical' data.

The Monte Carlo simulation method begins by describing the probability distributions of the component parts. In the case of a cost estimate, the component parts might usefully accord with the cost elements. (Some of the problems in using cost elements in this way, are

considered later in this paper.) As already suggested, it is possible to describe each elemental cost distribution in terms of its mean and standard deviation. This has several shortcomings in the context of Monte Carlo simulation. Instead, it is more usual to describe each distribution by the specification of three (triangulated) points: the minimum (or most optimistic) value, the most likely (or conventional estimate) value, and the maximum (or most pessimistic) value. The triangular distribution approximates well to the more general distribution shapes, and has the considerable benefit of simple computation. In this case, the minimum value is intended to indicate the point with a 97.5% likelihood of being exceeded. The maximum value indicates a 2.5% likelihood of being exceeded. These two limits mean that the resultant distribution covers 95% of cases. (Note: The extent of coverage can be chosen somewhat arbitrarily, since the choice is relatively unimportant in an absolute sense as long as the estimator is consistent. For further discussion, see Alberts, 1972).

The actual values for minimum, most likely and maximum can be determined in several ways. The most straightforward method is simply to select the values subjectively, relying upon the expertise of the estimator to determine resonable values. Where even limited data are available, the expertise of the estimator might very usefully be supplemented with graphs on which the data for each component has been plotted. In the unusual situation where sufficient data is available, a full statistical analysis can be used to predict the required values. Besides these three general approaches, other methods are also possible (see Smith, 1973).

Once triangular distributions have been determined for each component, simulation can begin. The Monte Carlo simulation method steps through each distribution, determining a single value from that distribution at random. These values are then used in a conventional way to calculate a total cost for that particular 'run'. At the end of each run the total cost estimate is recorded, prior to repeating the entire process over multiple runs. Typically there would be 50–200 (but up to 500) runs completed in any simulation. A larger number of runs than this gives only a marginal increase in accuracy, and may be wildly misrepresentative of the accuracy associated with the data used in its formulation.

The use of the term random here may also be misleading. A random number is used to select a value, but the selection process ensures that the frequency with which values are selected conforms to the appropriate triangular distribution. (See Appendix A for an algorithm to calculate these values.)

To complete a simulation, the results from all runs are aggregated into a frequency distribution (such as the one illustrated in Fig. 1). This frequency distribution forms the basis of a cumulative frequency graph from which probability (and thereby risk exposure) can be determined. To determine the risk exposure associated with a particular estimate value, a vertical line is extended upwards from the value point until it intersects the cumulative frequency line. The point of intersection is then extended horizontally until it meets the frequency axis, suitably converted to a risk exposure axis. The value at that location indicates the probability of that estimate value being exceeded, and therefore indicates the risk exposure.

Part three: practice

The problem

This part details specific data on eight high quality office buildings, constructed in Sydney

between June 1985 and April 1990. The eight cost analyses are used as the basis of an estimating process in which the inherent uncertainty is made explicit. The data (and our working definition of high quality offices) is taken from a report on the cost of prestige office buildings prepared by the Sydney office of Rider Hunt and Partners, quantity surveyors and construction cost consultants (Rider Hunt, 1990).

For the purposes of this exercise, the total cost is broken down into the 19 general elements that accord with current Australian practice. Statistical studies have established that elemental costs tend to be interdependent. If, for example, wall finishes are expensive, one might also anticipate expensive floor and ceiling finishes. Interdependence of this kind is termed correlation. Correlations can invalidate certain statistical methods, and can bias results in the Monte Carlo simulation method. This latter point is considered in more detail in the assessment of the Monte Carlo simulation method, given later in this paper.

The basic cost information is summarized in Table 1. All cost information given in this paper applies to high quality offices located in the Sydney CBD, and is current at December 1990. Figures are in Australian dollars per square metre of gross floor area.

Table 1. Elemental cost information for eight high quality office buildings constructed in Sydney, Australia between June 1985 and April 1990. Figures are in Australian dollars, expressed per square metre of gross floor area, and current at December 1990.

| Element | Office building | | | | | | | |
	1	2	3	4	5	6	7	8
Site preparation	48	7	9	25	32	9	2	16
Sub-structure	78	35	43	25	53	29	63	37
Super-structure								
Frame	467	291	278	322	400	528	287	242
Staircases	17	10	16	13	9	13	17	9
Roof	59	51	49	36	60	40	29	67
Extl. walls	677	316	340	407	578	495	349	245
Intl. walls	171	213	182	165	181	110	148	140
Finishes								
Wall	23	37	63	122	124	121	29	30
Floor	158	52	69	90	127	92	58	55
Ceiling	41	58	56	64	55	77	62	44
Fitments	24	18	12	9	25	19	19	8
Services								
Plumbing	104	55	53	66	85	80	50	56
Mechanical	290	204	214	222	252	207	216	177
Fire	78	55	63	61	38	58	55	58
Electrical	151	126	171	205	264	115	130	256
Lifts	223	149	200	159	174	179	171	167
Special	160	62	299	19	37	172	91	126
Site works	87	62	27	2	8	2	45	3
Sub-total	2856	1800	2144	2012	2502	2345	1821	1736
Preliminaries	522	297	421	339	627	547	399	540
as a %	18.3	16.5	19.7	16.8	25.1	23.3	21.9	31.1
Total	3378	2097	2565	2350	3129	2892	2220	2276

Source: Report on Cost of Prestige Office Buildings, Sydney Office of Rider Hunt, March 1990.

Against the predictions of cost uncertainty made in this paper, there are five comparable cost estimates (see Table 2). The first two comparable estimates come from published prices. Estimate 1 comes from Cordell's Building Cost Guide for New South Wales, and Estimate 2 comes from Rawlinson's *Australian Construction Handbook*. Estimate 3 is a model cost prediction prepared by Rider Hunt to accompany their report (Rider Hunt, 1990). The latter prediction, particularly, is intended for comparative studies rather than to

Table 2. The five cost estimates for high quality office buildings used as comparables. Figures are in Australian dollars, expressed per square metre of gross floor area, and current at December 1990.

Estimate 1	Cordell's	$2346
Estimate 2	Rawlinson's	$2333
Estimate 3	Base model (Rider Hunt)	$2008
Estimate 4	Typical model (Rider Hunt)	$2590
Estimate 5	Full model (Rider Hunt)	$3029

establish a specific budget. It is therefore presented as a general purpose office building that reflects the lower cost range of what might be achievable, and is accompanied by 'optional extra costs'. Estimate 4 is a more typical arrangement of the Rider Hunt model (or, Estimate 3 with certain of the optional extras). Estimate 5 is the Rider Hunt model with all possible optional extras. (Note: in the Rider Hunt document, other cost adjustments are also analysed and presented, but their consideration here is beyond the needs of this paper.)

The conventional statistical method

Given the cost data in Table 1, it is possible to make a statistical analysis of the risk exposure. Table 3 shows the calculated mean, standard deviation for the sample, standard deviation for the population and coefficient of variation for the sample. The figures clearly indicate where the uncertainty lies. For example, the coefficient of variation indicates several elements where there are considerable variations of cost relative to the mean. These are the Site Preparation, Wall Finishes, Special Services and Site Works. Note also the effects of 'swings-and-roundabouts' (or 'plusses offsetting minuses') in producing a relatively low coefficient of variation for the total costs.

To identify where the significant absolute variations occur, it is better to focus on the standard deviation. The large effects of variation in the elements Frame, External Walls and Preliminaries (those with a large standard deviation) are significant in determining the overall uncertainty. Particular attention should be given to any such significant elements should a more certain estimate be required.

The usual approach to translating a statistical analysis into some form of estimate is to produce a cost range. The upper and lower limits of the range are generally determined from a normal probability distribution, and aim to cover 90% of the cases. (Note: the 90% figure is largely arbitrary, but is fairly typical of the recommended values. See Beeston, 1983.) Based on Table 3, the range of total costs in this case would be approximately $1845–3381. That is, the mean ($2613) minus and plus (from normal curve areas) 1.64 × the standard deviation of the population (1.64 × $468 = $768).

Table 3. Statistical analysis of the cost data for eight high quality office
buildings constructed in Sydney, Australia between June 1985 and April
1990. Figures are in Australian dollars, expressed per square metre of
gross floor area, and current at December 1990.

Element	Mean	SD sam	SD pop	Coef. var (%)
Site preparation	19	15	16	82.89
Sub-structure	45	17	18	39.74
Super-structure				
Frame	352	95	102	28.99
Staircases	13	3	3	24.02
Roof	49	12	13	26.69
Extl. walls	426	137	146	34.32
Intl. walls	164	29	31	19.02
Finishes				
Wall	69	43	46	66.80
Floor	88	36	38	43.40
Ceiling	57	11	11	19.82
Fitments	17	6	6	38.09
Services				
Plumbing	69	18	19	27.93
Mechanical	223	32	34	15.37
Fire	58	10	11	19.03
Electrical	177	55	59	33.00
Lifts	178	22	24	13.34
Special	121	85	91	75.22
Site works	30	30	32	108.62
Sub-total	2152	368	393	18.27
Preliminaries	462	107	115	24.87
as a %	21.58	4.60	5	22.76
Total	2613	438	468	17.92

Importantly however, the interpretation of the range assumes a straight-line probability distribution. Any specific estimate is located within the range, proportionally. The wider the range, relatively, the greater the uncertainty. The cost range is therefore a very coarse indicator, and an alternative form of interpretation can give a considerably better impression of likelihoods.

One alternative approach is to assume a normal shaped probability distribution to both determine and interpret the cost distribution. The mean is then a value located so that 50% of values lie below and 50% of values lie above that figure. As the value is increased, more of the distribution lies below, and less above. As the value is decreased, less of the distribution lies below, and more above. The rate of change in the areas below and above is not proportional (as the cost range assumes), but is determined by the standard deviation. For example, at a point equal to the mean value plus one standard deviation, there would be 84% of the distribution below and 16% above. It is then possible to equate the areas with probabilities using normal curve values, and from the probabilities to determine the risk exposure.

Using the values from Table 3, it can be stated that an estimate of, say, $3081 (the mean, $2613, plus one standard deviation, $468) will be exceeded only 16% of the time. Alternatively, the risk exposure associated with an estimate of $3081 is 0.16 (where 0 represents little or no risk and 1 represents a very high risk).

Figure 2 shows the relative location of the five comparables from Table 2, with the mean from Table 3. Also shown is a 'Most Likely' figure. This is derived from Table 4, and will be explained later. To determine risk exposure, trace the vertical line from each comparable value to its intersection with the cumulative frequency line, and then horizontally across to the risk exposure axis. The location of these comparables highlights the potentially high risk exposure associated with certain estimates. For example, based on these calculations (which are specific to the high quality office data being used), the risk of being under the actual price with the published prices (Estimate 1 and Estimate 2) is in excess of 0.7 (rather high).
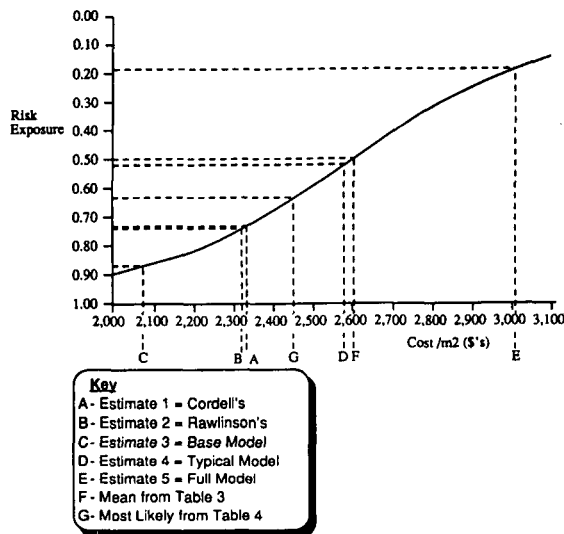

Fig. 2. The location of seven comparable cost estimates using a conventional statistical method. The distribution assumes a normal probability curve based on a mean of $2613 and a standard deviation of $468. Figures are in Australian dollars, expressed per square metre of gross floor area, and current at December 1990.

Assessment

The major concern in a statistical study such as this one is often the low number of data values on which it is based. That, at least, is the concern when viewed from the statistical point of view. However, when the size of the data sample is considered relative to the conventional estimating approach, the concern is considerably lessened. Conventional estimating is often based on only a single cost analysis and significant judgement. From this perspective even a small sample is an improvement. So the concerns about small data samples remain, but should not prohibit the application of statistical methods.

The list of coefficients of variation in Table 3 illustrates the scale of variation (uncertainty)

inherent in current cost data. In particular it shows that the 17.92% for total costs is low in comparison to those of the individual elements. Further, it is reasonable to surmise that the many studies of estimating accuracy, based as most of them are on total cost comaparisons, give an optimistic (too small) figure for estimating accuracy. The measure of accuracy for estimates at the elemental level is likely to be even greater than those for total costs. This pattern lends further weight to the need for risk exposure to be dealt with more explicitly. It also raises the possibility that the widespread application of generic rates per square metre of gross floor area might be intrinsically suspect.

From the statistical description given for total costs in Table 3, various indicators of uncertainty can be produced. The production of a mere cost range from such a description seems wasteful, when a more useful indication of probability (such as Fig. 2) can so readily be determined. The statistical probability indicated in Fig. 2 assumes a normal distribution curve, although the curve is known in reality to be skewed. The actual probabilites are therefore going to be somewhat different from those illustrated. However, the relative positions of the comparable estimate figures used here will not be affected by distribution shape. Use of the published cost data in this case will inevitably have a higher risk exposure. It is a simple exercise to determine the risk exposure for any proposed estimate figure.

The Monte Carlo simulation method

The Monte Carlo simulation method describes a probability distribution for each element. The description in this case will use three points: the minimum, most likely and maximum values. How are these three values determined?

The 'most likely' value is equivalent to the deterministic, single figure estimate of a conventional estimating approach. This figure may be determined using some objective data analysis technique, but it is more often arrived at based on expert judgement. Naturally, the data in Table 1 would be considered, but the 'most likely' value is inevitably determined subjectively. (There is simply not the data available to support an analytical approach.)

Similarly, it is not simply a case of selecting the minimum and maximum values for the distribution, to correspond to the minimum and maximum data values. For the purpose of this paper, the range between minimum and maximum is intended to cover 95% of cases (where all cases would include the data set of Table 1, and the more general experience of the costing expert). Again, a degree of expert interpretation is called for.

To help in each process, the data for an element can at least be graphed. This simply means, for each element, placing a mark for every data value on a range of possible values. Figure 3 shows this representation for the cost data relating to the element sub-structure. The minimum, most likely and maximum points are then selected to produce a distribution most representative of the element cost.

Relevant values for each element distribution are presented in Table 4. These values are input to the Monte Carlo simulation algorithm. After running the system 500 times and aggregating the results into 25 frequency groups, the graph in Figure 4 is produced. This graph details both the frequency distribution and the cumulative frequency distribution for the total costs of high quality office buildings. Notice that despite the skewed nature of the input distributions, the result for total costs is relatively normal.

To determine risk exposure, trace the vertical line from each comparable value to its intersection with the cumulative frequency line, and then horizontally across to the risk exposure axis. Thus, on this analysis, the published prices ('A', 'B' and 'C') have a high risk
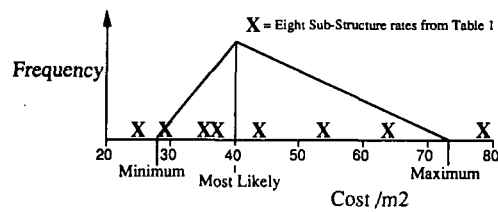
Fig. 3. Representation of data relating to the element sub-structures. Figures are in Australian dollars, expressed per square metre of gross floor area, and current at December 1990. The minimum and maximum are selected to include 95% of all cases. All cases including those indicated, and those of the more general experience of the costing expert.

Table 4. The minimum, most likely and maximum values determined for each element from cost data on eight high quality office buildings constructed in Sydney, Australia between June 1985 and April 1990. Figures are in Australian dollars, expressed per square metre of gross floor area, and current at December 1990. A statistical description of each distribution is also provided.

| Element	Min	Likely	Max	Mean	SD	Coef. Var (%)
Site preparation	2	10	42	14	10	71.43
Sub-structure	28	40	73	44	11	25.86
Super-structure						
Frame	250	310	500	332	63	18.84
Staircases	8	13	18	13	3	19.23
Roof	30	51	66	50	9	18.00
Extl. walls	260	370	620	393	90	22.88
Intl. walls	110	167	210	165	25	15.18
Finishes						
Wall	22	70	130	72	27	37.50
Floor	54	70	150	81	24	29.75
Ceiling	42	56	74	57	8	14.12
Fitments	8	19	25	18	4	23.39
Services						
Plumbing	50	58	94	63	11	17.55
Mechanical	182	216	284	222	26	11.50
Fire	40	57	75	57	9	15.31
Electrical	120	170	260	177	35	19.81
Lifts	151	172	218	176	17	9.51
Special	28	105	260	118	58	49.15
Site works	1	25	77	30	19	64.04
Sub-total	1386	1979	3176	2080	448	21.52
Preliminaries	300	500	620	487	80	16.44
as a %	16.50	20.00	29.00	21	3	14.94
Total (Abs. prel)		2479				
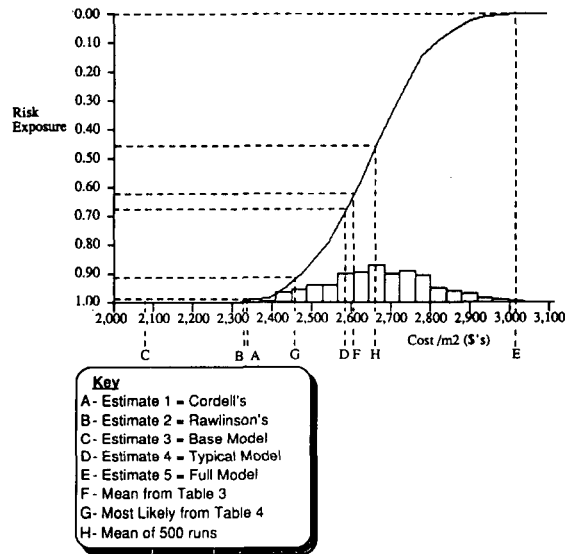
Fig. 4. The results of a Monte Carlo simulation. The graph indicates the location of comparable cost estimates for high quality office buildings. The results are plotted as a frequency histogram and cummulative frequency curve. The frequency axis has been adjusted to represent risk exposure. Figures are in Australian dollars, expresed per square metre of gross floor area, and current at December 1990.

exposure: there is a high probability that the estimates will be less than the actual price. A less risky estimate would be either 'D' (Estimate 4, the Rider Hunt basic value plus typical additions), 'F' (the mean derived from Table 3), or 'H' (the mean derived from the 500 simulation runs). To avoid all risk, the estimate 'E' (a figure in excess of $3000) would be required.

It is interesting that 'G' (the total of the 'Most Likely' column from Table 4, and therefore what would have been the conventional estimate) has a risk exposure in excess of 0.9. The high risk exposure indicates considerable skewness in the elemental probability distributions developed for Table 4. The central limit theory maintains that combining a number of probability distributions will tend to produce a normalized curve. This is apparent from the frequency distribution in Fig. 4, and because the mean of the distribution ('H') is located close to the 0.5 risk level.

The purpose of this exercise is not to determine a 'best' estimate directly. Rather, Fig. 4 indicates the risk exposure associated with any proposed estimate. The 'correct' estimate is then specific to the client's general attitude to risk, and should be assessed within the context of given values for all other sources of risk (i.e. it involves risk management). The exercise is not an end in itself.

Assessment

There are several problems with the Monte Carlo simulation method as presented in this paper. One of the most apparent is the use of a single form of probability curve. The triangular distribution used here approximates well with the most common forms of

distribution, but several alternatives do exist (Raftery, 1985). Fortunately, studies, have indicated that the exact choice of distribution shape for each element has only a small effect on the distribution produced for the total after several hundred runs (Pouliquen, 1970).

The same studies indicated that correlation between the variables (in our case, between the elements) had a far greater significance. However, those results may be slightly misleading, since the test variables used in that case were highly correlated. In our case the element rates may be correlated, but they do not appear to be as highly correlated as those in the test studies. Where high correlation is found to occur, use of Monte Carlo simulation is still possible, but the simulation process is rather more complex than the one described here.

Of more particular concern in this application of the technique is the highly subjective way in which the limits for the distributions given in Table 4 are determined. No two estimators are going to produce the same figures. Whenever there is insufficient data to determine the appropriate values objectively (which in cost estimation will be often), this limitation of the technique will remain. Note however, that the current alternative of a deterministic estimate tends also to be highly subjective. To date, there has been little study of the sensitivity of results to changes in the input values. Based on a small excercise undertaken for this paper, the sensitivity does not appear large.

One alternative to forcing subjective values into hard-edged numbers, is an approach to uncertainty called 'fuzzy sets' (Zadeh, 1973). Fuzzy set theory questions the premise on which conventional probability theory accounts for the uncertainty in an estimate. The principal change involves replacing the numerical values used conventionally, with so called 'lingusitic variables' (Zadeh, 1975). The approach has been implemented for risk analysis to some effect by Schmucker (1984), Turunen (1984) and Anderson (1988). However, the approach appears to introduce several problems of its own, and is rather circular in its theoretical basis (Newton, 1983).

Finally, the use of maximum and minimum limits at points covering only 95% of cases leads to a narrower range of values for the Monte Carlo simulation method than for the conventional statistical method. This narrowing is most apparent from a comparison of the corresponding coefficients of variation in Tables 3 and 4. The Monte Carlo simulation method is therefore expected to produce a more optimistic (narrower) range of probabilities.

When the results of the 500 runs are analysed, they are found to have a mean of $2663, a standard deviation for the sample of 132.2, for the population of 132.3, and a coefficient of variation of only 4.97%. This extremely low value for the coefficient of variation should be cause for some concern when compared to the corresponding figure of 17.92% given in Table 3. However, these figures are not strictly comparable. In its simplest terms, the value from Table 3 is determined by considering values across the table. Monte Carlo simulation determines the spread by combining the various elements from Table 4. In a sense, it considers values down the table.

Beeston (1986) has proposed a simple mechanism for determining the relevant statistical values in a way more comparable to the Monte Carlo simulation method. The mean for the total is determined from the sum of the elemental means. The standard deviation for the total is determined from the square root of the sum of the squares of the elemental standard deviations.

$$\text{Mean} = \text{sum of the elemental means} \tag{5}$$

$$\text{SD} = \sqrt{(\text{sum of squares of each element SD})} \tag{6}$$

The results of this computation are shown in Table 5. Computed in this way, the relevant values are far more comparable and less cause for concern.

Table 5. The statistical descriptions for total costs computed by combining elemental values. Figures are in Australian dollars, expressed per square metre of gross floor area, and current at December 1990.

Source	Mean	SD sam	SD pop	Coef. var (%)
500 runs	2663	132.2	132.3	4.97
Table 3	2613	239.0	255.5	9.78
Table 4	2566	163.7	n/a	6.38

Conclusion

Two methods of analysing the risk exposure in a cost estimate have been considered: a conventional statistical method and a Monte Carlo simulation method. Both methods analyse risk in terms of a probability distribution. The major distinction between the two is that the statistical approach depends on having adequate actual data, whereas the Monte Carlo simulation approach generates its own data set. In building cost estimating this tends to favour the Monte Carlo simulation method, since large data samples are rare.

A data sample of eight high quality office buildings has been used as the basis for this paper. The range of total costs in the sample is high. The figures themselves range from $2097 to $3378 (a range of 61% on $2907). Analysed statistically, the 90% limits range from $1845 to $3381 (a range of 83% on $1845). However, a coefficient of variation of 17.92% reflects the tighter grouping of the data overall – something the simple figures for range fail to identify. Unfortunately the corresponding coefficient of variations for the elemental costs clearly demonstrate the dependence of cost estimating accuracy on a 'plusses offsetting minuses' effect.

It is difficult to avoid terms such as 'accuracy' and 'error' when describing frequency distributions. However, while it may be useful to describe the variations in statistical terms, the variations themselves are caused in a very different way. Buildings from the same data sample have different specifications for the same elements, are constructed at different points in the economic cycle, with different expectations of cost, and often very different management and design teams. The point being, that differences in cost are not necessarily the result of 'error', and in preference to that term this paper uses the term 'mismatch'.

Conventional practice deals with the potential mismatch between estimate and final cost with an all-embracing contingency sum, or percentage addition. This approach has served well during times of relative calm and stability in the construction industry, but it is failing to meet current demands. The large and unpredictable variations in cost require a more explicit treatment of risk. A similar need is being recognized in the sister field of property investment, where the use of the all risks yield has been sternly challenged (MacLeary and Nanthakumaran, 1988).

Non-determinisitic methods have been proposed in the past, but generally as an alternative to conventional practice. The non-deterministic approaches presented in this paper make no such exclusion of conventional, deterministic practice. Indeed, the lack of suitable data would strongly suggest that conventional practice, based on expert judgement as well as actual data, is particularly well suited to cost estimating.

The simple techniques of statistics and Monte Carlo simulation being considered here, are not new either. Statistics has a long theoretical pedigree, and the notion of probabilistic cost estimating goes back at least to a paper by Spooner (1974). Statistics are already included within various cost estimating programmes (Brandon, 1985; Pegg, 1987). Monte Carlo simulation has been applied to resource models (Baxendale, 1984), network planning (Bennett and Ormerod, 1984) and bid analysis (Raftery, 1985), and proposed for cost modelling (Wilson, 1982) and cost planning (Mathur, 1982).

Genuine concerns remain however, when applying the simple techniques described here to real projects. In a sense, it is all too easy. Building cost exhibits few of the 'clean' features required for simple statistics. With the Monte Carlo simulation approach, for example, there is little firm evidence to support or challenge the degree of asymmetry assumed in the elemental probability distributions. Neither is there any real clarity regarding the extent and implications of correlations between element costs. Both concerns warrant some critical research attention. From a statistical point of view, they are potentially prohibitive to the application of such techniques in cost estimating. But there are other points of view. Relative to conventional practice, where asymmetry and correlation problems simply get ignored, the application of statistical and Monte Carlo simulation methods is to be encouraged.

An analysis of cost exposure is only one small part of risk management, but it is a necessary and foundational part. The possibility of conventional practice forming the platform for risk management has several attractions. First, the majority of existing cost data could be included in the process without modification. Second, there is a wealth of experience in the application of conventional estimating techniques. Third, conventional estimating is a recognized and well-developed expertise. Finally, and perhaps most importantly, estimators are already familiar with, and confident in, the production of a deterministic estimate.

Applied to the specific domain of high quality office buildings, in a very particular way, both methods of analysis have highlighted the significant exposure to risk attendant on 'typical' estimates. On one analysis, in one context, it would be incorrect to draw conclusions about the so-called 'book prices' or model prices. However, the potential risk exposure highlighted in this study should be cause for some concern. It does not negate the use of any particular estimate. Risk exposure is simply another aspect of an estimate to consider, but it is an important aspect. The dynamics of building cost now demand we move more fully to embrace the ideas of risk management.

References

Alberts, R.J. (1972) Probabilistic elements of cost estimating for buildings, *Research Report R72-4*, Department of Civil Engineering, Massachusetts Insitute of Technology, Boston.

American Association of Cost Engineers, (1983) *Cost Engineer's Notebook*. AACE, Morgantown, West Virginia.

Anderson, L. (1988) *The theory of possibility and fuzzy sets: new ideas for risk analysis and decision making*. Swedish Council for Building Research, Stockholm, Sweden.

Ashworth, A. and Skitmore, R.M. (1983) Accuracy in estimating, *CIOB Occasional Paper No. 27*. CIOB, London.

Baxendale, T. (1984) Construction resource models by Monte Carlo simulation, *Construction Management and Economics*, **2**, 201–17.

Beeston, D.T. (1983) *Statistical Methods for Building Price Data*. E. & F.N. Spon, London.

Beeston, D.T. (1986) Combining risks in estimating, *Construction Management and Economics*, **4**, 75–9.

Bennett, J. and Ormerod, R.N. (1984) Simulation applied to construction projects, *Construction Management and Economics*, **2**, 225–63.

Black, J.H. (1984) *Cost Engineering Planning Techniques for Management*. Marcel Dekker, New York.

Bowen, P.A., Wolvaardt, J.S. and Taylor, R.G. (1987) Cost modelling: a process-modelling approach. In *Building Cost Modelling and Computers* (edited by P.S. Brandon). E. & F.N. Spon, London, pp. 387–95.

Brandon, P.S. (1985) Data on tap, *Chartered Quantity Surveyor*, **7**, 300–2.

Flanagan, R. (1990) *Risk Management and Construction*. Blackwell Scientific, London. (In preparation).

Fong, S.W. (1987) Risk management, *The Cost Engineer*, **25**, 12–15.

Hillbrandt, P.M. (1985) *Economic Theory and the Construction Industry*. Macmillan, London.

MacLeary, A.R. and Nanthakumaran, N. (eds) (1988) *Property Investment Theory*. E. & F. N. Spon, London.

Mathur, K. (1982) A probabilistic planning model. In *Building Cost Techniques: New Directions*, (edited by P. S. Brandon). E. & F.N. Spon, London, pp. 181–91.

Morrison, N. and Stevens, S. (1981) *Construction Cost Databases*. PSA Report, HMSO, London.

Newton, S. (1983) Analysis of Construction Economics, Unpublished PhD Thesis, Department of Architecture and Building Science, University of Strathclyde.

O'Donnell, A.T. and Rhodes, T.E. (1983) *Risk, uncertainty and public sector investment appraisal*. Public Services Economic Division Report, PSA, London, May, 1983, (Draft).

Ogunlana, O. and Thorpe, T. (1987) Design phase cost estimating: the state of the art, *International Journal of Construction Management and Technology*, **2**, 34–46.

Pegg, I.A. (1987) Computerised approximate estimating from the BCIS on-line data base. In *Building Cost Modelling and Computers* (edited by P.S. Brandon). E. & F.N. Spon, London, pp. 243–9.

Perry, J.G. (1986) Risk management – an approach for project managers, *Project Management*, **4**, 211–16.

Perry, J.G. and Hayes, R.W. (1985) Risk and its management in construction projects, *Proceedings Institute of Civil Engineers*, Engineering Management Group, Part 1, Vol. 78, June, pp. 499–521.

Pouliquen, L.Y. (1970) Risk analysis in project appraisal, *World Bank Occasional Paper No. 11*. John Hopkins Press, New York.

Raftery, J. (1985) *Risk analysis in construction tendering – a guide for building contractors*, Laboratory of Building Economics Report, Technical Research Centre of Finland, August 1985.

Rider Hunt (1990) *Report on cost of Prestige Office Buildings*, Sydney Office of Rider Hunt, March 1990.

Schmucker, K.J. (1984) *Fuzzy Sets, Natural Language Computations, and Risk Analysis*. Computer Science Press, Maryland, USA.

Smith, V.K. (1973) *Monte Carlo Methods: Their Role for Econometrics.* Lexington Books, London.

Spooner, J.E. (1974) Probabilistic estimating, Proceedings of the American Society of Civil Engineers, *Journal of the Construction Division,* **100,** 65–77.

Turunen, I. (1984) Modern Methods for Cost Estimation and Decision-Making in Preliminary Plant Design, Unpublished PhD Thesis, Department of Chemistry, Helsinki University of Technology.

Wilson, A. (1982) Experiments in probabilistic cost modelling. In *Building Cost Techniques: New Directions* (edited by P.S. Brandon). E. & F.N. Spon, London, pp. 169–80.

Zadeh, L.A. (1973) Outline of a new approach to the analysis of complex systems and decision processes, *IEEE transactions on Systems, Management and Cybernetics,* **SMC-3,** 28–44.

Zadeh, L.A. (1975) The concept of a linguistic variable and its application to approximate reasoning, *Information Sciences,* **8,** 199–249, (Part I), 301–57, (Part II), **9,** pp. 43–80, (Part III).

Appendix

Algorithms to compute the standard deviation, and to select values from a triangular probability distribution in a Monte Carlo simulation.

The algorithms are expressed in 'computerese', a functional description intended to be applicable in various computer programming languages or spreadsheets.

To calculate the standard deviation as expressed in Equation 2 of the main text:

variables
    Sum {for the sum of all values}
    Sum_Squares {for sum of the squares of all values}
    No_Vals {for the number of values in the data set}
    Stnd_Dev {for the standard deviation}
begin algorithm for standard deviation
    Sum = add up all values;
    Sum_Squares = add up all (values*values);
    Stnd_Dev = SQRT ((Sum_Squares − ((Sum*Sum)/No_Vals))/No_Vals);
    {SQRT is a standard function to calculate the square root}
end algorithm for standard deviation

To select a value at random from a triangular distribution, so that the frequency with which values are selected conform to the appropriate distribution:

variables
    Min, Likely, Max {three values for minimum, most likely and maximum}
    Proportion {a calculated proportion of 'skewness' in the distribution}
    Random {a random variable with values between 0 and 1 − most computer systems provide standard functions for this}
    Value {the 'random' value we require, taken from a given distribution}
begin algorithm for 'random' value
    Proportion = (Max − Min)/(Likely − Min);
    if Random < Proportion **then**
        Value = Min + SQRT (Random*(Likely − Min)*(Max − Min));
    else
        Value = Max − SQRT((1 − Random)*(Max − Likely)*(Max − Min));
    {SQRT is a standard function to calculate the square root}
end algorithm for 'random' value