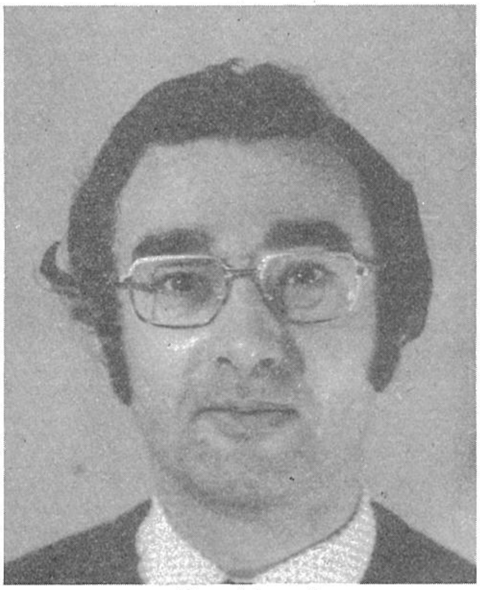
Cost Modelling for the Construction Industry

By A. Ashworth, MSc, AIQS, ACIArb



Allan Ashworth

INTRODUCTION

Many years ago quantity surveyors used to keep a "cube book" as an aid to providing cost advice for architects and clients. When a contract was signed the amount of the accepted tender was divided by the cubic content of the building and the resultant cost per cubic foot was entered in the book. In order to achieve some standardisation between projects rules prepared by the RIBA were adopted. One disadvantage of this method was that it did not relate to user requirements. The superficial area method was accepted as an alternative chiefly because of this factor. Other improvements to this method were developed using a combination of floor area and perimeter. In the mid-1950's a further advancement of these methods was the storey-enclosure method which calculated cost on a combination of floor, roof and wall areas. The unit method of approximate estimating has been applied successfully to those buildings where a relationship was designed to exist between the cost of the building and the number of functional units contained.

Each of these methods are in their simplest form, cost models, where cost is predicated against some formula. By the mid-1950's, however, the limitations of the single quantity rate approach was very apparent and improved methods of cost forecasting were introduced based upon approximate estimating and cost planning.

The disadvantage of the superficial area method as a model of cost is that cost is influenced by factors other than floor area alone. Examination of current data supplied by the Building Cost Information Service detail a considerable range of prices for each particular type of building analysed. The quantity surveyor in attempting to arrive at an approximate estimate of cost will subjectively amend these prices on the basis of his own experience and the particular circumstances surrounding the project concerned.

The new generation of cost models attempt to predict cost in a more objective manner largely eliminating the necessity for any subjective adjustment to the approximate estimate for the project. These models can be based upon data that have been inflation adjusted by using appropriate indices.

DEVELOPMENT OF A COST MODEL

1. Collection of Suitable Data

Every form of estimating relies heavily upon some type of suitable historic data. Before cost modelling can commence a sufficient quantity of accurate and reliable data needs to be obtained. The type of data selected may be for example; the manhours required per unit quantity of a particular construction operation, the prices from bills of quantities for elements or sections of work or the costs per square metre of floor area from a variety of building projects. The type of data collected will depend upon the type of model requiring to be constructed.

There are many difficulties associated with the collection of data, particularly when this requires the quantity surveyor to look beyond the projects he is controlling. Contractors are likely to be hesitant about supplying data because estimating is a sensitive area within their organisations.

In order to make the model statistically reliable a minimum amount of data will be required, and the more that can be procured the more satisfactory will be the final model in terms of reliability and accuracy. Because of the considerable variation in price between apparently identical bill items, this type of data is unsuitable for cost modelling purposes. The most fruitful areas therefore for the quantity surveyor working in private practice are either in respect of total building cost (not to be confused with costs-in-use although there is application here) or an elemental model such as reinforced concrete frames.

The type of data required will be in the form of historical costs such as tender sums, elemental analyses, etc. together with quantified variables that hopefully will describe these costs. For example, if we desired to build a model to predict the likely tender values of proposed projects we would need to assemble data from previously completed projects. This would include the tender sum together with those variable factors that influence the calculation of the tender sum. The variables considered might include for example; the gross internal floor area, roof areas, external wall area, shape factor, height, storey height etc., all of which can be quantified from initial sketch drawings. A minimum of thirty projects would be required for a suitable analysis to be achieved. The quantity surveyor would need to examine these projects in the same way that would do for traditional cost planning purposes. Factors such as regional location, contract conditions, type of competition, market factors, etc. would either be considered as suitable variables in the model or eliminated on the basis of the surveyors own knowledge. In order to keep the number of variables in the model to a minimum, factors such as inflation should be removed by use of indices. Many of the factors listed above can only be determined using the surveyors own expert knowledge of the projects concerned.

2. Types of Model

Many different types of cost models have already been constructed to suit cost predictions required at the different stages of the construction process. Some have been developed to forecast tender cost at the inception stage of the project, whereas others have been used in the cost planning process based upon the elemental sub-divisions. Others have been constructed to predict the contractors cost at the tender stage and have been monitored in parallel and as an alternative to pricing the traditional bill of quantities.

Cost models have not been confined to building projects alone. They have been applied in civil engineering to motorway projects both on behalf of the client and the contractor. They have also been used for cost engineering projects in the process plant industries. Several attempts have also been made in their application to specialist environmental engineering aspects of building projects.

Models based upon a single category of building project are likely to be easier to manipulate and more accurate at forecasting costs, than models seeking to embrace a wide variety of building types. The professional offices, therefore, undertaking large numbers of projects of a similar type are more likely to be able to prepare useful models, than a practice involved with projects across a wide spectrum of building types.

3. Techniques Used

There are several different methods that can be used for cost modelling purposes. The most popular, useful and applicable technique, however, is that of multiple regression analysis. This is a statistical technique that will find a formula or mathematical model that best describes the data available. The technique is used in those situations where the relationship between the variables is not unique, in the sense that a particular value of one variable always corresponds to the same value of other variables. This situation typically occurs when examining cost data. Simple linear regression analysis quantifies the relationship between two variables, such as floor area with total cost and can therefore be plotted graphically. Multiple linear regression analysis relates three or more variables and because of the complexities involved it is not suitable for graphical display and requires the use of the electronic computer. It is envisaged that this latter method has the greatest application to cost modelling, since cost is unlikely to be able to be described by a single variable.

The following points need to be considered in the construction of a cost model.

i. The choice of a single response variable. This is probably easily selected, e.g. the total cost of a proposed project, the elemental cost of a proposed project, etc.

ii. The selection of several regressor variables which are expected to determine the response. Many variables are considered initially, but where possible as few as necessary should be incorporated within the final model. The choice of these variables will depend upon what is being predicted, but might include floor and roof areas, wall areas, numbers of storeys, etc., all of which can be measured easily from the sketch plans.

iii. The number of sets of data used in the construction of the model must be greater than the number of variables fitted in the model. Ideally $2\frac{1}{2}$ times the number of variables should equal the number of sets of data required. For example, if our final model included twelve variables then we would require thirty sets of data for analysis purposes. Where the number is less than this force fitting of the model can occur leading to an incorrect analysis.

iv. Some of the regressor variables may be found to be more useful if they are combined to form new variables (termed derived variables). This also can have the effect of reducing the number of variables in the final model.

There are two opposing criteria in selecting the resultant model.

i. In order to make the equation as useful as possible as many regressor variables as possible should be included.

ii. Because of the costs of collecting the data and the fact that a large amount of data would be required, the model should include as few variables as possible.

There are a number of different methods and computer programs that can be used to isolate the most important variable. The compromise between (i.) and (ii.) above is termed "selecting the best regression equation". There is unfortunately no unique statistical procedure for carrying out this operation, and personal judgement and skill will be required. To add to the apparent confusion the alternative methods do not necessarily lead to the same solution.

A useful method that can be adopted for building a cost model is referred to as "stepwise regression" and is based upon the following procedures.

i. The regressor variables are added to the model one at a time and the influence on the results is noted. The order of insertion is determined by the partial correlation coefficient (i.e. the mathematical relationship between each of the regressor variables).

ii. Re-examination of each variable is incorporated in the cost model at every stage.

iii. A regressor variable which may have been a significant variable to enter at an early stage may become superfluous because of other regressor variables that have now entered the model. Such a variable can be entirely removed from the model. This whole process is referred to as backwards elimination and forwards selection.

4. Preparation for the Computer

The arithmetic involved in multiple linear regression analysis is considerable and in practical terms can only be undertaken where access to some form of computer is available. The advent of cheap computer power has made cost modelling a reality. Without it this process would have been too tedious to contemplate and the results could not be achieved within a reasonable period of time.

Equally so, cost models which a few years ago could only have been developed on a main frame computer can now be successfully constructed using either a mini-computer or a micro processor.

It is anticipated that the surveyor who wishes to consider developing cost modelling techniques will initially use a pre-written program for regression analysis. This is likely to be obtained from one of the standard statistical packages available to the particular computer being used. Eventually as the surveyor becomes more proficient with the technique and the program, he will be able to amend or write subroutines to suit his own independent needs.

Access to the computer will be via one of the established input devices, e.g. punched cards, paper tape, visual display unit. Each have their own advantages and selection will be a matter of personal choice. The surveyor will of course have to enter his data in the appropriate manner in order that the computer can process it. This is unlikely to prove to be a deterrent and even the inexperienced will soon grasp the basic facts of data entry. Alternatively there are facilities available where the data can be run on a computer by specialist software firms.

5. Analysing the Cost Model

Once the data has been processed through the computer, and a print out of the models obtained, preferably in hard copy format, the surveyor must then interpret the results. At this stage unless he is particularly conversant with statistical techniques he is likely to require the assistance of a statistician.

Certain important concepts regarding the representation of the data, the conformity of the data to a linear model and the correlation between the variables need to be understood. These concepts are, however, beyond the scope of this paper but reference to textbooks on regression analysis will provide the required explanations.

There are many statistics that can be used to measure the values in the model. The following are some of the most important.

1. Coeffient of variation

The coefficient of variation is the standard deviation (root-mean square of deviations from the arithmetic mean) expressed as a percentage of the mean. In regression analysis the standard deviation is referred to as the standard error of estimate.

$$\text{coefficient of variation} = \frac{\text{standard error of estimate}}{\text{mean cost of all schemes}} \times 100$$

As with the standard deviation it can be shown that 65% of all cases will fall within one standard error. For example, if the coefficient of variation was calculated as 10% then 65% of the cases should fall within ±10%. A model, therefore, with as small a coefficient of variation as possible is therefore desired. Most regression analysis computer programs will print out this information therefore there is no necessity for the surveyor to have to calculate it independently.

2. Correlation coefficient.

This is the quantitative factor which describes the ratio of explained variation to the total variation. In multiple regression analysis it is the multivariate counterpart to the simple correlation coefficient. The values vary between −1 and +1, −1 indicates perfect negative correlation and +1 perfect positive correlation. The value 0 indicates that no relationship exists at all. A value as near to +1 as possible is therefore required. It must be emphasised, however, that the correlation factor should not be examined in isolation, but a knowledge of the cause of this relationship should also be understood. A correlation coefficient of 0.9973 for example indicates that approximately 99% of the variation is explained by the regressor variables in the model.

3. Residuals

Residuals are the differences between the observed values and the values predicted by the regression equation. These differences are the amounts by which the regression equation has been unable to explain the variation within the data. Therefore the smaller the values of these residuals the better the fit of the data. In order to examine the residuals the computer programs besides tabulating the observed and predicted values can print the residual values in graphical form.

TESTING THE MODEL

When the best possible model has been constructed it is necessary to test this using further data. Best in this context would incorporate the following criteria:

i. The final model should explain a high percentage of the variation in the data (98%), otherwise its predictive powers will become restricted.

ii. The coefficient of variation should be as low as possible (10%). Previous research has indicated some deterioration in this value when models have been applied to new data.

iii. A maximum number of variables should be included in the model (12) in order to reduce the amount of data required initially and the time and cost required in collecting future data.

iv. There should be no discernible patterns in the residuals.

(NOTE: The figures in brackets represent an indication of the likely values to be achieved).

Two tests that can be adopted to justify the use of the model in practical project estimating might be as follows:

i. Comparison of the model's predictability against the actual values obtained from further projects.

ii. Comparison of these predicted values against estimates prepared using any of the traditional methods.

CONCLUSIONS

If the full potential of the computer is to be harnessed for the benefit of the quantity surveyor then it is likely that cost models will have some part to play. The time consuming task of calculating algebraic formula has in the past discouraged this approach. The computer, however, has an appetite to perform repetitive and complex arithmetic without effort and is particularly suited to this task.

The development of cost models and their application to the wider aspects of estimating have the following advantages.

i. Cost information can be provided more quickly.

ii. More information is generated so that more informed decisions can be made.

iii. The information will be more reliable introducing greater confidence in the decision making process.

iv. Suitable cost information is able to be produced at an earlier stage in the design process.

REFERENCES

1. DRAPER, N. and SMITH, H. "Applied Regression Analysis" Wiley-Interscience, 1966.
2. McCAFFER, R. "Some examples of the use of Regressional Analysis as an Estimating Tool" The Quantity Surveyor, 1975.
3. BUCHANAN, J. S. "Cost Models for Estimating" Royal Institution of Chartered Surveyors, 1972.
4. "Topographical Cost Models" Highway Economics Unit, Ministry of Transport, 1970.
5. HUGHES, G. A. "Estimating for Civil Engineering Works", The Quantity Surveyor, 1974.
6. ASHWORTH, A. "Regression Analysis for Contractors: An Assessment of its Potential" M.Sc. Thesis, Loughborough University of Technology, 1977.

'When will They Ever Learn?' (Re JCT 80)

By D. Phillips, FIQS

D. Phillips

Those members of the Construction Industry who have, like myself, spent a lifetime principally on the contracting side protecting Contractors' interests and/or preparing and pursuing claims on their behalf will, I feel sure, as and when they have studied the JCT 80 and considered the application of the Terms and Conditions relevant to the practical carrying out of the works and the consequential financial affect which will arise therefrom, be asking the question which is the heading of this article and having the sentiments expressed in the well known song from which this heading is a line.

Although many articles have been written and published and Seminars given, these have in the main been written and given by people with a legal background or other professionals who regard themselves as contract specialists, but both of these categories of persons who have expounded upon JCT 80 would appear to have failed to appreciate the practical rather than the theoretical application of these Conditions to the day to day running of the contract, and in particular the considerably increased responsibilities placed upon the Architect so that any failure by the Architect to strictly comply with same affords to the Contractor a much wider avenue for presenting claims to ensure the successful financial outcome to the contract. Contractors who are properly organised and who have learnt the lesson of what has been described as "THE HIGH COST OF UNDER-ORGANISATION" will not be slow to note that JCT 80 requires continuous notices to be given to the Architect, and that in respect of each of same the Architect must take specific action either without delay or within a certain specified period so that it can be anticipated that there will be a considerable increase in the amount of correspondence with which the Architect will be required to deal and take such action as is required by the Conditions, knowing always that unless action is taken in the manner and within the proscribed periods of time as laid down, the Employer can be placed in a situation of fault and thus be unable to operate any rights due to him under the contract. The article by Tay Moxley in "Building", 23rd January 1981, shows that the Architects have appreciated the difference between practical and theoretical applications of the Conditions of JCT 80, whilst the counterpart article by John Sims only further illustrates the difference of approach, and this can be seen quite clearly when he compares Clause 26 of JCT 80 with Clauses 11(6) and 24(1) of JCT 63, because the conclusion he draws therefrom does not take cognisance of the fact, as any Architect knows, that for a properly organised Contractor the giving of prior notice of likely disruption of regular working is no difficulty whereas for the Architect to take remedial action so as to avoid this situation may involve extra work in re-designing and/or production of drawings at a much earlier date than would have been anticipated, but in any case is likely to involve additional costs and time and may in fact cause delays on the contract.

The objections which have so far been put forward by the Architects to the use of this Form of Contract do not fully anticipate the difficulties which will be encountered in the operation of JCT 80, and I feel it is also in the interests of the Quantity Surveying Profession for the members thereof to be made more aware of what will be the practical application of this Form of Contract, rather than the theoretical interpretations which have been placed upon same in the many articles which have so far appeared in the various Trade Journals.

Whilst quite a lot of publicity has been given to and have highlighted the provisions in respect of nominated Sub-contractors, and it seems that the consensus of opinion is that there are unlikely to be ever correctly implemented, there is a very much wider reason for discontent with JCT 80 in that it incorporates too many changes at one time so that even those persons fully conversant with the intricacies of a building contract are unable to assess the relevance of each clause in relationship to the other clauses, and this factor together with the revised format whereby the Condition Clauses have a large number of sub-sections and sub-sub-sections makes this JCT 80 a Form of Contract which can only present very big problems for the Architect in the day to day running of the contract and if it is used will, without doubt,