THE MATURING CONCEPT OF ESTIMATING PROJECT COST CONTINGENCY - A REVIEW

David Baccarini
Department of Construction Management, Curtin University of Technology, PO Box U1987, Perth 6845, Western Australia, Australia.

ABSTRACT

Contingency is a ubiquitous component of project cost estimating. This paper provides a review of the literature pertaining to the estimating of project cost contingency. It describes the flaws of the tradition percentage method for estimating project cost contingency and sets out more robust estimation methods - regression analysis, Monte Carlo simulation and artificial neutral networks. In particular, the application of regression analysis for predicting project cost contingency is reviewed in detail as a prelude to the author’s research into the development and testing of a regression model for forecasting project cost contingency for engineering construction projects.

Keywords: project cost contingency, regression, cost modelling

INTRODUCTION

Accurate early cost estimates for engineering construction and building projects are extremely important to the sponsoring organisation (Overlander & Trost 2001). Sponsors require a budget estimate at the early stage of projects to manage project costs. This budget estimate has two main components - baseline estimate and cost contingency - which together represent the sponsor’s estimated final cost of the project. Therefore cost contingency is included within a budget estimate so that the budget represents the total financial commitment for the project sponsor, so the estimation of accurate cost contingency is of critical importance to projects.

Merrow & Schroeder (1991) highlighted the important link between predicting cost growth (i.e. difference between budget estimate and final actual cost) and project cost contingency by stating that cost growth can be viewed as inadequate contingency within cost estimates. Their research showed that there is no discernable relationship between cost growth and the level of contingency provided, although one might have been expected because contingency is emant to cater for cost growth. Reasonably accurate forecasts of final costs of construction projects are needed for justification of projects on economic grounds and for efficient capital planning and financing. These forecasts need to be as reliable as possible at relatively early project planning stages before capital commitments are made, which means an appropriate estimation for project cost contingency is critical.

CONTINGENCY – DEFINITION & ATTRIBUTES

Prior to reviewing the estimating methods for project cost contingency, the key attributes of the concept of project cost contingency need to be understood. Patrascu (1988) states that contingency is probably the most misunderstood, misinterpreted and misapplied word in project execution. Cost contingency has been broadly defined as “The amount of funds, budget or time needed above the estimate to reduce the risk of overruns of project objectives to a level acceptable to the organization” (PMI, 2004). The key attributes of project cost contingency are:
Reserve – Cost contingency is a reserve of money (PMI 2004). This is perhaps the most commonly understood component of project cost contingency (Baccarini, 2005).

Risk and Uncertainty – The need and amount for contingency reflects the existence of risk and uncertainty in projects (Thompson and Perry 1992). So the provision of project cost contingency is a risk management tool.

Total Commitment - The inclusion of contingencies within a budget estimate means that the estimate represents the total financial commitment for a project. Contingency should avoid the need to appropriate additional funds and reduces the impact of overrunning the cost objective.

Project Behaviour - Contingency can have a major impact on project outcomes for a project sponsor. If contingency is too high it might encourage sloppy cost management, cause the project to be uneconomic and aborted, and lock up funds not available for other organisational activities; if too low it may be too rigid and set an unrealistic financial environment, and result unsatisfactory performance outcomes (Dey et al 1994).

CONTINGENCY - ESTIMATION

A range of estimating techniques exists for calculating project cost contingency- see Table 1.

<table>
<thead>
<tr>
<th>Contingency Estimating methods</th>
<th>References (Examples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional percentage</td>
<td>Ahmad 1992, Moselhi 1997</td>
</tr>
<tr>
<td>Method of Moments</td>
<td>Diekmann 1983; Moselhi, 1997, Yeo 1990</td>
</tr>
<tr>
<td>Monte Carlo Simulation</td>
<td>Lorance &amp; Wendling 1999, Clark 2001</td>
</tr>
<tr>
<td>Factor Rating</td>
<td>Hackney 1985, Oberlander &amp; Trost 2001</td>
</tr>
<tr>
<td>Individual risks – expected value</td>
<td>Mak, Wong &amp; Picken 1998; 2000</td>
</tr>
<tr>
<td>Range Estimating</td>
<td>Curran 1989</td>
</tr>
<tr>
<td>Regression Analysis</td>
<td>Merrow &amp; Yarossi 1990; Aibinu &amp; Jagboro 2002</td>
</tr>
<tr>
<td>Artificial Neural Networks</td>
<td>Chen &amp; Hartman 2000; Williams 2003</td>
</tr>
<tr>
<td>Fuzzy Sets</td>
<td>Paek, Lee, &amp; Ock, 1993</td>
</tr>
<tr>
<td>Influence Diagrams</td>
<td>Diekmann &amp; Featherman 1998</td>
</tr>
<tr>
<td>Theory of Constraints</td>
<td>Leach 2003</td>
</tr>
<tr>
<td>Analytical Hierarchy Process</td>
<td>Dey, Tabucanon &amp; Ogunlana 1994</td>
</tr>
</tbody>
</table>

Traditional percentage is the most commonly used estimating method in practice (Baccarini, 2005). However, three estimating methods have gained prominence in recent times - Monte Carlo simulation, regression analysis and artificial neural networks. These will be reviewed with particular focus on regression analysis for cost modelling, which will lead to future research in this area by the author.

Traditional percentage

Traditionally cost estimates are deterministic i.e. point estimates for each cost element based on their most likely value (Mak et al 1998). Contingencies are often calculated as an across-the-board percentage addition on the base estimate, typically derived from intuition, past experience and historical data. Research indicates that this is the most common approach for estimating project cost contingency (Baccarini, 2005). This estimating method is arbitrary and difficult to justify or defend (Thompson and Perry 1992). It is an unscientific approach
and a reason why so many projects are over budget (Hartman 2000). A percentage addition results in a single-figure prediction of estimated cost which implies a degree of certainty that is not justified (Mak et al 1998). The weaknesses of the traditional percentage addition approach for calculating contingencies has led for a search for a more robust approach as evidenced by the range of estimating methods set out in Table 1.

Monte Carlo Simulation (MCS)

MCS is a quantitative technique for analysing risk and provides a structured way of setting the contingency value in a project cost estimate (Clark, 2001). The output of MCS when applied to estimating project cost is a probability distribution for the total final cost of the project. An example of its application is provided by Honeywell Performance Polymers and Chemical, which used MCS in 47 projects ranging from US$1.4m to US$505m (Clark, 2001), with contingency set at 50% probability level (median).

Artificial Neural Networks (ANNs)

ANNs are an information processing technique that simulates the biological brain and its interconnected neurons (Chen & Hartman, 2000). The structure of ANNs mimics the nervous system by allowing signals to travel thorough a network of simple processing elements (akin to neurons) by means of interconnections among these elements. ANNs employ a mechanism to learn and acquire problem-solving capabilities from ‘training’ examples by detecting hidden relationships among data and generalising solutions to new problems. ANNs are suitable for non-linear modelling of data, which contrasts with the linear approaches using regression. Over the past decade the use of ANNs for cost estimating has grown. ANN can be used to predict project cost overruns and thereby assist management in developing an appropriate contingency. For example, Chen & Hartman (2000) used ANN to predict the final cost of completed oil and gas projects from one organisation using 19 risk factors as the input data. It was found that 75% of the predicted final cost aligned with the actual variance i.e. where the ANN model predicted an overrun/underrun, an overrun/underrun actually occurred. The prediction accuracy of ANN outperformed multiple linear regression.

Regression

Regression models have been used since the 1970s for estimating cost and are a powerful statistical tool for analytical and predictive purposes in examining the contribution of variables to overall estimate reliability (Kim et al 2004). An extensive review of cost modelling techniques by Skitmore & Patchell (1990) found the use of regression analysis for cost modelling has primarily focused upon the search for the best predictors of tender price. This indicates a need for regression cost modelling of the client’s final cost of project. The application of regression analysis for cost modelling follows the principle of parsimony. That is, models should be sophisticatedly simple and fit the data adequately without using any unnecessary parameters, and generally produce better forecasts (Sonmez 2004). The development and selection of early stage cost forecasting models will have limited data available so the use of a complicated forecasting model will add unnecessary assumptions and thereby work against the principle of parsimony. Regression techniques can be applied to achieve the principle of parsimony (Cheung 2005). Furthermore regression models allow explicit relationships between dependent and independent variables to be analysed. There has been some research using regression analysis to forecast the cost of construction/engineering project – See Table 2.
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Project Type</th>
<th>Application</th>
<th>Dependent variable</th>
<th>Independent variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merrow, Phillips and Myers 1981</td>
<td>pioneer process</td>
<td>Early phases</td>
<td>Cost growth – early phase to completion</td>
<td>unproven technology, process impurities, complexity, inclusiveness, project definition.</td>
</tr>
<tr>
<td>Merrow and Schroeder 1991</td>
<td>hydroelectric</td>
<td>Early phases</td>
<td>Final construction cost</td>
<td>megawatt capacity, hydraulic head, type of project, height of dam, year of project appraisal.</td>
</tr>
<tr>
<td>Jahren &amp; Ashe 1991</td>
<td>naval facilities</td>
<td>Award Phase</td>
<td>Cost growth – construction phase</td>
<td>award-estimate difference (i.e. difference between winning bid and government estimate)</td>
</tr>
<tr>
<td>Dissanayaka &amp; Kumaraswamy 1999a</td>
<td>building</td>
<td>Award Phase</td>
<td>Cost growth – construction phase</td>
<td>project complexity, client characteristics, original cost estimate, contractor characteristics</td>
</tr>
<tr>
<td>Dissanayaka &amp; Kumaraswamy 1999b</td>
<td>construction</td>
<td>Award Phase</td>
<td>Cost growth – construction phase</td>
<td>client confidence in team; risk retained by client for quantity variations, level of construction complexity related new technology, payment modality</td>
</tr>
<tr>
<td>Bacon &amp; Besant-Jones 1998</td>
<td>power generation</td>
<td>Early phases</td>
<td>Cost growth - early phase to completion</td>
<td>thermal power: estimated cost, estimated schedule, project type, country and civil costs; hydropower: estimated costs, foreign exchange, station size, project type, hydraulic head, financing agency, country</td>
</tr>
<tr>
<td>Oberlander &amp; Trost (2001)</td>
<td>process plants</td>
<td>Early phases</td>
<td>Contingency early phase to completion</td>
<td>45 elements: who is involved in estimate; how estimate is prepared; what is known about project, other factors.</td>
</tr>
<tr>
<td>Setyawati, Sahirman &amp; Creese 2002</td>
<td>educational buildings</td>
<td>NR</td>
<td>Final construction cost</td>
<td>floor area, building height</td>
</tr>
<tr>
<td>Williams 2002, 2003</td>
<td>roads</td>
<td>Award Phase</td>
<td>Final construction cost</td>
<td>lowest bid price</td>
</tr>
<tr>
<td>Attalla and Hegazy 2003</td>
<td>reconstruction</td>
<td>Award Phase</td>
<td>Cost growth – construction phase</td>
<td>built drawings, unit prices, critical path method, prequalification of contractors, inspection by operator-maintenance/end-users.</td>
</tr>
<tr>
<td>Odeck 2004</td>
<td>roads</td>
<td>Early phases</td>
<td>Cost growth - early phase to completion</td>
<td>estimated cost, completion time</td>
</tr>
<tr>
<td>Kim, An &amp; Kang 2004</td>
<td>residential buildings</td>
<td>Award Phase</td>
<td>Final construction cost</td>
<td>gross floor area, storeys, total units, duration, roof types, foundation type, usage of basement, finishing grade</td>
</tr>
<tr>
<td>Sonmez 2004; 2005</td>
<td>care/retirement</td>
<td>Early phases</td>
<td>Final construction cost</td>
<td>project year, location, building area, percentage of health and common areas, total area per unit</td>
</tr>
<tr>
<td>Burroughs &amp; Juntima 2004</td>
<td>industrial facilities</td>
<td>Not Reported</td>
<td>Contingency</td>
<td>project definition, use of new technology, process complexity, contracting and execution strategy, equipment percentage.</td>
</tr>
</tbody>
</table>
Merrow, Phillips and Myers 1981

This research statistically analysed 44 pioneer process plants provided by 34 private sector firms in North America. They analysed the cost growth (i.e. ratio of forecast cost to actual cost) of these projects as a basis for gauging the reliability of an estimate and to assess the probable ultimate cost of a process facility. Regression analysis was used to ‘yield more realistic expectations of ultimate costs’ (p.1). The regression model for the dependent variable of cost growth used dependent variables measurable with some precision early in the project life cycle: degree of unproven technology, degree of process impurities, complexity, inclusiveness (i.e. completeness of estimate), and project definition. The coefficient of determination \( R^2 \) was 0.83 and 47% the estimates were predicted with +/-5% of the actual cost growth.

Merrow and Schroeder 1991

This research developed ordinary least squares multiple regression models to predict actual costs of hydroelectric projects by using independent variables measurable early in the project development cycle. The model was derived from 56 hydroelectric projects funded by the World Bank between 1971 and 1988. The regression equation for predicted total capital cost was based on five variables: number of megawatts of capacity, hydraulic head, type of project, height of dam and year of project appraisal. \( R^2 = 0.96 \). The researchers support use of regression for cost estimating: “The accuracy of the models suggests that the cost growth problems for World Bank supported hydroelectric projects ... are more a matter of appropriate estimating than significant deviations from an appropriate project cost. In other words, despite cost growth ..., most projects are probably not costing more than they should” (p. I.5.3)

Jahren and Ashe 1991

This research dealt with cost growth on construction projects during the construction phase for 1576 US naval facilities engineering command construction projects. Cost growth is defined as the percentage difference in cost between the final contract cost and the contract award amount. Regression analysis was conducted to determine if a linear relationship existed between the award-estimate difference (i.e. difference between winning bid and government estimate) and cost growth rate. It was found that contracts with award amounts less than the government estimate were more likely to have cost growth rates above 5%. Furthermore cost growth was found to more likely on larger projects than smaller ones.

Dissanayaka & Kumaraswamy 1999a

Multiple regression analysis was applied to analyse the data from 32 Hong Kong building projects to forecast the cost growth (i.e. ratio of final cost paid for construction to tender price) during the construction phase of projects. The model used cost growth as the dependent variable and four independent variables – project complexity, client characteristics, original cost estimate and contractor characteristics \( R^2 = 0.79 \).

Dissanayaka & Kumaraswamy 1999b

A survey on significant factors affecting cost performance of projects was completed by 30 respondents covering clients, contractors, consultants involved in the Hong Kong construction industry. Univariate analysis was used to identify independent variables having a significant association with the dependent variable of cost growth i.e. ratio of certified final account to initial contract value. Multiple linear regression analysis was performed to develop models for cost growth using these independent variables. The final model produced a \( R^2 \) of 0.84. The four significant independent variables affecting cost growth were: levels of client confidence.
in the construction team; risk retained by client for quantity variations, level of construction complexity related new technology and payment modality.

_Bacon and Besant-Jones 1998_

This research investigated the reliability of estimates for construction costs of 135 power generation projects funded by the World Bank between 1965 and 1986, and completed by 1994. The analysis is based on cost growth i.e. a comparison of the estimated costs at the time of project approval with the actual cost of implementing the project as determined after project completion. The costs cover capital costs, engineering services and local duties on imports. The research found that estimated values were significantly biased below actual values, with squared correlations between the actual and estimated costs of 76% and an average underestimation of 21% and standard deviation of 34% (around a mean of 117% for actual to estimated values). An investigation of World Bank project reports was undertaken for variables that might be expected to have some correlation with cost overrun and known at the inception of the project. The multiple regression model for thermal power costs found five significant variables: estimated cost, estimated schedule, project type, country and civil costs; \[R^2 = 0.523\]; and hydropower projects had eight significant variables: estimated costs, foreign exchange, station size, project type, hydraulic head, financing agency and country \[R^2 = 0.511\].

_Oberlander & Trost (2001)_

This research developed a model using regression analysis to predict the amount of cost contingency required based the quality/accuracy of the project cost estimate. The model was based on detailed analysis of 67 completed capital projects (US$5.6bn) from 22 companies in the process industry (16 owners and 6 contractors and engineering firms). The research identified 45 elements to measure the quality of early estimates, divided into four categories: who is involved in the estimate; how the estimate is prepared; what is known about the project; and other factors considered whilst preparing the estimate. Multivariate regression using ordinary least-squares fit through the 45 elements of the 67 projects provided the basis for predicting the accuracy of an estimate and the consequent contingency to improve this accuracy. The prediction model was \(y = mx + b\), where \(y\) represented the percentage contingency and \(x\) represented the estimate score, \(m\) is the slope and \(b\) is the intercept. For an estimate, each element is rated from 1 (best) to 5 (worst) and entered into the model for an overall score for estimate quality to be derived. This score then predicts the accuracy of the estimate; the greater the inaccuracy, the need for more contingency for a chosen confidence level. The results showed a significant correlation between the estimate score and the accuracy of the estimate.

_Setyawati, Sahirman and Creese 2002_

This research used building construction data of 41 new constructed educational buildings obtained from Design Cost Data publication between July 1998 and December 2001, Regression analysis was used to create a model for predicting construction cost, based on two independent variables: floor area and building height \[R^2 = 0.927\]. This model was further developed by sorting and rearranging data using regression analysis based on percentage error, which produced a regression equation with an improved \(R^2\) of 0.972.

_Williams 2002, 2003_

This research used regression analysis to predict the completed cost of competitively bid highways projects constructed by five highway agencies, based on 3444 projects. A natural log transformation and stepwise regression procedure yielded a best performing predictive model for completed cost based on one independent variable – the lowest bid price – resulting
in $R^2$ of the five regression models (one for each agency) between 0.893 and 0.992. Furthermore, the predicted final cost increases as an increasing percentage of the low bid as the project size increases. The regression models were tested using independent data sets that were not used to calculate the regression model and between 69% and 77% of projects were predicted within 10% of the actual costs.

Attalla and Hegazy 2003

This paper developed a predictive model of cost deviation for reconstruction projects. A survey of 32 construction organisations (mostly construction managers and project administrators) was conducted to obtain reasons for cost growth for 50 reconstruction projects, and the actual cost deviation from estimated values. 36 factors were identified as having an impact on the cost performance of reconstruction projects. A Cost Performance Index was formulated, based on two variables: increase of contract value due to for change orders, and the cost of rework for repairs paid for by the contractor in addition to the original contract value. Correlation coefficients between each of the 36 variables and the CPI resulted in 18 factors having a significance relationship (i.e. $r>0.5$) and considered for model development. Stepwise regression (backward-stepping) resulted in a predictive model for CPI based on five independent variables: as-built drawings, unit prices, critical path method, prequalification of contractors, and inspection of operator-maintenance/end-users. [$R^2 = 0.897$]. Nine case studies excluded during model development were used for validation purposes, and the correlation between predicted and actual CPI was an r value of 0.9031.

Odeck 2004

This research used regression analysis to investigate the statistical relationship between actual costs and estimated costs at the detailed planning stage of 620 road construction projects, using data from Norwegian road construction during 1992-1995 performed by the Norwegian Public Roads Administration. The findings revealed a discrepancy between estimated and actual costs, with a mean cost growth of 7.9% ranging from -59% to +183%. Cost growth appeared to be predominant among smaller projects compared to larger ones. Stepwise regression identified two significant independent variables: estimated cost and completion time [$R^2 =0.21$].

Kim, An and Kang 2004

Data of construction costs of 530 residential buildings built by general contractors between 1997 and 2000 in South Korea was used for regression analysis to predict actual final costs of residential construction projects. Regression analysis was performed on 490 projects, with the remaining 40 projects used to test the model for predicting the dependent variable of final construction costs. Independent variables were: gross floor area, storeys, total units, duration, roof types, foundation types, usage of basement and finishing grade. The mean absolute error was 6.95% when the model was applied to the 40 test projects.

Sonmez 2004; 2005

Regression analysis was used for conceptual cost estimating based on 30 continuing care retirement community projects built by a contractor in the United States in 14 different states during 1975-1995. Regression analysis was used to estimate project cost, containing five variables: project year, location, building area, percentage of health and common areas, and total area per unit [$R^2 = 0.949$].

Burroughs and Juntima 2004

Regression analysis was used to forecast the Contingency Performance Indicator [CPI] i.e. the absolute value of percent of contingency used minus the percent of contingency estimated.
For example, is 50% contingency is estimated but only 20% is consumed, then the CPI is 50-20 = 30%. The data used to formulate the regression equation was derived from of 1500 industrial facilities projects, with approximately half completed after January 2000, and ranging in size from less than US$1m to more than US$1.5bn. Through regression analysis, the significant independent variables were: project definition level, use of new technology, process complexity, contracting and execution strategy, and equipment percentage. The regression model produced a median CPI of 7%.

FINDINGS & CONCLUSIONS

Interestingly, despite the ubiquitous nature of cost contingency with projects there is a paucity of scholarly research into a deeper understanding of the concept of project cost contingency. The literature review of the concept of project cost contingency identified four key attributes; it is (i) a reserve of money that (ii) caters for risk within projects, (iii) indicating the organisation’s final total financial commitment, and (iv) affecting stakeholders’ behaviour within projects.

Traditionally, contingencies are often calculated as an across-the-board percentage addition on the base estimate, typically derived from intuition, past experience and historical data. The literature review highlighted several serious flaws with this estimating method. This judgmental and arbitrary method of contingency calculation is difficult for the estimator to justify or defend. A percentage addition results in a single-figure prediction of estimated cost, which implies a degree of certainty that is simply not justified. It does not encourage creativity in estimating practice, promoting a routine and mundane administrative approach requiring little investigation and decision making.

This paper briefly reviews more robust and justifiable approaches to estimating project cost contingency. In particular the application of regression analysis is explored and advocated as an approach for cost modelling for predicting the final cost for engineering construction and building projects. It is surprising that there has been very little research conducted into the application of regression for predicting the final cost of projects. There has been some application of regression analysis for predicting the tender prices of projects but it would be expected that project sponsors are more interested in their final cost commitment rather than the cost at tender stage. Furthermore, most of the limited research into the development and testing of a regression model for the final cost of construction projects has not taken into account the project variable of project cost contingency. The author is presenting researching the estimation of project cost contingency as a component of a regression model for predicting the final cost of construction engineering projects

REFERENCES


Fenton, R.E, Cox, R.A. and Carlock, P. (1999). Incorporating contingency risk into project cost and benefits baselines; a way to enhance realism. INCOSE Conference


