Abstract

Large transportation projects have been afflicted by poor cost estimates and large cost overruns. This paper uses transit projects in the United States as an example to develop a probabilistic methodology for establishing sufficient contingency. Contingency is a reserve budget for coping with risks and uncertainties and to help keep the projects on budget. The premise of the proposed methodology is that accuracy of cost estimate is affected by project complexity. Also, there is an inherent underestimate in cost estimates because the estimator estimates components’ costs based on the modes (most likely values) of cost distributions and then adds up these modes to arrive at total cost. In fact, the means of these cost categories have to add up to the mean of the total cost. The paper first suggests a correction for this error and then develops a methodology for establishing the level of complexity for the project, and based on these steps, develops a method for calculating an adequate contingency budget.

In order to establish the contingency budget, first, a rating system is used to determine the complexity of the project. It is assumed that the more complex projects are harder to estimate accurately, and the likelihood of cost overrun is higher for these projects. Coefficient of variation is used as an indication of level of uncertainty in the cost and is assumed to be a function of complexity. The rating system scores projects according to pre-determined factors that we believe contribute to projects’ complexity. Then different coefficients of variations are assigned in accordance with projects’ complexity scores. It is shown that project contingency can be computed based on the coefficient of variation and the desired confidence level specified by the user. A validation is performed using a set of actual project costs.

Keywords: Complexity, Contingency, Cost uncertainty, Probabilistic, Transit projects.

1. Introduction

Large transportation projects have been fraught with cost overruns all over the world [2]. In a Transportation Research Board (TRB)-sponsored project, Booz.Allen [1] analyzed a number of U.S. Transit projects and suggested several reasons for cost overruns. These included scope creep, optimistic original cost estimates, insufficient contingency, delays in project startup time, and items left out of the original cost estimates. In addition to these, Flyvbjerg contends that political pressure by project champions may cause underestimation of project costs and overestimation of its benefits [4]. Since early 2000s federal government in the United States started using probabilistic approaches to validate transportation project cost estimates and to establish contingency budgets to ensure adequate funding for these projects. In this approach a risk workshop is convened where experts in various areas help to identify project risk factors and then provide a range of possible costs for these risk factors. The outcome is a distribution for the total risks (or the total project cost depending on the approach) which will be used in establishing project budget based on a desired confidence level. One of the main issues in this approach is that the ranges established can be affected by the bias of the estimator. In other words, if the estimator is also a champion for the project, his or her risk estimates tend to be on the low side. Because of this issue, in many transit projects, the
range for the total costs was completely eclipsed by the actual costs because the actual project costs exceeded the most pessimistic cost range.

This paper proposes an approach that minimizes the effect of bias in the calculation of total costs. Our area of emphasis in this paper is transit projects in the United States. The approach is general however, and with appropriate modification can be applied to projects in other parts of the world. The premise of the proposed methodology is that the estimator estimates components’ costs based on the modes (most likely) of the cost distributions and then adds up these modes to arrive at total cost. In this approach it is assumed that any significant cost component is a random variable rather than a deterministic value. The estimator only considers one realization of this component when estimating. It seems reasonable to assume that the estimator is using the most likely value or the mode of the cost distribution, the same way that a scheduler in PERT approach estimates activity duration based on its most likely value [8]. The estimator eventually adds up these modal values to arrive at a total cost estimate. The issue is that means of these cost categories have to add up to the mean of the total cost. The sum of the modes does not equal to the mode of the sum. The paper first suggests a correction for this error by considering the probability distribution of cost components and then develops a methodology for establishing an adequate contingency budget.

1.1. Standard Cost Categories (SCC) for Capital Projects

According to the Federal Transit Administration (FTA) [3], all transit projects requesting federal support will have to submit a cost estimate for their project according to the format called the Standard Cost Categories (SCC). Standard Cost Categories is “a consistent format for reporting, estimating and managing of capital costs for New Starts projects” since 2005. It is believed the information collected from this format across the country could be beneficial to FTA in future cost estimate, because it allows cost comparison among transit projects and will allow the cost anomalies to be identified more easily.

Based on FTA’s Standard Cost Categories (SCC) for Capital Projects, a transit project’s cost is broken down into 10 categories (Table 1).

<table>
<thead>
<tr>
<th>SCC</th>
<th>Category</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Guideway &amp; Track Elements (Route Miles)</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Stations, Stops, Terminals, Intermodal (Number)</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Support Facilities: Yards, Shops, Admin. Bldgs</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>Sitework &amp; Special Conditions</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>Systems</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>Row, Land, Existing Improvements</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>Vehicles (Number)</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>Professional Services (Applies To Cats. 10-50)</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>Unallocated Contingency</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>Finance Charges</td>
<td></td>
</tr>
</tbody>
</table>

The traditional way the estimators use can be explained as following. When faced with a new transit project, estimators will examine what items are needed. For every detailed component, such as procurement of steel, concrete, etc, estimators usually keep a record of previous prices. Then from these historic records, estimators normally choose a value for use in the cost estimate. Given the uncertainty of costs especially in the future, it can be assumed that cost components are random variables following certain distributions. Past research shows that the cost for a detailed component in construction costs follows a lognormal distribution ([9] and [10]). This is intuitively acceptable although it has been verified quantitatively in previous research. Costs are positive and bounded on the lower side (cannot have negative cost), while at times the possibility of assuming very large costs are not unrealistic, hence a lognormal distribution makes sense. As the cost estimate is done deterministically, the estimator picks the most likely value of each cost to generate his or her estimate. In other words, the estimator is working with the modes of cost distributions [9]. The use of mode is well understood in the industry. For example, in PERT scheduling, the scheduler models uncertainty in project activity durations by providing pessimistic, most likely
(mode), and optimistic estimates of duration [8]. Furthermore, while in highway projects, because of the large number of projects, it is possible to use the mean values of unit prices for former projects from Departments of Transportation webpages, this is not possible for transit projects as there is no detailed unit prices available for transit projects and the number of projects are much fewer compared to highways. Furthermore, even in highway projects, the State DoT’s usually only use the lowest bids in calculation of the listed average unit prices on their webpages and that creates a bias in the estimate of the mean of the distribution.

Normally, the base cost of a transit project (excluding contingency and finance charges) can be divided into 8 categories from SCC10 to SCC80 as presented in Table 1. These 8 categories contain numerous sub-phases and these sub-phases may contain even more sub-components. The lognormal assumption discussed above is only valid when the most detailed category is concerned. At each detailed stage, estimators normally choose the most likely value for each item, i.e., the mode.

2. Basic Error in Total Cost Estimate

According to central limit theorem, when summing several lognormal distributions, the sum will converge towards a normal distribution, assuming independence between distributions. When summing distributions together, the mean after summation will simply be the sum of means of each lognormal distribution. However, the same cannot be said about the value of modes. Therefore as the summation of distributions proceeds, the mode of the final normal distribution will not be equal to the sum of modes of those original lognormal distributions. As the estimator is working with modes, the total cost is assumed to be the most likely cost of the project. Errors occur in this process as there is a difference between mode and mean of a lognormal distribution. The next section describes the correction of this error.

The mean, mode and variance of a lognormal distribution are:

\[ m = e^{\mu - \sigma^2} \]  
\[ E(x) = e^{\mu + \frac{\sigma^2}{2}} \]  
\[ \text{Var}(x) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2} \]  

where \( \mu \) and \( \sigma \) are mean and standard deviation of the underlying normal distribution, and \( m \) is the mode of the lognormal distribution. Now, if \( n \) lognormal random variables are summed up, we have:

\[ \text{Sum of modes} = \sum_{i=1}^{n} m_i = \sum_{i=1}^{n} e^{\mu_i - \sigma_i^2} \]  
\[ \text{Sum of means} = \mu_T = \text{sum of } E(x) = \sum_{i=1}^{n} e^{\mu_i + \frac{\sigma_i^2}{2}} \]  

Then the difference between sum of modes and \( \mu_T \):

\[ \Delta = \mu_T - \sum_{i=1}^{n} m_i = \sum_{i=1}^{n} e^{\mu_i + \frac{\sigma_i^2}{2}} - \sum_{i=1}^{n} e^{\mu_i - \sigma_i^2} \]

From Equations (2) and (3), \( \sigma^2 \) and \( \mu \) can be obtained.

\[ \sigma^2 = \ln [1 + \frac{\text{Var}(x)}{E(x)^2}] = \ln[1 + \text{c.o.v}^2(x)] \]  
\[ \mu = \ln(m) + \sigma^2 \]  

In Eq (7), c.o.v is coefficient of variation which is the ratio of standard deviation over mean. Using Eqs (7) and (8) we have:

\[ \mu_i = \ln(m_i) + \sigma_i^2 = \ln[m_i(1 + \text{c.o.v}^2)] \]  
\[ E(x) = e^{\mu_T + \frac{\sigma^2}{2}} = e^{\ln[m_i(1\text{c.o.v}^2)] + \frac{\ln(1+c.o.v^2)}{2}} = m_i(1 + \text{c.o.v}^2)^\frac{3}{2} \]
Where $\Delta$ is the difference between the total estimate ($\sum_m$) and the total expected value. The increase in cost estimate needed to bring the estimate to the mean total cost is calculated from Eq (13).

$$\text{Increase} = \frac{\text{Expected Value} - \text{Estimate}}{\text{Estimate}} = \frac{\sum \Delta}{\sum m} = \frac{[(1 + c.o.v^2)^{3/2} - 1]}{\sum m} = (1 + c.o.v^2)^{3/2} - 1$$

It can be seen that using the proposed approach, as long as projects have the same c.o.v, their percent increase will remain the same. This finding actually will help the future process of creating charts for estimators to use because c.o.v is the only determining factor.

3. Effect of Complexity on Project Cost

It can be argued that c.o.v is a representation of a project’s cost uncertainty. At least part of cost uncertainty (or the inaccuracy of cost estimates) can be attributed to the complexity of the project. Complexity of a project can originate from several aspects. Everything from scope to construction can influence a project’s complexity. Gidado [5] suggests that there seems to be two perspectives on project complexity in the industry: the managerial aspect, which involves the planning of bringing together numerous parts of work; and the operative and technical aspect. Wood and Ashton [11] suggest that high number of trades involved and long timescale will increase projects’ complexity. Hartmann, et al [6] discussed factors which can increase the complexity of a subway project. In this report, existing heavy traffic, having several construction components underground, public concern, tight site conditions, and multiple contractors all make the project complex.

In this research, we studied 28 transit projects completed in the past three decades in the United States and identified nine criteria to represent a project’s complexity ([1] and [7]). After further analysis, these criteria were further refined because it was felt that there were several criteria that correlated with others and double-counted a project’s complexity score. Eventually, the following five criteria were selected for measuring a project’s complexity: (1) Project Changes, (2) Duration of Project, (3) Multiple Contracts, (4) Underground Work/Complexity of Stations and (5) Utility Relocation. Each transit project in the database was rated according to these five criteria and each project was identified as having a low, medium, or high level of complexity. For a detailed description of this process please refer to [7]. Table 2 shows the ratings (scores) that were used for assessing the complexity of projects. Complexity score could be an integer between 1 and 5 and the c.o.v was selected as 0.1, 0.2, or 0.4 depending on the complexity.

<table>
<thead>
<tr>
<th>Score</th>
<th>Coefficient of Variation</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 1</td>
<td>0.1</td>
<td>Low</td>
</tr>
<tr>
<td>2 - 3</td>
<td>0.2</td>
<td>Medium</td>
</tr>
<tr>
<td>4 - 5</td>
<td>0.4</td>
<td>High</td>
</tr>
</tbody>
</table>

4. Estimating Contingency

For this research, we define contingency as the percentage that has to be added to the base estimate to arrive at the desired project budget (Eq 14).

$$\text{contingency} = \left( \frac{\text{Final budget} - \text{Total estimated value (mode)}}{\text{Total estimated value (mode)}} \right) \times 100\%$$
By re-writing Eq (14) and some manipulation, Eq (15) can be transformed into Eq (16) [7].

\[
\text{Contingency} = \left( \frac{x - \text{total expected value (mean)}}{\text{total estimated value}} \right) + \left( \frac{\text{expected values (mean) - total estimated value}}{\text{total estimated value}} \right)
\]

\phantom{= x - \text{mean mode}} + \frac{\text{mean mode}}{\text{mode}}

\text{(15)}

\[
\text{contingency} = Z \times c.o.v \times (1 + c.o.v^2)\frac{3}{2} + [(1 + c.o.v^2)^{\frac{3}{2}} - 1]
\]

\text{(16)}

In Eq (16), Z is the statistic for a standard normal distribution \([(x-\mu)/\sigma]\). Using different values for Z one can establish a desired confidence level for project budget. Figure 1 shows the outcome of this analysis. For example, if after careful consideration of project characteristics, the level of complexity of the project is determined to be medium, and if the owner is comfortable with a budget that ensures project completion within budget with a likelihood of 80% (a figure that is typically used in transportation projects in the United States), then they need to adjust the base estimate (estimator’s prepared estimate without any contingencies) by a factor of 1.239. A confidence level of 90% would require an adjustment of 1.332. An important factor to consider is that these values will depend on the level of scope definition. The less the scope is developed, the higher the uncertainty. Figure 1 values are produced based on information at the end of final design.

![Figure 1 – Percent Contingency vs Confidence Level for Projects with medium Complexity (c.o.v =0.2) at the end of Final Design](image)

5. Validation of Model Results

In order to test the validity of the model results, we reviewed the 28 transit projects in the database. For each project, the base cost was estimated. For this purpose, the estimate at the end of final design was reduced by 10% (the assumed contingency as the actual value of contingency was not reported) to arrive at base cost estimate for the project. This is assumed to be the sum of modes of project components. For each project, the level of complexity was established and then using charts similar to Figure 1 (depending on project complexity), the contingency percentage as defined in this paper was established for various confidence levels.

Table 3 gives the results of this analysis. For example, the analysis shows that applying the contingency percentage as suggested by this paper, 78.6% of projects (22 out of 28 projects) would have not exceeded their established budgets. This number is arrived by assuming a confidence level of 75%. Table 3 shows that there is reasonable agreement between the model output and the actual project performance. Because there were only 28
projects in the database, percent of projects with no overrun are more discrete compared to the continuous function that defines the confidence level resulting from the application of the model.

Table 3. Percentages of projects having no overrun

<table>
<thead>
<tr>
<th>Confidence level</th>
<th>Percentage of projects having no overrun</th>
</tr>
</thead>
<tbody>
<tr>
<td>75%</td>
<td>67.9%</td>
</tr>
<tr>
<td>80%</td>
<td>78.6%</td>
</tr>
<tr>
<td>85%</td>
<td>85.7%</td>
</tr>
<tr>
<td>90%</td>
<td>85.7%</td>
</tr>
</tbody>
</table>

6. Conclusion

A probabilistic method is presented to help in establishing capital budgets by considering the inherent underestimation of cost in traditional estimating approach and also the project complexity. A desired confidence level can be selected by the project owner and a budget calculated. The approach also helps remove the estimator’s bias in estimating project costs.

References