Journal of Risk Analysis and Crisis Response, Vol. 2, No. 4 (December 2012), 223-232

A New Approach for Contingency Determination in a Portfolio of Construction Projects

Payam Bakhshi*
Assistant Professor, Department of Construction Management, Wentworth Institute of Technology
550 Huntington Ave., AnxSo 008-F, Boston, MA 02115, USA
**E-mail: bakhship@wit.edu*

Ali Touran
Professor, Department of Civil & Environmental Engineering, Northeastern University
360 Huntington Ave., 400 SN Eng., Boston, MA 02115, USA
atouran@coe.neu.edu

Received 5 March 2012; accepted 4 August 2012

Abstract

In this paper, a new approach for contingency determination in a portfolio of construction projects is proposed. The model proposed helps an agency find the level of confidence needed for individual projects to ensure that the portfolio budget will meet the minimum level of confidence based on available funding and the agency's policy goals. The promise of this model is to protect a portfolio of projects against cost overrun by adjusting their original budgets. A Bayesian approach is employed to update the model on regular intervals. As more information becomes available in the future, the required adjustment in portfolio budget will be reduced, because the accuracy of estimating the contingency is improved. The proposed model is an effective tool for the agencies/owners to develop contingency budgets.

Keywords: Contingency, Risk, Probabilistic Model, Portfolio of Projects, Bayesian.

1. Introduction

Large transportation capital projects all around the globe have been experiencing cost and schedule overruns. Nearly 50% of the large active transportation projects in the United States overran their initial budgets (Sinnette 2004). To overcome the cost overrun issue, identifying cost escalation factors have been the subject of much research (Shane *et al* 2009; Flyvbjerg *et al* 2003; Pickrell 1990). For instance, Shane *et al* (2009) identified 14 risk factors classified in two categories: 1. Internal Sources such as bias, poor estimating, and contract document conflicts; 2. External Sources such as effects of inflation, market conditions, and unforeseen events/ conditions. Contingency is a reserve budget for coping with risks and uncertainties and to help keep the projects on budget. An owner agency usually adds contingency to the estimated project cost to account for the uncertainties. Risks and uncertainties associated with a project are impediments to reach an accurate cost estimate. Contingency is traditionally estimated as a predetermined percentage of project base cost depending on the project phase. In recent years, some agencies have started conducting formal probabilistic risk assessment to estimate contingency budget rather than deterministic approach (Touran 2010; Molenaar 2005). However, to establish the contingency budget, an agency must make all effort to set aside a budget which is optimized. This becomes more important when an agency is dealing with a portfolio of projects. Allocation of an excess budget for a project will use up the money that can be spent on other projects. For instance the current practice in the U.S. to estimate the contingency budget in transit projects called Top-down Model is based upon a probabilistic method using lognormal distributions for different cost categories in the project. Research shows the way that cost categories are ranged is very conservative resulting in a contingency budget far larger than what is indeed needed (Bakhshi and Touran 2009). Nevertheless, despite all claims regarding improved models, budget estimating for transit projects have been inaccurate for several decades (Flyvbjerg 2006). Also, for large capital programs consisting of several projects, establishing contingency has not been well studied.

2. Proposed Model for Calculating Contingency

This model is a continuation and major improvement on an earlier model developed by Touran (2010) for calculating contingency for a portfolio of projects. The new model uses a Bayesian approach for updating the calculations based on new data that becomes available. The application of the model is shown on a group of transit projects. Even though the application of this model in this paper is on transit projects, it is a mathematically flexible model that can be applied on any type of construction project.

In the model proposed here, the portfolio consists of projects with different owners who have requested funding from the same source. For example, in the case of transit in the United States, these are projects submitted by state agencies for obtaining federal funding. For each of these projects, it is assumed that a formal risk assessment has been conducted based on the specific risks affecting each project as required by the regulations. The objective of the model presented here is then to adjust the overall portfolio budget based on the historical data on budget shortfalls. In other words, it is assumed that a detailed risk assessment has been conducted at the individual project level and that the requested funding reflects that.

The model assumes normal distribution for the cost overruns/ underruns and truncated normal distribution for the cost of each project in the portfolio. These assumptions are based on the following factors: first, the cost overrun/underrun distribution will be used as a *prior* distribution in the Bayesian approach. As more information becomes available, the distribution becomes more refined and converges to the true distribution regardless of the initial assumption about *prior*, and second, the use of normal distribution allows the derivation of closed form solution for the calculation of contingency based on desired confidence levels. Furthermore, tests of goodness of fit showed that assumption of normality was adequate for the project cost data that was available.

To form truncated normal distribution of cost for each project, it is assumed that the probability of experiencing underrun m is α as the discrete portion of distribution. The parameter m is added because project owners have the tendency to spend most of the budget by enhancing and embellishing of the projects when they realize their projects will be completed under the budget. This parameter equips the model to consider the fact that the project may be completed under the initial budget by a certain percent. m is an arbitrary number based upon agency's objectives and α can be determined by reviewing the historical cost overruns/underruns. An agency may decide to input m equal to zero and find the α corresponding to that. Fig. 1 illustrates the truncated normal $X \sim N(\mu', \sigma')$, where the normal component of the distribution is $N(\mu, \sigma)$.

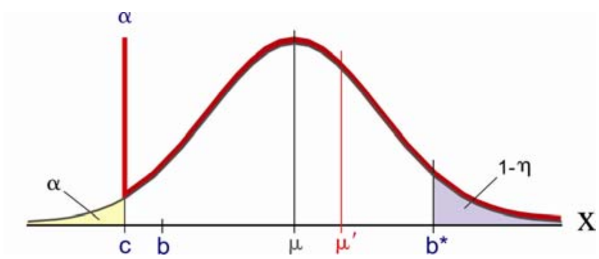

Fig. 1: Truncated Normal Distribution

We presume that there is a database of construction projects comprising of $i = 1, ..., n$ projects each with the initial budget of b_i. It is found through this historical data that there is α % chance to have m percent underrun and get the project done with $c_i = (1-m).b_i$. The model is constructed using the following parameters:

$b_i =$ Initial budget allocated for project i;
$x_i =$ Actual cost of project i;
$m =$ The maximum expected underrun;
$c_i =$ Minimum expected project i cost which is $(1-m).b_i$;
$\alpha =$ Percent of projects in the historical data having underrun more than or equal to m;

$\delta =$ Cost overrun/ underrun;
$\overline{\delta} =$ Average of cost overruns/underruns in the historical data;
$\overline{\sigma} =$ Standard deviation of cost overruns/ underruns in the historical data;
$\beta =$ Average rate of cost overrun/ underrun relative to b which is $1+\overline{\delta}$;
$\rho =$ Average rate of cost overrun/ underrun relative to c which is $\beta/(1-m)$;
$\mu_i =$ Mean of underlying normal distribution in project i;
$\sigma_i =$ Standard deviation of underlying normal distribution in project i;
$\mu_i' =$ Mean of hybrid normal distribution in project i;
$\sigma_i' =$ Standard deviation of hybrid normal distribution in project i;
$\varphi =$ A constant coefficient which is equal to σ_i' / c_i;
$b_i^* =$ Revised budget of project i;
$B =$ Sum of all individual initial budgets, $\sum b_i$;
$\eta =$ Percent of confidence that individual projects' cost will not be more than b^*;
$B^* =$ Sum of all revised individual budgets based on η, $\sum b_i *$;
$\gamma =$ Probability that portfolio of projects' cost will not be more than B^*;

Project cost X_i is defined as follows:

$$\begin{cases} P(X_i < c_i) = 0 \\ P(X_i = c_i) = \alpha \\ P(X_i > c_i) = 1-\alpha \end{cases} \quad (1)$$

The Probability Distribution Function (PDF) of project cost which is a truncated normal is:

$$\begin{cases} f(x) = \alpha & for\ x_i \le c_i \\ f(x) = \dfrac{1}{\sqrt{2\pi}\sigma_i}.e^{\frac{-(x_i-\mu_i)^2}{2\sigma_i^2}} & for\ x_i > c_i \end{cases} \quad (2)$$

The mean μ' and standard deviation σ' of truncated normal can be calculated using the following equations:

$$\mu_i' = E(X_i) = \alpha.c_i + x_i \times \int_{c_i}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_i}.e^{\frac{-(x_i-\mu_i)^2}{2\sigma_i^2}} dx_i = \alpha.c_i + (1-\alpha).\mu_i + \frac{\sigma_i}{\sqrt{2\pi}}.e^{\frac{-[\Phi^{-1}(\alpha)]^2}{2}} \quad (3)$$

$$\sigma_i'^2 = E(X_i^2) - \mu_i'^2 = \alpha.c_i^2 + x^2{}_i \times \int_{b_i}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_i}.e^{\frac{-(x_i-\mu_i)^2}{2\sigma_i^2}} dx_i - \mu_i'^2 = \alpha.c_i^2 - \mu_i'^2 + \frac{\sigma_i}{\sqrt{2\pi}}.[\sigma_i.\Phi^{-1}(\alpha) + 2\mu_i].e^{\frac{-[\Phi^{-1}(\alpha)]^2}{2}} + (1-\alpha).(\sigma_i^2 + \mu_i^2) \quad (4)$$

Where:

$$\Phi^{-1}(\alpha) = \frac{c_i - \mu_i}{\sigma_i} \quad (5)$$

$\Phi^{-1}(\alpha)$ is the inverse of cumulative function for standard normal distribution. β is the average rate of cost overruns/underruns and can be calculated from the historical data as:

$$\beta = 1+\overline{\delta} = 1+\frac{1}{n}.\sum_{i=1}^{n}\frac{x_i - b_i}{b_i} \quad (6)$$

Also, we know μ_i' is the expected value of the final cost of project i, so we can model it as a multiplier of its budget:

$$\mu_i' = \beta.b_i = \rho.c_i \quad (7)$$

and knowing that $c_i = (1-m).b_i$, we have:

$$\beta.b_i = \rho.c_i \rightarrow \rho = \frac{\beta.b_i}{c_i} \quad \Rightarrow \quad \rho = \frac{\beta}{1-m} \quad (8)$$

By rearranging Eq. (5), mean of the underlying normal distribution is calculated. Also, by substituting Eq. (7) and (9) in Eq. (3) and rearranging, the standard deviation of underlying distribution is found:

$$\mu_i = c_i - \sigma_i.\Phi^{-1}(\alpha) \quad (9)$$

$$\sigma_i = \frac{(1-\rho).(1-m).b_i}{(1-\alpha).\Phi^{-1}(\alpha) - \frac{e^{\frac{-[\Phi^{-1}(\alpha)]^2}{2}}}{\sqrt{2\pi}}} \tag{10}$$

Reviewing Eq. (4) shows that all terms of $\sigma_i'^2$ are comprised of a constant coefficient times $c_i^{\,2}$. Therefore, $\sigma_i'^2$ can be written in the form of:

$$\sigma_i'^2 = \varphi^2.c_i^{\,2} \xrightarrow{or} \sigma_i' = \varphi.c_i \tag{11}$$

Where φ is a constant coefficient for all values of σ_i' and c_i that can be computed using Eqs. (4) and (11). Referring to Fig. 1, if a budget $b_i^* > b_i$ for each project is selected, the chance of shortfall of budget would be limited to η. So:

$$\begin{aligned} P(X_i \le b_i^*) &= F(b_i^*) \\ &= \alpha + \Phi(\frac{b_i^* - \mu_i}{\sigma_i}) - \Phi(\frac{c_i - \mu_i}{\sigma_i}) \\ &= \alpha + \Phi(\frac{b_i^* - \mu_i}{\sigma_i}) - \alpha \\ &= \eta \end{aligned} \tag{12}$$

If we rearrange Eq. (12), we obtain:

$$b_i^* = \mu_i + \sigma_i.\Phi^{-1}(\eta) \tag{13}$$

We know that the original portfolio budget is:

$$\begin{aligned} B = \sum b_i = \sum \frac{c_i}{(1-m)} = \frac{1}{(1-m)} \sum c_i \\ \xrightarrow{Then} \sum c_i = (1-m).B \end{aligned} \tag{14}$$

Using Eq. (13), (14) and substituting μ_i using Eq. (9), the new portfolio budget can be computed as follows:

$$\begin{aligned} B^* &= \sum b_i^* \\ &= \sum[\mu_i + \sigma_i.\Phi^{-1}(\eta)] \\ &= \sum[c_i - \sigma_i.\Phi^{-1}(\alpha)] + \sum[\Phi^{-1}(\eta).\sigma_i] \\ &= (1-m).B + [\Phi^{-1}(\eta) - \Phi^{-1}(\alpha)].\sum \sigma_i \end{aligned} \tag{15}$$

Substituting Eq. (10) in Eq. (15), the ratio of B^*/B is found as follows:

$$\frac{B^*}{B} = (1-m).\left[1 + \frac{(1-\rho).[\Phi^{-1}(\eta) - \Phi^{-1}(\alpha)]}{(1-\alpha).\Phi^{-1}(\alpha) - \frac{e^{\frac{-[\Phi^{-1}(\alpha)]^2}{2}}}{\sqrt{2\pi}}}\right] \tag{16}$$

Eq. (16) gives the required portfolio budget increase with respect to η. We now assume that project costs are statistically independent and the total actual cost of all projects in the portfolio is T. The assumption of independence is reasonable because these projects are scattered throughout the country and the owners and management structure are different, as these are various state agencies using the federal funds. However, it should be noted that pairwise correlation between costs of any two concurrent projects may exist when there is a belief that they are using common resources, common management, or being affected by other common factors such as statutory/ regulatory constraints, political conditions, or unemployment.

Based on Central Limit Theorem, T will follow an approximate normal distribution with the mean μ_T and the standard deviation σ_T. Therefore:

$$T = \sum X_i \xrightarrow{CLT} T \sim N(\mu_T, \sigma_T) \tag{17}$$

$$\mu_T = \sum \mu_i' = \sum(\rho.c_i) = \rho.(1-m).B \tag{18}$$

$$\begin{aligned} \sigma_T^2 = \sum \sigma_i'^2 = \sum(\varphi.c_i)^2 = \varphi^2.\sum c_i^2 \\ \xrightarrow{Then} \sigma_T = \varphi.\sqrt{\sum c_i^2} \end{aligned} \tag{19}$$

Defining γ as the percent of confidence that portfolio of projects cost will not be more than B^*, we have:

$$P(T < B^*) = \Phi(\frac{B^* - \mu_T}{\sigma_T}) = \gamma \tag{20}$$

Rearranging Eq. (20) and using Eq. (18) and (19), we obtain:

$$\Phi^{-1}(\gamma)=\frac{B^*-\mu_T}{\sigma_T} \xrightarrow{Then} B^*=\rho.(1-m).B+\varphi.\sqrt{\sum c_i^{\,2}}.\Phi^{-1}(\gamma) \tag{21}$$

By equating Eq. (15) and (21) and substituting σ_i from Eq. (10), γ can be found as follows:

$$\gamma=\Phi\{\frac{(1-m).(1-\rho).B}{\varphi.\sqrt{\sum c_i^{\,2}}}\times [1+\frac{[\Phi^{-1}(\eta)-\Phi^{-1}(\alpha)]}{(1-\alpha).\Phi^{-1}(\alpha)-\frac{e^{-[\Phi^{-1}(\alpha)]^2}}{\sqrt{2\pi}}}]\} \tag{22}$$

Rearranging Eq. (22) gives:

$$\eta=\Phi\{\Phi^{-1}(\alpha)+ [(1-\alpha).\Phi^{-1}(\alpha)-\frac{e^{-[\Phi^{-1}(\alpha)]^2}}{\sqrt{2\pi}}]\times [\frac{\varphi.\sqrt{\sum c_i^{\,2}}.\Phi^{-1}(\gamma)}{(1-m).(1-\rho).B}-1]\} \tag{23}$$

Eqs. (16) and (23) are used to calculate the require percent increase in portfolio budget and probability of overrun for each individual project in the portfolio based on probability of having sufficient budget for the portfolio of projects. In the next section we will see how these values are updated when new completed projects become available.

3. Fundamentals of Bayesian Approach

An agency employing the proposed model is expecting to experience less or even no cost overrun in the newly funded projects. Since the model has been constructed based on limited observed data, it is conceivable that the overrun will not be eliminated in the first attempt. Therefore the model needs to be updated on a yearly or bi-yearly basis, depending on the number of completed projects. To this end, a Bayesian approach is utilized to update the model as the information regarding the costs of new projects become available. When the observed data are limited and making decision on the available information is required, the Bayesian approach can help update the system as more data is acquired. The fundamentals are based on Eq. (24) (Ang and Tang 2006):

$$f''(\bar{\delta})=k.L(\bar{\delta}).f'(\bar{\delta}) \tag{24}$$

where:

$k=\left[\int_{-\infty}^{\infty}P(\delta_j=\delta_1,...,\delta_k|\bar{\delta}).f'(\bar{\delta}).d(\bar{\delta})\right]^{-1}$ is the normalizing constant;

$L(\bar{\delta})=$ the likelihood of observing the new cost overruns/ underruns assuming a given $\bar{\delta}$, mean of the distribution; $f'(\bar{\delta})=$ prior distribution of $\bar{\delta}$; and $f''(\bar{\delta})=$ posterior distribution of $\bar{\delta}$.

Eq. (24) can be used to update the proposed model in light of new information acquired through newly completed projects during a certain period of time. $\bar{\delta}$ is the average of cost overruns/underruns required to calculate the parameters β and ρ of the model.

4. Bayesian Approach for *k* Completed Projects

Let's assume that k new projects are recently completed. Further, assume that these k new projects with the cost overrun/underrun of δ_j are statistically independent of each other. The probability of observing $\delta_1,...,\delta_k$ coming from a population Δ having an underlying normal distribution $N_\Delta(\bar{\delta},\sigma)$ is:

$$P(\delta_j=\delta_1,...,\delta_k|\bar{\delta})=\prod_{j=1}^{k}N_\Delta(\delta_j|\bar{\delta}).d(\delta) \tag{26}$$

Therefore, the likelihood function can be written as:

$$L(\bar{\delta})=\prod_{j=1}^{k}N_\Delta(\delta_j|\bar{\delta},\bar{\sigma}) =\prod_{j=1}^{k}\frac{1}{\sqrt{2\pi}.\sigma}\exp\left[-\frac{1}{2}.\left(\frac{\delta_j-\bar{\delta}}{\sigma}\right)^2\right] \tag{27}$$

It should be noted that Eq. (27), the joint likelihood of the sample, is the product of k individual observed normal likelihoods. It is known that the product of k normal likelihoods has the shape of a normal distribution as follows (Bolstad 2007):

$$L(\bar{\delta})\propto N\left(\frac{\delta_1+...+\delta_k}{k},\frac{\sigma}{\sqrt{k}}\right)=N(\delta_L,\sigma_L) \tag{28}$$

The posterior distribution is now the product of likelihood $L(\bar{\delta})$ and prior $f'(\bar{\delta})$. If the prior is a normal PDF such as $f'(\bar{\delta}) \sim N(\delta', \sigma')$, then the posterior has also a normal PDF with the mean and standard deviation as follows (Ang and Tang 2006):

$$f''(\bar{\delta}) = k.L(\bar{\delta}).f'(\bar{\delta}) \sim N(\delta'', \sigma'') \tag{29}$$

$$\begin{cases} \delta'' = \dfrac{\delta_L.(\sigma')^2 + \delta'.(\sigma_L)^2}{(\sigma')^2 + (\sigma_L)^2} \\ \sigma'' = \dfrac{\sigma'.\sigma_L}{\sqrt{(\sigma')^2 + (\sigma_L)^2}} \end{cases} \tag{30}$$

By finding the updated distribution of cost overruns/ underruns, the main model is revised to calculate new required percent increase in portfolio budget B^*/B. This is done by substituting α, β, and ρ parameters of the model with the following values:

$$\begin{cases} \alpha_{new} = P(x < -m) = \\ \qquad P(Z < \dfrac{-m-\delta''}{\sigma''}) = \Phi(\dfrac{-m-\delta''}{\sigma''}) \\ \beta_{new} = 1 + \delta'' \quad \Rightarrow \quad \rho_{new} = \dfrac{\beta_{new}}{1-m} \end{cases} \tag{31}$$

5. Numerical Example

5.1. Main Model

To show the application of the model, a set of 28 transit projects (Booz Allen Hamilton 2005) is selected (Table 1). These projects have been funded by Federal Transit Administration (FTA) of the U.S. Department of Transportation in the past 20 years. Cost overrun/underrun of each project is defined to be the percent of difference between actual final cost (as-built cost) and estimated cost at the end of final design (FD), when in transit projects Full Funding Grant Agreement (FFGA) is established by the FTA. Reviewing Table 1 shows that 22 out of 28 projects have been completed before 2004, five projects in 2004 and one project in 2005. We set aside the 22 projects as historical data to find the primary values for α and β used in the model. Then the model is applied to the set of five projects completed in 2004 to see the effect of model on cost overruns/underruns. The project completed in 2005 is not considered in the application process of the model as it would fall in another fiscal year. From this point on, for consistency and ease of referencing, we call the set of 22 projects "Historical Dataset", the set of five projects "First Dataset".

Table 1: Cost Overrun/underrun of 28 Transit Projects (Booz Allen Hamilton 2005)

Proj. ID	Project Name	Year Completed	Cost at the FFGA (in M$)	Actual Cost (in M$)	Cost Overrun/ Underrun
1	Atlanta North Line Extension	1999	$381.3	$472.7	23.97%
2	Boston Old Colony Rehabilitation	1997	$551.9	$565.0	2.37%
3	Boston Silver Line (Phase 1)	2004	$413.4	$604.4	46.20%
4	Chicago Southwest Extension	1989	$350.9	$474.6	35.25%
5	Dallas South Oak Cliff Extension	2002	$517.2	$437.2	-15.47%
6	Denver Southwest Line	1999	$176.3	$177.1	0.45%
7	Los Angeles Red Line MOS 1	1991	$960.3	$1,490.1	55.17%
8	Los Angeles Red Line MOS 2	1994	$1,524.6	$1,921.6	26.04%
9	Los Angeles Red Line MOS 3	1995	$1,345.6	$1,227.6	-8.77%
10	Minneapolis Hiawatha Line	2004	$675.4	$715.3	5.91%
11	New Jersey Hudson-Bergen MOS1	2000-2002	$992.1	$1,113.0	12.19%
12	New York 63rd Street Connector	2001	$645.0	$632.3	-1.97%
13	Pasadena Gold Line	2003	$693.9	$677.6	-2.35%
14	Pittsburgh Airport Busway (Phase 1)	2000-2002	$326.8	$326.8	0.00%
15	Portland Airport MAX Extension	2001	$125.0	$127.0	1.60%
16	Portland Banfield Corridor	1984	$286.6	$246.8	-13.89%
17	Portland Interstate MAX	2004	$314.9	$349.4	10.96%
18	Portland Westside/Hillsboro MAX	1998	$910.2	$963.5	5.86%
19	Salt Lake North-South Line	1999	$312.0	$311.8	-0.06%
20	San Francisco SFO Airport Ext.	2003	$1,167.0	$1,550.2	32.84%
21	San Juan Tren Urbano	2005	$1,250.0	$2,250.0	80.00%
22	Santa Clara Capitol Line	2003	$159.8	$162.5	1.69%
23	Santa Clara Tasman East Line	2001	$275.1	$276.2	0.40%
24	Santa Clara Tasman West Line	1999	$332.5	$280.6	-15.61%
25	Santa Clara Vasona Line	2004	$313.6	$316.8	1.02%
26	Seattle Busway Tunnel	1990	$395.4	$611.1	54.55%
27	St Louis Saint Clair Corridor	2000	$339.2	$336.5	-0.80%
28	Washington Largo Extension	2004	$433.9	$456.0	5.09%

To verify the assumption of normality, a test of goodness of fit using @Risk (Palisade Corp. 2008) software is conducted on 22 cost overruns/ underruns of Historical Dataset. The test using the Chi-squared statistic passed at 1% level of significance (P-value= 0.0219). Fig. 2 depicts the superposition of the normal distribution on the original data histogram.

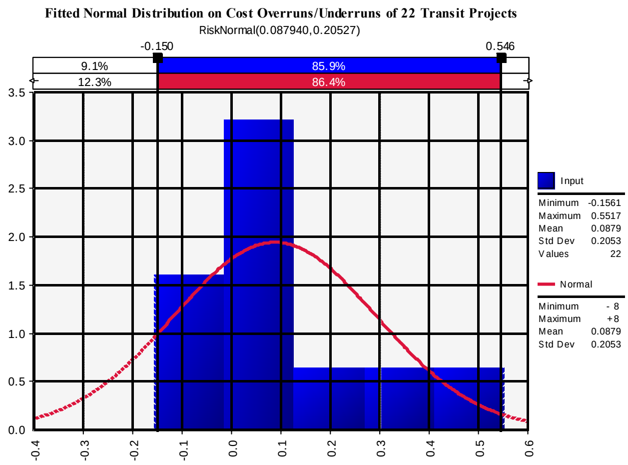

Fig. 2: Fitted Normal Distribution on Cost Overruns/Underruns of 22 Transit Projects

Fig. 2 demonstrates the limitation in the cost underrun values. It means that in the real world we are dealing with projects that their costs would not be less than a certain value. This certain value can be approximated using historical data. Reviewing the historical data, we assume that the FTA defines $m = 15\%$ as the maximum expected underrun. Using Fig. 2, it is found that the value of α corresponding with $m = 15\%$ is $\alpha = 9.1\%$ and the average of cost underruns/overruns is $\bar{\delta} = 8.79\%$; thus $\beta = 1.0879$ and $\rho = 1.0879/(1-0.15) = 1.2799$. After estimating values of α and β from the historical data, the model is ready to be applied on any prospective set of projects which here is the First Dataset.

From Historical Dataset (Fig. 2), $\alpha = 9.1\%$, $\beta = 1.0879$, and $\rho = 1.2799$ were estimated. Using Eqs. (4) and (11), $\varphi = 0.1875$ is calculated. Then by the means of Eq. (23), the corresponding η s for different γ s are calculated. This is done for γ between 5% and 95% and the result is depicted in Fig. 3.

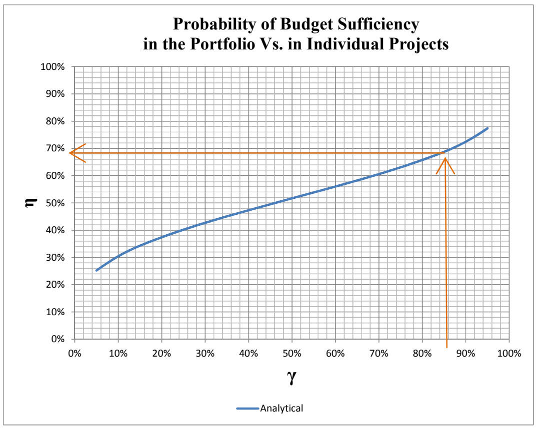

Fig. 3: Probability of Budget Sufficiency in the Portfolio of Independent Projects (γ) vs. in Individual Projects (η)

Then, Eq. (16) is employed to compute the required percent increase in portfolio budget based on the η values (probability that individual projects are sufficiently funded) found from Eq. (23). The required percent increase in budget is graphed versus γ and shown in Fig. 4. In order to make sure that the results are accurate, we simulated the model to find increasing factor which is superimposed on the analytical curve found using the analytical approach. These two curves are very similar.

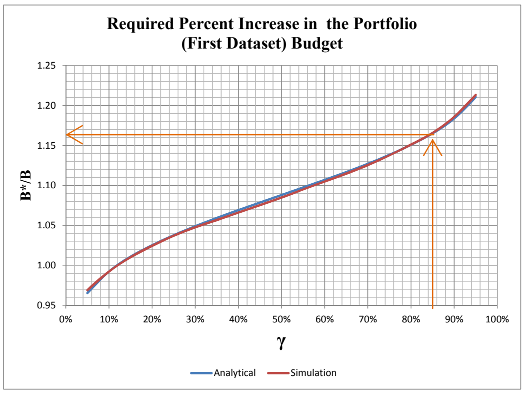

Fig. 4: Required Percent Increase in Budget (B^*/B) *vs.* Probability of Budget Sufficiency in the Portfolio of Independent Projects (γ)

For example, one can see in Fig. 3 that if the FTA wants to have 85% confidence that allocated budget for the portfolio of projects will not fall short, it needs to consider a minimum level of confidence of 68.78% $\cong$ 69% in each individual project risk assessment. Also, Fig. 4 illustrates that the FTA needs to increase the portfolio budget by 16.52% in order to have 85% level of confidence that the budget for the portfolio is sufficient. This finding is significant because the proposed methodology provides a method for calculating the percent increase over existing portfolio budget levels to achieve a certain confidence level in individual projects.

In Table 2, a comparison is made between the actual cost overrun/underrun of projects in the First Dataset and cost overrun/underrun if the budget had been adjusted with the estimated increasing factors. Even though the required budget increase in the portfolio can be distributed differently between the projects, we assume all will be increased proportionally by multiplying the required increase factor ($B^*/B = 1.1652$) by the cost at the FFGA to reach Adjusted Cost at the FFGA.

Table 2: Comparison of Cost Overrun/Underrun of Projects in the First Dataset Using the Proposed Model

Proj. ID	Cost at the FFGA (in M$)	Adj. Cost at the FFGA (in M$)	Actual Cost (in M$)	Cost Overrun/ Underrun	
				Actual	Adjusted
3	$413.40	$481.71	$604.40	46.20%	25.47%
10	$675.40	$786.99	$715.30	5.91%	-9.11%
17	$314.90	$366.93	$349.40	10.96%	-4.78%
25	$313.60	$365.42	$316.80	1.02%	-13.30%
28	$433.90	$505.59	$456.00	5.09%	-9.81%
Total	$2,151.20	$2,506.64	$2,441.90	13.84%	-2.31%

In the second column (from the left) of Table 2, the original costs at the FFGA of all five projects in the portfolio (First Dataset) are presented. As it was stated earlier, the proposed model considering the performance of past projects in the Historical Dataset suggests the FTA to increase the total portfolio budget of the First Dataset by 16.52% in order to have 85% confidence that the budget for the portfolio (First Dataset) is sufficient. Then the costs at the FFGA of all projects in the portfolio are adjusted by the increase factor of 1.1652 and are shown in the third column of Table 2. This column represents the budget of each project if the FTA had used the proposed model. The actual costs of projects are given in the fourth column of Table 2. The fifth column shows the actual cost overruns/ underruns considering the original budgets at the FFGA and the sixth column gives the cost overruns/ underruns if the FTA had used the model. Table 2 depicts that the model could alleviate cost overrun of some projects in the First Dataset. However, the promise of this model is to protect a portfolio of projects against cost overrun. The last row of Table 2 shows that if the FTA had used the proposed model to allocate budget for five new projects, they could prevent occurring cost overrun of 13.84% with experiencing -2.31% cost underrun. We expect by updating the model and considering the performance of the recently completed projects, we reach more accurate and optimized increasing factor for budgeting of future projects.

5.2. *Bayesian Updating*

In this step, we use the information collected from completed projects (actual costs) in the First Dataset to update the model. The cost overruns/ underruns of five projects are considered new observations and serve to form the underlying distribution. The prior distribution is the normal distribution fitted to the histogram of 22 cost overruns/ underruns in the Historical Dataset with a mean of 8.79% and standard deviation of 0.2053.

Considering 85% confidence as a reasonable level, we found that 16.52% increase on the total budget was required. By means of Bayesian updating and recent performance of the transit projects sponsored by the FTA, the α and β of the model can be updated. The prior distribution comes from the Historical Dataset as follows:

$$f'(\bar{\delta}) = N(0.0879, 0.2053) = \frac{1}{\sqrt{2\pi}(0.2053)} \exp\left[-\frac{1}{2}.\left(\frac{\bar{\delta}-0.0879}{0.2053}\right)^2\right] \quad (32)$$

Five new observations are the cost overruns/underruns of projects with adjusted cost at the FFGA using 16.52% increasing factor shown in Table 2. Using Eq. (28), the joint likelihood function, the product of five individual normal PDFs, is calculated as:

$$L(\bar{\delta}) \propto N\left(\frac{\delta_1+...+\delta_k}{k}, \frac{\sigma}{\sqrt{k}}\right) = N(\frac{0.2542-0.0911-0.0478-0.1330-0.0981}{5}, \frac{0.1582}{\sqrt{5}}) = N(-0.0231, 0.0708) \quad (33)$$

To find the posterior distribution, Eq. (30) is used:

$$\begin{cases} \delta'' = \dfrac{\delta_L.(\sigma')^2 + \delta'.(\sigma_L)^2}{(\sigma')^2 + (\sigma_L)^2} = \\ \dfrac{-0.02313\times(0.2053)^2 + 0.0879\times(0.0708)^2}{(0.2053)^2+(0.0708)^2} = \\ -1.13\% \\ \sigma'' = \dfrac{\sigma'.\sigma_L}{\sqrt{(\sigma')^2+(\sigma_L)^2}} = \\ \dfrac{0.2053\times(0.0708)}{\sqrt{(0.2053)^2+(0.0708)^2}} = 0.0669 \end{cases} \quad (34)$$

Prior, likelihood, and posterior distributions of cost overrun/ underrun are shown in Fig. 5.

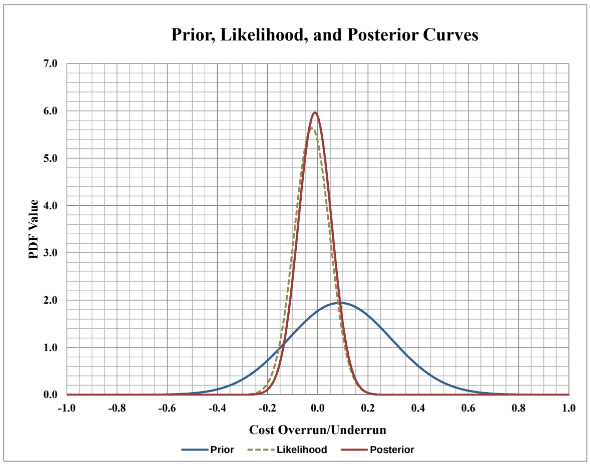

Fig. 5: The prior, likelihood, and posterior distributions of cost overrun/ underrun Using Projects in the First Dataset

The posterior distribution parameters can now be used to update α, β and ρ parameters considering $m = 15\%$:

$$\begin{cases} \alpha_{new1} = P(x < -m) = P(Z < \dfrac{-m-\delta''}{\sigma''}) = \\ \Phi(\dfrac{-0.15+0.0113}{0.0669}) = 1.90\% \\ \beta_{new1} = 1+\delta'' = 1+(-0.0113) = 0.9887 \Rightarrow \\ \rho_{new1} = \dfrac{\beta_{new1}}{1-m} = \dfrac{0.9887}{0.85} = 1.1632 \end{cases} \quad (35)$$

Replacing the new values of parameters (α_{new1}, β_{new1} and ρ_{new1}) in the model, it is ready and updated to be applied to any prospective dataset. One can see in Fig. 5 that the posterior curve has become narrower and moved to the left compared to the prior curve. This means when the performance of the recently completed projects came into consideration, the average cost overruns/ underruns ($\overline{\delta}$) from 8.79% in the Historical Dataset (prior) decreased to 1.13% underrun in the posterior curve. Moreover, as an advantage of Bayesian updating, when more data becomes available, the dispersion (standard deviation) of the parameter under consideration diminishes and that the posterior curve becomes narrower. It should be noted that if one integrates the newly completed projects into Historical Dataset and calculates the parameters of the model, this will give an equal weight to all projects. In other words, this approach will not distinguish between a project completed in 1984 and a project completed in 2004. Using Bayesian approach, the performance of recently completed projects (First Dataset) will have more influence on updating the parameters of the model than projects in the Historical Dataset. In summary, this procedure enables the model to suggest the required increase in the portfolio budget of the future projects considering mostly the performance of the most recent projects used for the updating as well as the performance of historical projects.

6. Conclusion

In this paper, a model is proposed which uses a truncated normal distribution and utilizes historical data to assist the agencies to estimate the required confidence level for risk assessment in the project-level in order to get a desired confidence for the sufficiency of portfolio budget. It also calculates the required increase in the portfolio budget based on the desired confidence level. For instance, in the given numerical example, 22 transit projects (Historical Dataset) were used to initialize the model and calculate the primary parameters of the model. Then the model was applied to five transit projects (First Dataset) to estimate the necessary project-level confidence level for risk assessment and the required increase in the portfolio budget while the FTA has 85% confidence that the portfolio budget will not fall short. It was found that using the model, in order to have 85% confidence that the portfolio budget for the First Dataset is sufficient, each individual project in the portfolio should have a contingency such that $\eta = 69\%$; moreover, the FTA needs to increase the original portfolio budget by 16.52%. This model considers the recent performance of the newly completed projects and is updated as new project data becomes available employing a Bayesian approach. The example provided here is updated with the actual costs of five projects in the First Dataset. The primary parameters (α, β and ρ) of the model are updated and it becomes ready to be used for any upcoming portfolio of projects. This process can be repeated on regular intervals (*e.g.* every two years) so that the future projects can be budgeted considering the most recent performance of projects. The proposed model can be used by agencies such as the FTA which is funding a portfolio of transit projects every year as an effective tool to develop contingency budget.

7. References

1. Ang, A., and Tang, W. (2007). *Probability Concepts in Engineering: Emphasis on Applications in Civil and Environmental Engineering*, 2nd Edition, John Wiley and Sons, Inc, Hoboken, NJ.
2. Bakhshi, P. and Touran, A. (2009). "Comparison of Current Probabilistic Approaches for Budget Estimating for Transportation Projects." 7th International Probabilistic Workshop, Delft, The Netherlands, 25-26 November.
3. Bolstad, W. (2007). *Introduction to Bayesian Statistics*, 2nd Edition, John Wiley and Sons, Hoboken, NJ.
4. Booz Allen Hamilton Inc. (2005). "Managing Capital Costs of Major Federally Funded Public Transportation Projects." *web-only* Document 31, Transit Cooperative Research Program (TCRP), Washington, D.C.
5. Flyvbjerg, B. (2006). "From Nobel Prize to Project Management: Getting Risk Right." *Project Management Journal*, 5-15, August.
6. Palisade Corporation. (2008). "@Risk: Risk Analysis Add-in for Microsoft Excel." Version 5.0.1: Professional Edition, Ithaca, New York.
7. Pickrell, D. H. (1990). "Urban Rail Transit Projects: Forecast Versus Actual Ridership and Cost." DOT-T-91–04, U.S. Dept. of Transportation, Washington, D.C.
8. Molenaar, K. R. (2005). "Programmatic Cost Risk Analysis for Highway Mega-projects." *Journal of Construction Engineering and Management*, ASCE, Vol. 131, No. 3, 343–353, March.
9. Sinnette, J. (2004). "Accounting for Megaproject Dollars." *Public Roads,* Federal Highway Administration, Vol. 68, No. 1, July/August, Available at:

   http://www.fhwa.dot.gov/publications/publicroads/04jul/07.cfm (Accessed on Feb 08, 2010).
10. Shane, J., Molenaar, K., Anderson, S. and Schexnayder, C. (2009). "Construction Project Cost Escalation Factors." *Journal of Management in Engineering,* ASCE, Vol. 25, No. 4, 221-229, October.
11. Springer, M. (1979). *The Algebra of Random Variables*, 1st Edition, John Wiley and Sons, Inc, Hoboken, NJ.
12. Touran, A. (2010). "A Probabilistic Approach for Budgeting in a Portfolio of Projects." *Journal of Construction Engineering and Management,* ASCE, Vol. 136, No. 3, 361-366, March.