Estimation of Risk Contingency Budget in Projects using Machine Learning

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 55-10 (2022) 3238–3243

Estimation of Risk Contingency Budget in Projects using Machine Learning

C. Capone*, T. Narbaev**

**Kazakh-British Technical University, Almaty, Kazakhstan (e-mail: c_capone@ise.ac).*

***Kazakh-British Technical University, Almaty, Kazakhstan (e-mail: t.narbaev@kbtu.kz).*

Abstract: To manage risks against unexpected cost overruns, project teams use Contingency Budget (CB). Its accurate estimation has been a subject of multiple studies proposing either deterministic or probabilistic models. In this study, we propose a deterministic Machine Learning-based approach to estimate CB. Based on the k-means clustering, our model integrates the Expected Monetary Value (EMV) method and binomial distribution concepts. We test our methodology using 20 risk registers containing 25 risks with associated probabilities and impacts. Using Monte Carlo simulation, we compare our model's estimates with the ones by the traditional EMV. The model provided more accurate CB estimates and is more straightforward in use than the Monte Carlo simulation.

Keywords: Contingency Budget, Machine Learning, Project Management, Risk management, Risk probability, and impact.

1. INTRODUCTION

Since projects are unique business endeavors, risks are inherent to their nature. A variety of risks can impact the projects, from its start to finish, making them depart from their planned requirements (e.g., time, cost, and quality) (Thamhain, 2013; De Marco et al., 2017; Hamzeh et al., 2020). Project teams use contingency reserves to manage such risks and associated deviations from the project plan. According to the Project Management Institute (2021), contingency reserve is time or money allocated in the schedule or cost baseline for known risks with active response strategies. To address identified risks, should they occur, the project budget should include contingency reserve funds, i.e., Contingency Budget (CB).

CB is a reserve set aside in the project budget to cover approved risk response actions against unexpected cost overruns (Barraza and Bueno, 2007; Project Management Institute, 2021). Effective use of CB is crucial since it is used to protect the interests of project stakeholders (e.g., sponsors, contractors, and investors) from risks. Also, CB can be used as a potential opportunity to improve the relationship of the project partners (e.g., return unspent CB to a client or use it to improve a project outcome) (Baccarini 2004; Narbaev and De Marco, 2014).

CB is defined during the project planning phase, and various methods exist to estimate it. Conventional methods include taking a given percent of the project budget or a fixed value (Xie et al., 2012). Such approaches are deterministic and do not consider uncertainty. They are usually a top-down estimate defined at the aggregate project level and often based on the organization's experience (e.g., analogous or benchmark estimate) in managing projects. These estimates do not involve a detailed quantitative analysis of the risks inherent to a given project. Furthermore, being every project a unique endeavor, they do not consider the project's unique characteristics.

Alternatively, more advanced approaches have been proposed in the literature to quantify risk events and their impact. Standard features of such models are comprehensive risk representation and ranking, risk probability calculation, risk impact calculation, and iteration capabilities (Project Management Institute, 2019). Most such probabilistic models are based on Decision Tree analysis, probability-impact analysis, Expected Monetary Value (EMV), Program Evaluation and Review Technique (PERT), Buffer management, Bayesian Belief Network, and Monte Carlo simulation approaches (Cagliano et al., 2015; Hu et al., 2015; Nunez et al., 2016; Qazi et al., 2022). In projects, such models quantify risk probabilities and impacts using a risk register or risk breakdown matrix (Project Management Institute, 2019), which detailed explanation is given in Hillson et al. (2006). The following section introduces a few such studies pertinent to our proposed approach.

We offer a Machine Learning (ML) based model to estimate CB in projects in this study. Our model belongs to the family of unsupervised ML techniques based on k-means clustering. The model integrates the EMV approach and the Binomial distribution. It overcomes the limitation (i.e., inaccuracy) of the EMV approach and produces more accurate CB estimates through the clustering of risks' probabilities and impacts. Then, we compare the proposed model's estimates with the estimates obtained by the Monte Carlo simulation.

In the next section, we briefly introduce the literature. Then, we present our model and its application results. Finally, we

10.1016/j.ifacol.2022.10.140

conclude our paper with the main findings, limitations, and expected future research.

2. PERTINENT LITERATURE

Numerous studies have proposed models to assess risks and manage appropriate CB in projects. First, we briefly review the pertinent studies relevant to the risk assessment and CB management during the project planning phase. Then, we present the studies that addressed CB during the project execution phase, including its re-evaluation.

Risk assessment with a CB estimation is one of the crucial processes in the project planning phase. It helps project teams manage potential risk impacts on the project's scope, schedule, cost, and quality (Moreno-Cabezali and Fernandez-Crehuet, 2020) and avoid schedule delays and budget overruns (De Marco and Narbaev, 2021). In this respect, Touran (2003) developed a probabilistic model that considered uncertainties in CB calculations based on the confidence level (CL) defined by the project client. His model estimated the change rates in project costs, the average cost of such changes, and the variation coefficient of the cost of such changes. On the other side, Cioffi and Khamooshi (2009) proposed a method to assess the total risk impact with a given certainty. Their method was based on the probability and impact values of risks from a risk register, and they used CL to calculate CB. To address the traditional project time-cost tradeoff problem, Nunez et al. (2016) aimed to minimize the time reduction cost by optimizing the penalty risk exposures from the expected delays. Based on this, they proposed efficient heuristics to readjust resource allocation during project status reviews.

A few studies addressed the budget allocation issues in managing risks. Zhang et al. (2018), using the bow-tie risk assessment, developed an efficient budget allocation model that considered various disruption risks and, based on their impacts, offered solutions either to prevent or mitigate the risks. Their model can assess the impacts of disruption risks on the supply chain performance. With the emphasis on transportation projects, Dadsena et al. (2019) proposed a fuzzy logic-based model using the failure mode and effect analysis. Under the given risk impact (criticality) value and associated mitigation budget, they suggested the model to develop a set of risk mitigation strategies. Alternatively, Moreno-Cabezali and Fernandez-Crehuet, (2020) developed a fuzzy logic-based model utilizing the opinions of 90 experts (through a survey) about risk probabilities and risk impacts on the performance of additive manufacturing Research & Development (R&D) projects. As stated, their model could improve managerial decision-making by prioritizing the risks (e.g., defective design, insufficient financing) that are critical to such R&D projects.

We note that the risks and their associated CB should not be only estimated during the project planning phase but effectively managed and re-evaluated during the project execution phase. In this respect, using Monte Carlo simulation, Barraza and Bueno (2007) offered an approach to allocating the portions of the total CB to the project activities. Assuming that the activities had the same normal distribution patterns of CB, the contingency could be updated based on the past project performance. Xie et al. (2012) presented a model for updating CB during the project execution. Based on the past periodic information about identified risks in a sample project and using the Value at Risk technique, they produced the range of the CB forecasts for the remaining life of the project. Using second-moment Bayesian statistics, Kim (2015) proposed a model for CB-adjusted cost forecasting in projects. His method considered the changes in past cost performance, which could update the project's budget estimates. This allowed for necessary risk response actions through such contingency control. Narbaev and De Marco (2017) proposed another model for CB-adjusted cost forecasting. Their approach considered different patterns of contingency spending that were grouped based on the project manager's (aggressive, neutral, and passive) risk response actions. Under given risk, the project's total budget was adjusted using nonlinear regression modeling. In De Marco et al. (2017), the researchers extended the validity of their model by testing it at the early, middle, and late stages of the project development. Overall, the reliability criteria for their model were accuracy and stability of the CB-based budget forecasts that could capture the inherent relationship between schedule, cost, and risk of a project. Recently, Kuo et al. (2019) developed a model based on the risk breakdown matrix and quantified CB using a Particle Swarm Optimization algorithm. Their model optimized the contingency distribution while keeping the project's total budget minimum. This allowed the project team to choose the optimal risk response actions (i.e., risk acceptance, mitigation, avoid, or transfer). Alternative to the CB optimization, Guan et al. (2021) developed a budget allocation method and offered managerial solutions for risk prevention and risk protection separately. Based on numerical experiments, they found that the relationship between such risk response actions (measured by their cost) and associated impact could be either linear (by the continuous linear curve) or nonlinear (by the convex function).

On the one side, as a summary of the above review, we note the continuing research that proposes advanced probabilistic methods to estimate CB. One reason that motivates future studies is the issue of inaccuracy in CB estimations due to the lack of information about the project's outcome. Therefore, multiple studies on risk management application to construction, production, and information technology projects investigate this issue.

On the other side, to cope with the risk and uncertainties of the dynamic business environment, we see a substantial increase in using ML methods in the construction, production, and information technology management literature. Nevertheless, we note that their practical applications are still limited (Pospieszny et al., 2018; Kanakaris et al., 2019). Their applications to manage risks and estimate associated CB in projects are also scant in the literature.

3. METHODOLOGY AND RESULTS

3.1 Model background

The simulations, such as the Monte Carlo simulation, are widely used in risk analysis and CB estimation. The reason is that the deterministic methods cannot be easily applied for analysis under uncertainty (Bakhshi and Touran, 2014; Touran, 1993). Monte Carlo Simulation is without any doubt very accurate but also complex for practitioners to be applied. Keeping this in mind, we propose a new deterministic approach and compare its CB estimation results to the ones found by the Monte Carlo simulation. Our methodology extends two non-simulation methods to calculate CB: the EMV method (Touran, 2006) and binomial distribution (Cioffi and Khamooshi, 2007). Both methods consider the information coming from the project risk register with associated probability and impact values. We use ML to group the project risks in different clusters, homogenous by risk likelihood, and then apply the binomial distribution for each cluster.

3.2 Using the EMV approach to determine CB

The calculation of CB using the EMV approach (CB_{EMV}) is straightforward: CB is the sum of the product of each risk impact with its probability (Touran and Lopez, 2006). The model is presented in (1):

$$CB_{EMV} = \sum_{i=1}^{n} CB_i = \sum(impact_i \cdot probability_i) \quad (1)$$

Where $impact_i$ represents the impact of the i^{th} risk in the risk register and $probability_i$ its probability. CB_i represents the CB due to the i^{th} risk, and CB_{EMV} is the sum of all contingencies for the whole project. However, the limitation of this approach is that it considers CB as the expected impact value among all the possible impacts that can occur once the risks are materialized, as per the definition of the expected monetary value. Instead, it says nothing regarding CL (i.e., a range estimate) for CB to cover the risks present in the risk register. Indeed, managers often want to understand the CB under a given CL (e.g., in 90% of the cases that CB will be enough). Indeed practitioners need to know which CB should cover the materialized risks under a certain probability level. To represent this limitation more straightforwardly, let us consider a simple fictional risk register with three risks and their associated impacts and probabilities, as shown in Table 1.

Table 1. A sample risk register with three risks

Risk #	Impact (in USD)	Probability
1	1000	30%
2	1500	45%
3	2500	20%

According to (1), CB_{EMV} is equal to \$1475. With the aid of Monte Carlo simulation, we calculate that the probability of covering project risks using a contingency of \$1475 is only 43.94%, which is not sufficient in most cases. The simulation was set with one million iterations, and only on 43.94% of them, \$1475 was sufficient to cover the overall risks' impact.

3.3 Using the binomial distribution to determine CB

A discrete probability distribution can be used to overcome this EMV limitation, i.e., the impossibility of calculating CB with a given CL. The probability distribution allows us to identify the CB under a CL, i.e., the probability that the risks' impact is less or equal to the CB once the risks occur. In their study, Cioffi and Khamooshi (2007) approximate the discrete probability distribution that describes the probability of the occurrence of the risks with the binomial distribution. The binomial distribution represents the probability distribution of the number of successes of a series of experiments, with the premise that the probability of success (p) every time (n) we repeat the experiment is the same. A successful outcome in this scenario is the occurrence of a risk. With the binomial distribution, it is possible to calculate the probability that, among n risks, only k risks are happening if all the risks have the same probability of happening. The distribution returns a good approximation of this probability if the range of the risk probabilities is within 20% (Cioffi and Khamooshi, 2007), e.g., risks probabilities in the risk register are between 10% and 30%, and we use p as the average risks' probabilities. With a CL, e.g., 90%, we can calculate k, i.e., the number of risks that can happen within this CL, and from it, we find CB by multiplying k by the average of the risk register impacts. We have thus 90% of confidence that k risks will happen, and therefore we have 90% confidence that CB, calculated by k multiplied by average risks impact, will cover the project risks.

The binomial distribution Expected Value (EV) represents the number of risks that we expected to happen, and it is given by (2):

$$EV(n,p) = n \cdot p \quad (2)$$

Where p is the average probability in the risk register, and n is the number of risks present in the risk register.

To use (2), we assume that the risks present in the risk register have the same or very similar probabilities. Further, if we add the condition that all the risk's impacts are the same or at least very close to each other's, the expected CB can be approximated by (3):

$$CB \approx impact_{average} \cdot EV(n, p_{average}) \quad (3)$$

Where $impact_{average}$ is the average of the risk impacts and $p_{average}$ is the average risk probabilities.

To calculate CB within a given CL, we use the binomial probability mass function by (4):

$$S(x,p) = \frac{x!}{k!\,(x-k)!}\, p^k \cdot (1-p)^{x-k} = CL \quad (4)$$

Where x represents the number of risks under a given CL.

Solving it by x, the calculation of CB can be rewritten as in (5):

$$CB \approx impact_{average} \cdot x \quad (5)$$

3.4 Proposed ML model for CB estimation

Equation (5) is more accurate than (1) since we can calculate CB under a specific CL. However, it requires more complex calculations and a risk register with risks' probabilities very close to each other's and similar risks' impacts.

Instead, we propose a simplification of (5) using the binomial distribution variance (*Var*), given by (6):

$$Var(n,p) = n \cdot p \cdot (1-p) \tag{6}$$

The following equation can therefore approximate CB:

$$CB_{CL} \approx impact \cdot \Big(EV(n,p) + Z\_score(CL) \cdot \sqrt{Var\ (n,p)} \Big) \tag{7}$$

Where *Z_score(CL)* is the z-score associated with a specific CL. In most cases, CL is 90% (Rothwell, 2005), then the *Z-score(90%)*=1.2816

Equation (7) is an extension of (5). However, (7) still has the limitation present in (5); it is necessary to consider that the risks' probabilities and their associated impacts are the same or very close to each other's. We note that using the average probabilities and impacts might lead to inaccurate estimations.

Our methodology increases the CB estimation accuracy, and we compare the accuracy results of our model with the ones of the Monte Carlo simulation, using it as a benchmark. Fig. 1 in Appendix A presents our calculations. Our methodology uses a k-means ML algorithm to identify risk clusters and applies (7) to each cluster. Indeed, each cluster contains risks with similar probabilities, and therefore we can use the cluster's average probability (i.e., the cluster probability centroid) instead of the average risks probability. We maintain the approximation to consider risk impacts close to each other. In this approximation, however, we do not undermine the application of (7) since it is not related to the binomial distribution. Using our methodology, which considers the clusterization, we calculate CB for each cluster as per (8):

$$CB_i = ICC_i \cdot \Big(EV_i(n_i, PCC_i) + zscore(CL\%) \cdot \sqrt{Var(n_i, PCC_i)} \Big) \tag{8}$$

Where ICC_i represents the impact of the cluster centroid, calculated as the average impact of the risks included in the i^{th} cluster; PCC_i represents the probability of the cluster centroid, calculated as the average probability of the risks included in the i^{th} cluster; n_i is the number of risks in the i^{th} cluster, and CB_i is the CB of the i^{th} cluster under a given CL.

Then, the total CB for a given project is given by (9):

$$CB_{TOT} = \sum_{i=1}^{N} CB_i \tag{9}$$

We validate our methodology using a sample risk register that contains 25 risks with random probabilities and impacts, simulating with 3 clusters. In our simulation, the cluster centroid's probabilities are 11%, 51%, and 93%, and cluster centroid's impacts are $13,017.40, $696.8, and $205.83. With 100,000 Monte Carlo trials, we identified a CB ($CB_{TOT\text{-}MC}$) of $35,374 with CL equal to 90.97%. We note that $CB_{TOT\text{-}MC}$ represents our benchmark.

Applying (7) to the full risk register (i.e., considering all risks in one cluster), we calculate the deterministic CB of $79,565 with a theoretical CL of 90% but the actual CL of 99.99%. The actual CL is calculated using the Monte Carlo trials. We see that CB calculated using (7) is far from CB calculated using Monte Carlo ($CB_{TOT\text{-}MC}$). We note that the (7) inaccuracy is due to the use of the risks average probability.

Then we repeat this calculation, but with three clusters, and using our proposed model with (8). Fig. 1 in Appendix A presents CB estimation results. CB_{TOT} by (9) is $36,949 with its actual CL of 90.97% (close to what we got with the Monte Carlo simulation). This contingency value is much closer to $35,374 (obtained by the Monte Carlo simulation) compared to $79,565 found by (7), with a percentage of variation of only 4% against a percentage of variation of 139%, as can be seen in Fig. 1.

4. FINDINGS AND DISCUSSIONS

We tested our methodology using 20 random risk registers, each containing 25 random risks (Table 1 in Appendix A), and we benchmarked our results with the results obtained by the Monte Carlo Simulation on the same random risk registers. We summarize the following findings.

The accuracy value of the CB estimation by the conventional model (5) (that does not consider clustering) differs significantly from the CB obtained by the Monte Carlo simulation. The average difference between these two models is 142%, with a pick of 207%. On the contrary, as per our proposed ML-based model (9) (which considers the clustering of risks by their probabilities), the CB estimation results are closer to those obtained by the Monte Carlo simulation. The average difference is 6%, with a pick of 12%.

The first contribution of our methodology is in its ability to provide more accurate CB estimates. The second contribution is in its simplicity: the approach involves only algebraic operations. In addition, if the risk register contains a limited number of risks, the clustering can be done manually, grouping the risks into three risk categories (low, medium, and high likelihood). The average probability for each of these three categories (alias clusters) can be easily calculated. Then the application of (8) on each cluster is straightforward. Finally, the CB_{TOT} is the sum of the cluster CB, calculated using (9).

Project management practitioners can use our methodology to have a practical, simple, and deterministic framework to calculate CB under uncertainty with higher accuracy. It can be represented in a spreadsheet to ease the CB quantification. The Monte Carlo simulation continues to represent one of the best approaches in this context, but its application is limited due to its intrinsic complexity for practitioners.

5. CONCLUSION

In this paper, we offered an ML-based approach to estimate CB. Our model integrates the concepts of the EMV technique, the binomial distribution, and k-means clustering. The model provided more accurate CB estimates compared to the traditional EMV approach and is simpler in use than the Monte Carlo simulation.

One limitation of our approach lies in how the risk impacts are found: they should be deterministic (i.e., not uncertain) and with a low dispersion around the averages of the impacts. A way to overcome this limitation could be using fuzzy logic to determine risk impacts since they are usually quantified using incomplete and imprecise information. However, this limitation does not affect the application of our methodology in practice since, typically, the cluster impacts range in the risk register is narrow.

Due to the size limit of this paper, we couldn't be able to demonstrate its extended analysis with more evaluation criteria for CB estimates and comparison with other ML approaches. Also, we couldn't test the model by tailoring it to the project type, e.g., construction, production, R&D, or information technology. We aim to address these limitations in the extended version of our paper in the future.

ACKNOWLEDGMENTS

This research was funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP09259049).

REFERENCES

Baccarini, D. (2004). Estimating project cost contingency – A model and exploration of research questions. *In: Khosrowshahi, F. (Ed.), 20th Annual ARCOM Conference.* Association of Researchers in Construction Management, Volume 1, 105–113.

Bakhshi, P. and Touran, A. (2014). Calculation of contingency in construction projects. *IEEE Transactions on Engineering Management.* Volume 50, Issue 2, May 2003, 135-140

Barraza, G. A., and Bueno, R. A. (2007). Cost Contingency Management. *Journal of Management in Engineering*, 23(3), 140–146.

Cagliano, A. C., Grimaldi S., Rafele, C. (2015) Choosing project risk management techniques. A theoretical framework, *Journal of Risk Research*, 18:2, 232-248, https://doi.org/10.1080/13669877.2014.896398

Cioffi, D. F., Khamooshi, H. (2007). A practical method of determining project risk contingency budgets. *Journal of the Operations Research Society*, 60(4), 565–571.

Dadsena, K. K., Sarmah, S. P., Naikan, V., N., A. (2019) Risk evaluation and mitigation of sustainable road freight transport operation: a case of trucking industry, *International Journal of Production Research*, 57:19, 6223-6245. https://doi.org/10.1080/00207543.2019.1578429

De Marco, A., Mangano, G., Narbaev, T. (2017). The influence of risk on the equity share of build-operate-transfer projects. *Built Environment Project and Asset Management*, 7(1), 45–58

De Marco, A., Narbaev, T., Rafele, C., Cagliano, A. C. (2017). Integrating Risk in Project Cost Forecasting. Conference: XXII Summer School "Francesco Turco" – Industrial Systems Engineering. Volume 22.

De Marco, A., Narbaev, T. (2021). Factors of Schedule and Cost Performance of Tunnel Construction Megaprojects. *The Open Civil Engineering Journal*. 15 (1). http://dx.doi.org/10.2174/1874149502115010038

Guan, X., Servranckx, T., Vanhoucke, M. (2021). An analytical model for budget allocation in risk prevention and risk protection, *Computers & Industrial Engineering*, Volume 161, 107657. https://doi.org/10.1016/j.cie.2021.107657.

Hamzeh, A. M., Mousavi, S. M., Gitinavard, H. (2020). Imprecise earned duration model for time evaluation of construction projects with risk considerations. *Automation in Construction*, 111, 102993.

Hillson, D., Grimaldi, S., Rafele, C. (2006). Managing Project Risks Using a Cross Risk Breakdown Matrix. *Risk Management*, 8, 61–76.

Hu X., Cui, N., Demeulemeester, E. (2015). Effective expediting to improve project due date and cost performance through buffer management. *International Journal of Production Research*, Volume 53, 2015 - Issue 5, 1460-1471

Kanakaris, N., Karacapilidis, N., Kournetas, G., Lazanas, A. (2019). Combining Machine Learning and Operations Research Methods to Advance the Project Management Practice. *In Proceedings of International Conference on Operations Research and Enterprise Systems*, 135-155.

Kim, B. C. (2015). Integrating risk assessment and actual performance for probabilistic project cost forecasting: A second moment Bayesian model. *IEEE Transactions on Engineering Management*, 62(2), 158–17

Kuo, R. J., Nugroho, Y., Zulvia, F. E. (2019). Application of particle swarm optimization algorithm for adjusting project contingencies and response strategies under budgetary constraints. *Computers & Industrial Engineering*, 135, 254-264.

Moreno-Cabezali, B. M., Fernandez-Crehuet, J. M. (2020). Application of a fuzzy-logic based model for risk assessment in additive manufacturing R&D projects, *Computers & Industrial Engineering*, Volume 145, 106529, https://doi.org/10.1016/j.cie.2020.106529.

Narbaev, T., De Marco, A. (2017). Earned value and cost contingency management: A framework model for risk adjusted cost forecasting. *Journal of Modern Project Management*, 4(3), 12–19

Narbaev, T., De Marco, A. (2014). An Earned Schedule-based regression model to improve cost estimate at completion. *International Journal of Project Management*, 32(6), 1007–1018.

Nunez, M.A., Kuo, L. & Chiang, I.R. (2016). Managing risk-adjusted resource allocation for project time-cost tradeoffs. *Ann Oper Res*. https://doi.org/10.1007/s10479-016-2122-7

Pospieszny, P., Czarnacka-Chrobot, B., Kobylinski, A. (2018). An effective approach for software project effort and duration estimation with machine learning algorithms. *Journal of Systems and Software*, 137, 184-196.

Project Management Institute. (2019). *The standard for risk management in portfolios, programs, and projects*. Project Management Institute, Newtown Square, PA.

Project Management Institute. (2021). *A Guide to the Project Management Body of Knowledge (PMBOK Guide)*. 7th edition. Project Management Institute, Newtown Square, PA.

Qazi, A., Simsekler, M.C.E. & Formaneck, S. (2022). Supply chain risk network value at risk assessment using Bayesian belief networks and Monte Carlo simulation. *Ann Oper Res*. https://doi.org/10.1007/s10479-022-04598-3

Rothwell, G. (2005). Cost Contingency as the Standard Deviation of the Cost Estimate. *Cost Engineering*. Volume 47, No. 7, July 2005.

Thamhain, H. (2013). Managing risks in complex projects. *Project Management Journal*, 44(2), 20–35.

Touran, A. (1993). Probabilistic cost estimating with subjective correlations. *Journal of Construction Engineering and Management*, Volume 119, Issue 1, 587

Touran, A. (2003). Probabilistic model for cost contingency. *Journal of Construction Engineering and Management*, 129(3), 280-284.

Touran, A. and Lopez, R. (2006). Modeling Cost Escalation in Large Infrastructure Projects. *Journal of Construction Engineering and Management,* Volume 132, No. 8, 853-860

Xie, H., AbouRizk, S., and Zou, J. (2012). Quantitative Method for Updating Cost Contingency throughout Project Execution. *Journal of Construction Engineering and Management*, 138(6), 759–766.

Zhang, Y., Zhao, C., Pang, B. (2018). Budget allocation in coping with supply chain disruption risks, *International Journal of Production Research*, 56:12, 4152-4167, https://doi.org/10.1080/00207543.2018.1430905

APPENDIX A

Table 2. CB estimation accuracy results from the traditional EMV approach versus the proposed ML model.

RR#	EMV, %	ML, %		RR# (cont.)	EMV, % (cont.)	ML, % (cont.)
1	139	4		11	22	11
2	100	12		12	169	8
3	29	12		13	194	6
4	130	9		14	100	7
5	201	6		15	105	4
6	167	5		16	142	3
7	156	6		17	143	4
8	207	9		18	181	3
9	151	3		19	157	6
10	183	4		20	155	7

Note: RR# – The number of risk registers simulated; EMV, % - The results by the traditional EMV approach with one cluster; ML, % - The results by the proposed ML approach with three k-means clusters, %. The percentage values represent the variation with regard to the Monte Carlo simulation used as a benchmark.

Figure 1. Comparison between the application of a binomial distribution (with Bernoulli trials) to the whole risk register ("Bernoulli 1 Cluster") and to the three clusters separately ("Bernoulli 3 Cluster")

```
Confidence 0.9000 (z-score: 1.2816)

25 risks, EV=18840 (5507.24,0.45)
# cluster 1: 10 (40%) (13017.40,0.11)
# cluster 2: 9 (36%) (696.89,0.51)
# cluster 3: 6 (24%) (205.83,0.93)
Max Impact: 137681

MONTE CARLO SIMULATION (100000 trials)
Contingency mean: 18844
Contingency mean confidence: 0.6591
Contingency standard deviation: 12897
Contingency at 0.90 (μ+σ): 35374
Contingency confidence: 0.9097

BERNOULLI 1 CLUSTER
Contingency expected value: 62008
Contingency standard deviation: 13700
Contingency at 0.90: 79565
Contingency confidence: 0.9999
Bernoulli vs Monte Carlo: 1.39

BERNOULLI 3 CLUSTERS
Contingency expected value: 18841
Contingency standard deviation: 14129
Contingency at 0.90: 36949
Contingency confidence: 0.9097
Bernoulli vs Monte Carlo: 0.04
>>>
```