Reference Class Forecasting and Machine Learning for Offshore Oil & Gas megaproject planning: Methods, Application, and Interpretation

This dissertation will study methods to improve the accuracy of within-project forecasts of cost and schedule using empirical benchmarking data from similar projects. Reference Class Forecasting (RCF) will be used for this purpose and the potential for adapting Machine Learning (ML) to RCF will be demonstrated by application. Furthermore, the results from the research will be interpreted both for relative performance and for deriving hypotheses regarding the effect of broader underlying behavioral, cognitive and sensemaking issues. The results will also be used to discuss the potential for ML to bridge the gap between the two distinct approaches to characterize heuristics for decision-making in complex projects, one of which emphasizes cognitive limitations from heuristics and biases, and the other which importance of these heuristics. The focus will be on the prediction of Oil & Gas (O&G) Offshore megaproject performance. These projects have a track record of underperformance on measures of schedule, cost and benefits (Merrow, 2012). While they have accounted for investments amounting to several hundred billion US$ per year and underpin critical global energy supply, studies of the causes for their underperformance are relatively scant (Rui, et al., 2017).

This research will build on the work of Prof. Flyvbjerg (Flyvbjerg, 2006), who adapted Kahneman & Tversky’s (1979) “outside view” approach to project planning as Reference Class Forecasting. The outside view approach recommends the use of the “distribution of outcomes under similar situations” for forecasting. This will complement “inside view” forecasting which typically uses DCF, NPV, Mont Carlo simulation etc. Megaprojects are often bespoke endeavors by meta-organizations or independent firms connected through contractual relationships (Lundrigan, et al., 2015). Usually a JV comprised of O&G Companies uses EPC contractors with specialist subcontractors, vendors etc. to deliver an offshore project (Lee, 2019). These challenging projects combine high upfront capital requirements with anticipation of long-term returns in an uncertain complex socio-economic-technical environment. Therefore, accurate cost and schedule estimation are challenging, especially across contractual networks with principal-agent conflict being a key governance problem (Müller, 2009). Principal-agent issues, optimism and behavioral biases often result in significant inside view underestimations of project costs and schedule (Flyvbjerg, 2014). RCF uplifts in-project estimates using the actual track record of similar projects, these comparable projects constituting the reference class. The track record essentially defines a probability distribution. RCF assumes that this distribution is relevant to any project in that class and computes empirically uplifted estimates to fall within a desired certainty percentile, as opposed to predicting specific estimates.

A survey of offshore project track record from existing data and literature will first quantify and establish the extent of the cost and schedule underestimation problem in Oil and Gas projects. This work proposes to apply RCF to estimate project cost and schedule uplifts using projects in their class drawing on existing literature on RCF application to megaprojects. Oil and Gas projects will be subdivided into classes using factors identified from studies into their performance, such as location, capital budget, planned duration etc. Additionally, questionnaires may be used to elicit answers regarding factors that influence project outcomes from experienced project planners or consultants to interpret and classify the collected project data. However, these questionnaires may not be used for the research. The cost and schedule records of projects in each class will be randomly subdivided into a training and testing dataset. Following this, distributions will be constructed for each class of projects from the training data. Several cost and schedule uplifts for each project with different certainty percentiles will be estimated. The forecasted percentile uplifts will be compared with the track record in the test data.

We will also take a predictive approach to forecast project specific uplifts. A specific cost and schedule uplift will be forecast for each project. The predictive approach will use data analytics and Machine Learning (ML) algorithms. This approach can utilize “inside view” project information such as the effect of the pre-EPC phase shaping process, while keeping the “outside view” approach of using a reference class. A prediction model will be used downstream of a classification model. The goal is to apply a rigorous approach using predictive models developed from Machine Learning to learn from the distributional data. Outliers in a reference class may be extreme but are significant to statistical analysis of cost overrun (Flyvbjerg, et al., 2018). Outliers and “fat tails” in project distributional data can be corelated with factors that affect their occurrence by a trained model, rather than just being part of the reference class probability distribution where they have equal probability of occurrence for every project in the class. The mean of a reference class distribution can change from the effect of factors that could be discovered by the trained model. Factors affecting performance and forecasting accuracy will be identified using a combination of clustering models and statistical assessment of the data informed by pertinent extant literature. “Outside” factors (roughly factors predominantly outside the project organization’s control after the appraisal phase) may include location, political, socio-economic challenges etc. The first step, using models from Data Analytics such as Classification or Clustering, will use these outside factors to “fingerprint” projects and group them into reference classes. Project specific “inside” factors identified in previous literature include size, Front End Loading, technical novelty, continuity etc. Behavioral factors identified for megaprojects such as optimism bias and strategic misrepresentation are also relevant and can be related to relationships and cross-national or cross-cultural factors across the project contractual & stakeholder network. The second step will use ensemble models to aggregate models that may use Neural Networks, Regression Trees, or Bayesian Networks. These models will use the inside factors as inputs and the distributional data in the project’s reference class as training data to predict specific cost and schedule uplifts. Projects in a reference class will be subdivided into training and test data as before.

The 2 approaches of using conventional RCF to forecast uplifts within desired certainty percentiles, and Machine Learning using a reference class to make a specific prediction, will be compared to conventional estimation processes in the Oil and Gas Industry. They will also be compared with each other, and their use in conjunction will be explored. For instance, fat tails in the distributional data could be weighted using outputs from the ML models. It is observed that the availability of larger data sets can make the use of ML approaches more accurate in the future. But this work will provide a model for adapting proven RCF methods with ML methods, and lead to further research on methods and application. This work will result in models and methods to apply RCF using Machine Learning algorithms as well as hypothesis to be tested, validated, and extended in further work.

A screenshot of text

Description automatically generated

Fig-1: Dissertation Plan

# References

Flyvbjerg, B., 2006. From Nobel Prize to Project Management: Getting Risks Right. *Project Management Journal,* 37(3), pp. 5-15.

Flyvbjerg, B., 2014. What you should know about megaprojects and why: An overview. *Project Management Journal,* 45(2), p. 6–19.

Flyvbjerg, B. et al., 2018. Five things you should know about cost overrun. *Transportation Research Part A: Policy and Practice,* Volume 118, pp. 174-190.

Kahneman, D., 2012. *Thinking, Fast and Slow.* London: Penguin.

Kahneman, D. & Tversky, A., 1979. Intuitive Prediction: Biases and Corrective Procedures. *TIMS Studies in the Management Sciences,* Issue 12, pp. 313-327.

Lee, C.-Y., 2019. *Integrative Trust-Based Functional Contracting: A Complementary Contractual Approach to BIM-Enabled Oil And Gas EPC Project Delivery.* [Online]   
Available at: https://espace.curtin.edu.au/bitstream/handle/20.500.11937/75059/Lee%20CY%202019.pdf?sequence=1&isAllowed=y  
[Accessed 4 October 2019].

Müller, R., 2009. *Project Governance.* 1st ed. London: Routledge.

Merrow, E. W., 2011. *Industrial megaprojects: Concepts, strategies, and practices for success.* 1st ed. Hoboken, NJ: John Wiley & Sons.

Rui, Z. et al., 2017. Investigation into the performance of oil and gas projects. *Journal of Natural Gas Science and Engineering,* 1(38), pp. 12-20.