FINAL REPORT

O18: Exploring the use of artificial intelligence (AI) solutions to improve the accuracy of project delivery forecasts

ARRB Project No.: 015402

Author/s: Alex Aldana, Kieran Hay, Catherine LeGrand, Paul A Schmidt and Mathew Bereni.

Prepared for: Queensland Department of Transport and Main Roads

April 2021

AN INITIATIVE BY:

SUMMARY

The Queensland Department of Transport and Main Roads (TMR), in collaboration with the Australian Road Research Board (ARRB), has investigated the suitability and performance of artificial intelligence (AI) solutions to enhance decision making through improved accuracy and precision of project cost and duration forecasting. Machine learning (ML) models use reference class forecasting to identify predictive patterns in large datasets.

With appropriate instructions, suitable data and calibration, ML models provide improved final cost forecasting accuracy that may assist TMR's budgeting capabilities for individual projects and their work program. The project scope entailed an evaluation of the suitability of TMR project data for use in ML models as well as the effectiveness and accuracy of ML models in predicting the forecast final costs and forecast completion dates of projects using provided datasets.

Machine learning models use ample amounts of high-quality project data that have minimal missing values. A collection of 116 TMR projects that demonstrated high levels of adherence to machine learning suitability factors formed the basis of the model's dataset. This data comprised 89 small (< $10m) and 26 large ($10m – < $100m) budget projects that encompassed 16 different work types and 15 delivery programs.

Predictive analytics were carried out on TMR's dataset, which presented evidence-based predictive and descriptive observations. Predictive observations were thereafter identified in the parameters of cost/duration forecasting accuracy and budget adjustment. The parameters of project capital expenditure, construction year, wet season start month, work type and delivery program were identified as descriptive observations and therefore cannot be applied to the overall TMR portfolio. Additionally, these analytics found that projects typically return capital budget and/or contingency late in the project life cycle.

Machine learning modelling for project cost and duration forecasting was developed and tested, resulting in the following observations:

For individual projects:

- The model improved the final construction costs by reducing the prediction percentage error by 21% (15.9% error for the business versus 12.6% for the model), forecasting an average savings of $148,000.
- The model improved forecasting final construction duration by a reduction in the prediction error by 11% (8.5 days) (26% error for the business versus 23% for the model).
- The model improved the early warning of duration underrun and overrun by an average of 4 and 2 months, respectively.
- The model improved the early warning of cost underruns by an average of 4 months.

For the aggregate portfolio of all projects:

- The model improved the cost forecast predictions by reducing prediction error by 87% (from 10.7% error for the business versus 1.4% for the model).
- The model measured the aggregate portfolio value and tracked the actual value more accurately than the current TMR approach.

Improving the predictability of project cost and duration under/overruns is a challenging problem affecting major projects. Artificial intelligence technologies such as machine learning models offer a faster, cheaper, and more powerful way to conduct many thousands of experiments to build evidence-based predictive models for tackling these challenges. TMR's substantial project database has demonstrated applicability for ML modelling to improve the accuracy and precision of project cost and duration forecasting.

The ML model identified significant financial and delivery opportunities for the re-distribution of capital budget and/or contingency back into their portfolio, given that forecast project costs and contingency may be predicted earlier and more accurately. Overall, AI technologies have demonstrated capabilities that enhance capital productivity, portfolio performance, early warning capability and a decrease in monitoring costs.

ML predictions may be a valuable and effective tool for Project and Program Managers to validate traditional cost and duration forecasts.

ACKNOWLEDGEMENTS

ARRB acknowledges the effort and participation of the Paul Schmidt (TMR PM), TMR team and Endeavour Programme to collect the projects data and evaluate the data according to project scope.

CONTENTS

TABLES

FIGURES

1 INTRODUCTION

1.1 BACKGROUND

The Queensland Department of Transport and Main Roads (TMR) is one of the major recipients of government funding for the planning, development, implementation, operations, and maintenance of transport infrastructure in Queensland. TMR has an extensive amount of data on the delivery of infrastructure projects. Carefully calibrated analysis and interpretation of such data by incorporating artificial intelligence (AI) solutions have been tested and considered to be an effective tool that, when implemented correctly, has the capacity to improve project delivery forecasts and affords the user additional time for proactive decision making.

TMR, in conjunction with the Australian Road Research Board (ARRB) and under the National Asset Centre of Excellence program (NACOE), explored the suitability and performance of AI's Machine Learning (ML) solutions to facilitate 'data driven' decision making through improved accuracy and precision of project cost and duration forecasting. The ML model assessed in this project uses dynamic reference class forecasting by finding predictive patterns in large datasets and Weak Signals dynamic theory to identify discontinuities. Reference class forecasting is a method of predicting the future by using the outcomes of similar historical situations to make more accurate predictions. Based on ML from historical project management data, a reference class forecasting model has been shown to reduce cognitive biases, such as overconfidence, to deliver more accurate and earlier forecast predictions than conventional/traditional applications.

NACOE is exploring the potential benefits of the application of this specific type of AI technology to enhance the delivery management of infrastructure projects and portfolios.

This report provides a summary of the project findings, including the identification of value adding to TMR project management system.

1.2 ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

AI combines ideas from many disciplines, including philosophy, computer science, mathematics, psychology, neuroscience, and the social sciences. There are few modern activities that are untouched by AI today, yet its potential is only just being realised.

ML is one prominent branch of AI that is used in a plethora of varying complex applications; it is a way of programming computers by 'training' them to learn by example, rather than by 'knowledge engineering', which involves programming hand-crafted rules. Therefore, ML is an application of AI that provides systems with the ability to learn and improve from experience without being explicitly programmed automatically.

AI learning starts with a set of observations of data undertaken by computer analysis that searches for specific patterns. Observable patterns are then implemented to 'train' the machine to make self improved decisions that would assist future analyses. Machine learning focuses on the development of computer programs and algorithms that enable accessibility and rapid interpretation of data to make predictions from historical data, at a scale and speed that are not possible with manual methods (Expert System Enterprise 2020).

This report provides high-level objective research into the conceptual application of ML's ability to make economical predictions in regards to project timeframe and final cost on a large portfolio of transport projects. Improving the cost and time predictability of these portfolios has the potential to stretch state budgets to be used more effectively.

1.3 PROJECT AIM AND OBJECTIVES

The purpose of this project was to explore the suitability and performance of ML solutions to improve the accuracy of project forecast final costs and completion date, compared with the Project Manager's estimates.

The aim of this project was to assess the implementation of AI technology, data analytics and reference class forecasting, using a ML model to:

- provide early warning of changes (underrun and overrun) to out turn cost forecasts
- identify likely factors affecting cost forecast change
- provide a safe and transparent environment for continuous improvement and sophistication of new AI technology adoption
- provide a platform for future AI tools (such as scenario modelling)
- identify opportunities for ML model application within TMR's portfolio.

This project's objective was to provide a report summarising the application of the ML model to TMR projects that evaluated the advantages and drawbacks for project forecasting on cost and duration purposes. Additionally, this report explores opportunities and recommendations for the application of ML within TMR's business processes.

1.4 PROJECT SCOPE

The scope of the project includes:

- evaluation of the suitability of TMR project data for use in ML models, with a focus on data from the construction phase (TMR milestones KM11-Construction Start and KM12-Practical Completion)
- evaluation of the effectiveness and accuracy of ML models in predicting the forecast final costs and forecast completion dates of projects using provided datasets
- identification of both limitations and opportunities of the ML models and recommendations of whether the models could be implemented in the future.

1.5 PROJECT METHODOLOGY

The project was undertaken as per the methodology summarised in Table 1.1.

Table 1.1: Project Methodology

Stage no. – Title	Task description
Stage 0 – Preliminary Study	A preliminary study was undertaken to determine the requirements of an ML model for forecasting through the following: • Gathering an initial set of data for one of TMR's projects based on minimum data requirements. This project data was then used in a live modelling demonstration. • Undertaking a live ML modelling demonstration which was based on the supplied TMR project data. Project data was provided to an existing industry trained civil development ML model. • Comparing resulting ML forecasts against TMR's existing cost projection baseline, actuals and forecast metrics for the selected project. • Providing a formal report for review, which outlined the results of the preliminary study and recommendation on the continuation of the project.
Stage 1 – Data Suitability	A data suitability study and review of TMR's data structure was undertaken through the following: • Undertaking a workshop to discuss and agree on business goals, project organisation/structure and to identify suitable project datasets and information. • TMR provided project financial and schedule data in order to identify suitability and any gaps for the training of the ML demo model. • Reviewing and analysing TMR's project data for reference class forecasting analytics and ML suitability.

Stage no. – Title	Task description
	• Pre-processing the supplied TMR project data in preparation for ML (data validation, cleaning and transformations). This was to ensure there were no issues regarding quality, structure, and quantity of data prior to the project's next stages. • Providing a formal report for review which outlined the results of the data suitability and review of data structure.
Stage 2 – Data Analytics	Advanced data analysis of TMR data was undertaken through the following: • Conducting statistical analysis on pre-processed project data to understand, and make explicit, underlying patterns, trends and relationships in the data. This was focused on identifying drivers of project success and failure that could be discovered in the project data. • Reviewing and exploring the use of qualitative information made available by TMR (e.g. risk register data) to advise NACOE on its usefulness to improve scheduling forecasts using a ML model. • Providing a formal report for review, which outlined the results of the data analytics and identification of trends and drivers.
Stage 3 – ML Model Development and Application	The training and testing of predictive ML models based on the findings of previous stages were undertaken. The performance of the model based on the accuracy of forecasted outcomes was determined through the following: • Undertaking the ML model training using the 'leave one out cross validation' (LOOCV) method. This method is suitable for when an accurate estimate of model performance is critical (Brownlee 2020). • Undertaking the prediction of forecast final costs and forecast final completion dates using the newly trained ML model. The results were assessed for accuracy against the existing TMR data. • Providing a formal report for review which outlined the results of the ML model training and the accuracy of the model.
Stage 4 – Final report	• Delivery of a summary report (this report) which outlines the results and findings from Stages 0 to 3, provides recommendations for the application of ML, and provides lessons learned regarding the use of AI for project costing purposes.

1.6 STRUCTURE OF REPORT

This report is structured as follow:

- Section 2: Description of a pilot test evaluating the functionality of the ML.
- Section 3: Discussion of Stage 1 – Data Suitability findings with descriptive results. The section also includes challenges with the data collection and, finally, a description of each data set results.
- Section 4: Description of results from Stage 2 – Data Analytics highlighting the predictive and descriptive observations from the selected data set.
- Section 5: Review of results from Stage 3 – Provides review on the ML model development, including assessment, performance and limitations of ML.
- Section 6: Provides a high level identification of the TMR project management process, including system, controls, reports and forecasting. Additionally, this section identifies possible applications inside the TMR project management process to apply ML.
- Section 7: Summarise the project results and provides recommendations with further possible opportunities for ML improvement.

2 STAGE 0 – PRELIMINARY STUDY

Stage 0 was undertaken as an initial scoping exercise to determine whether an established pre-trained ML model would be able to predict final forecast costs of a project on a monthly basis and to what level of accuracy. The pre-trained ML model's reference class was developed from a civil industry project database (not TMR's) that comprised 500 projects across Australia, spanning up to 7.5 years and budgets primarily less than $30 million. The scoping exercise's purpose was to determine if ML models could make the predictions with accuracy similar to humans and determine whether this project would proceed to the next stages.

A demonstration of a pre-trained ML model was undertaken using the monthly financial data from the construction phase of a TMR project. The demonstration used the TMR milestones KM-11 to KM-12, which correspond to Construction Start and Practical Completion, respectively.

The model was used to predict the final forecast cost for each month of the project. This was done by providing 25 data points of financial data for the project. The first five data points were individually supplied to the model to test its ability to produce the forecast final cost for each data point and ensure the model used single data points to generate forecasts. Once this was confirmed, the remaining data points were supplied in groups of five as a batch. After each batch, the forecast final cost from the model was reviewed for changes. The generated forecast was then compared against TMR's current cost projection tools for baseline, actuals and forecast for the chosen project.

The model predicted the forecast final cost (FFC) using P40 and P60 models, where a P40 and P60 value means there is a 40% and 60% chance (respectively) that the actual final cost is no higher than the forecasted final cost. The results of the pre-trained ML model's forecasts can be seen in Table 2.1.

Table 2.1: Stage 0 pre-trained ML model predictions

No	Original budget	Current budget	Actual cost	TMR's FFC – ML FFC at P60	TMR's FFC – ML FFC at P40
1	$74,700,000.00	$74,700,000.00	$30,000.00	−$6,865,323.00	−$4,136,413.00
2	$74,700,000.00	$74,700,000.00	$30,000.00	−$6,865,323.00	−$4,136,413.00
3	$74,700,000.00	$74,700,000.00	$1,440,588.00	−$6,865,323.00	−$4,136,413.00
4	$74,700,000.00	$74,700,000.00	$1,988,521.00	−$6,865,323.00	−$4,136,413.00
5	$74,700,000.00	$74,700,000.00	$2,730,178.00	−$6,865,323.00	−$4,136,413.00
6	$74,700,000.00	$74,700,000.00	$4,104,476.00	−$6,865,323.00	−$4,136,413.00
7	$74,700,000.00	$74,700,000.00	$6,249,619.00	−$6,865,323.00	−$4,136,413.00
8	$74,700,000.00	$74,700,000.00	$9,102,718.00	−$20,408,477.00	−$18,473,973.00
9	$74,700,000.00	$74,700,000.00	$12,786,447.00	−$20,408,477.00	−$18,473,973.00
10	$74,700,000.00	$74,700,000.00	$14,755,403.00	−$20,408,477.00	−$18,473,973.00
11	$74,700,000.00	$74,700,000.00	$18,349,483.00	−$20,408,477.00	−$18,473,973.00
12	$74,700,000.00	$74,700,000.00	$21,681,481.00	−$20,408,477.00	−$18,473,973.00
13	$74,700,000.00	$74,700,000.00	$26,453,929.00	−$20,408,477.00	−$18,473,973.00
14	$74,700,000.00	$74,700,000.00	$30,091,126.00	−$20,408,477.00	−$18,473,973.00
15	$74,700,000.00	$74,700,000.00	$32,307,812.00	−$20,408,477.00	−$4,136,413.00
16	$74,700,000.00	$74,700,000.00	$36,360,566.00	−$20,408,477.00	−$18,473,973.00
17	$74,700,000.00	$74,700,000.00	$40,711,564.00	−$6,865,323.00	−$18,473,973.00
18	$74,700,000.00	$74,700,000.00	$43,736,316.00	−$20,408,477.00	−$18,473,973.00
19	$74,700,000.00	$74,700,000.00	$46,711,557.00	−$20,408,477.00	−$18,473,973.00
20	$74,700,000.00	$64,700,000.00	$51,287,584.00	−$30,408,477.00	−$28,473,973.00
21	$74,700,000.00	$64,700,000.00	$54,791,198.00	−$5,864,443.00	−$1,310,384.00
22	$74,700,000.00	$64,700,000.00	$61,957,766.00	−$5,864,443.00	−$1,310,384.00
23	$74,700,000.00	$64,700,000.00	$61,304,593.00	−$5,864,443.00	−$1,310,384.00

No	Original budget	Current budget	Actual cost	TMR’s FFC – ML FFC at P60	TMR’s FFC – ML FFC at P40
24	$74,700,000.00	$64,700,000.00	$61,389,975.00	−$5,864,443.00	−$1,310,384.00
25	$74,700,000.00	$64,700,000.00	$61,872,672.00	−$5,864,443.00	−$1,310,384.00

The resulting P40 and P60 forecasts demonstrated that the pre-trained model predicted overruns at approximately $1.3 million and $6 million, respectively, compared with the actual TMR project that finished with an approximate $3 million underrun.

Cases for why the results were significantly different from the actuals included:

- difference between typical project behaviour to the current project
- data used to train the pre-trained model was likely not an exact match for this project
- supplied data only contained a small feature set
- the pre-trained model had not been tuned or calibrated to TMR data
- model calibration fitting the model to production.

It was identified that using TMR specific data for the ML model training would likely improve the forecasting accuracy. Additionally, the reference data used to train the model consisted of projects that primarily overran the budget in the civil industry database, leading to a bias in that direction.

These results suggest that with appropriate instructions, data, and calibration, the ML model could assist TMR in enhancing their budgeting capabilities by providing more accurate forecasted final cost predictions for individual projects and the whole work program.

ARRB’s recommendation after the demonstration was to continue with the next stages of the project to investigate whether a newly trained ML model could more accurately predict forecasted costs.

3 STAGE 1 – DATA SUITABILITY

Stage 1 was aimed at understanding TMR's data and evaluating this data for its suitability to be used in ML model training. In this stage of the project, an analysis of a subset of TMR's project data was undertaken.

3.1 DATA COLLECTION

The initial database comprised a total of 48 TMR projects, which included financial and milestone data attributes. TMR's database contained approximately 600 data columns for each project, including information that was not relevant for this study. For each project, the data was screened to capture only the required and relevant data attributes. It was noted that including all available data attributes might cause an overfitting modelling error where the features exceed the sample size.

The data screening provided the ML model with information including:

- Project ID
- Project Type (i.e. road, infrastructure, civil, others)
- Project Status (Active or Closed)
- Project Starting Date
- Reporting Month
- Forecast Project End Date (updated according to monthly progress)
- Original Project Budget
- Current Budget (Variation of original budget)
- To Date Actual Cost
- Forecast Cost To Complete.

The data suitability assessment of the initial database in regard to the 48 TMR projects proved to be insufficient. The training of the ML model required larger sample size (refer to Section 3.2.1).

Therefore, a secondary data collection was initiated to provide a more robust database with 116 historical TMR projects. This database consists of 89 small budget (< $10m) and 26 large budget ($10m – < $100m) projects that encompassed 16 different work types and 15 delivery programs as outlined in Appendix B - Table B.1.

The secondary data assessment demonstrated high suitability for ML applications, including high compliance to the data dictionary, and contained minimal missing data (refer to Section 3.2.2).

3.2 DATA SUITABILITY ANALYSIS

To achieve an appropriate level of performance for machine learning, data used for the model construction needed to be of a high-quality, with sufficient quantity, and contain minimal missing data. Additionally, the data needed to comply with the data dictionary where the information provided was in a specific format that meaningfully represented the 'features' that the ML directory called for.

The methodology of the data suitability analysis implemented was as follows:

1. perform high-level exploratory data analytics on the data set
2. process and quality assurance including statistical analysis,
3. validate data through compliance with the data dictionary and S-curve analyses
4. determine the implication for ML model suitability.

Additionally, throughout the analysis, the data needed to demonstrate high levels of adherence to machine learning suitability factors included:

- completeness
- interpretability
- realistic project behaviour
- high compliance to a data dictionary.

3.2.1 INITIAL RESULTS – INITIAL DATA COLLECTION

The initial 48 TMR projects provided had complete cost and time data that spanned the entire project life cycle and demonstrated realistic project behaviour. Initial data suitability analysis of these 48 projects showed:

- Five projects were identified to show behaviour trends as these projects were representative of 'typical' historical project data (refer to Appendix A). These trends were assessed for their suitability in reference class forecasting, analytics and ML, and their high levels of adherence to ML suitability factors.
- Only 18 projects were suitable for the ML cost prediction model:
  - 14 projects were active, which did not allow for a full life cycle assessment of the project's cost.
  - 16 projects showed atypical project cost behaviours on S-curves compared with the typical five project behaviour trends.
- Only 34 projects were suitable for the ML time prediction model:
  - 14 projects were unusable due to the forecasted end dates that remained static throughout the project's life cycle.

These results indicated that the majority of data provided was not suitable for ML modelling as TMR captured different information at different times. Furthermore, pre-2009 data contained less detail in comparison to data from recent years.

As TMR's organisation evolves, so does their data systems, allowing for a more robust and better system that provides stable data to support decision making with clear project management. Robust and stable data are preferable for a ML model's training, which led to undertaking a second data collection from more recent projects.

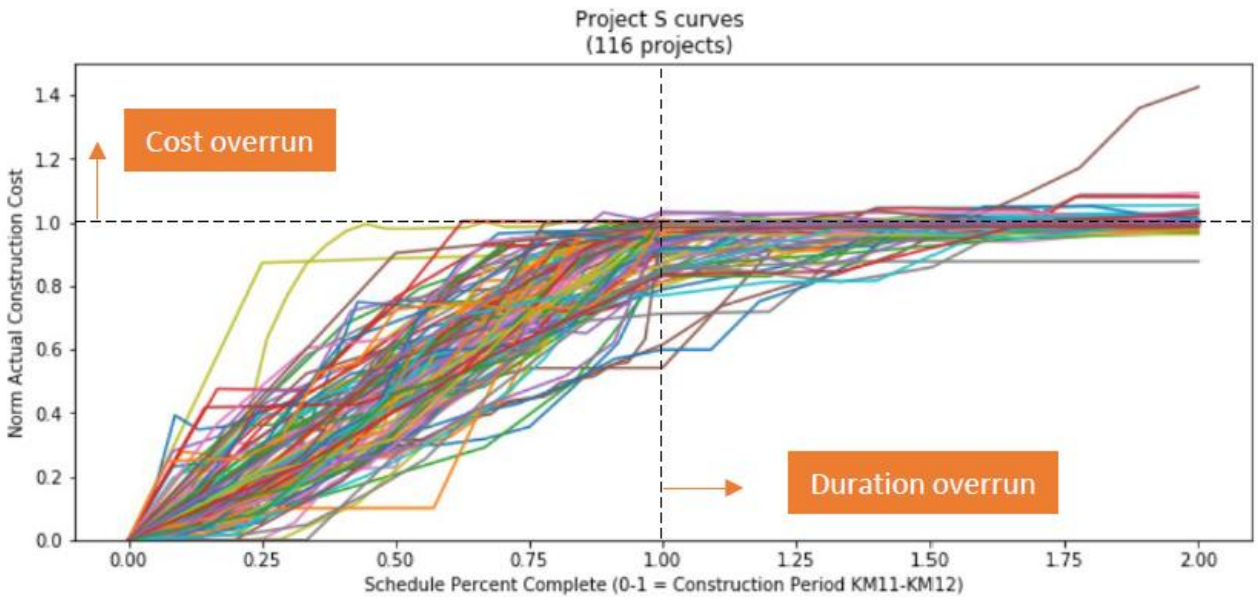
3.2.2 SECONDARY RESULTS – SECONDARY DATA COLLECTION

The second TMR database comprised of 116 projects. Cost and time performance analyses both revealed that these projects contained a combination of underruns and overruns for both cost and duration. Additionally, more project underruns compared to overruns were identified in this database. These observations were a result of the following evaluations:

- Cost: final cost to the original budget ratio for construction stages between the start and practical completion
- Duration: final construction to the original construction duration ratio for time performance.

Figure 3.1 compares the project progress over time using a normalised actual construction cost on the Y-axis and a normalised elapsed schedule (time) on the X-axis. Cost and time overruns occurred when the normalised actual construction cost and normalised elapsed schedule exceeded 1.0. Therefore, the project data could be assessed for realistic project behaviour based on their project progress S-curves as shown in Figure 3.1.

Figure 3.1 Project S-curves for completed projects (n=116)



Source: Stage 1: Data Suitability Report (Endeavour Programme, 2020b).

In conclusion, the data suitability analyses of TMR's 116 projects indicated that the dataset was highly suitable for usage in ML to dynamically predict cost and time outcome performance. The analyses also highlighted the following:

- An established corporate data system that provides large quantities of project data that are easily extracted and analysed is very important.
- The project data was high quality as it demonstrated realistic project behaviours. This ensured that the machine learning outcomes would meaningfully depict real project outcomes.
- Projects were identified at portfolio-level, and the outcomes were highly suitable for use in ML modelling.

4 STAGE 2 – DATA ANALYTICS

Stage 2 involved a statistical analysis of the TMR's 116 project dataset to identify and understand any trends detected within the data. This section outlines the analyses of data and the identification of predictive and descriptive observations of the ML model.

4.1 DATA STRUCTURE AND ANALYSES

Predictive analytics was undertaken on TMR's 116 project dataset to provide evidence-based predictive and descriptive observations. These observations were intended to improve the predictability of project performance parameters within TMR's entire project portfolio.

The selected performance parameters of interest to TMR included:

- cost forecasting
- duration forecasting
- key influences
- budget adjustments.

The predictive analysis classified performance parameters as being either predictive or descriptive observations.

The resulting predictive observations included the following parameters:

- forecasting accuracy
  - final cost
  - final duration
  - current budget during and near project completion
- budget adjustment.

The resulting descriptive observations included the following parameters:

- wet season start month
- work type and delivery program (Table B.1) S-curves
- project performance by start year
- project capital value.

4.1.1 PREDICTIVE OBSERVATIONS

Regardless of a project's capital value, predictive observations were identified in the following performance parameters:

- Forecast accuracy – conservative trend identified.
- Budget adjustment behaviours – no trend identified.

An increase in forecasting errors for both cost and duration parameters was identified. This was based on the comparison between small (< $10m, n = 89) and large ($10m< – < $100m, n = 26) budget projects.

Furthermore, these observations inferred that the TMR project managers are hesitant to release the contingency cost reserves and are less reluctant to report time underruns even when project conditions are well known both during and near project completion.

Statistically significant testing of this data further predicted that these trends are likely to be found throughout the broader TMR project portfolio.

4.1.2 DESCRIPTIVE OBSERVATIONS

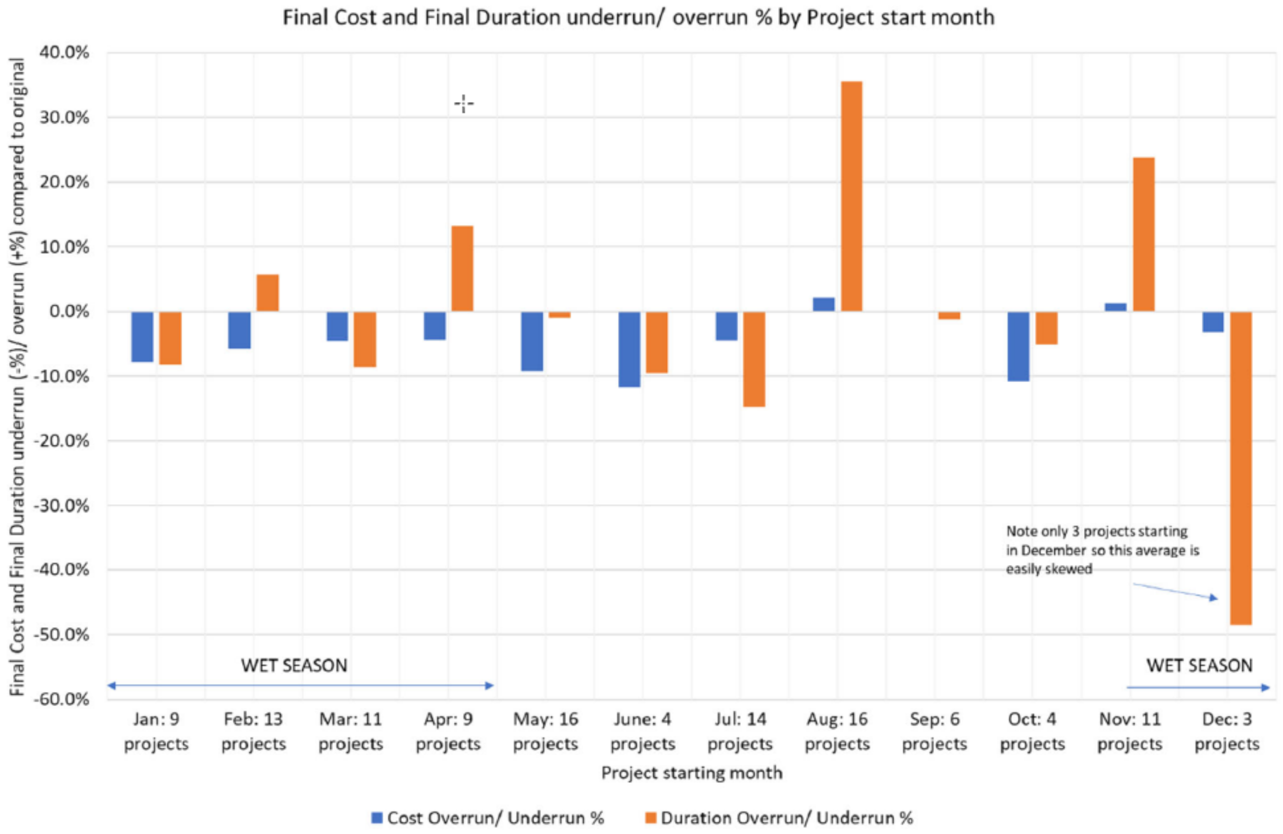
Descriptive observations were identified and modelled in Figure 4.1:

- **Starting month** – impacts both cost and duration forecasting, particularly during Queensland's wet season (November to April). However, this observation failed to satisfy the statistically significant test criteria.
  - **Cost data** – overall data skewed towards producing underruns.
    - Wet season: data primarily produced underruns except for November, which produced a marginal cost overrun.
    - Outside the wet season: only the month of August produced a cost overrun while the remaining months produced cost underruns.
  - **Duration data** – outturn performance fluctuated from month to month.
    - Wet season: duration overruns were present in the start months of November, February and April. The November overrun was followed by a substantial underrun in December.
    - Outside the wet season: duration overruns were present in the start month of August whilst the remaining months produced underruns.

Project performance by start year – a comparison of projects that started in either 2017–18 or 2019–20 produced non-statistically significant differences in both cost and duration overruns (Figure 4.2).

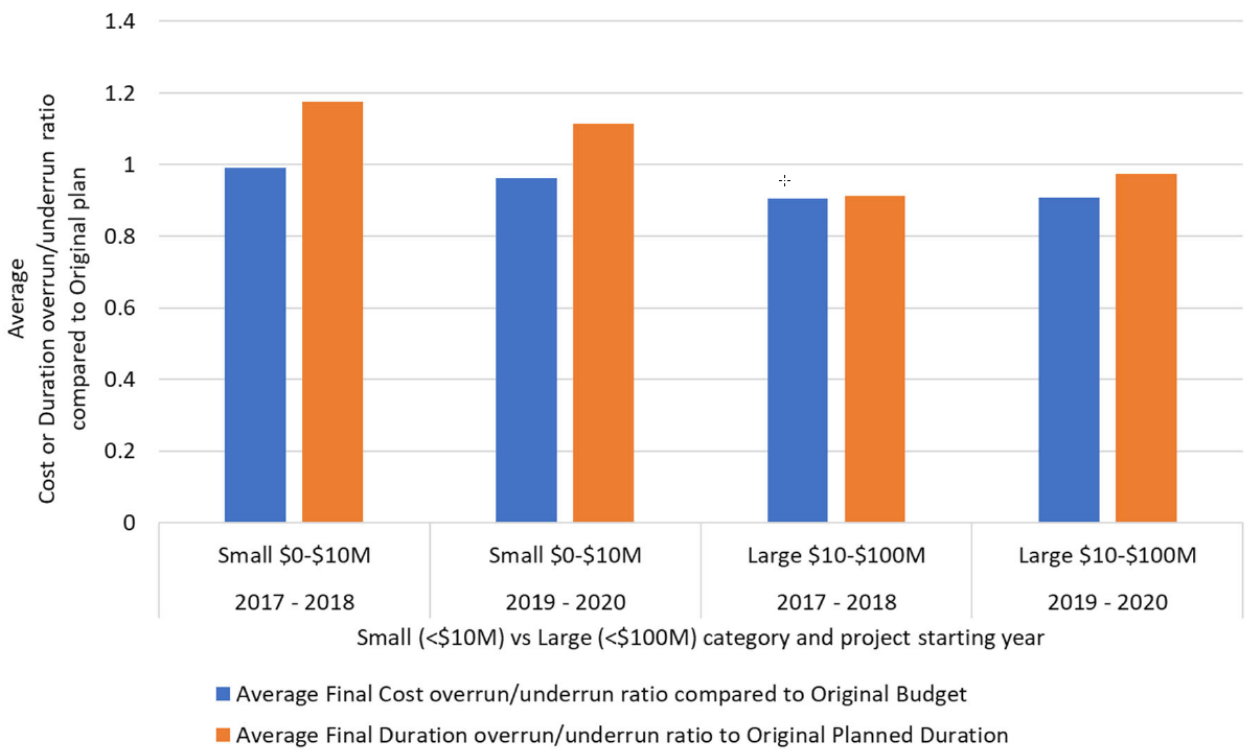
Project capital value – irrespective of a project's capital expenditure (i.e. the funds used to undertake new projects), non-statistically significant differences in both cost and duration overruns were produced (Figure 4.2).

Figure 4.1: Final cost and duration underrun/overrun % by project start month



Source: Stage 1: Data Suitability Report (Endeavour Programme, 2020b).

Figure 4.2: Cost and duration overrun/underrun by capital expenditure by starting year and project size



Source: Stage 1: Data Suitability Report (Endeavour Programme, 2020b).

4.2 SUMMARY

The predictive analytics demonstrated that several factors were able to be statistically assessed and provided an indication of what this data would look like in an ML model. However, the descriptive observations did not meet the statistically significant test criteria and therefore, could not be generalised to the broader TMR project portfolio.

The resultant observations from Stage 2 were as follows:

- Predictive observations on forecast duration and budget can be applied to the broader TMR portfolio.
  - Project final cost and duration forecasting on averages were conservative.
  - Forecast budgets (including near project completion) on averages were conservative.
- Descriptive observations on project capital expenditure, construction year, wet season start month, work type and delivery program were fairly small. Therefore, it could not be applied to the overall TMR portfolio.

Other observations included:

- Projects typically returned capital budget and/or contingency late in the project life cycle.
- There were insufficient samples to forecast behaviours on many performance factors.

It should be noted that in some of the analyses, the data averages were sensitive and could be easily skewed as a direct result of the small number of projects being assessed (i.e. less than 13 projects in each month during the wet seasons). As a result, the primary restriction limiting the degree to which these observable trends in the performance parameters could be substantiated was directly related to the sample size assessed. An increase in the number of projects for analyses would definitively increase the accuracy and confidence of these observations obtained.

Furthermore, the conservative trend identified in project forecasting may be unnecessarily prohibiting the significant financial opportunities to be realised for releasing capital budget and/or contingency back into the portfolio sooner, thus improving overall capital productivity and portfolio performance.

5 STAGE 3 – ML MODEL DEVELOPMENT AND APPLICATION

Stage 3 was aimed at training and testing a predictive ML model and evaluating the performance based on the accuracy of forecasted outcomes. This stage used TMR's 116 project dataset and took into account the learnings from the earlier stages.

5.1 ML MODEL TRAINING

To ensure the true 'out-of-sample' performance was estimated, the model utilised the 'leave one out cross validation' (LOOCV) method for testing/training. The LOOCV method meant that 116 experiments were performed, where for each experiment, the model was trained on 115 projects and the remaining 'left out' project was used as the 'test' project to assess predictive accuracy. This method is suitable for when an accurate estimate of model performance is critical (Brownlee, 2020).

The reported result was the average performance observed across the 116 experiments for the 'left out' project in each case. This method allows for as much data as possible to be used for model training while ensuring that there is no 'leakage' of information from the test data. As a result, there is confidence that the reported predictive accuracy will match what would be achieved if the models were deployed and used on other projects that they have never seen (e.g. currently active projects).

5.2 ML MODEL ASSESSMENT

The accuracy of the ML models was assessed using the following metrics:

- **Mean Absolute Error (MAE):** the average absolute deviation from the true value. MAE is non-negative, having a minimum value of 0 (no error at all). The lower the MAE is better.
- **Mean Absolute Percentage Error (MAPE):** the average absolute deviation from the true value divided by the true value (i.e. as a proportion of the true value). MAPE is non-negative, having a minimum value of 0 (no error at all). A lower MAPE is better.

Additionally, the following supplementary assessments were performed:

- Model optimisation feature selection and hyperparameter optimisation, allow the ML model to adapt to the provided data and ensure optimal performance.
- Predictive performance of final cost and duration models for individual projects when compared against the original business forecast performance.
- Aggregate (summed) project portfolio predictive performance (cost only).
- Calibration of model uncertainty estimates (confidence intervals).
- Estimation of early warning for construction cost and duration under/overruns used the following data to simulate a scenario:
  - **Cost underruns:** 41 projects (35%) with construction cost less than 0.9x original cost.
  - **Cost overruns**: 29 projects (25%) with construction cost greater than 1.0x original cost.
  - **Duration underruns:** 37 projects (32%) with construction duration less than 0.85x original duration.
  - **Duration overruns**: 36 projects (31%) with construction duration greater than 1.1x original duration.

The statistical distributions of cost and durations differ. This is why cost/duration underrun (0.9x versus 0.85x respectively) and overrun thresholds (1.0x versus 1.1x respectively) are slightly different. These thresholds were calculated to provide enough project samples for the ML models to distinguish underrun from overrun.

5.3 ML MODEL KEY PERFORMANCE RESULTS

The ML models' results for predictions and accuracy of final construction cost and final construction duration are summarised in the subsequent sections with corresponding data presented in Appendix C.

5.3.1 FINAL CONSTRUCTION COST

TMR’s forecasted final construction cost for individual projects was improved using the ML model. The model reduced the cost prediction error by 21% (15.9% error for the business versus 12.6% for the model) an average difference of a 3.3% or $148,000 improvement. Refer to Table C.1 for additional details. Additionally, the model consistently predicted an error rate less than the current approach by approximately 3% (Figure C.1) throughout the project’s life cycle.

The estimated practical value of these improvements are:

- **Cost underruns** were predicted by the model approximately four months before the current approach by TMR (Figure C.2 and Figure C.3). The model’s predictions were accurate 76% of the time compared with TMR’s being correct 99%. However, it is noted that reports of underruns by the business were identified on an average of 90% through the project's construction duration.
- **Cost overruns** were predicted by the model approximately five months after the current approach by TMR (Figure C.4 and Figure C.5). The model’s predictions were found to be accurate 91% of the time compared with TMR’s being correct 53%. The late reporting from the model is likely attributed to the fact that the business is systematically biased towards over-estimating forecast costs (by about 9%) and also hesitant to forecast underruns until 90% through a project's construction (Figure C.5).

TMR’s early ‘warning’ validity remains questionable, with a precision rate of 53% correctly forecasting overruns, while the model’s warning is more credible with a 91% precision rate.

5.3.2 FINAL CONSTRUCTION DURATION

TMR’s forecasted final construction duration for individual projects was improved using the ML model. The model reduced the duration prediction error by 11% (26% error for the business versus 23% for the model), an average difference of 3.1% or approximately 8.5 days (Table C.2). The model demonstrated consistent improvements in duration forecasting and was consistently below TMR’s forecast for the entire construction period except at project close (T = 1.0) (Figure C.6). This is indicative of construction durations being consistently updated throughout the project.

The estimated practical value of these improvements are:

- **Duration underruns:** warnings were accurately predicted by the model an average of 4 months before TMR’s current approach (Figure C.7). These claims were found to be accurate 68% of the time compared with TMR’s approach being accurate 64% of the time.
- **Duration overruns:** warnings were accurately predicted by the model an average of 2 months before TMR’s current approach (Figure C.8). These claims were found to be accurate 83% of the time compared with TMR’s approach being accurate 81% of the time.

These results demonstrate that the model and TMR’s current approach are predicting duration underruns and overruns with approximately the same level of accuracy as each other. However, in regard to underruns, the model predicts this four months before TMR, and in regards to overruns, it predicts this two months before TMR.

5.3.3 AGGREGATE PORTFOLIO COST

The ML model demonstrated improved predictions when assessing the aggregate portfolio of all the projects (n = 116) over a normalised project schedule. The aggregate cost forecast performance improved with an 87% reduction in error from the business’ existing 10.7% down to the model’s 1.4% (Figure C.9).

The model more accurately measured the portfolio value (Figure C.10) and tracked the true value (Figure C.11), while TMR's cost forecasts were consistently greater for both total and true costs.

5.3.4 SIGNIFICANT EFFECTIVE FACTORS

The ML model uses effective factors to adjust TMR's forecast to obtain its prediction. The overall median effective factor from all the data is approximately 0.91. This signifies that the model typically forecasts final costs that are approximately 91% of TMR's final costs and that TMR's current forecasts tend to be approximately 10% higher than the true final cost.

Project factors that have multipliers larger than the median overall multiplier require less downward adjustments and are more accurate or carry fewer contingencies. Multipliers that are less than the median overall multiplier tend to over-estimate the final cost more than normal or carry more contingencies.

Two project factors that were found to have greatly affected the cost prediction accuracy as they were significantly different to the median overall multiplier of 0.91 were the Delivery Program (n = 4, Figure C.12) and Investment Program (n = 2, Figure C.13). The other ten project attributes were found not to be significant in improving the model's prediction accuracy.

5.4 LIMITATION OF ML MODELS

TMR has large quantities of high-quality project data suitable for predictive time and cost modelling using machine learning. However, limitations within this project that have been shown to affect the final output from the ML model include:

- Historical data provided by TMR. The analyses do not include commentary on similar historical data or other data available elsewhere within the client's organisation(s).
- Data covers a time-period between KM-11 and KM-12 only. Analyses outside of this period are out of scope, and therefore data were not provided.
- To conduct statistically meaningful analyses, a minimum of 50 projects within each category (as per TMR suggested values i.e. less than < $5m, < $10m, < $25m, < $50m and < $100m) should be targeted.
- The ML model's poor performance in predicting cost overruns was due to the fact that TMR is systematically biased towards over-estimating forecast costs (by about 9%) and hesitant to forecast underruns until 90% through a project (Figure C.2). This results in few cases of project overrun and hence fewer samples for the ML models to learn from. Their early 'warning' validity remains questionable, with a low precision rate of 53% correctly forecasting overruns (Figure C.4). In comparison, the model's warning is more credible with a 91% precision rate.

6 TMR PROJECT CONTROL FRAMEWORK

TMR plans, manages and oversees the delivery of a safe, efficient and integrated transport network. To ensure effective management of their projects and programs, implementing program and project controls ensures planning and delivery areas have a clear and consistent approach to meet delivery expectations.

Controls are integral in the gathering, managing and analysing of information for performance monitoring, as well as to support management in making informed decisions. These apply to all stages of a program or project life cycle, from planning and development, through ongoing review cycles, to finalisation and closure.

This section describes the TMR procedures for project management control and how the ML model can assist TMR project managers in their process.

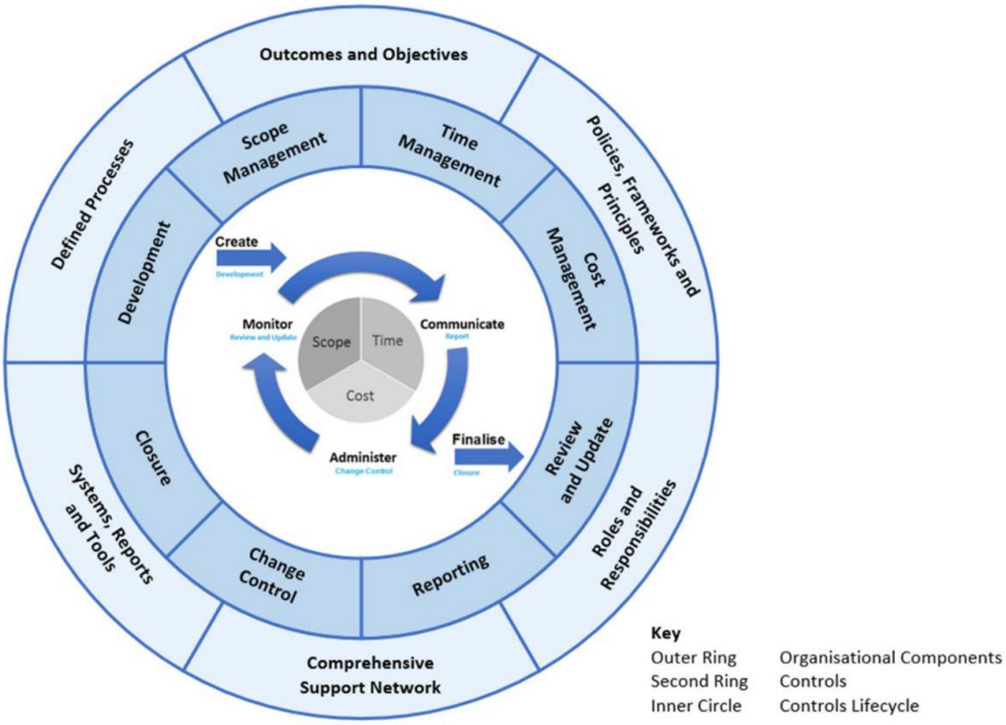
6.1 TMR CONTROL FRAMEWORK

The TMR Controls Framework consists of three principal components that are:

1. **Controls Overview –** defines the purpose, scope and governance.
2. **Detailed Guidelines –** defines the processes and business rules.
3. **Procedural Documents and Tools –** provides the tools and work instructions.

These components are integrated and support TMR's Control Model, as illustrated in Figure 6.1. The framework integrates the Organisational Components (outer ring) with the Controls (second ring) and the Controls Life cycle (inner ring). Within the second ring, Controls contains the controls processes that ensure scope, time and cost are managed and monitored.

Figure 6.1 TMR's Controls Framework

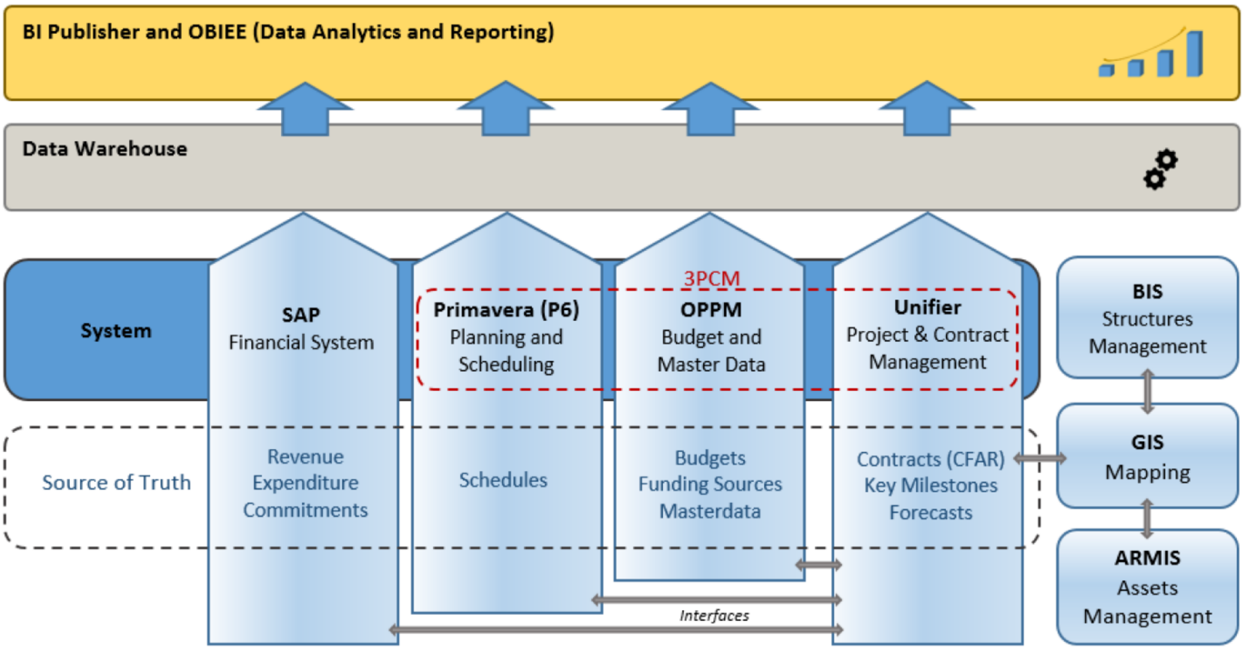


Source: Control framework part A: controls overview' (Queensland Department of Transport and Main Roads, 2020a).

6.2 SYSTEMS AND CONTROLS

The control model, project management process, and guidelines are supported by a series of software tools, which are integrated to manage the information, which supports project managers in understanding complex and relevant data for decision making. The diagram below (Figure 6.2) demonstrates the relationships between the systems and their primary function. Interfaces are used to transfer information between the systems and into the data warehouse to allow for reporting to be designed and delivered.

Figure 6.2 Control system integration



Source: Control framework part A: controls overview' (Queensland Department of Transport and Main Roads, 2020a).

Project managers are the key personnel responsible for implementing the Controls for delivering projects on time and within budget. Therefore, project managers are accountable for ensuring controls processes are undertaken, and information is up to date in all TMR Controls systems. However, the use of different software helps to support project managers decisions.

6.3 REVIEW, UPDATE AND REPORT

TMR undertakes three major program events each year, with planning areas to develop and commit to delivering their program of works. Additionally, the events provide a formal opportunity to communicate issues and opportunities to senior management. At each event, a baseline for future performance and external reporting is taken, and funding is aligned to the program estimates.

The annual Queensland Transport and Roads Investment Program (QTRIP) cycle of events are:

- **QTRIP Development** – Critical program exercise in the infrastructure investment annual calendar. It establishes the department's committed delivery program on which their performance is judged. The annual QTRIP publication identifies a two-year approved program of work and an additional two-year indicative program. A baseline of time and cost is taken to facilitate external reporting and monitor external program and project performance for the year.
- **October Review** – A review that re-aligns the program and the distribution of funds from bulks (where possible), post the end of financial year outcomes and budget carryover process. The program is aligned with delivery expectations, and a new baseline is taken for future internal performance to be monitored.
- **January Review** – A review process where the business determines which programs and projects are required to be taken into consideration.

These periodic reviews allow for the updating and planning of program and infrastructure investment by identifying needs and risks. The reviewing and planning of different business portfolio programs and expenditures are based on structured reports from data and information extracted from the project management system and control.

Additionally, the key focus of structured reporting is the monitoring of progress against planned positions to facilitate conversations about potential impacts to the program and project deliverables. By ensuring structured and timely reporting processes are in place at different levels throughout the department, and that information flows to all levels of management, informed decisions and targeted responses can be made.

6.4 FORECASTING

Accurate forecasting within the system is critical to ensure that TMR's reporting, both internally and externally, reflects the current expected delivery position. Maintaining accurate forecasts also provides all levels of management with confidence that the Planning and Delivery areas understand the financial situation of their programs and projects and are proactively managing their budget requirements.

TMR follows the below forecasting principles for their internal system (Queensland Department of Transport and Main Roads, 2020b):

- Review and update forecasts monthly.
- Forecast to expected delivery.
- Past performance informs future forecasts.
- Actively manage contingency, savings, bulk.
- Coordination and communication are essential.
- Assign and share responsibility.

The overarching Forecasting Principles provide an efficient and effective way to ensure better forecasting outcomes for Planning and Delivery Areas whilst upholding policies and guidelines. Project managers are integral to the forecasting process as they are the conduit between works on the ground and the organisation, therefore continuous communication with all those involved is vital to the success of the forecasting process.

Maintaining and updating accurate forecasts and financial information such as Value of Work Completed (VoWC), Budget Expenditure Flow (BEF), and the Estimate Final Cost (EFC) provides certainty around TMR's program delivery and allows management to make informed decisions. Additionally, these allow TMR to identify ongoing project budget requirements, including early warning of variations, potential savings and budget requirements.

6.5 ML AND TMR PROJECT MANAGEMENT

The ML modelling results have highlighted several areas within TMR's project framework which could benefit from ML technology.

From the review of TMR's project management system, there is an opportunity to utilise AI for cost forecasting and project completion as a tool for project and portfolio/program managers to use in their monthly reports and planning. As the AI application can provide earlier warnings compared to traditional methods, project managers can review and identify project risks and countermeasures to guarantee project performance.

AI can also be applied in the Control circle (refer to Figure 6.1) on the Time and Cost Management. Furthermore, the AI tool may be applicable in parallel to the Control System Integration (refer to Figure 6.2) as a check for a project's behaviour and performance. Additionally, the AI can consider the information provided by the different software support systems (such as SAP and 3PCM – Oracle Primavera's Portfolio Management (OPPM), Unifier, P6 and BI Publisher), providing the project managers and planners relevant information for checks and review.

There is also the opportunity to use the AI application as an additional tool that provides information to be included on the reports evaluated during the annual QTRIP events, as it could provide additional insight for future project investments and opportunities.

7 SUMMARY AND RECOMMENDATIONS

Improving the predictability of project cost and duration underruns and overruns is a challenging problem affecting many TMR projects. AI technologies such as ML models offer a faster, cost-efficient and effective way to conduct multiple concurrent iterations to build evidence-based predictive models for tackling these challenges.

TMR currently has a substantial database of historical project data. Using this resource, the suitability and application of ML modelling were demonstrated, showcasing improvements to accuracy and precision of project cost and duration forecasting.

Additionally, it was identified that TMR uses a series of procedures and control tools for project management and program forecasting that support decision-makers in understanding the requirements at a project and portfolio level.

For this reason, the integration of ML into TMR's existing control framework is supported, as it would act as a validation tool and provide evidence-based data to support decisions.

7.1 STUDY RESULTS

This project has investigated the suitability and performance of ML solutions to facilitate 'data-driven' decision making through improved accuracy and precision of project and portfolio/program cost and duration forecasting. From this investigation, the following key results were identified:

- TMR's current project financial data is sufficient to meet the data suitability requirements for ML models.
- Predictive and descriptive observations can be made from the fields available in TMR's project financial data.
- ML model training can be undertaken using TMR's project financial data.
- A trained ML model is able to predict project final construction cost and final construction date with the following results:
  - Final cost prediction error reduced from 15.9% (business) to 12.6% (ML model), improvement of 3.3% or an average of $148,000.
  - ML model predicted cost underruns four months before the business. ML model's accuracy was 76% compared to business' 90%.
  - ML model predicted cost overruns five months after the business. ML model's accuracy was 91% compared to business' 53%.
  - ML model predicted duration underruns four months before the business. ML model's accuracy was 68% compared to business' 64%.
  - ML model predicted cost overruns two months before the business. ML model's accuracy was 83% compared to business' 81%.
  - Aggregate cost forecast performance improved with a reduction in error from the business' existing 10.7% down to the model's 1.4%.
- Limitations exist for the ML model that includes:
  - ML model only looked at the project's construction phase between KM-11 (Delivery/Construction start) and KM-12 (Practical completion).
  - Additional project data should be included with a minimum of 50 projects within each category (as per TMR suggested values i.e. less than < $5m, < $10m, < $25m, < $50m and < $100m).
  - Predictions made by the model may be affected by the business' bias towards over-estimating forecast costs and hesitation to forecast underruns until 90% through the project.

7.2 RECOMMENDATIONS

AI technology has unanimously demonstrated its capability to enhance capital productivity and portfolio performance, act as an early warning tool and decrease monitoring costs.

However, for broader application within TMR’s network, ARRB recommends the following for consideration.

7.2.1 PILOT STUDY

A pilot study should be undertaken on the application of AI forecast for one or multiple projects from start to completion. This would aim at evaluating the monthly performance of the ML model. The results will include, but not necessarily be limited to, only numerical forecasting details and feedback from the project managers to identify if the AI provides benefits to the project.

The project or projects selected for the trial will be compared with other comparable projects where no AI application has been included and where the data is sufficient to establish a like-for-like comparison (including lessons learned and risk assessment).

The AI application trial would need to include a visualisation tool to identify trends and performance throughout the project. This would need to be comparable with the current dashboard that TMR use in their project management.

7.2.2 EXPAND THE DATASET FOR FURTHER REFINEMENT

Increasing the project dataset size will allow the ML model to refine its findings and therefore improve the cost and schedule forecasting.

The additional data to increase the machine learning capability to forecast cost and time can be collected in parallel with the pilot trial.

7.2.3 DEVELOPMENT OF INFORMATION PIPELINE AND DATA ACCESS

For further refinement and application of the ML, data collection needs to be undertaken in a specific format matching the data dictionary. This can be challenging depending on the project management system and the level of access for data collection and compilation.

Therefore, data access and collection procedures need to be reviewed for clarity to ensure effective and efficient data transfer. These procedures need to align with the minimum requirements for the ML model’s data dictionary.

7.2.4 DEVELOP AN IMPLEMENTATION STRATEGY

To implement ML into TMR’s current business practices a roll-out strategy will need to be developed. This strategy should include the following:

- Identification of short and long-term goals for the implementation of ML technologies. This should outline how successful implementation is measured.
- Plan for the short and long term implementation of ML technologies into TMR’s current software packages. Will this require modification of existing software packages or require new software packages entirely?
- Procedures for user access and interaction with the ML tool.
- Implementation of a training program for TMR’s project managers to ensure they input the appropriate data, understand outputs, and can effectively utilise the ML tool's outputs.

7.3 FURTHER ML DEVELOPMENT

This exercise has only offered a small glimpse of the predictive and descriptive insights available from the ML model, trained on TMR's project data. Further data available on TMR's data systems indicate that other AI modelling capabilities are possible. For instance, the application of dynamic cash flow predictions on a monthly, quarterly, annual or financial year basis.

The ML model identified significant financial and delivery opportunities for the re-distribution of capital back into TMR's portfolio that can be applied pending these costs are realised earlier and more accurately. Further applications with ML models can improve areas like portfolio cash management, project preparation and portfolio monitoring.

Additionally, some of the future activities that can be pursued to improve and refine the ML model are shown in Table 7.1, including potential benefits to the model.

Table 7.1: Future development

<table>
<tr><th>Proposed items</th><th>Potential benefits</th></tr>
<tr><td>1. Expand this analysis dataset to be greater than the current model sample (more than 200 projects).</td><td rowspan="7">• Improved capital productivity through greater precision at cost and time forecasting.
• Increase the evidence-base for portfolio benchmarking and decision making.
• Increase real-time monitoring and early warning capability.
• Increase the breadth of early warning and analytic capability earlier in the project life cycle.
• Improve monitoring and evaluation of asset operations and benefits realisation.
• A richer understanding of historical risks, lessons learned and project contractual behaviour.</td></tr>
<tr><td>2. Expand this analysis dataset to include larger project sizes (i.e. > $100m).</td></tr>
<tr><td>3. Build an enhanced S-Curve tool that allows for multiple live-project overlays and real-time data feeds.</td></tr>
<tr><td>4. Apply these analyses to data earlier than KM-11 (Construction start). For example, at KM-1 (Planning started), KM-4 (Planning final deliverable completed), KM-6 (Business case approved), KM-10 (Main contract award) or to concept/final business case. It is noted that these milestones have high levels of variability and may not enable consistent trends for statistical analysis.</td></tr>
<tr><td>5. Apply these analyses to data after KM-12 (Practical completion). For example, to operations and benefits realisation. This could be more beneficially applied to the aggregate portfolio rather than individual projects.</td></tr>
<tr><td>6. Use natural language processing to apply these statistical analyses to other qualitative project data such as lessons learned, project closeout reports, project evaluations, risk management reports, and other contractual data.</td></tr>
<tr><td>7. Application to areas of cash management, project preparation and portfolio monitoring (details on implication and requirements will need to be explored).</td></tr>
</table>

REFERENCES

Brownlee, J 2020, *LOOCV for evaluating machine learning algorithms,* webpage, Machine Learning Mastery, Vermont, Vic, viewed 14 April 2021, <https://machinelearningmastery.com/loocv-for-evaluating-machine-learning-algorithms/>.

Endeavour Programme 2020a, *Stage 1: Data Suitability Report*, REP-ARRB-TMR-OCT-001, Rev 0, Endeavor Programme, Newstead, Qld.

Endeavour Programme 2020b, *Stage 1: Data Suitability Report*, REP-ARRB-TMR-OCT-001, Rev A, Endeavor Programme, Newstead, Qld.

Endeavour Programme 2020c, *Stage 2: Analytics Report*, REP-ARRB-TMR-OCT-001, Rev D, Endeavor Programme, Newstead, Qld.

Endeavour Programme 2020d, *Stage 3: Modelling Report*, REP-ARRB-OCT-003, Rev A, Endeavor Programme, Newstead, Qld.

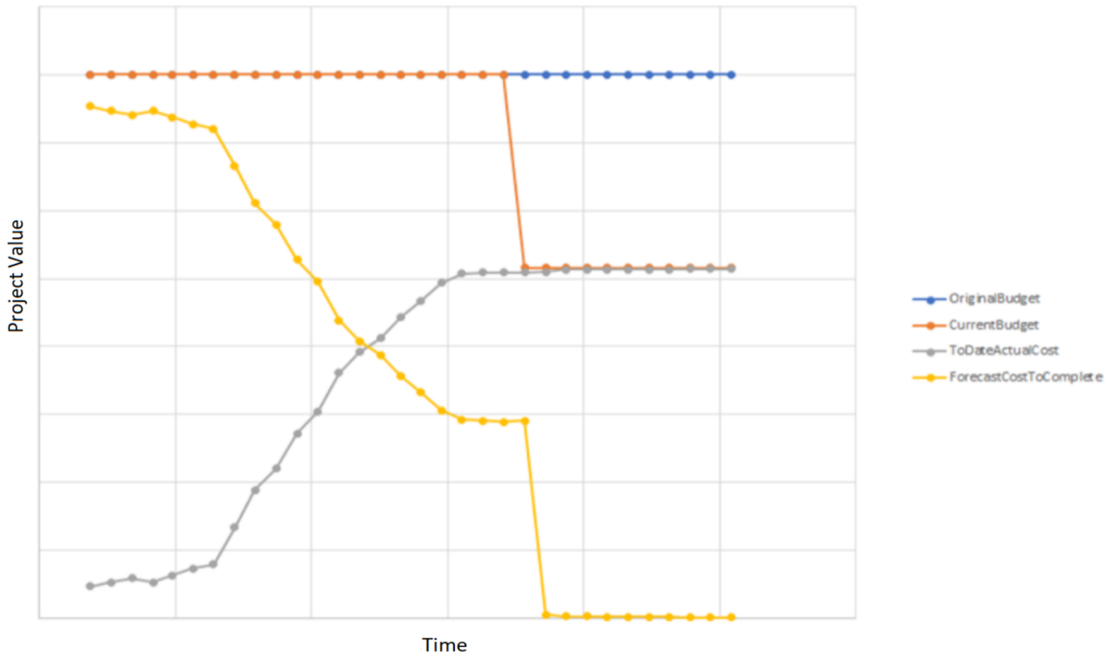
Expert.ai 2020, *What is machine learning? a definition*, webpage, Expert.ai, Rockville, MD, USA, viewed 14 April 2021, <https://www.expert.ai/blog/machine-learning-definition/>.

Queensland Department of Transport and Main Roads, 2020a, ‘Control framework part A: controls overview’, TMR, Brisbane, Qld.

Queensland Department of Transport and Main Roads, 2020b, ‘Control framework: project financial forecasting guideline part B’, TMR, Brisbane, Qld.

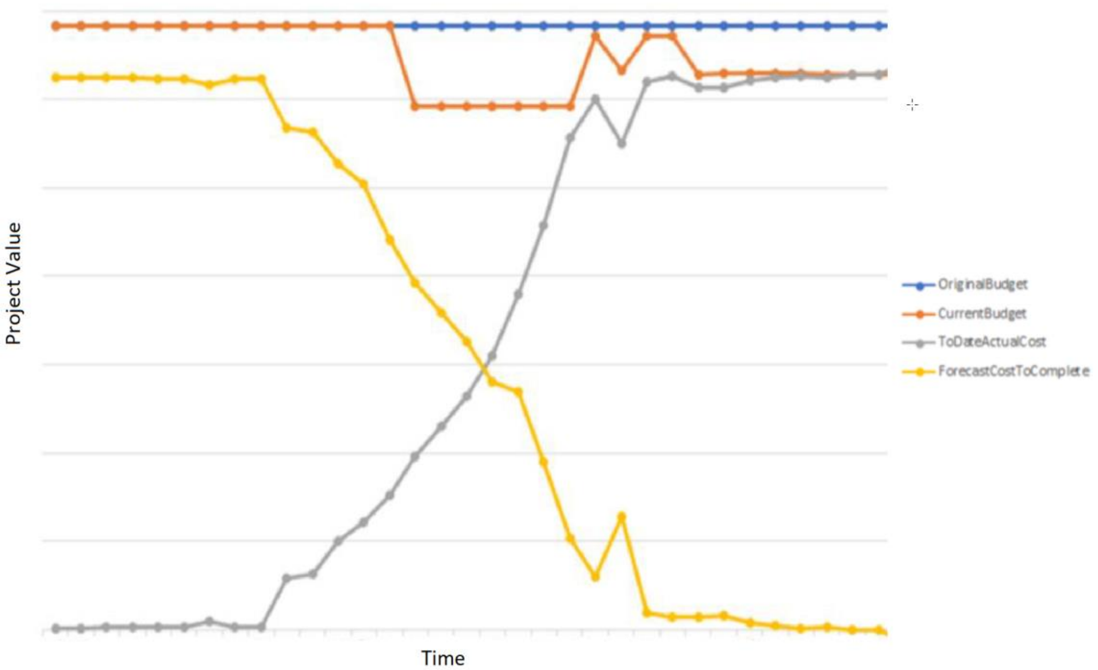
APPENDIX A TYPICAL PROJECT BEHAVIOUR BASED ON S-CURVE

Figure A.1: S-Curve behaviour A



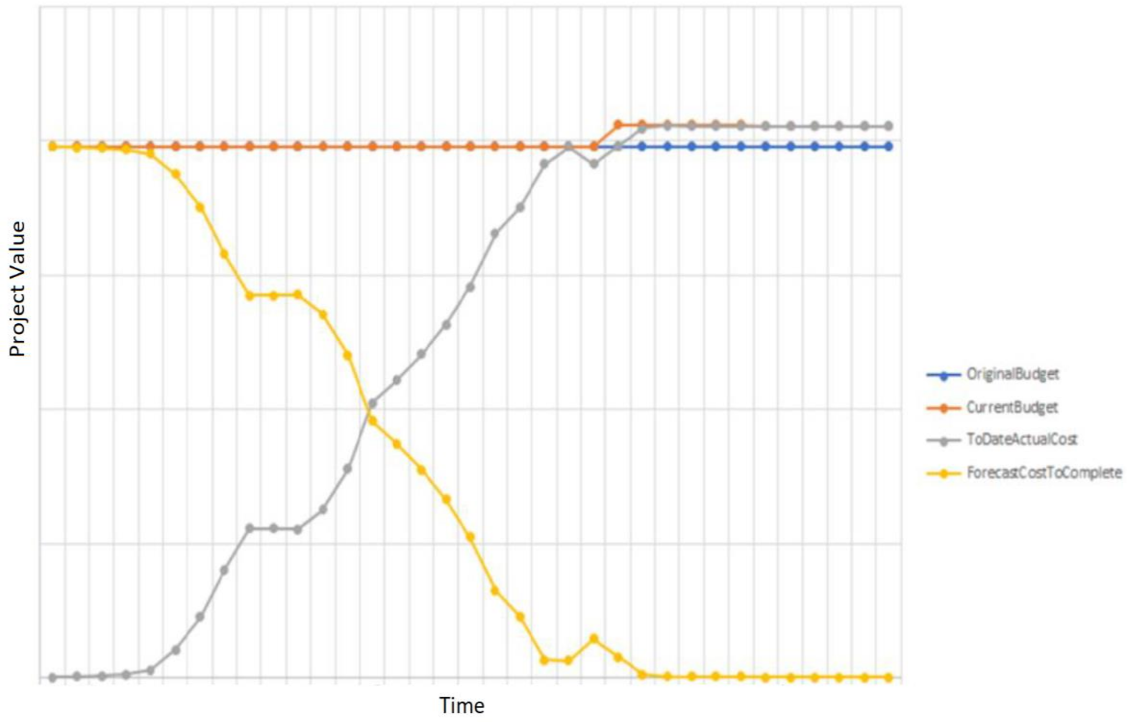
Source: Stage 1: Data Suitability Report (Endeavour, 2020a).

Figure A.2: S-Curve behaviour B



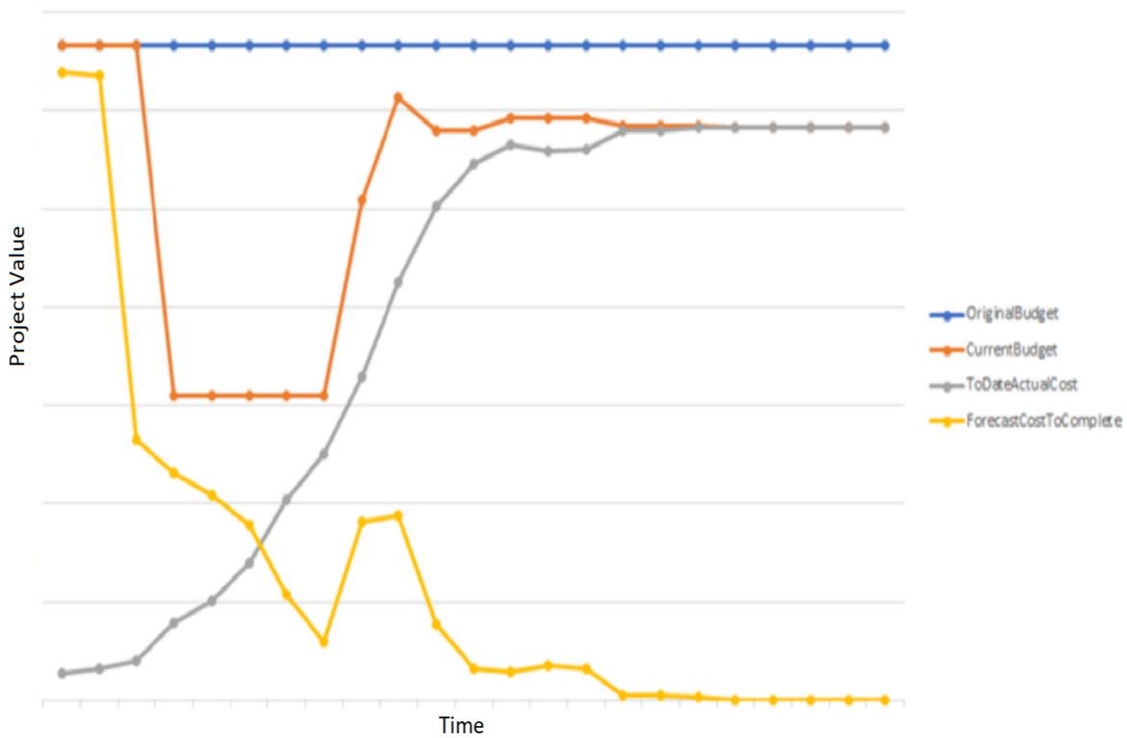
Source: Stage 1: Data Suitability Report (Endeavour, 2020a).

Figure A.3: S-Curve behaviour C



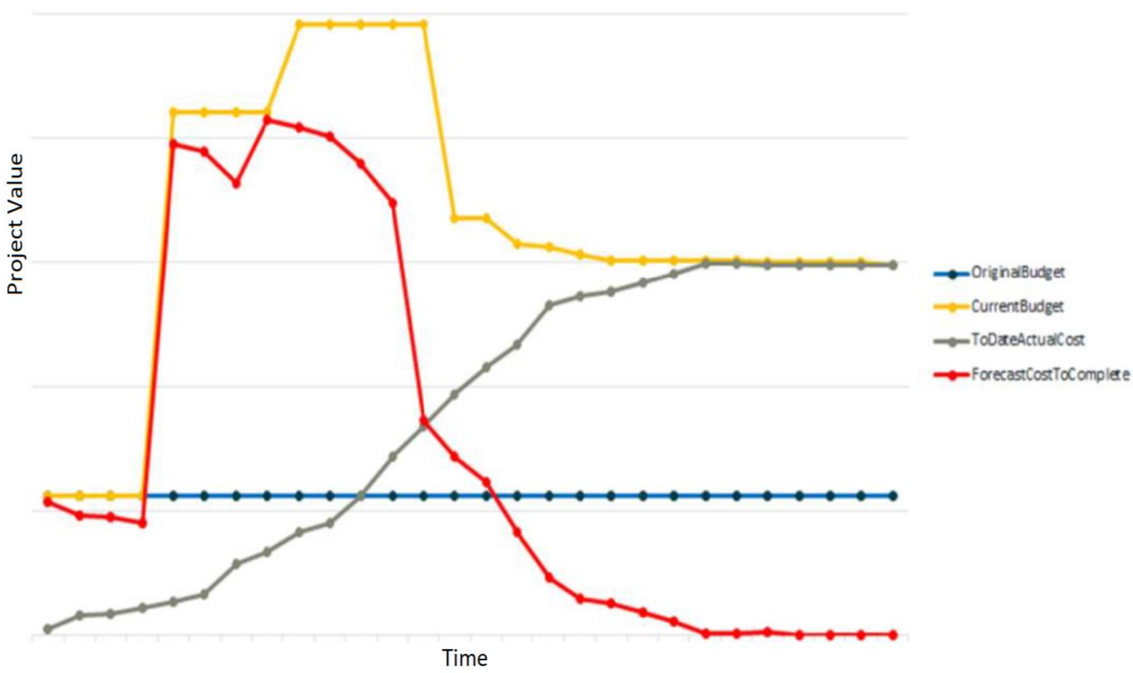
Source: Stage 1: Data Suitability Report (Endeavour, 2020a).

Figure A.4: S-Curve behaviour D



Source: Stage 1: Data Suitability Report (Endeavour, 2020a).

Figure A.5: S-Curve behaviour E



Source: Stage 1: Data Suitability Report (Endeavour, 2020a).

APPENDIX B WORK TYPE AND DELIVERY PROGRAMS

Table B.1: Work type and delivery programs of TMR's projects (n = 116)

Work type	# Projects	Delivery program	# Projects
Grids, guidance & delineation enhancement works	3	Darling Downs District	11
Maritime major enhancement works	7	Far North District	20
Miscellaneous operational activities	1	Fitzroy District	6
New or upgraded drainage FAC	1	Mackay Whitsunday District	6
New route-cons to seal STD	1	Metropolitan District	8
Passenger transport enhancement works	1	North Coast District	10
Passenger transport priority works	4	Northern District	1
Prov of other FAC. In road res	2	North West District	8
Realignment	1	Road Infrastructure	1
Road safety enhancement works	19	South Coast District	13
Sealing previously unsealed road	16	South West District	3
Traffic-flow improvements	1	Service planning and infrastructure	2
Upgrade existing bridges	3	Wide Bay Burnett	10
Widening existing pavement	18	Boating Infrastructure	7
At-grade intersection or level crossing	15	Central West District	8
Auxiliary lane construction	4		
Bridges – construction new and replace	4		
Construction of additional lanes	6		
Cycleways construction	5		
Duplication	2		

APPENDIX C SECONDARY DATA COLLECTION RESULTS

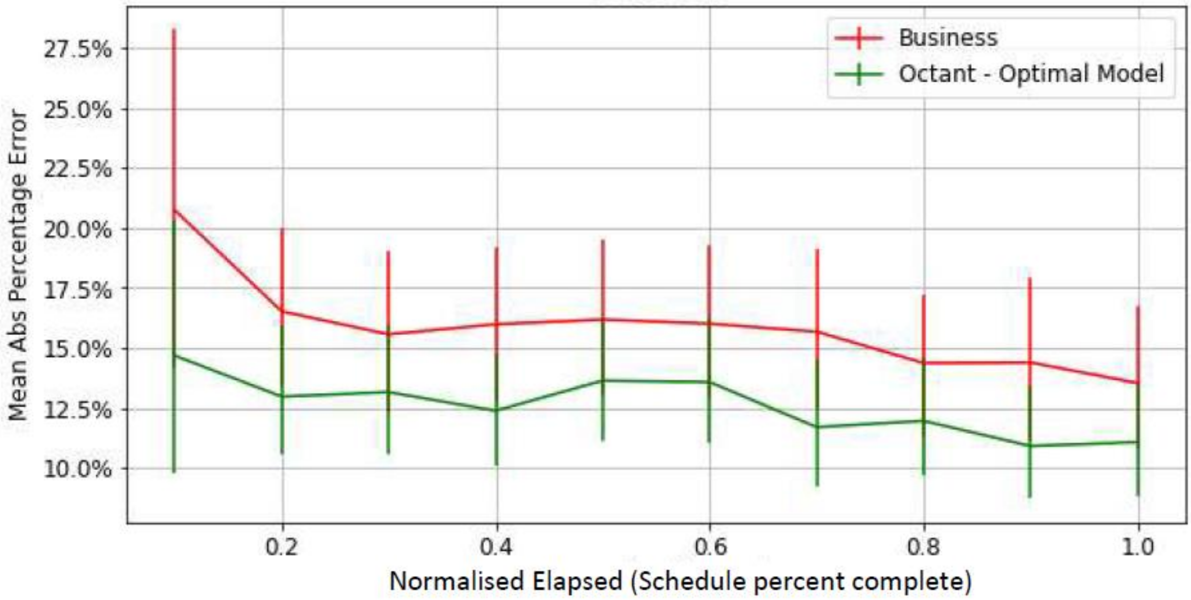
C.1 FINAL CONSTRUCTION COST DATA

Table C.1: Construction cost prediction error of the business vs ML model

Project progress (normalised elapsed threshold upper)	Mean Absolute Error (MAE)			Mean Absolute Percentage Error (MAPE)		
	Business ($)	ML Model ($)	Improvement (Business-ML) ($)	Business (%)	ML Model (%)	Improvement (Business-ML) (%)
0.1	2,169,060	1,929,889	239,171	20.7	14.7	6.1
0.2	1,296,156	1,185,523	110,633	16.5	13.0	3.5
0.3	1,330,770	1,284,044	46,726	15.6	13.2	2.4
0.4	1,231,329	1,146,130	85,199	16.0	12.4	3.6
0.5	1,149,548	1,104,703	44,845	16.2	13.6	2.5
0.6	1,126,901	1,087,768	39,133	16.0	13.6	2.4
0.7	1,319,849	1,154,223	165,626	15.7	11.7	4.0
0.8	1,275,821	1,065,752	210,069	14.4	12.0	2.4
0.9	1,296,949	983,948	313,000	14.4	10.9	3.5
1.0	1,020,080	787,325	232,755	13.5	11.1	2.5
MEAN	1,321,646	1,172,931	148,716	15.9	12.6	3.3

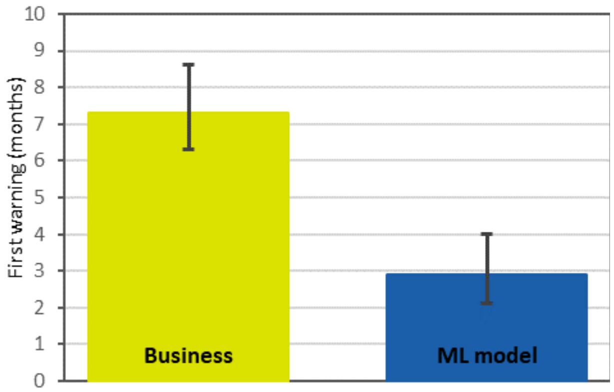
Source: Stage 3: Modelling Report (Endeavour Programme, 2020d).

Figure C.1: Construction cost prediction error (MAPE) vs normalised elapsed time



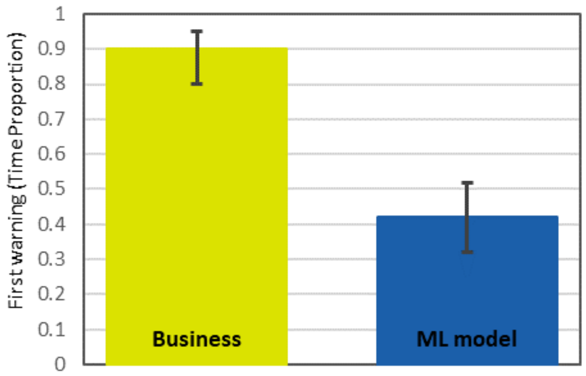
Source: Stage 3: Modelling Report (Endeavour Programme, 2020d).

Figure C.2: Time to first warning (months) – construction cost underruns



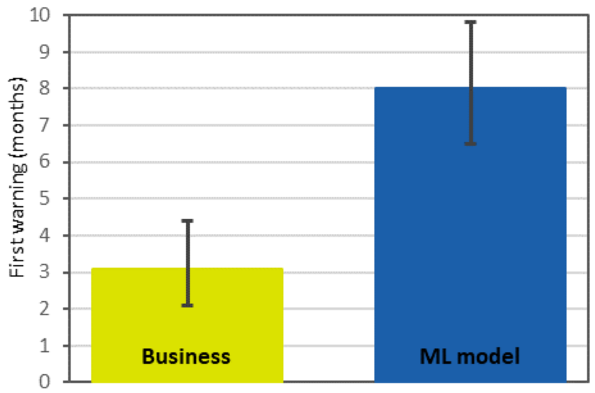
Source: Stage 3: Modelling Report (Endeavour Programme, 2020d).

Figure C.3: Time to first warning (proportion of construction duration) – construction cost underruns



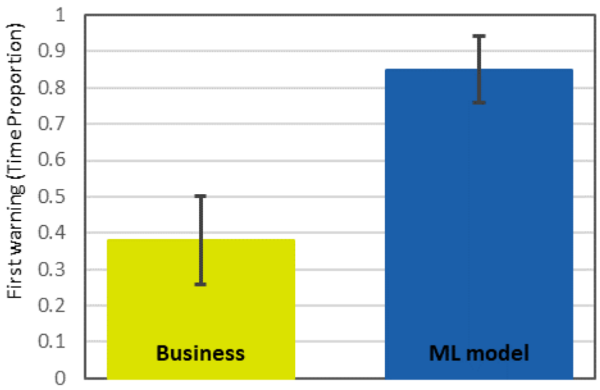
Source: Stage 3: Modelling Report (Endeavour Programme, 2020d).

Figure C.4: Time to first warning (months) – construction cost overruns



Source: Stage 3: Modelling Report (Endeavour Programme, 2020d).

Figure C.5: Time to first warning (proportion of construction duration) – construction cost overruns



Source: Stage 3: Modelling Report (Endeavour Programme, 2020d).

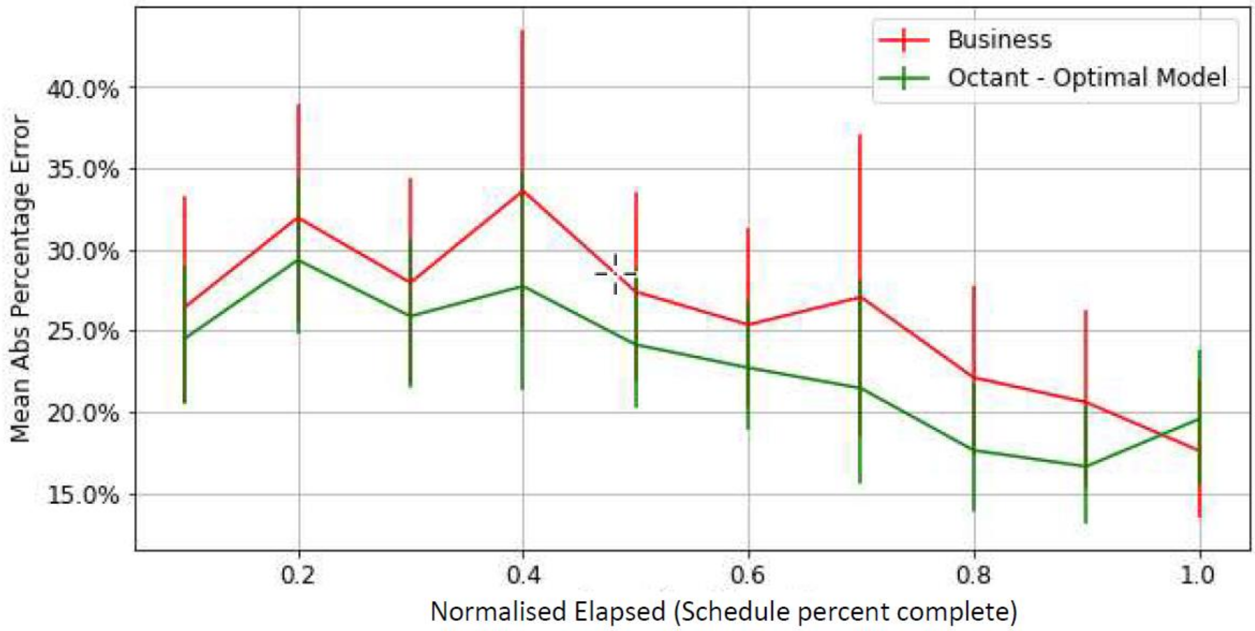
C.2 FINAL CONSTRUCTION DURATION DATA

Table C.2: Construction cost prediction error of the business vs ML model

	Mean Absolute Error (MAE)			Mean Absolute Percentage Error (MAPE)		
Project progress (normalised elapsed threshold upper)	**Business (days)**	**ML Model (days)**	**Improvement (Business-ML) (days)**	**Business ($)**	**ML Model ($)**	**Improvement (Business-ML) ($)**
0.1	115.2	106.6	8.6	26.5%	24.5%	1.9%
0.2	93.4	85.8	7.7	32.0%	29.3%	2.6%
0.3	92.5	80.4	12.0	28.0%	25.9%	2.1%
0.4	86.4	70.5	15.9	33.6%	27.7%	5.9%
0.5	75.5	64.8	10.7	27.4%	24.1%	3.3%
0.6	70.7	64.7	5.9	25.4%	22.7%	2.6%
0.7	70.6	62.4	8.2	27.1%	21.5%	5.6%
0.8	63.5	53.4	10.1	22.1%	17.6%	4.5%
0.9	60.5	52.1	8.3	20.6%	16.6%	4.0%
1.0	41.6	44.1	-2.5	17.6%	19.5%	−1.9%
MEAN	77.0	68.5	8.5	26.0%	23.0%	3.1%

Source: Stage 3: Modelling Report (Endeavour Programme, 2020d).

Figure C.6: Construction duration error (MAPE) vs normalised elapsed time



Source: Stage 3: Modelling Report (Endeavour Programme, 2020d).

Figure C.7: Time to first warning (months) – construction duration underruns

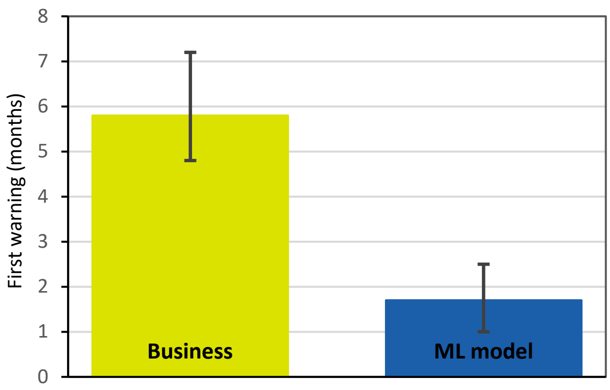
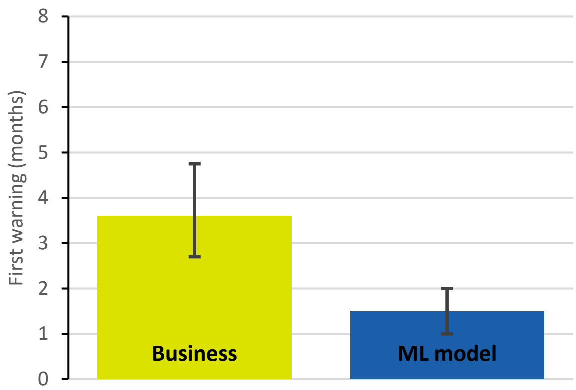


Figure C.8: Time to first warning (months) – construction duration overruns

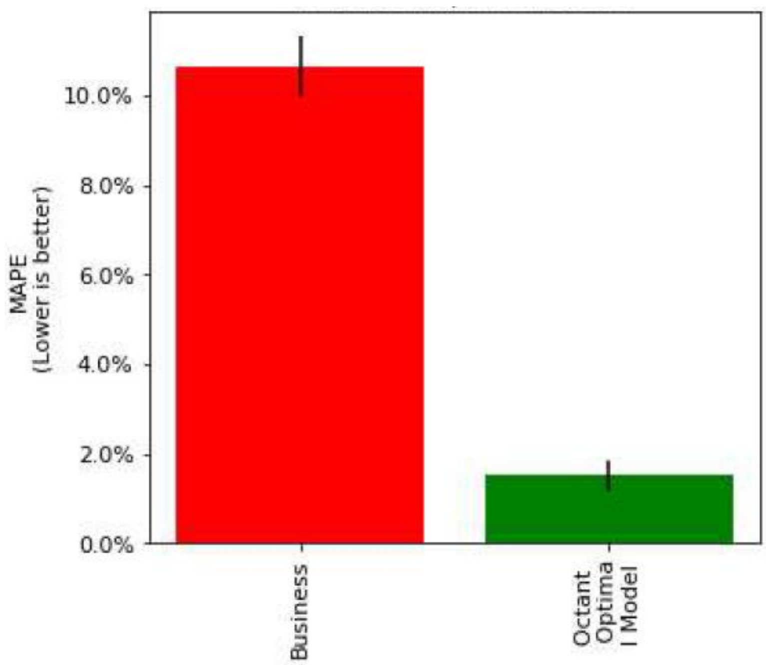


Source: Stage 3: Modelling Report (Endeavour Programme, 2020d).

Source: Stage 3: Modelling Report (Endeavour Programme, 2020d).

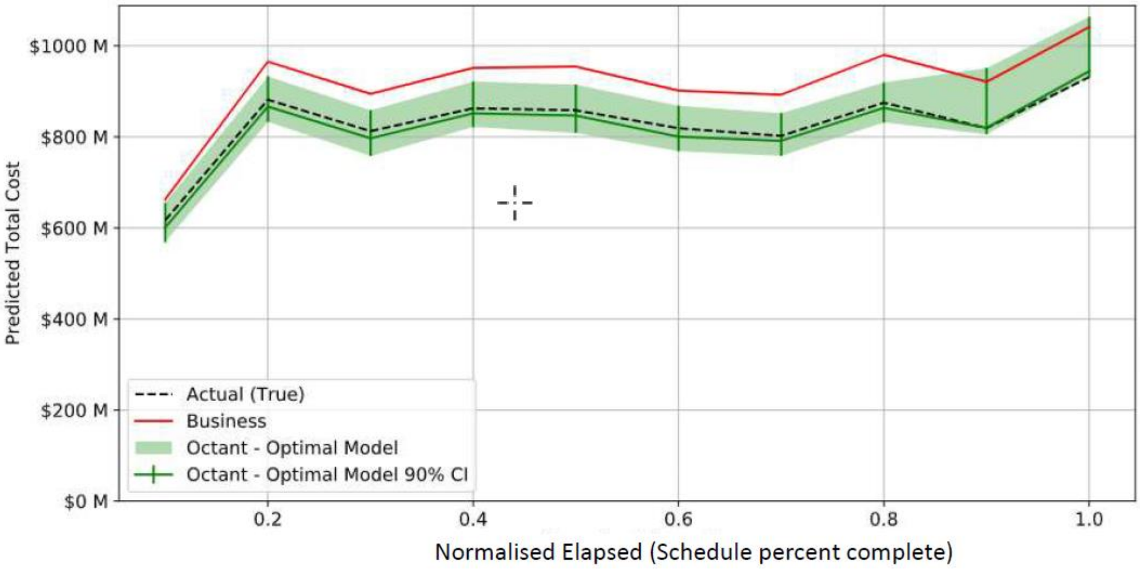
C.3 AGGREGATE PORTFOLIO COST DATA

Figure C.9: Aggregate total cost prediction (MAPE)



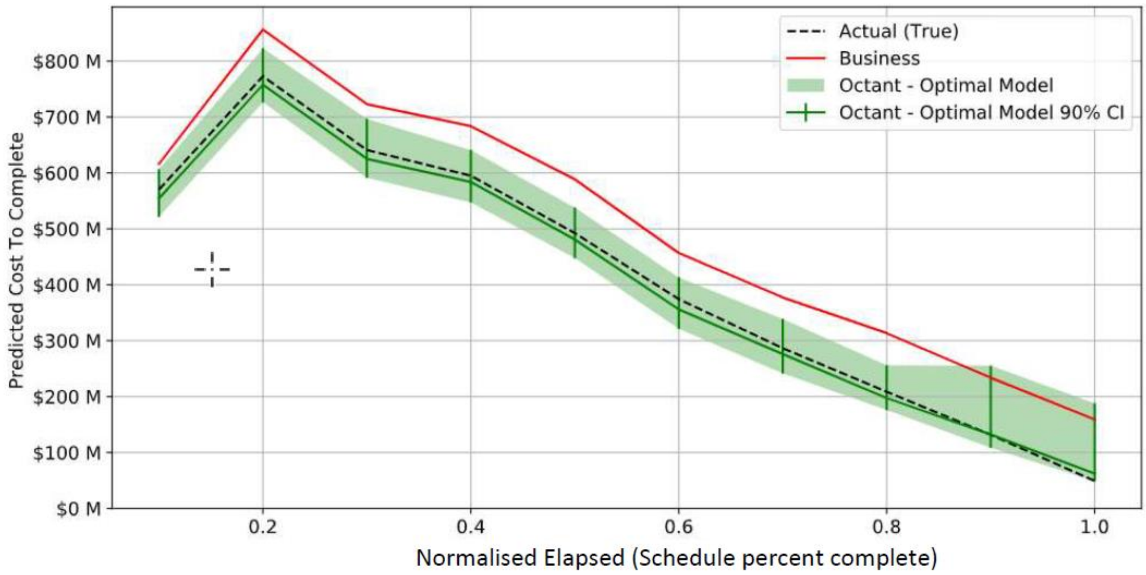
Source: Stage 3: Modelling Report (Endeavour Programme, 2020d).

Figure C.10: Aggregate total cost prediction



Source: Stage 3: Modelling Report (Endeavour Programme, 2020d).

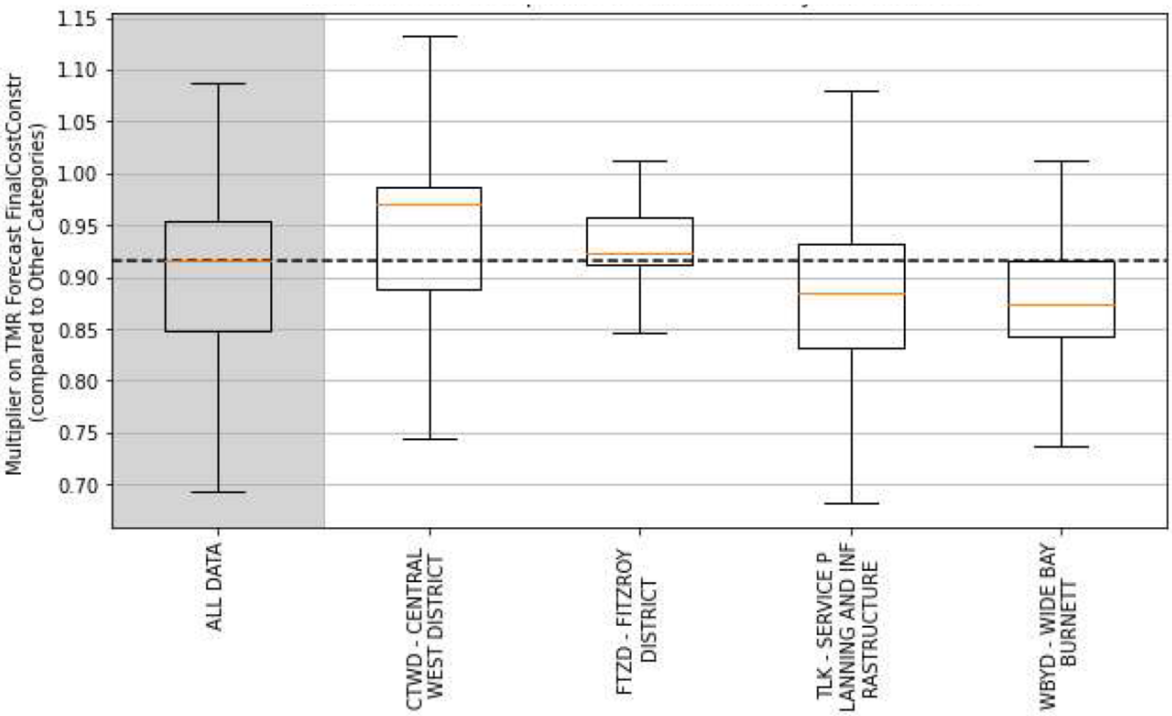
Figure C.11: Aggregate cost to complete prediction



Source: Stage 3: Modelling Report (Endeavour Programme, 2020d).

C.4 SIGNIFICANT EFFECTIVE FACTORS DATA

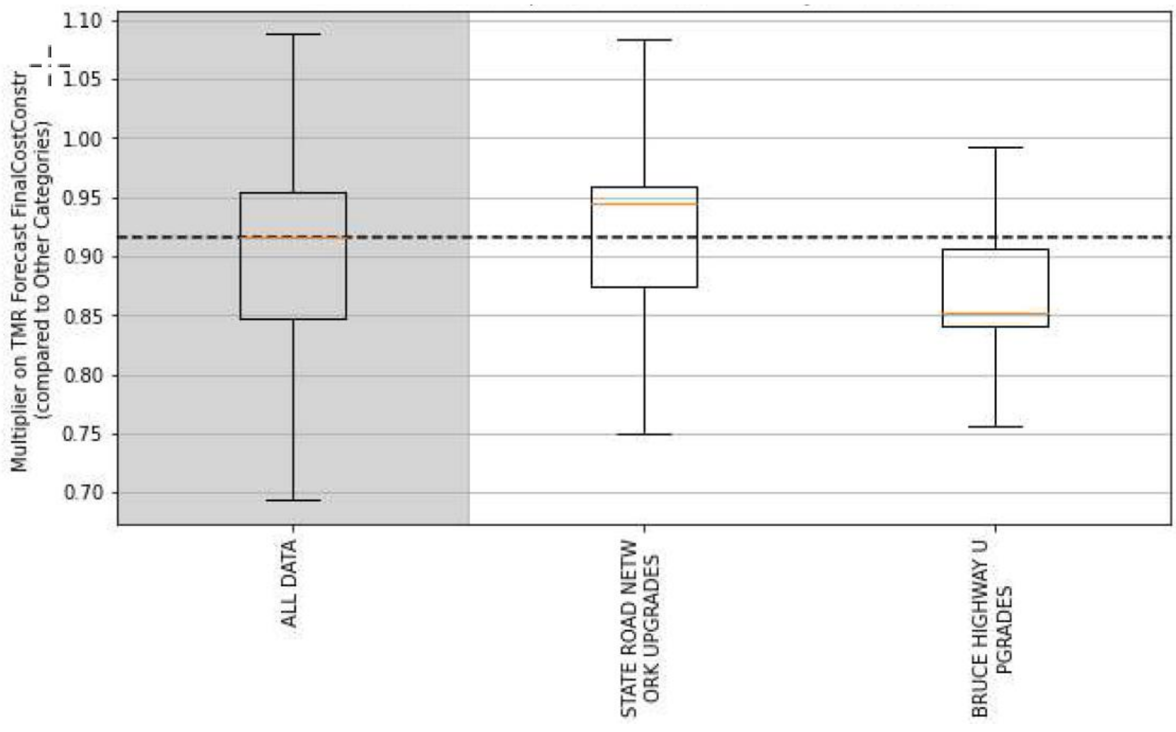
Figure C.12: Delivery program multipliers on final cost forecast



Source: Stage 3: Modelling Report (Endeavour Programme, 2020d).

Notes: Boxes represent lower quartile (25th percentile) to upper quartile (75th percentile), with median indicated by the orange lines and 'whiskers' extending to 1.5 x inter-quartile range.

Figure C.13: Investment program multiplier on final cost forecast



Source: Stage 3: Modelling Report (Endeavour Programme, 2020d).

Notes: Boxes represent lower quartile (25th percentile) to upper quartile (75th percentile), with median indicated by the orange lines and 'whiskers' extending to 1.5 x inter-quartile range.