Journal of Construction Project Management and Innovations, 16(1), 1-11
DOI: http://doi.org/10.36615/jcpmi.v16i1.4567
ISSN: 2223-7852

RESEARCH PAPER

Predicting the Likelihood of Cost Overruns in Tunnel Projects in Australia

Ishmael Adams[1,2*], Vishnuvardhan Reddy[1], Prithivi mukhiya[1], Godslove Ampratwum[1]

[1]Swinburne university of Technology, Parramatta, 2150, New South Whales, Australia
[2]Lead Institute of Higher Education, Parramatta, 2150, New South Whales, Australia

Correspondence
Ishmael Adams
Swinburne University of Technology, Parramatta, 2150, New South Whales, Australia
Email address: ishmaeladams@swin.edu.au; ishradam@gmail.com

Received: 05-10-2025; Revised: 08-12-2025; Accepted:04-02-2026

Abstract

Cost overruns remain a persistent challenge in tunnel construction projects, yet their predictive assessment, particularly within the Australian context, has received limited scholarly attention. The purpose of this study is to develop a probabilistic, distribution-based framework for understanding and predicting the likelihood and magnitude of cost overruns in Australian tunnel projects, while explicitly examining the influence of project contract size on cost overrun risk. Using historical data from 27 completed tunnel projects, the study applies descriptive and inferential statistical techniques, goodness-of-fit testing, and probability modelling to characterise cost overrun behaviour. The analysis addresses two key questions: (1) What are the statistical properties and best-fit probability distribution of cost overruns? (2) How does contract size influence the likelihood and magnitude of overruns? The results indicate an average cost overrun of 46.60%, accompanied by pronounced variability and positive skewness (2.50). The Dagum distribution provides the best fit for the overall dataset, while the Generalised Extreme Value (GEV) and Gamma distributions best represent projects valued below and above AUD 1 billion, respectively. Although ANOVA results reveal no statistically significant difference in mean cost overruns between small and large projects (p-value = 0.89), smaller projects exhibit substantially greater variability, reflecting distinct underlying risk structures. Probability analysis further demonstrates a 77.34% likelihood of exceeding a 5% cost overrun threshold. By shifting the focus from mean-based assessment to distribution-driven risk modelling, this study contributes a quantitative, tunnel-specific approach for predicting cost overruns. The findings support more informed contingency allocation, probabilistic cost estimation, and risk-aware decision-making for infrastructure stakeholders involved in tunnel project planning and delivery in Australia.

Keywords: Australian tunnel projects, Cost overruns, Probability Distribution Function (PDF), Cumulative Distribution Function (CDF), ANOVA, Descriptive statistics.

1. INTRODUCTION

Tunnel projects are critical components of modern infrastructure, facilitating transportation, utilities, and water systems while minimising surface disruption in urban areas. Despite their importance, these projects are notoriously susceptible to cost overruns, leading to significant financial and managerial challenges (Flyvbjerg, 2010). Globally, tunnel construction has a history of budget exceedances, with prominent examples such as the Channel Tunnel (80% overrun) (Global Infrastructure Hub, 2021) and Boston's Big Dig (200% overrun) (Flyvbjerg, 2010) highlighting the severity of the issue. In

Australia, projects like the Sydney Metro City and South-west (64% overrun) (Sydney Metro, 2016) and Epping to Chatswood Rail (84% overrun) (Thiess Hochtief, No Date) demonstrate that cost escalations remain a persistent problem. These overruns strain public funds, delay project completion, and undermine stakeholder confidence, emphasising the need for better predictive models and risk management strategies (Vickerman, 1997; Altshuler and Luberoff, 2004).

The causes of cost overruns in tunnel projects are multifaceted, ranging from geological uncertainties and design changes to inflation, labour shortages, and contractual disputes (Frimpong, Oluwoye and Crawford, 2003; Assaf and Al-Hejji, 2006; Gómez-Cabrera, Gutierrez-Bucheli and Muñoz, 2024). Traditional cost estimation methods often fail to account for these complexities, leading to overly optimistic budgets. While probabilistic approaches, such as Monte Carlo simulations and reference class forecasting, have been proposed to improve accuracy, their application in tunnel projects - particularly in the Australian context - remains unexamined. Existing research has examined cost over- runs in general construction, but tunnel-specific studies are unknown, leaving a gap in understanding the unique statistical behaviour of overruns in this high-risk sector (Love et al., 2013, 2014; El-Kholy, 2015; Plebankiewicz, 2018).

This study aims to address this gap by analysing the statistical characteristics and probability distributions of cost overruns in Australian tunnel projects. By leveraging historical data from 27 projects across five states, the research seeks to identify the best-fit probability models for different project scales and assess the influence of contract size on overrun likelihood. The findings will provide a quantitative framework for predicting cost deviations, enabling project managers to allocate contingency reserves more effectively. Additionally, the study explores whether smaller projects (below AUD 1 billion) exhibit different overrun patterns compared to larger ones (above AUD 1 billion), offering insights into risk stratification based on project size.

The objectives of this study are twofold: first, to analyse the statistical properties and identify the best-fit probability distribution for cost overruns in tunnel projects; and second, to examine the relationship between contract size and overrun trends. By achieving these objectives, this study contributes a quantitative, tunnel-specific approach for predicting cost overruns. The findings support more informed contingency allocation, probabilistic cost estimation, and risk-aware decision-making for infrastructure stakeholders involved in tunnel project planning and delivery in Australia.

2. LITERATURE REVIEW

Predicting cost overruns in construction projects, particularly tunnels, has been a focal point in infrastructure research due to their financial and operational risks. Studies highlight that cost overruns stem from factors such as poor planning, geological uncertainties, design changes, and inflation (Flyvbjerg, Bruzelius and Rothengatter, 2003; Frimpong, Oluwoye and Crawford, 2003; Assaf and Al-Hejji, 2006; Ullah et al., 2017; Gómez-Cabrera, Gutierrez-Bucheli and Muñoz, 2024).

Infrastructure projects that experience no cost overruns typically benefit from a combination of realistic early cost estimation, strong governance, and disciplined project execution. Studies show that projects using reference class forecasting and data-driven estimates are less affected by optimism bias and therefore set more achievable budgets (Flyvbjerg, Skamris Holm and Buhl, 2003; Flyvbjerg, 2014). Stable project scope and mature designs before construction further limit costly change orders, which are widely identified as a primary driver of overruns (Lind and Brunes, 2015). In addition, systematic risk identification, quantified risk allowances, and active risk management enable projects to absorb unforeseen events without exceeding approved budgets (Ahiaga-Dagbui and Smith, 2014). Effective procurement strategies and appropriate risk allocation under clear contractual arrangements reduce claims and disputes, particularly when competent contractors are selected through rigorous tender processes (Müller and Lecoeuvre, 2014). Finally, strong project governance, experienced project management teams, and continuous cost monitoring allow early detection of potential deviations and timely corrective action, significantly improving the likelihood of cost certainty (Odeck, 2014; Flyvbjerg, 2021).

Traditional estimation methods often underestimate these risks, leading to significant budget deviations. To address this, researchers have explored probabilistic and data-driven approaches for more accurate forecasting.

Love et al. (2013) analysed 276 Australian engineering projects and found the Fréchet distribution best modelled cost overruns, while later work from them (Love et al., 2014) identified the Generalised Logistic distribution for transportation projects. These studies emphasise that statistical distributions can quantify overrun likelihoods, aiding contingency planning. Similarly, Plebankiewicz and Edyta (Plebankiewicz, 2018) proposed a fuzzy logic model to predict overruns in building projects, achieving 53% accuracy in case studies, though its applicability to tunnels remains untested.

Machine learning and regression techniques have also gained traction. El kholy (El-Kholy, 2015) used case-based reasoning and regression to predict overruns in Egyptian projects, identifying material costs and procurement delays as key predictors. Asiedu and others (Asiedu, Frempong and Alfen, 2017) developed a multiple linear regression model for Ghanaian projects, finding that project scope and contractor classification explained 30% of cost deviations.

Despite advancements, gaps persist in tunnel-focused predictive models, particularly in Australia. This study fills this gap by analysing 27 tunnel projects to determine optimal probability distributions for different contract sizes, offering a tailored framework for risk assessment applying Dagum, GEV, Gamma. By integrating these findings, project managers can enhance cost predictability and mitigate financial uncertainties in tunnel construction.

3. RESEARCH APPROACH

The methodology for analysing cost overruns in tunnel projects begins with comprehensive data collection from completed projects. The key variables gathered include the original contract value (OCV), which represents the initial budget at project award, and the actual construction cost (ACV), reflecting the final expenditure upon completion. The cost overrun percentage is calculated by comparing these two values, providing a standardised measure of budget deviation. To ensure robust statistical analysis, the dataset must include a diverse range of tunnel projects, varying in size, complexity, and geographic location. This diversity helps capture different scenarios that influence cost overruns, from geological challenges to contractual variations, ensuring the findings are representative and reliable.

Descriptive statistics are then employed to summarise and interpret the dataset. These statistics include measures of central tendency such as the mean, median, and mode, which provide insights into typical cost overrun values. Variability is assessed through metrics like range, variance, and standard deviation, highlighting the spread of data points around the mean. Additionally, skewness and kurtosis are analysed to understand the distribution's shape. Skewness indicates whether the data is asymmetrical, with positive values suggesting a right-tailed distribution and negative values a left-tailed one. Kurtosis measures the tailedness of the distribution, with higher values indicating more outliers. These analyses help identify anomalies and patterns, forming a foundation for deeper statistical investigation (Montgomery and Runger, 2010; Ott and Longnecker, 2010).

To determine whether cost overruns differ significantly across project categories - such as size, location, or procurement method - Analysis of Variance (ANOVA) is conducted. ANOVA tests the null hypothesis that all group means are equal against the alternative hypothesis that at least one mean differs. The F-statistic, calculated as the ratio of between-group variance to within-group variance, is used to assess significance. A resulting p-value below 0.05 indicates statistically significant differences among groups, suggesting that certain project characteristics may systematically influence cost overruns. This step is crucial for identifying which factors contribute most to budgetary deviations, enabling targeted risk management strategies (Sahai and Ojeda, 2004).

The next phase involves fitting the cost overrun data to theoretical probability distributions using maximum likelihood estimation (MLE) (Myung, 2003). This technique estimates distribution parameters that best align with the observed data. Goodness-of-fit tests, including the Kolmogorov-Smirnov, Chi-Square, and Anderson-Darling tests, are then applied to evaluate how well each distribution matches the empirical data. The Kolmogorov-Smirnov test measures the maximum discrepancy between the empirical and theoretical cumulative distribution functions, while the Chi-Square test compares observed and expected frequencies across data bins. The Anderson-Darling test, more sensitive to tail behaviour, is particularly useful for detecting extreme deviations. The distribution with the highest p-value or lowest test statistic is selected as the best fit, providing a reliable model for further analysis (D'Agostino, 1986).

Once the optimal distribution is identified, its probability density function (PDF) and cumulative distribution function (CDF) are derived. The PDF describes the likelihood of specific cost overrun values, while the CDF provides the probability that a cost overrun will not exceed a given threshold. The PDF for a continuous distribution is expressed as:

$$f(x) = \frac{dF(x)}{dx}$$

where F(x) is the CDF.

The CDF for a continuous distribution is expressed as:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

These functions enable precise probability calculations, such as determining the chance of a cost overrun surpassing a critical level. This probabilistic approach allows stakeholders to quantify risks, allocate contingencies, and make informed decisions, ultimately enhancing the financial planning and execution of future tunnel projects (Thode, 2002; Klugman, Panjer and Willmot, 2012; Schervish and DeGroot, 2014).

By systematically applying these statistical techniques, the methodology offers a rigorous framework for understanding and mitigating cost overruns in tunnel construction. From initial data collection to advanced probability modelling, each step contributes to a comprehensive analysis that supports better budgeting, risk assessment, and project management. The insights gained can guide policymakers, engineers, and contractors in optimising resource allocation and minimising financial uncertainties, ensuring more successful and economically viable infrastructure development.

4. RESULTS

The study's secondary data, which covers 27 tunnel projects in Australia and is broken down by state, was gathered from several reliable sources, such as government websites, industry papers, academic journals, and official project reports. These sources provide detailed information on project costs, contract sizes, and cost overrun, ensuring a comprehensive dataset for analysis.

The data for the 27 tunnel projects, including both Original Contract Value and actual construction costs, has been gathered, and the percentage of cost overruns for each project has been calculated. Based on the calculated cost overruns, descriptive statistics such as mean, standard deviation, kurtosis, and skewness were computed using Excel software, and the results are presented in Table 1.

4.1 Descriptive Statistics

The descriptive data as shown in Table 1 reveal significant variance in the percentages of cost overruns among the projects under examination. Typically, projects exceeded their initial budgets by nearly half, with an average cost overrun percentage of 46.56%. The distribution has been severely affected by a few extreme outliers, as evidenced by the median, which is much lower at 13.21%. This is further supported by the biggest cost overrun of 327.35% and the high skewness score of 2.50, both of which depart greatly from the mean. The mode of 0% indicates that certain projects had no cost overruns, which increased the skewness of the distribution. The large range of cost overrun percentages is highlighted by the 75.96% standard deviation, which illustrates how unpredictable and variable project planning can be. The data features heavy tails and a strong peak, indicating a leptokurtic distribution, as indicated by the kurtosis value of 7.00, which highlights the existence of extreme outliers. The wide disparity between the minimum (-1.08%) and highest overruns is highlighted by the range of 327.35%.

Overall, a very diverse and skewed dataset resulted from some projects staying below budget while others had significant cost increases. This implies that there is a considerable chance of cost overruns in infrastructure projects, with certain instances being particularly serious.

Table 1. Descriptive Statistics

Descriptive Statistics for %cost overrun	
Mean	46.56
Standard Error	14.62
Median	13.21
Mode	0
Standard Deviation	75.96
Sample Variance	5765.29
Kurtosis	7.00
Skewness	2.50
Count	27

4.2 ANOVA Analysis

The data was divided into two categories according to the original contract value: one group with values less than 1 billion and the other with values greater than 1 billion. ANOVA was conducted between two groups categorised by the project's initial original contract value. The null hypothesis (H_0) assumes that the means of both groups are equal, indicating no significant difference, while the alternative hypothesis (H_1) proposes that at least one group mean is different (Sahai and Ojeda, 2004). The ANOVA analysis was performed using Microsoft Excel software to compare the cost overruns between tunnel projects costing less than 1billion(<1Bn) and those costing more than 1 billion (>1Bn). The summary statistics in Excel show that the <1Bn group has a higher average cost overrun (48.96) compared to the >1Bn group (44.64), with a significantly higher variance for the <1Bn group (8686.31) than for the >1Bn group (3880.565). This indicates greater variability in cost overruns for smaller projects. The ANOVA table as shown in Table 2 generated by Microsoft Excel reveals an F-value of 0.020706 and a p-value of 0.886735. Since the p-value is much greater than the typical significance level of 0.05, we fail to reject the null hypothesis. This suggests that there is no statistically significant difference in the mean cost overruns between the two groups.

Table 2. ANOVA Analysis

Anova: Single Factor						
SUMMARY						
Groups	**Count**	**Sum**	**Average**	**Variance**		
<1Bn	12	587.4934	48.95779	8686.313		
>1Bn	15	669.6395	44.64263	3880.565		
ANOVA						
Source of Variation	**SS**	**df**	**MS**	**F**	**P-value**	**F crit**
Between Groups	124.1368	1	124.1368	0.020706	0.886735	4.241699
Within Groups	149877.4	25	5995.094			
Total	150001.5	26				

Given the high p-value and low F-value, Excel's ANOVA analysis indicates that the cost overruns for both <1Bn and >1Bn tunnel projects likely follow similar probability functions. The differences in average cost overruns and variances observed are not statistically significant, implying that the scale of the project (below or above $1 billion) does not significantly impact the cost overrun distribution in this dataset. Excel's tools effectively facilitated this statistical comparison, providing clear insights into the data.

4.3 Best-fit Distributions

A goodness-of-fit test was carried out for 27 tunnel projects in Australia using Easy Fit software to determine the most suitable probability distributions for modelling cost overruns. The dataset was divided into three categories: all tunnel projects, projects costing less than 1 billion, and projects costing more than 1 billion. The Kolmogorov-Smirnov (K-S) Test, Anderson-Darling (A-D) Test, and Chi-Square Test were performed to assess the goodness of fit for each category. The results of these tests, including the best-fit distributions, sample sizes, and statistical values, are tabulated in Table 3.

The Maximum Likelihood Estimation (MLE) method was used to estimate the parameters of these distributions, ensuring the best possible fit to the observed data(Myung, 2003).

For all tunnel projects, the Dagum distribution was found to be the best fit with a sample size of 27. The K-S test yielded a p-value of 0.25288 and a statistic value of 0.18952, suggesting a moderate fit. However, the Anderson-Darling test statistic was 4.9649, indicating some deviation from the expected distribution, and the Chi-Square test was not applicable (NA) for this category (D'Agostino, 1986). These results from **Table 3.** suggest that while the Dagum distribution reasonably represents cost overruns across all projects, further refinement may be necessary to improve model accuracy.

Table 3. Goodness of Fitness test results

Variable	**Best fit Distribution**	**Sample size**	**Kolmogorov-Smirnov Test**		**Anderson-Darling Test**	**Chi-Square Test**
			P-Value	**statistic value**	**statistic value**	**statistic value**
All tunnel projects	Dagum Distribution	27	0.25288	0.18952	4.9649	NA
<1Bn tunnel projects	Generalised Extreme Value Distribution	12	0.58713	0.21113	0.78013	0.46463
>1Bn tunnel projects	Gamma Distribution	15	0.64961	0.18034	1.1397	1.9671

The general form of the PDF f(x) for the Dagum distribution is given by:

$$f(x) = \frac{\alpha k \left(\frac{x-\gamma}{\beta}\right)^{\alpha k - 1}}{\beta \left(1 + \left(\frac{x-\gamma}{\beta}\right)^{\alpha}\right)^{k+1}}$$

The CDF *F(x)*, which describes the probability that a cost overrun is less than or equal to a given value, for Dagam distribution is given by(Hastings, 2011):

$$F(x) = \left(1 + \left(\frac{x-\gamma}{\beta}\right)^{-\alpha}\right)^{-k}$$

Where,
k - Continuous shape parameter = 1.6457
β - Continuous scale parameter = 2.2159
α - Continuous shape parameter = 0.52902
γ - Continuous location parameter = -1.08

From Table 3, for projects costing less than 1 billion, the Generalised Extreme Value (GEV) distribution was identified as the best fit with a sample size of 12. The K-S test p-value was 0.58713, and the statistic value was 0.21113, indicating a better fit than the Dagum distribution for all tunnel projects. The Anderson-Darling test statistic was 0.78013, and the Chi-Square test statistic was 0.46463, both of which suggest that the GEV distribution effectively represents cost overruns in this category. The relatively low statistic values indicate that the GEV distribution captures both the central tendency and tail behaviour of cost overruns in smaller tunnel projects.

The PDF of the GEV distribution is:

$$f(x) = \begin{cases} \frac{1}{\sigma} exp\left(-(1+kz)^{-1/k}\right) \cdot \left(1 + k_{z'}^{-1-1}\right)k, & for\ k \neq 0 \\ \frac{1}{\sigma} exp(-z - exp(-z)), & for\ k = 0 \end{cases}$$

The CDF of the GEV distribution is(Hastings, 2011):

$$F(x) = \begin{cases} exp\left(-(1+kz)^{-1/k}\right), & for\ k \neq 0 \\ exp(-exp(-z)), & for\ k = 0 \end{cases}$$

Where,
k - continuous shape parameter = 0.67239
σ - continuous scale parameter = 16.586
μ - continuous location parameter = 6.3377
$\mathrm{z} = \frac{\mathrm{x} - \mu}{\sigma}$

From Table 3, For projects costing more than 1 billion, the Gamma distribution was found to be the most suitable, with a sample size of 15. The K-S test p-value was 0.64961, which was the highest among the three categories, and the statistic value was 0.18034, indicating an excellent fit. The Anderson-Darling test statistic was 1.1397, while the Chi-Square test statistic was 1.9671, both suggesting a reasonable approximation of cost overrun distributions for large tunnel projects. These results confirm that the Gamma distribution is well-suited for modelling cost overruns in high-cost projects, as it effectively captures the variability and distribution pattern.
The PDF for the Gamma distribution is:

$$f(x) = \frac{(x-\gamma)^{\alpha}}{\beta^{\alpha}\Gamma(\alpha)} - (exp(-(x-\gamma)\ /\ \beta)$$

The corresponding CDF is(Hastings, 2011):

$$F(x) = \frac{\Gamma_{(x-\gamma)/\beta}(\alpha)}{\Gamma(\alpha)}$$

Where,
α - continuous shape factor = 0.36091
β - continuous scale parameter = 98.601
γ - continuous location parameter = -1.08
Γ - Gamma function

4.4 Cost Overrun Probabilities

The Table 4 presents statistical probabilities related to cost overruns in tunnel projects, categorised based on overall projects, projects costing less than 1 billion (<1Bn), and those exceeding 1 billion (>1Bn). The best-fit distributions Dagum, Generalised Extreme Value, and Gamma distributions model the probability of cost overruns for these categories.

The probability values P(X<X1) represents the likelihood that the cost overrun percentage is below a specified threshold, while P(X>X1) indicates the probability of exceeding that threshold. Additionally, the table provides P(X1<X<X2), which shows the probability of cost overruns falling within a specified range (Klugman, Panjer and Willmot, 2012; Schervish and DeGroot, 2014).

For example, for all tunnel projects, the probability that the cost overrun is below 5% is 22.656%, while the probability of exceeding this threshold is 77.344%. Similarly, for projects under 1 billion, the probability of an overrun under 5% is 33.741%, and for those over 1 billion, it is 40.448%. These values indicate that cost overruns are a common occurrence, with higher probabilities for exceeding lower thresholds, highlighting potential financial risks in tunnel projects.

Table 4. Probabilities

Variable	**Best fit Distribution**	**Sample size**	**probability for cost overrun %**	**P(X<X1)**	**P(X>X1)**	**probabilities of cost overrun % between**	**P(X1<X<X2)**
All tunnel projects	Dagum Distribution	27	5%	0.22656	0.77344	1% and 5%	0.15716
			10%	0.55678	0.44322	5% and 10%	0.08919
			20%	0.64601	0.35399	10% and 20%	0.8923
			30%	0.69483	0.30517	20% and 30%	0.04882
			40%	0.72713	0.27287	30% and 40%	0.0323
			50%	0.75066	0.24934	40% and 50%	0.02353
<1Bn tunnel projects	Generalised Extreme Value Distribution	12	5%	0.33741	0.66259	1% and 5%	0.0998
			10%	0.44311	0.55689	5% and 10%	0.1057
			20%	0.59501	0.40499	10% and 20%	0.15189
			30%	0.69227	0.30773	20% and 30%	0.09726
			40%	0.75726	0.24274	30% and 40%	0.06499
			50%	0.80273	0.19727	40% and 50%	0.04546
>1Bn tunnel projects	Gamma Distribution	15	5%	0.40442	0.59558	1% and 5%	0.12688
			10%	0.49569	0.50431	5% and 10%	0.09128
			20%	0.60943	0.39057	10% and 20%	0.11374
			30%	0.68395	0.31605	20% and 30%	0.07452
			40%	0.73846	0.26154	30% and 40%	0.05451
			50%	0.78052	0.21948	40% and 50%	0.04206

5. DISCUSSIONS

The descriptive statistics presented in section 4.1 reveal substantial uncertainty and asymmetry in cost overruns for Australian tunnel projects. Although the mean overrun is high at 46.56%, the much lower median of 13.21% indicates that a small number of extreme cases disproportionately influence the average. This is further supported by the strong positive skewness (2.50) and high kurtosis (7.00), confirming a right-skewed, heavy-tailed distribution with significant outliers. The large standard deviation (75.96%) reflects considerable variability in cost performance across projects. Collectively, these characteristics demonstrate that reliance on average values alone can be misleading and reinforce the need for probabilistic, distribution-based modelling to adequately capture extreme cost risk in tunnel construction.

The ANOVA results presented in section 4.2 represent a critical finding of this study. The analysis shows no statistically significant difference in the mean cost overrun between tunnel projects valued below and above AUD 1 billion (p-value = 0.8867). This non-significant p-value indicates that, on average, project size does not materially influence the magnitude of cost overruns. From a mean-based perspective, smaller and larger tunnel projects therefore exhibit comparable budgetary performance, despite differences in scale and complexity.

However, this result should be interpreted alongside the distribution-fitting outcomes, which reveal a more nuanced picture of cost overrun behaviour. While the mean overrun is similar across project sizes, the best-fit probability distributions differ substantially. Projects below AUD 1 billion are best modelled by the Generalised Extreme Value (GEV) distribution, whereas projects exceeding AUD 1 billion follow a Gamma distribution. This contrast strongly suggests that the underlying risk structure differs by project size, even when average outcomes are statistically indistinguishable.

The GEV distribution identified for smaller projects reflects heavier tail behaviour and greater exposure to extreme cost overruns, consistent with the higher variance observed in this category. This may be attributable to weaker governance structures, reduced contingencies, and greater sensitivity to localised uncertainties. Conversely, the Gamma distribution associated with larger projects suggests a more stable and cumulative cost overrun process, reflecting stronger planning frameworks and formalised risk controls typical of megaproject environments.

This divergence between mean equivalence and distributional heterogeneity reinforces the need for size-specific probabilistic modelling. Relying solely on average overruns risks masking critical tail risks, particularly in smaller projects. Accordingly, these findings underpin the study's conclusion that effective contingency planning in tunnel projects should be informed by distributional characteristics rather than mean performance alone.

Examining specific cost overrun ranges, Table 4 provides further insights into the distribution of overruns. For example, in all tunnel projects, the probability of overruns between 10% and 20% is 89.23%, suggesting that moderate overruns are quite frequent. Similarly, for smaller projects (<1Bn), the likelihood of a cost overrun between 20% and 30% is 9.726%, whereas for larger projects (>1Bn), the probability for the same range is 7.452%. These values show that while overruns are prevalent across all tunnel projects, their extent and likelihood vary with project size.

The data in Fig. 1 and Fig. 2 suggest that larger tunnel projects (over 1 Bn) have a higher probability of lower cost overruns compared to smaller projects (under 1 Bn). This could be due to more rigorous planning, better resource allocation, and more experienced management in larger projects. However, all projects exhibit a significant risk of cost overruns, emphasising the need for effective risk management strategies and contingency planning in tunnel construction in Australia.

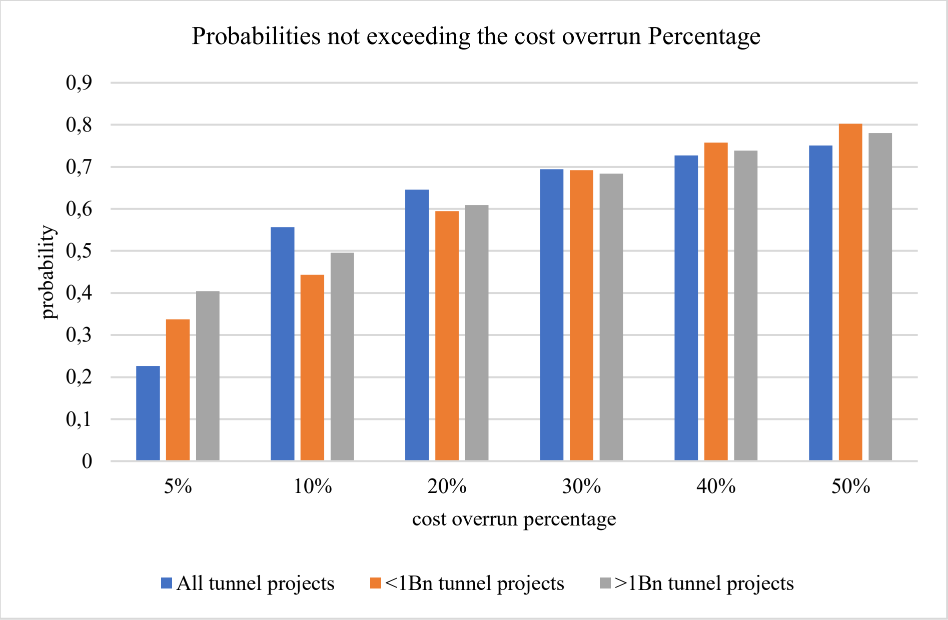

Fig. 1. Cost overrun percentage vs probability

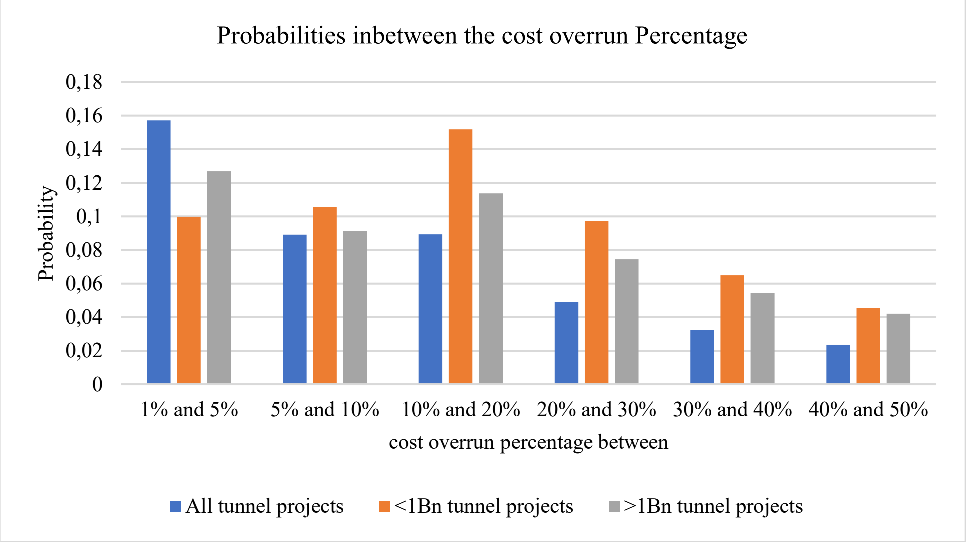

Fig. 2. Cost overrun between percentages vs Probability

Both Love and others (Love *et al.*, 2013, 2014) and the present study underscore that cost overruns in infrastructure projects are systemic and best explained through probabilistic modelling using heavy-tailed statistical distributions. While Love's research examined a broad sample of Australian engineering and transportation projects and identified the Fréchet and Generalised Logistic distributions as the best fit, this study narrows the focus to Australian tunnel projects, a sector largely overlooked in prior research. By analysing 27 tunnel projects, it was found that cost overruns are more severe, averaging 46.6% with extreme variability, compared to the more moderate ranges reported by Love. Moreover, while Love did not emphasise project size, this study applied ANOVA and demonstrated no significant difference in mean overruns between projects above and below AUD 1 billion, though smaller projects exhibited greater variability. The best-fit distributions also differed: Dagum for all tunnel projects, Generalised Extreme Value for smaller projects, and Gamma for larger projects. These findings extend Love's general framework by offering tunnel-specific probability estimates, such as a 77.34% chance of exceeding a 5% overrun, thereby providing infrastructure stakeholders with more tailored tools for contingency planning and risk management in tunnel construction (Love *et al.*, 2013, 2014).

This study strengthens theoretical understanding of cost overruns in tunnel projects by demonstrating that cost risk is best interpreted through probabilistic and distribution-based analysis rather than mean-based comparison alone. Although the ANOVA results show no statistically significant difference in mean cost overruns between projects below and above AUD 1 billion, the strong positive skewness, high kurtosis, and heavy-tailed behaviour observed in the descriptive statistics indicate that cost overruns in tunnel construction are inherently asymmetric risk events. The identification of different best-fit distributions demonstrates that project size influences the underlying structure of cost risk rather than its average magnitude. This finding supports the theoretical argument that relying on mean values can obscure critical tail risks and reinforces the application of distribution-driven and extreme value frameworks in infrastructure cost overrun research. From a practical perspective, the findings highlight the limitations of traditional deterministic cost estimates and fixed contingency allowances in tunnel projects. The high likelihood of cost overruns, including a 77.34% probability of exceeding a 5% overrun threshold, underscores the need for probabilistic contingency planning. Smaller tunnel projects, which exhibit greater variability and heavier tail behaviour, require more conservative contingencies and strengthened early-stage risk controls despite their lower capital value. In contrast, larger projects display more stable accumulation of cost overruns, supporting structured and staged risk management approaches. The probability distributions and cumulative risk estimates developed in this study provide stakeholders with a transparent, data-driven basis for budgeting, business case evaluation, and financial risk management in Australian tunnel projects.

6. CONCLUSION

This study presented a detailed statistical investigation of cost overruns in Australian tunnel projects, addressing a critical gap in infrastructure research by focusing on tunnel-specific cost behaviour through probabilistic modelling. Using historical data from 27 completed tunnel projects across multiple Australian states, the research examined the statistical characteristics of cost overruns and assessed the influence of contract size on cost performance using descriptive statistics, ANOVA, and distribution fitting techniques.

The findings confirm that cost overruns are pervasive in Australian tunnel construction, with an average overrun of approximately 46.6% and a strongly right-skewed, heavy-tailed distribution. The Dagum distribution was identified as the best fit for the combined dataset, while project size segmentation revealed that smaller projects (< AUD 1 billion) are best modelled using a Generalised Extreme Value distribution and larger projects (> AUD 1 billion) by a Gamma distribution. Although ANOVA results indicated no statistically significant difference in mean cost overruns between project size categories, smaller projects exhibited noticeably greater variability, highlighting distinct underlying risk structures that are not captured by mean-based analysis alone. This reinforces the importance of distribution-based risk assessment for effective contingency planning.

It is acknowledged that the dataset of 27 tunnel projects, while reasonable given the rarity, scale, and confidentiality constraints associated with tunnel construction in Australia, represents a limitation of this study. The relatively small sample size reduces the statistical power of the ANOVA test, increasing the likelihood that subtle differences between groups may remain undetected. Consequently, the non-significant ANOVA findings should be interpreted with appropriate caution. In addition, the limited sample constrains the broader generalisability of the results to all future tunnel projects, particularly those delivered under different procurement models, geological conditions, or governance frameworks. Nonetheless, the dataset is sufficient for probabilistic distribution fitting and exploratory risk analysis, and the results provide valuable initial insights into the statistical behaviour of cost overruns in Australian tunnel projects. Future research incorporating larger datasets or international comparisons is recommended to validate and extend these findings.

REFERENCES

Ahiaga-Dagbui, D. D., & Smith, S. D. (2014). Rethinking construction cost overruns: Cognition, learning and estimation. *Journal of Financial Management of Property and Construction, 19*(1), 38–54. https://doi.org/10.1108/JFMPC-06-2013-0027

Altshuler, A. A., & Luberoff, D. E. (2004). *Mega-projects: The changing politics of urban public investment*. Rowman & Littlefield.

Asiedu, R. O., Frempong, N. K., & Alfen, H. W. (2017). Predicting likelihood of cost overrun in educational projects. *Engineering, Construction and Architectural Management, 24*(1), 21–39.

Assaf, S. A., & Al-Hejji, S. (2006). Causes of delay in large construction projects. *International Journal of Project Management, 24*(4), 349–357.

D'Agostino, R. B. (1986). *Goodness-of-fit techniques*. CRC Press.

El-Kholy, A. M. (2015). Predicting cost overrun in construction projects. *International Journal of Construction Engineering and Management, 4*(4), 95–105.

Flyvbjerg, B. (2010). Policy and planning for large-infrastructure projects: Problems, causes, and curses. In *Dialogues in Urban and Regional Planning* (pp. 243–268). Routledge.

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

Flyvbjerg, B. (2021). Top ten behavioral biases in project management: An overview. *Project Management Journal, 52*(6), 531–546.

Flyvbjerg, B., Bruzelius, N., & Rothengatter, W. (2003). *Megaprojects and risk: An anatomy of ambition*. Cambridge University Press.

Flyvbjerg, B., Skamris Holm, M. K., & Buhl, S. L. (2003). How common and how large are cost overruns in transport infrastructure projects? *Transport Reviews, 23*(1), 71–88.

Frimpong, Y., Oluwoye, J., & Crawford, L. (2003). Causes of delay and cost overruns in construction of groundwater projects in developing countries: Ghana as a case study. *International Journal of Project Management, 21*(5), 321–326.

Global Infrastructure Hub. (2021). *Global practices for cross-border infrastructure projects*.

Gómez-Cabrera, A., Gutierrez-Bucheli, L., & Muñoz, S. (2024). Causes of time and cost overruns in construction projects: A scoping review. *International Journal of Construction Management, 24*(10), 1107–1125.

Hastings, N. (2011). *Statistical distributions* (C. Forbes, M. Evans, N. Hastings, & B. Peacock). John Wiley & Sons, Inc.

Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2012). *Loss models: From data to decisions*. John Wiley & Sons.

Lind, H., & Brunes, F. (2015). Explaining cost overruns in infrastructure projects: A new framework with applications to Sweden. *Construction Management and Economics, 33*(7), 554–568.

Love, P. E. D., et al. (2013). Determining the probability of project cost overruns. *Journal of Construction Engineering and Management, 139*(3), 321–330.

Love, P. E. D., et al. (2014). Overruns in transportation infrastructure projects. *Structure and Infrastructure Engineering, 10*(2), 141–159.

Montgomery, D. C., & Runger, G. C. (2010). *Applied statistics and probability for engineers*. John Wiley & Sons.

Müller, R., & Lecoeuvre, L. (2014). Operationalising governance categories of projects. *International Journal of Project Management, 32*(8), 1346–1357.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology, 47*(1), 90–100.

Odeck, J. (2014). Do reforms reduce the magnitudes of cost overruns in road projects? Statistical evidence from Norway. *Transportation Research Part A: Policy and Practice, 65*, 68–79.

Ott, R. L., & Longnecker, M. (2010). *An introduction to statistical methods and data analysis*. Cengage Learning Inc.

Plebankiewicz, E. (2018). Model of predicting cost overrun in construction projects. *Sustainability, 10*(12), 4387.

Sahai, H., & Ojeda, M. M. (2004). *Analysis of variance for random models*. Birkhäuser Boston.

Schervish, M. J., & DeGroot, M. H. (2014). *Probability and statistics*. Pearson Education.

Sydney Metro. (2016). *City & southwest - Final business case summary*.

Thiess Hochtief. (n.d.). Epping to Chatswood rail line. Retrieved from http://www.awedwards.com.au

Thode, H. (2002). Testing for normality. In *Marcel Dekker, Inc.* (pp. 99–123). New York.

Ullah, K., et al. (2017). Theoretical framework of the causes of construction time and cost overruns. In *IOP Conference Series: Materials Science and Engineering* (p. 012032). IOP Publishing.

Vickerman, R. (1997). High-speed rail in Europe: Experience and issues for future development. *The Annals of Regional Science, 31*, 21–38.