Understanding the Effects of Correlation in Modelling Construction Cost Uncertainty

by

Jace David Thibault

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Construction Engineering and Management

Department of Civil and Environmental Engineering
University of Alberta

Abstract

Estimating costs for construction projects is a complex and often uncertain process. Given the inherent uncertainty of estimating and the unique risk profile of each construction project, many owners use cost uncertainty analysis to understand a project's range of potential costs. Since results of the analysis can determine whether a project proceeds or how a project's budget is set, it is essential that model results represent the actual range of costs that exists for a project. Various studies of actual and estimated costs have found cost underestimation, which occurs when estimates fail to adequately anticipate potential cost overruns, to be a persistent issue. While the causes of cost underestimation are numerous, this thesis focuses on two contributing methodology deficiencies: the inaccurate representation of uncertainty in model inputs and failing to account for correlation between model inputs. To address the former deficiency, the research examines elicitation biases, explores leading elicitation protocols, and demonstrates the application of the Sheffield elicitation framework (SHELF) protocol to cost uncertainty analysis for a construction project. To address the latter deficiency, this research proposes a Monte Carlo simulation experiment to test the hypothesis that since cost uncertainty (measured by standard deviation) is affected by correlation between model inputs, modelling correlated inputs independently results in underestimation of cost uncertainty. The experiment involves generating correlated cost item data, collecting statistics and generating independent samples, applying three correlation modelling methods, and evaluating those methods relative to the correlated data under a variety of scenarios. Experimental results found that independent modelling yielded an underestimation of cost uncertainty in most cases and that correlation modelling methods generated less error on average. A case study is also presented for the application of correlation modelling methods to cost uncertainty analysis of a light rail transit (LRT) project.

Acknowledgments

I am extremely grateful to my supervisors, Dr. Simaan AbouRizk and Dr. Yasser Mohamed, for their guidance and support throughout this process. Dr. AbouRizk, thank you for all that you have taught me these past few years, both inside and outside of the classroom. I truly appreciate your generosity and mentorship. Dr. Mohamed, thank you for your wisdom and reassurance.

I am also grateful to the team at SMA (especially to Holly, Matt, Steven, Jesse, Mohammed, and Hala) for granting me the flexibility to pursue my studies, providing me with the knowledge and context I needed to understand my research, and offering help along the way.

I could not have done this without the support of my friends and family. To my incredible husband, Jordon, for his love and support throughout my journey as a student (which has spanned the entirety of our relationship thus far). Your patience and encouragement have meant the world to me. To my parents for nurturing my curiosity and confidence. Thank you for always believing in me.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Explanation
AACE	Association for the Advancement of Cost Engineering
CDF	Cumulative distribution function
CLT	Central limit theorem
CM	Classical model
CPM	Critical path method
CS	Cost risk
CV	Coefficient of variation
DND	Department of National Defence
EW	Equal weighting
FMCS	Fuzzy Monte Carlo simulation
GOC	Government of Canada
KITC	Knee-in-the-curve
KL	Kullback-Leibler
LRT	Light rail transit
LRV	Light rail vehicle
MCMC	Markov chain Monte Carlo
MOLE	More-or-less elicitation
MSAT	Multiple simulation analysis technique
NASA	National Aeronautics and Space Administration
NGT	Nominal group technique
OMSF	Operations, maintenance, and storage facility
PE	Point estimate
PS	Protect scenario
QOI	Quantity (or quantities) of interest
QRA	Quantitative risk analysis
RAIC	Royal Architectural Institute of Canada
RBS	Risk breakdown structure
RIO	Rational impartial observer
SEM	Standard error of the mean
SHELF	Sheffield elicitation framework
SSQ	Sum of squared differences
WBS	Work breakdown structure

Chapter 1 – Introduction

1.1 Background and Problem Statement

Estimating costs for construction projects is a complex and often uncertain process. Owners rely on cost estimates to assess the viability of projects, make decisions on whether to advance them, and establish budgets and set contingencies. Given the inherent uncertainty of estimating and the unique risk profile of each construction project, many owners use cost uncertainty analysis to understand a project's range of potential costs. Cost uncertainty analysis, also referred to as cost risk or probabilistic cost estimating, can be performed using a variety of approaches including analytics and simulation. Simulation-based cost uncertainty analysis is one of the most commonly used forms of Monte Carlo simulation in construction practice (Shaheen et al. 2007). The popularity of simulation-based cost uncertainty analysis is largely attributed to the availability of commercially available risk analysis software packages and the insights afforded to decision makers by examining risk and uncertainty.

Probabilistic cost estimating refers to a subset of cost uncertainty analysis which relies on the use of probability distributions. Probabilistic cost estimating involves (1) determining cost drivers and identifying risks, (2) establishing probability distributions for uncertain inputs, (3) accounting for correlation between inputs, (4) calculating the project cost, and (5) analyzing results to assess potential variability in project cost. For the simulation-based approach, the fourth activity involves sampling uncertain variables, such as the cost for each cost item in the project, and combining values using mathematical relationships to estimate the overall cost for the project.

Using the outputs from the analysis, decision makers can assess the potential range of project costs associated with estimate variability and risk, and determine whether a project should proceed. If a project is to proceed, decision makers set the budget for the project by establishing a level of confidence, typically expressed as a percentile of possible results, based on their tolerance for uncertainty and risk. For example, setting a project budget at the 80th percentile value for simulated project cost would mean that 80% of observations would be at or below the budget set. Conversely, 20% of observations would be above the budget set, representing a 20% chance of cost overrun for the project.

Since results can inform whether a project proceeds or how a project's budget is set, it is essential that model results represent the actual range of costs that exists for a project. Cost underestimation occurs when estimates fail to adequately anticipate potential cost overruns. Various studies of actual and estimated costs have found underestimation to be a persistent issue (Hollmann 2012). The misrepresentation of costs is likely to lead to the misallocation of scarce resources which yields a poor outcome for those that finance and use infrastructure such as taxpayers (Flyvbjerg et al. 2002). While the causes of underestimation are numerous, systems engineering literature points to methodology deficiencies as a major source of cost underestimation (Garvey et al. 2016). This thesis focuses on two methodology deficiencies causing cost underestimation in particular: (1) the inaccurate representation of uncertainty in model inputs and (2) failing to account for correlation between model inputs.

Probability distributions are often established subjectively using expert opinion, i.e., through elicitation, and used to represent the uncertainty in costs for each cost item. While the elicitation process is intended to be informed by experience and knowledge of the project as well as an understanding of the factors affecting cost uncertainty, it remains subjective and is therefore susceptible to various biases. Since the elicited probability distributions represent the uncertainty of costs for the cost items and are the basis of probabilistic cost estimating, inaccuracies can lead to misstatement of the possible range of results. In many different fields, protocols have been developed to reduce biases during the elicitation of subjective probability distributions and therefore improve the quality of model inputs, however such a protocol has not been developed specific to cost uncertainty analysis in construction practice. To investigate the representation of uncertainty in model inputs, this research examines elicitation biases, explores leading elicitation protocols, and demonstrates the application of such a protocol to cost uncertainty analysis for a construction project.

Applying classical statistics, a project's cost uncertainty is derived from the cost uncertainty of its components. Specifically, the variance of a project's cost is the sum of the variance of its cost items plus the sum over all covariation between each cost item pair. Correlation between cost items affects the extent of cost uncertainty for a project's cost because covariance is a function of correlation. Thus, correlation is a critical consideration for analyzing cost uncertainty as ignoring it "can be equivalent to setting its value to zero, which can significantly underestimate a program's total cost risk" (Garvey et al. 2016). This research proposes a Monte Carlo simulation experiment to test the hypothesis that since cost uncertainty (measured by standard deviation) is affected by correlation between model inputs, modelling correlated inputs independently results in underestimation of cost uncertainty. The experiment involves the application of three methods for modelling correlation: the independence method, the knee-in-the-curve method, and the random assessment method. Each iteration of the model generates a population of costs which are used for comparison against sample costs produced by the different correlation modelling methods. The outcomes of the experiment are used to understand the strengths and weaknesses of each correlation modelling method in a variety of scenarios and provide recommendations for modelling correlation in cost uncertainty analysis.

This research also presents a case study for the application of correlation modelling methods to cost uncertainty analysis of a light rail transit project. The case study applies correlation modelling methods from the Monte Carlo simulation experiment and assesses sensitivity to different values of correlation in establishing or validating a project's budget.

1.2 Purpose of Study

The main objectives of this study are to:

- understand potential biases in elicitation and the use of an elicitation protocol to reduce such biases;
- examine the effects of correlation on project costs and assess various methods for modelling correlation;
- provide a set of recommendations for modelling correlation in cost uncertainty analysis; and,
- demonstrate the application of correlation modelling methods to an actual construction project.

1.3 Expected Contributions

Academic contributions from this study include:

- developing an understanding of the impact of the number of cost items and extent of correlation on the range of project costs produced by cost uncertainty analysis; and,

- developing a methodology for analysts to assess the ability of different correlation modelling methods to estimate cost uncertainty in an unbiased and reproducible model that can be expanded by creating additional scenarios (e.g., manipulating other model parameters).

Industrial contributions from this study include:

- demonstrating the application of an elicitation protocol to cost uncertainty analysis in construction practice; and,
- providing analysts with a guide for valuing and capturing correlation between cost items in cost uncertainty analysis.

1.4 Research Methodology

The research methodology for this thesis is as follows:

- Conduct a literature review to identify (1) the various forms of cost uncertainty analysis developed by the cost analysis community, (2) potential biases in elicitation and elicitation protocols used in other fields to reduce such biases, and (3) methods of valuing and assigning correlation.
- Using an illustrative example, apply an elicitation protocol to cost uncertainty analysis for a construction project.
- Design a Monte Carlo simulation experiment to test the hypothesis that since cost uncertainty (measured by standard deviation) is affected by correlation between model inputs, modelling correlated inputs independently results in underestimation of cost uncertainty.
- As part of the experiment, generate a population of correlated costs and apply different correlation modelling methods to independently generated cost samples.
- Repeat the experiment under a variety of model scenarios and evaluate the performance of correlation modelling methods by comparing results from each method with the corresponding population of costs.
- Analyze the outcomes of each model scenario and provide recommendations for correlation modelling based on the analysis.

- Apply correlation modelling methods from the Monte Carlo experiment to an actual project and assess the impact of correlation on results of the analysis.

1.5 Organization of Thesis

This thesis consists of six chapters organized as follows: **Chapter 2** includes a review of the literature relating to cost estimation and cost uncertainty analysis. **Chapter 3** presents a summary of potential elicitation biases as well as the application of an elicitation protocol for establishing subjective probability distributions to probabilistic cost estimating in construction engineering. **Chapter 4** presents a Monte Carlo simulation experiment including the process of generating cost item data, collecting statistics and generating samples, applying correlation methods, and evaluating each method relative to the generated data. **Chapter 5** documents a case study for the application of correlation modelling methods to cost uncertainty analysis of a light rail transit project. **Chapter 6** summarizes the research, provides recommendations for modelling correlation, and highlights topics for further research.

Chapter 2 – Literature Review

2.1 Introduction

By addressing the uncertainty and risk inherent in estimating and executing capital construction projects, cost uncertainty analysis allows owners to make more informed decisions. Given the benefits of viewing project budgets in terms of probability and risk, probabilistic estimating techniques have been recommended by the Committee on Budget Estimating Techniques (National Research Council 1990). The range of project costs generated by cost uncertainty analysis is intended to communicate the robustness of the project estimate and schedule, the degree of uncertainty and risk in the project based on current information, and the budget amount required to achieve a specific probability of not being overrun. In addition to understanding the range of project costs, the cost uncertainty analysis allows mitigation measures to be identified. For example, cost uncertainty analysis can establish which aspects of a project have the most uncertainty so that efforts can be directed to achieving more certainty. This chapter introduces cost estimate classification, the factors that affect estimate accuracy, and the issue of cost underestimation in construction projects. Next, processes for cost uncertainty analysis are discussed. This chapter concludes with a summary of the literature and a description of limitations in the literature.

2.2 Definitions

This section includes definitions for risk, uncertainty, and cost uncertainty analysis which provide context for the use of these terms throughout this thesis.

2.2.1 Definition of Risk

There are many different meanings for the term 'risk' but the following definitions are useful in the context of cost uncertainty analysis:

- AACE International (2008) defines risk as "an undesirable potential outcome and/or its probability of occurrence".
- The Project Management Institute (2017) defines risk as "an uncertain event or condition that, if it occurs, has a positive or negative effect on one or more project objectives".
- Garvey et al. (2012) define risk as "exposure to loss" and state that risk is both probability and consequence.

Moussa (2013) presents several definitions of risk which emphasize the concept of probability as well as positive and negative impacts associated with risk. While definitions in the literature vary as to whether the term ‘risk’ contemplates desirable outcomes or positive effects, they tend to agree that (1) a risk may or may not occur and (2) if a risk does occur that there is a consequence.

2.2.2 Definition of Uncertainty

Uncertainty can be defined in broad terms as “knowledge incompleteness due to inherent deficiencies with acquired knowledge” (Ayyub 2003). There are various types and sources of uncertainty which have been described in the literature (Ayyub 2003; Morgan and Henrion 1990); uncertainty may arise from incomplete information, inconsistency, linguistic imprecision, variability, simplifications, approximations, and more. In the context of risk analysis, AACE International (2008) defines uncertainty as the sum of risks and opportunities, where opportunities are similar to risks except outcomes are desirable rather than undesirable. Garvey et al. (2012) describe uncertainty as “the indefiniteness or variability of an event” whether favourable or unfavourable. In this thesis, the term ‘uncertainty’ is used to refer to the aspects of a project (particularly numerical values) that lack certainty and can thus be described using a range. Examples of such aspects include the accuracy of a cost estimate, future escalation rates, and pessimistic versus optimistic activity durations.

2.2.3 Definition of Cost Uncertainty Analysis

Cost uncertainty analysis is the process of identifying, measuring, and interpreting how uncertainty affects a system’s cost (Garvey et al. 2016). The term ‘uncertainty’ in this definition is to be interpreted in the broad sense and thus refers to both risk and uncertainty as defined in this thesis. Terminology for ‘cost uncertainty analysis’ in the literature varies; for example, the National Aeronautics and Space Administration, or NASA (2008), refers to cost uncertainty analysis as a ‘cost risk assessment’. The term ‘probabilistic cost estimating’ is commonly used in the literature (Isidore and Back 2002; Touran 1993; Diekmann 1983) and refers to cost uncertainty analysis which relies on the use of probability distributions. As shown in Figure 1, cost uncertainty analysis can be performed using a variety of approaches including analytics and simulation. In this thesis, the term ‘cost uncertainty analysis’ is used broadly to refer to analytical and simulation-based approaches (described in Section 2.4.1) as well as alternative methods (described in Section 2.4.2) unless otherwise stated.

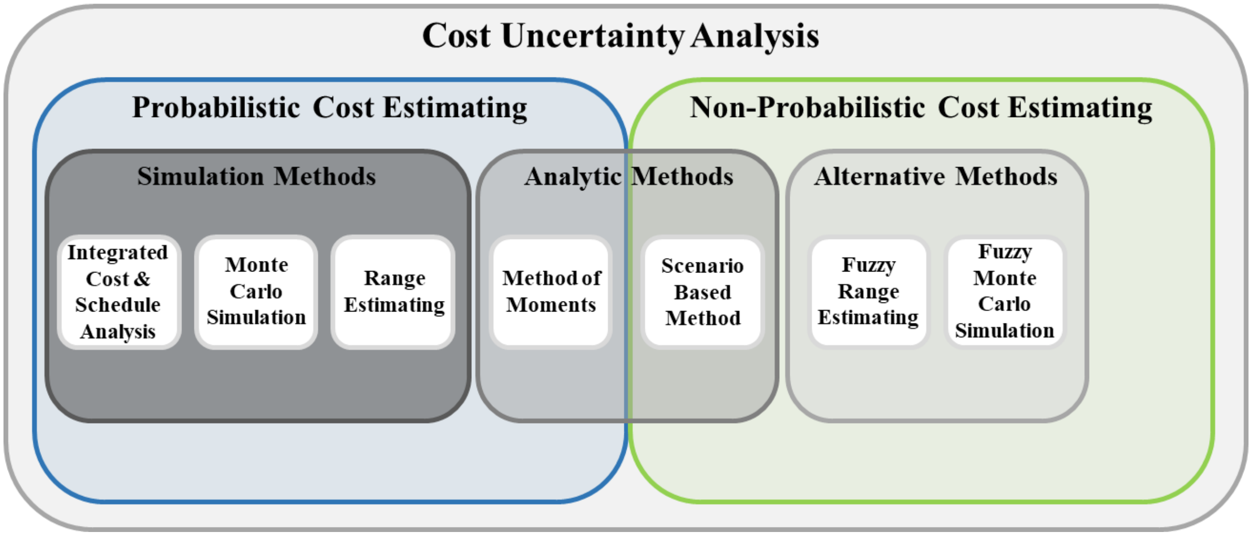


Figure 1: Conceptual representation of various cost uncertainty analysis categories and methods.

Range estimating is a particular form of simulation-based cost uncertainty analysis that "combines Monte Carlo sampling, a focus on the few critical items, and heuristics (rule of thumb) to rank critical risks and opportunities" (AACE International 2008). Michael Curran, considered one of the most influential pioneers of range estimating, similarly defines range estimating as a "decision technology which synergistically combines Pareto analysis, heuristics and Monte Carlo Simulation" (Curran 1990). Range estimating is considered "one of the most commonly used forms of Monte Carlo simulation in construction practice" (Shaheen et al. 2007). The term 'range estimating' is used in a variety of contexts including as a part of the risk assessment and analysis process (W. Willson, Making realistic and understandable Infrastructure Commitments through risk based estimating, presented at 2009 CSVA Conference, 2009). The National Research Council (1990) describes range estimating as a "technique that permits an estimator to quantify his confidence/uncertainty about an estimate". As the definitions above indicate, some consider range estimating to be a form of Monte Carlo risk analysis, while others consider it to be a process concerning estimate uncertainty, i.e., probabilistic cost estimating. In this thesis, the term 'range estimating' is used to refer to a Monte Carlo simulation in which probability distributions are assigned exclusively to critical items and the output is a distribution of a project's costs caused by uncertainty and risk.

There are significant differences between range estimating and other forms of simulation-based cost uncertainty analysis which will be addressed throughout this section. Despite the differences between range estimating and other simulation-based methods, the National Research Council

(1990) describes several common features among probabilistic models: “(1) they require that an estimate be made of the potential variability of each element in an estimate; (2) they employ the laws of probability to determine the impact of possible variations in the cost of individual elements on overall costs; (3) they require the use of a computer; and (4) they present the results in the form of a histogram or a cumulative distribution showing either the probability of various estimates proving to be the actual cost of the project, or the probability of cost overruns of various magnitudes”.

2.3 Cost Estimating

Cost estimating can be described as the approximation of monetary resources needed to complete project work (Project Management Institute 2017). While different owners use different terminology and processes for construction cost estimating, most construction projects require the preparation of numerous and increasingly detailed cost estimates throughout the project lifecycle (National Research Council 1990).

2.3.1 Cost Estimate Classification Systems

Cost estimate classification systems are used to categorize cost estimates by one or more characteristics including project maturity level, end usage, and expected accuracy. The Joint Federal Government / Industry Cost Predictability Taskforce (Guide to Cost Predictability in Construction: An Analysis of Issues Affecting the Accuracy of Construction Cost Estimates, unpublished report, 2012) compared cost estimate classification systems developed by the Association for the Advancement of Cost Engineering (AACE), the Department of National Defence (DND), the Royal Architectural Institute of Canada (RAIC), and the Government of Canada (GOC) as shown in Figure 2.

COST ESTIMATE CLASSIFICATION SYSTEMS					
AACE	**Class 4**	**Class 3**		**Class 1**	**Class 1**
DND		Indicative		Substantive	
RAIC	Sketch Design	Design Development		Construction Documents	Tender Documents
GOC	D	C	B		A
Design Documentation % Complete	**12.5%**	**25.0%**		**95.0%**	**100.0%**
Cost Estimate Accuracy (+/- %)	**30.0%**	**25.0%**		**15.0%**	**10.0%**

Figure 2: Cost Estimate Classification Chart adapted from the Joint Federal Government / Industry Cost Predictability Taskforce (Guide to Cost Predictability in Construction: An Analysis of Issues Affecting the Accuracy of Construction Cost Estimates, unpublished report, 2012).

In ABCD estimate classification, a Class A estimate is the closest to full project definition and maturity and a Class D estimate has the lowest maturity level of project definition.

Cost estimate classification is particularly useful in the context of cost uncertainty analysis as it provides a preliminary indication of the expected accuracy (or range) of an overall cost estimate. AACE International (2020) cautions that while estimate accuracy is generally correlated with estimate classification, the accuracy of any given estimate is not fixed or determined by its classification category. Significant variations in estimate accuracy are possible if the factors affecting accuracy vary which is why estimate accuracy must be evaluated "on an estimate-by-estimate basis in conjunction with some form of risk analysis process" (AACE International 2020).

2.3.2 Factors Affecting Cost Estimate Accuracy

While the primary driver of cost estimate accuracy for most estimate classification systems is the percentage of design definition, many factors affect the accuracy of cost estimates. AACE International (2020) states that in addition to the percentage of project definition and project-specific risks, estimate accuracy is also driven by systemic risks such as:

- "Level of non-familiar technology in the project;
- Complexity of the project;
- Quality of reference cost estimating data;

- Quality of assumptions used in preparing the estimate;
- Experience and skill level of the estimator;
- Estimating techniques employed; and,
- Time and level of effort budgeted to prepare the estimate".

Ahuja et al. (1994) note that construction estimates utilize historical data that reflect average conditions and that, while many factors can cause a variance, the factors affecting productivity are the most difficult to predict. Such factors can be grouped into economic and project-specific factors; examples of the former and latter are labour rates and work methods, respectively (Ahuja et al. 1994).

Numerous studies have been conducted in recent years on the factors affecting cost estimating accuracy for construction projects (Hatamleh et al. 2018; Mahamid 2015). A study conducted by Hatamleh et al. (2018) considered 50 factors frequently investigated by researchers for the 13 years prior and prioritized the top 10 factors affecting estimate accuracy. Excluding the factors already listed above, key factors identified by Hatamleh et al. (2018) include:

- Clear and detail drawings and specification;
- Equipment (cost/availability/performance);
- Site constraints (access, storage, services);
- Material availability; and,
- Financial capabilities of the client.

The factors affecting estimate accuracy are numerous and vary in their degrees of subjectivity and interdependence. These considerations, coupled with the notion that the factors affecting estimate accuracy can vary from one cost item to another within a given project, indicate that establishing the variability of cost items for cost uncertainty analysis is not trivial.

2.3.3 Cost Underestimation

Comparisons of actual and estimated costs have been studied since the 1950s (Linder 2005). Reviews of such studies indicate that infrastructure projects are "frequently overrunning our funding estimates and by very large margins" (Hollmann 2012). The authors of the first statistically significant study of cost underestimation in transportation infrastructure put the issue more bluntly: “the cost estimates used to decide whether [transportation infrastructure projects] should be built are highly and systematically misleading” (Flyvbjerg et al. 2002). The causes of cost underestimation are debated but assertions include strategic misrepresentation (i.e., lying), facilitating the biases of management, and a lack of adjustment in estimating practices to fit the difficulty that projects present (Hollmann 2012). Garvey et al. (2016) assert that in addition to optimistic assumptions, inadequate attention has been paid to correlation between WBS element costs.

Regardless of the causes, cost uncertainty analysis cannot solely overcome the issue of cost underestimation. Rather, cost uncertainty analysis should be performed according to recommended best practices, with an awareness of the potential for bias, and with an understanding of the importance of correlation. Finally, results from cost uncertainty analysis, particularly worst case analysis, should be compared against empirical reality (Hollmann 2012).

2.4 The Process of Cost Uncertainty Analysis

Cost uncertainty can be analyzed using a variety of approaches and such approaches can be categorized differently. NASA (2008) refers to cost uncertainty analysis as a ‘cost risk assessment’ and divides methods into analytic and simulation approaches. In this thesis, cost uncertainty analysis is divided into two broad categories: methods that rely on the use of probability distributions (known as probabilistic methods) and methods that do not. Since the focus of this thesis is on probabilistic cost estimating, particular attention has been given to probabilistic methods. The methods described in this section are not exhaustive but they are among the most common for analyzing cost uncertainty and thus act as a point of reference for this thesis.

2.4.1 Analyzing Cost Uncertainty using Probabilistic Methods

This section describes two probabilistic methods for analyzing cost uncertainty: the method of moments which is an analytic approach and the Monte Carlo simulation method which is a simulation approach. The analytic approach provides analytic alternatives for analyzing cost uncertainty without the use of simulation. While analytic and simulation approaches vary in their details, they both rely on the use of probability distributions (this explains why they are called ‘probabilistic’ methods) and generally follow the same main steps. Garvey et al. (2016) summarizes the process developed by the cost analysis community as follows:

1. Determine cost drivers and identify risks.
2. Establish probability distributions to model cost uncertainty.
3. Account for correlation between inputs.
4. Perform the cost uncertainty analysis using simulation or analytical approaches, or a combination of methods.
5. Identify the probability level associated with the point estimate cost.
6. Establish a budget sufficient to achieve desired level of confidence.
7. Allocate portions of the budget per the WBS and identify key risks and areas of uncertainty as candidates for mitigation efforts.

The remainder of this section elaborates on each step of the process.

2.4.1.1 Determining Cost Drivers and Identifying Risks

Cost items or parameters that can have a critical effect on the project cost are known as 'critical items' or 'cost drivers'. Once a point estimate cost has been developed, critical items can be determined using a variety of approaches including sensitivity analysis or the use of a critical variance threshold as contemplated by AACE International (2008). Curran (1988) provides thresholds relative to overall estimate cost (0.5% in conceptual estimates and 0.2% in detailed estimates) and to profit (5% in conceptual estimates and 2% in detailed estimates). Ahuja et al. (1994) similarly suggests concentrating on the components thought to vary such that the overall project cost would change by a specified threshold. AACE International (2008) emphasizes the following when determining critical items:

- The magnitude of an item in a cost estimate does not determine whether it is critical, rather "an item is critical only if it can change enough to have a significant effect on the bottom line";
- Uncertainty is concentrated in a select number of critical items, typically 20 or less; and,
- Items that are strongly related must be linked or combined.

One of the simulation-based methods, range estimating, requires critical items to be determined as probability distributions are applied solely to critical items (AACE International 2008; Ahuja et al. 1994; Curran 1990). Regardless of the approach used, determining critical items allows decision makers to focus on areas of high cost uncertainty and sensitivity.

Before establishing probability distributions, risks must be identified. Risk identification involves the use of practices and tools such as brainstorming, interviews, and checklists to identify risks for a project (AACE International 2012). For large-scale capital construction projects, a 'risk register' is typically established to document project risks along with their probability of occurrence and potential impacts. Once identified, risks can either be accounted for in the ranging of cost items or quantified and modelled separately; refer to Section 3.2 for more information.

2.4.1.2 Establishing Probability Distributions

Probability distributions are used to represent the potential range of costs for each cost item. High-quality data are considered central to successful cost uncertainty analysis and whenever possible historical data should be used to derive probability distributions (U. S. Government Accountability Office 2020). When available, historical data can be compared with a proposed distribution to evaluate the fit of the proposed distribution using the chi-squared test, the Kolmogorov-Smirnov test, or probability plots and correlation tests (Morgan and Henrion 1990).

2.4.1.2.1 Subjective Probability Assessments

In the context of probabilistic cost estimating, distributions are often specified by expert opinion as historical data are often not available or of low quality (Garvey et al. 2016). Subjective probability distributions are statistical representations of an expert's judgment in the distribution of probabilities for a particular variable. While theories of subjective probability date back to at least 1926 (Ramsey 1926), "[t]he idea of formally representing uncertainty using subjective probability judgments to inform decision-making began to be taken seriously in the 1960s and 1970s" (O'Hagan 2019).

2.4.1.2.2 Elicitation

The process of obtaining subjective probabilities from experts is called elicitation. More formally, elicitation can be defined as "the facilitation of the quantitative expression of subjective judgement" (Dias et al. 2018). Elicitation involves many areas of research including psychology, statistics, and risk assessment. Various reviews of elicitation have been published from the 1980s to 2000s; O'Hagan et al. (2006) built on a number of such reviews to introduce the fundamentals of subjective probability and elicitation, outline the general elicitation process, and present psychological theories about expert judgment of uncertainty. More recently, the state of the art and science concerning elicitation was summarized by Dias et al. (2018). Elicitation is an important tool used in many industries but it can be difficult to do and is subject to cognitive and motivational biases (U. S. Air Force 2007). O'Hagan et al. (2006) encountered "an enormous number of case studies in which researchers have sought to elicit subjective probabilistic information" including studies in medicine, agriculture, economics, and engineering.

Cognitive biases in elicitation primarily arise from a reliance on heuristics which are mental shortcuts used to make judgments quickly. Tversky and Kahneman (1974) identified three heuristics that are employed when making judgments under uncertainty: representativeness, availability, and anchoring. Availability and anchoring are of particular interest for establishing probability distributions as they can introduce bias during elicitation. O'Hagan (2019) explains the anchoring heuristic as follows:

> "[W]hen asked to make a numerical judgment, people start with a readily available value and adjust from that point to assign a judgment. The initial value is called the anchor, because typically the adjustment is insufficient and so the resulting judgment is biased toward the anchor".

The availability heuristic means that an individual's judgement of a probability is "driven by the ease with which they can think of previous occurrences of the event, or the ease with which they can imagine the event occurring" (Morgan and Henrion 1990). Another potential bias relevant to elicitation is overconfidence. Overconfidence is the tendency for people to provide ranges which are too narrow, meaning that a greater percentage of the true values lie outside the range than is indicated (Lichtenstein et al. 1982). Studies have shown that experts identify 70% of the possible uncertainty range at best (U. S. Government Accountability Office 2020).

Montibeller and von Winterfeldt (2015) define motivational biases as "those in which judgments are influenced by the desirability or undesirability of events, consequences, outcomes, or choices". Examples of motivational biases include optimism bias and group think. Optimism bias stems from the desirability of an attractive outcome. Optimism bias is commonly thought to occur in cost estimation (Dillon et al. 2002), although Flyvbjerg et al. (2002) consider optimism calculated on the basis of incentives to be deception rather than genuine optimism. Group think manifests when a cohesive group strives for consensus to the detriment of accurately assessing the topic being elicited. Subjective probability assessments are subject to a number of other biases, but formal elicitation protocols, discussed in Section 3.3.4.3, have been established in an effort to reduce them.

Best practices for elicitation have been well documented (O'Hagan et al. 2006); one of the most frequently stated practices is the development or use of a protocol, particularly when tailored to the subject matter being elicited (Morgan 2014). There are many protocols for expert assessment described in the literature and while there are differences in their detail, there is "broad consensus as to the components of the elicitation process" (O'Hagan et al. 2006). According to Morgan and Henrion (1990), one of the most influential elicitation protocols described in the literature is the Stanford/SRI protocol. The Stanford/SRI protocol includes five phases: motivating, structuring, conditioning, encoding, and verifying. The following is a brief summary of each phase for the Stanford/SRI protocol (Morgan and Henrion 1990):

- Motivating involves describing the reason for the elicitation, explaining and justifying the idea of probabilistic assessment, and searching for (and ideally overcoming) motivational bias.
- Structuring involves establishing a clear definition of the quantity being elicited stated "in the form in which the expert will most likely be able to provide reliable judgements".
- Conditioning involves encouraging the expert to think carefully about their judgment and avoid cognitive biases by having them explain how they will make their probability judgments.
- Encoding involves the actual obtaining of probabilities, starting with extreme values and their probabilities and then moving inward.
- Verifying involves testing the judgments provided by the expert to increase the accuracy of the assessments.

The Stanford/SRI protocol, like many protocols published in the literature, establishes a probability distribution by eliciting a number of probabilities (or alternatively a number of quantiles) from an expert and then fitting a suitable distribution to the elicited values. The beta distributions are commonly fit to elicited values as they are "standard distributions in statistics that are well known and relatively easy to work with" (O'Hagan et al. 2006).

As an alternative to eliciting probabilities or quantiles, probability distributions can be established by eliciting parameter values for probability distributions. When eliciting and fitting a parametric distribution, two parameters are typically required: the measure of location (e.g., mean or median) and the measure of spread (e.g., variance). In the context of probabilistic cost estimating, cost ranges are bounded (as quantities and costs cannot be negative) making the beta distribution a suitable choice for representing cost uncertainty. When modelling opinion with a beta distribution, values can either be elicited to determine the two shape parameters α and β or the shape parameters can be chosen by an expert directly, i.e., through visual inspection of distribution shape (AbouRizk et al. 1991).

Shape parameters can be determined through various elicitation methods including assessments of location statistics, quantiles, sample size, and mean absolute deviation about the mean; O'Hagan et al. (2006) presented 20 elicitation methods for determining α and β, several of which use only two assessments. Garvey et al. (2016) illustrate how the beta distribution can be specified from subjective assessments on the shape parameters α and β directly plus any two fractiles x_i and x_j. In this case, the minimum (a) and maximum (b) possible values of the beta distribution are as shown in Equation 1 and Equation 2, respectively.

Equation 1:

$$a = \frac{x_i y_j - x_j y_i}{y_j - y_i}$$

Equation 2:

$$b = \frac{x_j(1 - y_i) - x_i(1 - y_j)}{y_j - y_i}$$

Another alternative to eliciting probabilities or quantiles is the More-or-less elicitation (MOLE) technique which "relies on the heuristically based ability of humans to choose which of two alternatives is more likely" (Welsh et al. 2004). Rather than having experts make estimates directly, the MOLE technique has experts repeatedly choose between options using a computerized tool and then constructs a range logically consistent with their choices. The MOLE technique has been compared with numerous elicitation processes requiring the direct estimation of values and is shown to reduce overconfidence and improve accuracy in generated ranges (Welsh and Begg 2018).

2.4.1.3 Accounting for Correlation

Correlation is considered a very important aspect of cost uncertainty analysis as it can significantly affect results (Garvey et al. 2016; National Aeronautics and Space Administration 2008; Touran and Wiser 1992). Ignoring positive correlation between cost items can lead to the significant underestimation of cost variance which is a particular concern for cost uncertainty analysis. Garvey et al. (2016) demonstrates that underestimation worsens exponentially as the number of WBS cost items increases.

In the context of probabilistic cost estimating, correlation has typically been accounted for using one of two methods. In the first method, ranges are applied to all WBS items and correlation is accounted for by either using functional relationships within the model or explicitly assigning correlation between cost items. In the second method, range estimating, ranges are applied only to critical items and correlation is addressed by combining items that are strongly related and otherwise modelling critical items independently. The first method is commonly used in the context of systems engineering (Garvey et al. 2016) and has been documented in numerous guides and handbooks applicable to space systems, weapons systems, software systems, and more (U. S. Government Accountability Office 2020; National Aeronautics and Space Administration 2008). The first method has also been demonstrated for construction projects (Touran and Wiser 1992) however correlation has been historically neglected (Curran 1990). Range estimating is commonly used in construction practice (Ahuja et al. 1994) and has been documented in a recommended practice (AACE International 2008) as well as numerous articles (Curran 1990, 1989, 1988, 1976).

2.4.1.3.1 Ranging WBS Items

Where all WBS items are ranged and combined, correlation between WBS elements must be accounted for (Garvey et al. 2016; National Aeronautics and Space Administration 2008). Monte Carlo simulations automatically capture functional correlation between WBS items where relationships between WBS items are defined in the model. Relationships between variables can be created in a variety of ways. The U.S. Air Force (2007) offers an example where the cost of training is modelling by using a scaled factor of the cost of prime mission equipment which leads training costs to be correlated in the simulation. AACE International (2019) describes a method for correlating activity durations by applying a risk driver to multiple activities; the same principle can be applied to activity costs.

Where functional relationships between WBS items are not captured by the model, correlation values need to be specified. Garvey et al. (2016) describe three approaches for specifying correlations between cost items:

1. Assigning correlations subjectively.
2. Deriving correlations mathematically from the structures of the cost estimating relationships defined within the WBS.
3. Determining correlations empirically using Monte Carlo simulation.

Deriving correlations can take significant time and access to data. A project with n cost items has $n(n-1)/2$ cost item pairs (Garvey et al. 2016), meaning a WBS with 20 cost items has 190 pairs while a WBS with 50 cost items has 1225 pairs.

The first approach is typically employed where time or information is not available to assess or derive correlations between some or all WBS items or where correlations between some WBS items are not accounted for through functional relationships, e.g., a Monte Carlo simulation. The knee-in-the-curve method is an example of the first approach where a correlation, ρ, is chosen on the interval $0.10 \leq \rho \leq 0.30$, since for $\rho > 0.30$ there is little change in the percent that a project's cost range is underestimated by not capturing positive correlation when present (Garvey et al. 2016). Another example of the first approach is the use of subjective correlation coefficients where correlations between variables are estimated by experts. Touran (1993) illustrates the qualitative assessment of correlation between variables as either weak, moderate, or strong corresponding to correlation coefficients of 0.15 (mid-point of 0 to 0.3), 0.45 (mid-point of 0.3 to 0.6), and 0.8 (mid-point of 0.6 to 1), respectively. Similarly, NASA (2008) evaluates variables to be either uncorrelated, correlated by a small amount, or correlated by a large amount corresponding to correlation coefficients of 0, 0.3, or 0.75, respectively.

The second approach is suitable where relationships between WBS items are specified by mathematical functions. As an example, functional correlations can be derived analytically using the algebra of random variables. Garvey et al. (2016) show that if $Y = aX + Z$ where a is a real number and X and Z are independent random variables, then the correlation between Y and X

and the correlation between Y and Z are as shown in Equation 3 and Equation 4, respectively. The variables σ_X, σ_Y, and σ_Z represent the standard deviation of X, Y, and Z, respectively.

Equation 3:

$$\rho_{Y,X} = a\frac{\sigma_X}{\sigma_Y}$$

Equation 4:

$$\rho_{Y,Z} = a\frac{\sigma_Z}{\sigma_Y}$$

Where the mathematical functions describing WBS item relationships are too complex to analytically derive correlation, the third approach is useful. Monte Carlo simulation allows the correlation between WBS items to be determined empirically by programming relationships between WBS items, randomly sampling values for variables, and computing their correlations. Refer to Section 2.4.1.4.2 for further discussion of Monte Carlo simulation.

2.4.1.3.2 Ranging Critical Items

In range estimating, only critical items are ranged and correlation is accounted for by linking or combining items that are strongly related. After linking or combining related items, the critical items are considered independent, i.e., covariation between each pair of critical items is assumed to be negligible. Range estimating is current practice in construction engineering despite being sensitive to the number of inputs being ranged and potentially underestimating uncertainty by ignoring weak correlation between inputs. AACE International (2008) cautions against ranging non-critical items as “the inevitable result will be a far narrower predicted range of possible project costs than actually exists, misstatements of risk and opportunity, and understatement of required contingency”. According to AACE International (2008), dependencies between critical items typically emerge either when determining cost drivers or establishing probability distributions. If such dependencies are missed, then the covariance which contributes to the spread of overall results is also missed.

2.4.1.4 Performing Cost Uncertainty Analysis

Once probability distributions and correlation between inputs have been established, the actual cost uncertainty analysis calculations can be performed. Where the method of moments and the

Monte Carlo simulation method differ is in their computation of the project cost distribution as described in Section 2.4.1.4.1 and Section 2.4.1.4.2 below.

2.4.1.4.1 Method of Moments

The method of moments is a direct analytical technique wherein the mean and variance of project cost are derived as functions of the means and variances of its components. The equations for mean and variance of project cost as shown in this section are based on equations described by Diekmann (1983). For a project comprising n cost items, the project cost, $cost_{proj}$, is represented by Equation 5. The mean of project cost is as shown in Equation 6.

Equation 5:

$$cost_{proj} = X_1 + X_2 + X_3 + \cdots + X_n$$

Equation 6:

$$Mean(cost_{proj}) = Mean(X_1) + Mean(X_2) + Mean(X_3) + \cdots + Mean(X_n)$$

If the cost items within a project are not independent, then the variance of project cost is as shown in Equation 7.

Equation 7:

$$\begin{aligned} Var(cost_{project}) \\ &= Var(X_1) + Var(X_2) + Var(X_3) + \cdots + Var(X_n) + Covar(X_1, X_2, X_3, \dots, X_n) \end{aligned}$$

If the cost items within a project are independent, then the covariance between cost items is zero and the variance of project cost is simplified as shown in Equation 8.

Equation 8:

$$Var(cost_{project}) = Var(X_1) + Var(X_2) + Var(X_3) + \cdots + Var(X_n)$$

Since each cost item, $X_1, X_2, X_3, \dots, X_n$, is represented using a probability distribution, the mean and variance of each cost item are relatively simple to calculate using distribution parameters (that is once parameters have been set).

There are many probability distributions described in literature, but the uniform, triangular, beta, normal, lognormal, and Weibull distributions are among the most commonly used in construction. This thesis focuses primarily on the triangular and beta distributions as they have long been considered the distributions of choice for the type of subjective assessments often conducted as part of probabilistic cost estimating (Garvey et al. 2016). The mean and variance of a triangular distribution are as shown in Equation 9 and Equation 10, respectively. The variables $a, m,$ and b represent the minimum, most likely, and maximum values of a triangular distribution, X, respectively.

Equation 9:

$$Mean(X) = \frac{(a + m + b)}{3}$$

Equation 10:

$$Var(X) = \frac{(m - a)(m - b) + (b - a)^2}{18}$$

The mean and variance of a beta distribution are as shown in Equation 11 and Equation 12, respectively. The variables α and β represent shape parameters, while the variables a and b represent the minimum and maximum of a beta distribution, X, respectively.

Equation 11:

$$Mean(X) = a + (b - a)\frac{\alpha}{\alpha + \beta}$$

Equation 12:

$$Var(X) = (b - a)^2 \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}$$

Covariance between cost items, if applicable, can be assigned, derived, or determined using the approaches discussed in Section 2.4.1.3.1. Once the mean and variance for project cost have been calculated, they are used to specify a probability distribution of the project cost thus representing the range of potential project costs and their associated probabilities.

Proponents of the method of moments highlight its basis in probability theory as well as its ability to "provide insight into problem structure and subtleties not always apparent from empirically based methods, such as Monte Carlo simulations" (Garvey et al. 2016).

2.4.1.4.2 Monte Carlo Simulation Method

Monte Carlo simulation is a modelling technique involving the use of random numbers (Law 2015). Monte Carlo simulation has been used to estimate cost uncertainty since the 1960s (Dienemann 1966) and remains a popular method for modelling and measuring cost uncertainty.

Monte Carlo simulations rely on repetition of an experiment (each repetition is called an iteration) where random variates are sampled from probability distributions for use as a variables within the model and the desired output parameters of a model are calculated. Once a set of iterations has been performed, statistical analysis is performed on output parameters. In the context of probabilistic cost estimating, the steps in a Monte Carlo simulation are (Garvey et al. 2016):

1. For each variable in the model that is uncertain, randomly sample a value from its distribution function.
2. Once a set of samples has been established, combine these values according to the mathematical relationships specified by the model to produce a single value for the project's total cost.
3. Repeat the two steps above n-times to obtain n-values for the project's total cost.
4. Develop a frequency distribution from the n-values for the projects total cost.

Where correlation between variables is not captured in model input relationships or otherwise addressed, it must be valued and incorporated into the simulation model. Correlation can be induced in a simulation study using a variety of approaches. Iman and Conover (1982) introduced a distribution-free approach to inducing rank correlation among inputs that is employed by commercially available products such as @Risk. Touran and Wiser (1992) suggested a methodology for generating correlated random numbers using a multivariate lognormal distribution. Lurie and Goldberg (1998) presented an algorithm for generating Pearson-correlated random numbers. All of the approaches mentioned above require the development of a correlation matrix to specify the degree of correlation between input variables.

2.4.1.5 Identifying Probability Associated with Point Estimate

Once a distribution of project costs has been established, the probability associated with the point estimate can be identified. If the distribution of project costs has been determined using the method of moments, first the cumulative distribution function (CDF) can be evaluated by a numerical integration procedure (Garvey et al. 2016). The CDF, informally called the S-curve, is the probability that a variable (e.g., project cost) will exceed a particular value (e.g., point estimate or project budget). The probability associated with the point estimate can then be found by mapping the point estimate value to its corresponding percentile on the CDF.

If the distribution of project costs has been determined using Monte Carlo simulation, then the probability associated with the point estimate is simply the fraction of simulated observations that are less than or equal to the point estimate value. Many of the commercially available Monte Carlo simulation software packages can also generate a CDF for project cost; in this case the probability associated with the point estimate can also be found by mapping as described above for the method of moments.

2.4.1.6 Establishing and Allocating Budget

To establish an overall budget for a project using the results of probabilistic cost estimating, an owner must establish their desired level of confidence. The chosen level of confidence is typically expressed as a percentile of possible results. For example, setting a project budget at the 80th percentile value for simulated project cost would mean that 80% of observations would be at or below the budget set. Conversely, 20% of observations would be above the budget set, representing a 20% chance of cost overrun for the project. There are no set rules for choosing the level of confidence as this can depend on many factors including the risk tolerance of the owner and/or funding partners, the availability of funding, the ability to pool funding across multiple projects, and budgetary or other constraints. AACE International (2008) describes the 50^{th} percentile as a “risk neutral approach” whereas the 80^{th} percentile is considered more conservative and risk-averse. The U.S. Government Accountability Office (2020) explains that high risk programs can adopt “a higher confidence level estimate (70 or 80 percent)” to increase owner confidence of success within budget, make provision for unknown risks, and reduce the likelihood that additional funds are required. Conversely, there is consensus in the literature that setting a high level of confidence can result in a maldistribution of funds (AACE International 2008) and unaffordable portfolio budget (U. S. Government Accountability Office 2020).

Once a budget is set at the desired level of confidence, contingency is typically calculated as the difference between the budget at the desired level of confidence and the point estimate (U. S. Government Accountability Office 2020). Contingency is calculated as a lump sum but is often allocated to cost items or throughout the WBS. Reasons for allocating contingency include allowing for contingency to be managed based on a drawdown plan (AACE International 2008), organizational preference for splitting contingency between appropriations, and in some cases organizational requirements that prevent the separate display of contingencies (U. S.

Government Accountability Office 2020). The literature emphasizes that the allocation of contingency to cost items should not be interpreted as a commitment to fund such cost items as contingency should be retained by management in a separate control account (U. S. Government Accountability Office 2020; AACE International 2008).

The allocation of contingency is not trivial as each cost item contributes differently to the project cost distribution; cost items with more risk and uncertainty require more contingency than their counterparts (U. S. Government Accountability Office 2020). Additionally, the sum of costs for lower-level elements plus their contingencies must sum to the cost and contingency of their parents. Since each cost item has a different cost distribution, lower-level elements at a specified percentile will not sum to the percentile value of the aggregated element.

Several methods of allocation address the challenges described above. Two popular methods include allocation by standard deviation and allocation by need (U. S. Government Accountability Office 2020). The needs based allocation method was proposed by Book and published in the 2008 NASA Cost Estimating Handbook (National Aeronautics and Space Administration 2008). Book's algorithm assigns contingency to cost items based on the difference between their most likely estimate and the estimate at the specified percentile while adjusting for correlation between elements (Garvey et al. 2016). Given the dynamic nature of risk and uncertainty, AACE International (2008) recommends conducting periodic analysis to reassess risk and to reassign and/or release contingency.

2.4.2 Analyzing Cost Uncertainty using Non-Probabilistic Methods

This section describes alternative methods for analyzing cost uncertainty without using probability distributions: the scenario based method and alternative range estimating techniques that are grounded in fuzzy set theory.

2.4.2.1 Scenario Based Method

The scenario based method is a simplified analytic method for analyzing cost uncertainty which was formally developed and completed for the US Air Force Cost Analysis Agency in 2006 and formally published in 2008 (Garvey 2008). The scenario based method involves defining a set of conditions called the protect scenario (PS) that a project manager would want to protect against by holding enough budget to cover the occurrence of any or all conditions. Once the PS has been defined, its cost is established. Cost risk (CS) is then calculated as the difference between the PS and the point estimate cost, x_{PE}. Next, two values are established subjectively: the probability of the point estimate (PE) not being exceeded, α_{PE}, and the coefficient of variation (CV). The coefficient of variation is defined as the ratio of a distribution's standard deviation to its mean as shown in Equation 13 (Garvey 2008).

Equation 13:

$$CV = \frac{\sigma}{\mu}$$

Assuming the project cost, $cost_{proj}$, is normally distributed, the mean and standard deviation are as shown in Equation 14 and Equation 15, respectively (Garvey 2008). The variable D is the CV and the variable z_{PE} is the value such that $P(Z \leq z_{PE}) = \alpha_{PE}$ where Z is the standard normal random variable. Values for z_{PE} can be found in z-score tables for the standard normal distribution.

Equation 14:

$$\mu_{cost_{proj}} = x_{PE} - z_{PE}\frac{Dx_{PE}}{1 + Dz_{PE}}$$

Equation 15:

$$\sigma_{cost_{proj}} = \frac{Dx_{PE}}{1 + Dz_{PE}}$$

Once the mean and standard deviation have been computed, the distribution function of $cost_{proj}$ can be specified along with the probability associated with a particular value, such as the PS cost.

Since the scenario based method was introduced, an enhanced scenario based method has been published in the literature (Garvey et al. 2012). The enhanced scenario based method integrates historical cost performance data into its algorithms. The key benefits of the enhanced scenario based method are as follows:

- Avoids the need to develop probability distributions for various cost items which can be cumbersome;
- Captures correlation indirectly through the CV; and,

- Encourages discussion of scenarios and focuses attention on key risk events that can drive cost overruns.

2.4.2.2 Fuzzy Range Estimating

Shaheen et al. (2007) presented an alternate approach to range estimating called fuzzy range estimating that is based on fuzzy set theory rather than Monte Carlo simulation. A fuzzy set is a set characterized by a membership function which assigns a degree of membership to each element within the set (Zadeh 1965). According to Shaheen et al. (2007), fuzzy set theory allows for the elicitation of cost ranges without sacrificing accuracy as elicitation is foundational to fuzzy set theory. Fuzzy range estimating addresses a potential shortcoming of Monte Carlo based range estimating, namely the difficulty associated with establishing probability distributions based on subjective data, by modelling uncertainty using fuzzy numbers. A fuzzy number is a fuzzy membership function that satisfies several properties including being normal and convex. The fuzzy range estimating model proposed by Shaheen et al. (2007) is as follows:

1. Identify major cost packages and their subgroups.
2. Consult each expert and establish fuzzy numbers for each uncertain item.
3. Consolidate expert inputs for each uncertain item.
4. Calculate the project's total cost as a fuzzy number using fuzzy summation.

Shaheen et al. (2007) state that modelling uncertainty using fuzzy arithmetic is not particularly sensitive to moderate changes in the shapes of input distributions and does not require assumptions to be made with respect to correlation among inputs, however results are conservative and may overestimate uncertainty. The overestimation of uncertainty is a concern as this can prevent a viable project from being approved due to being perceived as having more risk or uncertainty than it actually has.

2.4.2.3 Fuzzy Monte Carlo Simulation

Sadeghi (2009) proposed a fuzzy Monte Carlo simulation (FMCS) framework as a generalized form of Monte Carlo simulation and used the FMCS framework to develop a range estimating template. Compared to the approach developed by Shaheen et al. (2007), the framework considers both fuzzy and probabilistic uncertainty in model inputs. Starting with a model containing both random variables (represented by probability distributions) and subjective variables (as fuzzy sets), sample sets are produced from probabilistic distributions leaving only fuzzy input variables. Fuzzy arithmetic is used to calculate the output in the form of a fuzzy set and the process is repeated n-times, resulting in a number of fuzzy sets with random variation. Fuzzy arithmetic can then be applied to the outputs to yield statistical results including mean, variance, quantiles, and the probability associated with a specific threshold.

Sadeghi (2009) found that use of the range estimating template indicated reasonable behaviour of the FMCS framework but noted that testing needs to be conducted on actual projects to justify the benefits that the FMCS framework contributes to the construction industry.

2.5 Discussion and Summary

The body of scholarship on the subject of cost uncertainty analysis is substantial, particularly in the context of systems engineering. There are numerous methods for analyzing cost uncertainty; this literature review is not exhaustive but it outlined the most commonly used methods and highlighted the similarities and differences between simulation-based methods used in construction engineering (e.g., range estimating) and those used in systems engineering. There appears to be consensus in the literature across industries with respect to the overall process of analyzing cost uncertainty using probability distributions and Monte Carlo simulation, however limitations have been found in the literature specific to construction engineering.

A gap was found in the construction engineering literature related to the development of subjective probability distributions, and more specifically elicitation protocols, within the context of probabilistic cost estimating. There is consensus in the literature on the key components of a good elicitation protocol but such protocols are numerous and varied in their detail. In Chapter 3, this thesis attempts to demonstrate that the elicitation protocols used in other fields are applicable to probabilistic cost estimating in construction engineering.

Prevailing modelling practices in construction engineering (e.g., range estimating) are sensitive to the number of inputs being sampled and have the potential to underestimate cost uncertainty[1]. The systems engineering perspective on probabilistic cost estimating involves capturing correlation between inputs; doing so allows for the ranging of all cost items within a construction

[1] In range estimating, correlation is accounted for by linking or combining items that are strongly related and otherwise assuming that critical items are independent. Thus, covariation between each pair of critical items is zero.

project while avoiding the potential for underestimation of project cost uncertainty. In Chapter 4, this thesis showcases the application of various correlation modelling methods to simulation-based cost uncertainty analysis for construction projects and contrasts this with assumptions of independence.

Chapter 3 – Establishing Subjective Probability Distributions using an Elicitation Protocol

3.1 Introduction

In probabilistic cost estimating, the uncertainty of each cost item is represented using probability distributions by specifying the type of distribution (e.g., beta vs. triangular distribution) and either statistical measures (e.g., mean and standard deviation) or parameters (e.g., minimum, maximum, and various shape parameters). The method of moments, previously described in Section 2.4.1.4.1, is a procedure in classical statistics that demonstrates a project's cost uncertainty is derived from the cost uncertainty of its components. While there are other inputs to cost uncertainty analysis that can affect results (refer to Section 3.2 for a detailed list), probability distributions used to represent cost uncertainty are the basis of cost uncertainty analysis and should be established as carefully and scientifically as possible. The literature is clear that if available, historical data should be used to establish probability distributions, however as previously mentioned, historical data are often not available or of low quality for cost uncertainty analysis (Garvey et al. 2016); particularly in construction practice, high-quality cost data are hard to find. In the absence of quality data, probability distributions are often established subjectively using expert opinion, i.e., through elicitation. Psychologists have identified various biases that can affect the elicitation process and structured elicitation protocols have been developed to avoid such biases as much as possible. This chapter describes the inputs being ranged and the probability distributions used to represent them, examines the pitfalls of elicitation and how to overcome them, explores leading elicitation methods and protocols, and demonstrates the application of such a protocol to cost uncertainty analysis for a construction project.

3.2 Probabilistic Cost Estimating Inputs

The primary input for probabilistic cost estimating is the cost uncertainty of cost items as represented by a probability distribution. Depending on the structure of the model, the uncertainty of cost items can represent either (1) the full range of potential costs inclusive of risks and compounding effects or (2) the expected estimate accuracy and general uncertainty related to quantities and unit rates with discrete risks and/or risk drivers modelled separately. Range estimating involves ranging inputs to represent the former (AACE International 2008), while other forms of cost uncertainty analysis involve ranging inputs to represent the latter

(AACE International 2019; AbouRizk 2013). For the purposes of this thesis, discrete risks are considered separately from cost item uncertainty in a process called quantitative risk analysis (QRA). The variables involved in range estimating and QRA generally include (AACE International 2019; AbouRizk 2013):

- For each risk:
  - Probability (p);
  - Cost impact (I_c);
  - Schedule impact (I_S); and,
  - Cost items affected if the risk occurs (i).
- For each cost item:
  - Quantity (q);
  - Unit rate (u);
  - Duration (d); and,
  - Escalation rates (e).

Depending on the expected duration of the project, susceptibility to market volatility, and the importance of completion timelines, range estimates can be static (i.e., simulation of costs at a discrete point in time) or can include integrated cost-schedule analysis. An in-depth description of cost-schedule integration is beyond the scope of this thesis but various techniques are described in the literature. For example, Isidore and Back (2002) created the multiple simulation analysis technique (MSAT) which combines discrete event simulation, regression, and numerical analysis to relate cost estimate and schedule data obtained from probabilistic cost and schedule estimating. AACE International (2019) provide a guideline for integrated cost-schedule risk analysis using risk drivers and Monte Carlo simulation of a critical path method (CPM) schedule. Generally speaking, schedule impact due to risk as well as cost item duration and escalation rates are necessary for integrated cost-schedule analysis models but are not required for static range estimates. There are other inputs that can be form part of a range estimate model, such as materials and productivity rates, however these are typically used to generate the variables above and can be ranged similarly if desired (AbouRizk 2013). In some cases, cost items are expressed as a lump sum, in which case the quantity is equal to one and the unit rate is equal to the lump sum amount.

There are many different families of probability distributions that can be used to represent uncertainty however, as previously mentioned, the uniform, triangular, beta, normal, and lognormal distributions are among the most commonly used in construction. Variables in probabilistic cost estimating tend to be bounded which make bounded families of distributions preferable to those that are unbounded. Depending on the characteristics of the input being ranged, some types of distributions are more suitable than others. Where data are available for construction simulation, AbouRizk (1990) recommends identifying a family of distributions that is flexible and has tractable parameters. Beta distributions are commonly fit to elicited values as they are "standard distributions in statistics that are well known and relatively easy to work with" (O'Hagan et al. 2006). As previously mentioned, this thesis focuses primarily on the triangular and beta distributions as they are generally considered the distributions of choice for elicited inputs.

3.3 Capturing Expert Knowledge

Expert knowledge has been used to support decision-making under uncertainty since at least the 1960s; elicitation is the process of capturing such knowledge in the form of probability distributions (O'Hagan 2019). Elicitation plays an important role in uncertainty analysis but it is susceptible to numerous biases at individual and group levels. Various elicitation protocols have been developed to reduce the biases that experts are prone to when making probability judgments. Before exploring the leading elicitation methods and protocols, it is important to understand the cognitive and motivational biases relevant to elicitation as well as common debiasing techniques. Debiasing refers to "attempts to eliminate, or at least reduce, cognitive or motivational biases" (Montibeller and von Winterfeldt 2018).

3.3.1 Cognitive Biases

As previously mentioned, individuals often make judgments on the basis of heuristics which are mental or cognitive shortcuts used to make judgments quickly. While heuristics are useful in everyday life, their application in more complex tasks (e.g., expert probabilistic assessment) can produce cognitive bias. A cognitive bias is a discrepancy between an expert's assessment and the correct answer in a judgmental task (von Winterfeldt and Edwards 1986). The psychology literature has identified many heuristics and biases, but some are of particular relevance to elicitation in the context of probabilistic cost estimating. Montibeller and von Winterfeldt (2018) distinguish between cognitive biases that are relevant to decision and risk analysis and those that are less or not at all relevant, with the former being more difficult to correct than the latter. Similarly, the remainder of this section focuses on relevant heuristics and cognitive biases.

3.3.1.1 Availability

The availability heuristic was originally identified by Tversky and Kahneman (1974) and concerns the ease with which an individual can recall or imagine an event or outcome. Research has found that the probability of an outcome that is easily recalled (i.e., available) tends to be overstated. Conversely, where recall or imagination are difficult, probabilities tend to be underestimated. When experts are asked to assess the probability of an outcome, their judgment is influenced by recent evidence, first-hand evidence, and dramatic events (O'Hagan 2019). The following are illustrative examples of where availability bias can arise in the context of probabilistic cost estimating:

- an estimator or quantity surveyor has seriously underestimated costs for a given discipline previously (in this case probability of underestimation could be overstated);
- experts have been involved in projects that experienced rare but memorable issues related to cost that far exceeded estimates (in this case probability of underestimation could be overstated); and,
- an individual lacks the domain-specific experience necessary to recall cost overruns (in this case probability of overestimation could be understated).

Bias related to availability can be reduced by training experts on probability and providing statistical and other relevant background information in advance of seeking probability assessments (Montibeller and von Winterfeldt 2018) as well as providing counter examples and challenging expert inputs against historical experience (U. S. Air Force 2007).

3.3.1.2 Anchoring

The anchoring heuristic, also originally identified by Tversky and Kahneman (1974), concerns attachment to an initially available value and the insufficiency of adjustments from such value. In a classic demonstration of anchoring, subjects were asked to estimate the percentage of African countries in the United Nations by first indicating whether a number (obtained by spinning a wheel of fortune in the subject's presence) was higher or lower than the true percentage and then estimating the percentage by adjusting upward or downward from the given number. The randomly determined values had a marked effect on the estimates: "the median estimates of the percentage of African countries in the United Nations were 25 and 45 for groups that received 10 and 65, respectively, as starting points" (Tversky and Kahneman 1974). Anchoring has since been demonstrated in a variety of domains including probability assessments (Furnham and Boo 2011). The following are illustrative examples of where anchoring bias can arise in the context of probabilistic cost estimating:

- an expert is asked to provide probabilities associated with numerous given cost item amounts rather than being asked to provide values associated with numerous probabilities, i.e., the elicitation of probabilities is used rather than the elicitation of quantiles (in this case the initial cost item amount can act as an anchor for subsequent judgments); and,
- an expert is told the point estimate value for a given cost item (e.g., $500,000) and is then asked to provide dollar values associated with various probabilities (e.g., "what cost value is associated with a 90% chance of not being exceeded") rather than using percentages (in this case the point estimate value can act as an anchor).

Debiasing techniques for anchoring include avoiding anchors altogether, providing multiple or counter-anchors, considering the opposite, and eliciting values from multiple experts each with different anchors (Montibeller and von Winterfeldt 2018; Mussweiler et al. 2000).

3.3.1.3 Overconfidence

As previously stated, overconfidence is the tendency for people to provide ranges which are too narrow, meaning that a greater percentage of the true values lie outside the range than is indicated (Lichtenstein et al. 1982). While overconfidence is not a heuristic itself, some theorize that it may be related to anchoring. Slovic (1972) posit that people anchor to their best estimate and adjust (insufficiently) upward and downward to determine their upper and lower limits, respectively. O'Hagan (2019) also explains that experts are often asked for an interval after providing or receiving an estimate which can serve as an anchor. Alternatively, Yaniv and Foster (1995) theorize that overconfidence arises as a result of people's preference for ranges which are informative versus accurate. Research shows that overconfidence is widespread and worsens for tasks of greater difficulty (Lichtenstein et al. 1982). Overconfidence is of particular concern as the elicitation of ranges is fundamental to probabilistic cost estimating.

Overconfidence can be reduced by training and feedback (O'Hagan et al. 2006), avoiding anchors or starting elicitations with extreme estimates (Montibeller and von Winterfeldt 2018), and asking experts to provide contradictory evidence (Koriat et al. 1980). Montibeller and von Winterfeldt (2018) state that using fixed value elicitations (in which experts provide probabilities for fixed values) instead of fixed probability elicitations (in which experts provide values corresponding to fixed probabilities) can reduce overconfidence, however evidence in the literature is conflicting as to whether eliciting probabilities or quantiles leads to better calibration (O'Hagan et al. 2006). Seaver et al. (1978) found the fixed value method to be superior to the fixed probability method however the fixed value method is arguably more prone to anchoring as discussed in Section 3.3.1.2. Murphy and Winkler (1974) found results from fixed probability elicitation to be more reliable.

Adjustments or 'corrections' are also described in the literature but their use is debated. One adjustment method involves assuming that the experts' minimum and maximum represent the 15^{th} and 85^{th} percentile of the distribution, respectively (U. S. Government Accountability Office 2020). O'Hagan et al. (2006) caution against adjustments or corrective procedures for two reasons: (1) the presence and extent of bias is difficult to predict and appropriately correct for and (2) experts may provide different assessments than they otherwise would have if they anticipate any adjustments will be made. Lichtenstein et al. (1982) similarly warned that an expert may adjust their answers if they suspect their judgments are being calibrated and also noted that since miscalibration depends on task difficulty, any events used for calibration purposes would need to match the difficulty of the topic actually being elicited.

3.3.1.4 Other Relevant Cognitive Biases

Montibeller and von Winterfeldt (2018) identifies a number of other cognitive biases, the following are relevant to probabilistic cost estimating:

- Certainty Effect: individuals tend to favour certain outcomes over merely probable outcomes with greater utility.

- Gain-Loss Bias: individuals evaluate probabilities differently depending on how outcomes are framed, i.e., in terms of gain or loss.
- Splitting Biases: individuals evaluate probabilities differently depending on how events are grouped.

The biases listed above are situational but should be considered and reduced where they have the potential to occur. The certainty effect occurs in the elicitation of judgments that employ probability versus certainty equivalent methods which could be relevant where experts (or even modellers) are presented with the choice of using cash or contingency allowances versus cost item ranging.

Gain-loss bias occurs in various types of judgments including where there are choices of risky options, a singular option is evaluated, or a choice is required based on a description of consequences. Risk quantification is susceptible to gain-loss bias since the likelihood of a risk or opportunity is typically elicited in the context of a particular consequence or impact.

Splitting biases have been observed in the elicitation of probabilities for event and fault trees. However, similar issues could arise in the elicitation of subjective probability distributions, as well as during quantitative risk analysis, since costs and risks are often represented (and sometimes aggregated) using a WBS or risk breakdown structure (RBS).

3.3.2 Motivational Biases

Motivational biases influence judgments by the benefit or detriment of events or outcomes regardless of whether such biases are conscious. While motivational biases are considered important and pervasive in decision and risk analysis, research on such biases in this context is sparse (Montibeller and von Winterfeldt 2015). This section focuses on motivational biases which occur at the individual level; motivational biases at the group level will be discussed in Section 3.3.4.

3.3.2.1 Confirmation Bias

Confirmation bias is the unconscious tendency for an individual to selectively recall knowledge or evidence when seeking to validate their beliefs. Confirmation bias has been demonstrated experimentally as well as in real-world contexts including medicine, law, and science (Nickerson 1998); theoretical explanations are numerous in the literature but the most persuasive among them is that individuals favour a confirmational strategy of hypothesis testing rather than a falsifying one.

Debiasing techniques include training experts on the nature of confirmation bias, requesting alternative explanations or hypotheses early in the process, and encouraging individuals to think of reasons why their findings may be wrong (Nickerson 1998). Montibeller and Winterfeldt (2018) indicate that using multiple experts with different points of view on explanations or hypotheses can also reduce confirmation bias.

3.3.2.2 Optimism Bias

Optimism bias refers to an increase in an individual's expectations about an outcome based on its desirability. Optimism bias, also referred to as wishful thinking, has been demonstrated in a variety of contexts such as evaluating odds in games of chance, health risk perception, and predicting future life events. For example, students tend to believe that they are more likely than their peers to experience positive events (Weinstein 1980). As previously mentioned, optimism bias is often said to occur in the estimation of project costs (Dillon et al. 2002), although there is disagreement about whether widespread cost underestimation is caused by genuine optimism bias or willful deception (Flyvbjerg et al. 2002). Montibeller and Winterfeldt (2018) indicate that using decomposition as well as eliciting values from multiple experts with alternate views on outcomes can reduce optimism bias.

3.3.2.3 Pessimism Bias

If optimism bias concerns the desirability of a positive event or consequence, then pessimism bias concerns the undesirability of a negative event or consequence. More formally, pessimism bias is the tendency to "be cautious, prudent, or conservative in estimates that may be related to harmful consequences" (Montibeller and von Winterfeldt 2018). Pessimism bias has been demonstrated in a number of contexts including perceptions of adverse health effects (Dolinski et al. 1987) and assessments of likelihoods of reaching life goals (Chapin 2001), however, the psychological literature has not covered common sources of pessimism to the same extent as common sources of optimism (Blanton et al. 2001). Estimates that are said to be 'conservative' typically involve pessimism bias and require careful consideration before being incorporated into a probabilistic cost estimate. Debiasing techniques applicable to optimism bias, discussed in Section 3.3.2.2, are also applicable to pessimism bias.

3.3.2.4 Affect-Influenced Bias

The affect heuristic concerns the reliance on positive and negative feelings to guide judgments and decision making (Slovic et al. 2007). The affect heuristic has been used to explain the commonly observed inverse relationship between perceived risk and perceived benefit (Finucane et al. 2000; Alhakami and Slovic 1994). Alhakami and Slovic observed that the perceived risk-benefit relationship is associated with an individual's affective evaluation. For activities or outcomes considered 'good', individuals tend to value the risk as low and the benefit as high, whereas for 'bad' activities or outcomes, individuals tend to value the risk as high and the benefit as low. In reality, the relationship between risk and benefit is not inverted and can even be positively related, i.e., an increase in risk is associated with an increase in benefit (Finucane et al. 2000).

Skagerlund et al. (2020) found the affect heuristic to be stable across different domains (e.g., social, health, and economic domains) and elicitation methods (i.e., it is reproducible in both joint and separate conditions). Debiasing techniques include avoiding loaded descriptions of consequences, cross-checking judgments with alternative elicitation protocols, and using multiple experts with alternative points of view (Montibeller and von Winterfeldt 2018).

3.3.2.5 Other Relevant Motivational Biases

The U. S. Air Force (2007) identifies a number of other motivational biases relevant to probabilistic cost estimating including misunderstanding, competitive pressures, project advocacy, and seeking personal benefit (e.g., fulfilling career goals). To avoid some of these more obvious forms of motivational bias, experts with a perceived or actual conflict of interest should be called upon sparingly if at all. In any case, using multiple experts with competing points of view will help reduce motivational bias. The general debiasing techniques discussed in Section 3.3.3 are relevant to motivational biases.

3.3.3 Individual-Level Debiasing Techniques

In addition to the bias-specific techniques discussed in Section 3.3.1 and 3.3.2, Fischhoff (1982) identifies general debiasing techniques each separated according to the underlying reasons for the corresponding bias at the individual level, namely: faults in the process, faults in the expert(s), and misalignment between the process and expert(s). Unfair tasks can be addressed by offering incentives for good performance, providing clear instructions, and asking fewer questions. Faults in the expert(s) can be divided into two broad groups: experts with faults that can be improved and experts with faults that cannot be improved. Strategies for the former include describing and warning of biases as well as extensive training, while strategies for the latter include finding alternative experts, recalibrating responses, and accepting error. Misalignment between the process and expert can be addressed by encouraging experts to recall contradictory evidence, decomposing complex problems, and contemplating alternative outcomes.

3.3.4 Group Biases

While groups benefit from information pooling, error checking, and motivation gains (Kerr and Tindale 2004), group elicitation is susceptible to a number of unique group-level biases. Additionally, Kerr et al. (1996) found that groups can amplify or attenuate individual biases depending on several factors including the size of the group, the magnitude of individual bias, and the nature of the group process. The trend of facilitation for group elicitation highlights the importance of understanding group-level biases and how to reduce or avoid them (Montibeller and von Winterfeldt 2018). This section focuses on the group biases that are relevant to probabilistic cost estimating as well as common debiasing techniques.

3.3.4.1 Groupthink

Janis (1972) describes groupthink as “a mode of thinking that people engage in when they are deeply involved as a cohesive in-group, when the members’ strivings for unanimity override their motivation to realistically appraise alternative courses of action”. According to Janis, groupthink requires high cohesiveness, i.e., close colleagues participating in a group elicitation exercise could be prone to groupthink. While groupthink may not be a conscious behaviour, it is considered to be a form of motivational bias since the judgments of affected group members are hindered by a desire to achieve consensus over accuracy.

3.3.4.2 Group Overconfidence

Group overconfidence is similar to individual overconfidence (discussed in Section 3.3.1.3) and concerns the tendency for groups to provide ranges which are too narrow. Fewer studies have been conducted on group overconfidence compared to individual overconfidence, but nonetheless have included groups in organizational settings, with social prediction tasks, and with simple frequency tasks (Joyner 1992). In some instances, group overconfidence can cause a group to be more confident in the accuracy of their judgments than the individual confident members, particularly following group interaction (Plous 1995). Explanations for individual overconfidence, including anchoring and the informative versus accuracy trade-off, are thought to apply to group overconfidence.

3.3.4.3 Group Polarization

Group polarization was derived from the “risky-shift” phenomenon observed by Stoner (1961) which suggested groups were on the whole more risky than their average individual member. The term ‘risky-shift’ is now considered a misnomer as it has been well established that a shift to greater caution can also be demonstrated for certain choice-dilemmas (Myers and Lamm 1976). Group polarization is a more general concept that concerns the enhancement of the tendency initially held by the majority of group members following group discussions. Joyner (1992) posits that group polarization and group overconfidence are related and can magnify one another when a group shifts to whatever view was taken at the beginning of the discussion and increases in their confidence despite being no more accurate.

3.3.4.4 Group-Level Debiasing Techniques

According to Montibeller and von Winterfeldt (2018), debiasing techniques are similar for the group biases discussed herein and include:

- using multiple experts with alternative points of view (and from different organizations);
- encouraging different perspectives; and,
- using structured elicitation procedures and facilitated decision processes.

Joyner (1992) describes the following approaches for reducing group overconfidence specifically:

- asking group members to consider alternative reasons or outcomes;
- training experts in the form of warnings about overconfidence; and,
- assigning one or more group members the role of devil’s advocate as a critical evaluator of alternative outcomes or options.

The underlying reasons presented by Fischhoff (1982) for biases at the individual level can be extended to biases at the group level, e.g., faults in the process, faults in the group(s), and misalignment between the process and the group(s).

3.4 Elicitation Methods and Protocols

When expert knowledge is elicited, evaluations are commonly made by more than one expert and supported by a facilitator (O'Hagan 2019; Montibeller and von Winterfeldt 2018). The notion that collective judgment can be an improvement on individual judgment is explicit in several contemporary writings for scientific and general audiences and implicit in the reliance on collective judgment in many settings, e.g., representative democracy, juries, and boards of directors (Kerr and Tindale 2011). Conversely, historic and experimental evidence has been published demonstrating that individual judgment can outperform collective judgment in some situations. However, the general consensus in the literature is that, on average, groups outperform individuals on decision tasks (Kerr et al. 1996) and that groups can be superior to individuals for many forecasting tasks (Kerr and Tindale 2011). In the context of probabilistic cost estimating, eliciting probability distributions from multiple experts can improve decision makers' confidence in outcomes of the analysis.

Based on the advantages and justifications discussed above, this thesis focuses on elicitation methods and protocols involving multiple experts. Since it is usually necessary to represent each uncertain variable using a single probability distribution, evaluations from multiple experts are aggregated using one of three approaches: mathematical aggregation, behavioural aggregation, or a combination thereof. The remainder of this section explores the leading elicitation methods and protocols for each approach.

3.4.1 Mathematical Aggregation

Mathematical aggregation is the process of combining probability distributions separately elicited from a number of experts into a single aggregated distribution (O'Hagan et al. 2006). Probability distributions can be mathematically aggregated by pooling (e.g., simple weighted average, weighted geometric mean, etc.) or by applying Bayes' theorem. In either case, mathematical aggregation is not susceptible to the group biases discussed in Section 3.3.4 since experts participate individually.

While a singular subjective probability distribution represents an expert's judgment in the distribution of probabilities, a pooled distribution does not represent the belief of any particular expert. Contrarily, Bayesian approaches can be interpreted to represent the decision maker's beliefs since each expert's distribution is treated as data and these data are used to update the decision maker's prior distribution (Clemen and Winkler 1999). This section explores both Bayesian and opinion pooling approaches to mathematical aggregation.

3.4.1.1 Bayesian Approaches

Many Bayesian approaches have been proposed in the literature (Cooke 1991); in general these approaches have three things in common:

1. they require a decision maker to establish a prior distribution;
2. they consider expert judgments to be observations (i.e., data);
3. they apply Bayes' theorem in some form to update the prior distribution on the basis of the elicited observations.

On the origin of Bayesian aggregation approaches, Clemen and Winkler (1999) write:

> Winkler (1968) provides a Bayesian framework for thinking about the combination of information and ways to assess differential weights. Building on this framework, Morris (1974, 1977) formally establishes a clear Bayesian paradigm for aggregating information from experts.

Morris' approach requires a prior distribution from the decision maker, a probability distribution from each expert, and a likelihood function associated with the experts' information. As Clemen and Winkler (1999) explain, assessing the likelihood function is difficult as it must account for the precision and bias of each expert's distribution as well as dependence among the expert's distributions. O'Hagan et al. (2006) describe Bayesian approaches as "complex and difficult to implement". Given the difficulties associated with assessing a suitable likelihood function, various models have been created to aggregate single probabilities as well as probability distributions (Clemen and Winkler 1999). More recently, focus has shifted towards Bayesian nonparametrics, copulas, and Markov chain Monte Carlo (MCMC) methods (Hartley and French 2018).

3.4.1.2 Opinion Pooling Approaches

Opinion pooling is a widely used form of mathematical aggregation that involves eliciting separate judgments from experts, fitting a probability distribution to each expert's judgments, and combining the fitting distributions into an aggregate distribution using a mathematical formula (O'Hagan 2019). There are a number of mathematical formulas that can be used to combine distributions, each formula grants different properties to the resulting combined distribution.

3.4.1.2.1 Linear Opinion Pooling

The linear opinion pool is a weighted linear combination of expert probabilities, as shown by Equation 16 (Clemen and Winkler 1999), where n is the number of experts, $p_i(\theta)$ represents expert i's probability distribution for variable θ, $p(\theta)$ represents the combined probability distribution, and w_i to w_n are the non-negative weights which sum to one.

Equation 16:

$$p(\theta) = \sum_{i=1}^{n} w_i p_i(\theta)$$

In addition to being easily understood and calculated, the linear opinion pool satisfies several axioms, namely the unanimity property and marginalization property (Clemen and Winkler 1999). The weights in Equation 16 can be used to reflect the importance, validity, or quality of the different experts. If the experts are considered equivalent then Equation 16 becomes the arithmetic average whereas if some experts are considered better (through formal scoring or otherwise) then those experts would receive a higher weighting. Refer to Section 3.4.1.2.4 for more information on determining weights. One drawback of linear opinion pooling is that it is not externally Bayesian which means in light of new data, a decision maker is left with two different resulting combined probability distributions to choose from without any justification for doing so (O'Hagan et al. 2006).

3.4.1.2.2 Logarithmic Opinion Pooling

The logarithmic opinion pool uses multiplicative averaging, as shown by Equation 17, where n is the number of experts, k is a normalizing constant, and the weights w_i to w_n are restricted such that $p(\theta)$ is a probability distribution (Clemen and Winkler 1999).

Equation 17:

$$p(\theta) = k \prod_{i=1}^{n} p_i(\theta)^{w_i}$$

If the experts are considered equivalent (i.e., all weights are equal to $1/n$) then Equation 17 becomes the geometric mean of the individual distributions. The logarithmic opinion pool, unlike the linear opinion pool, satisfies the principal of external Bayesianity which allows for the updating of either the combined distribution or the experts' individual probability distributions (Clemen and Winkler 1999). Conversely, the logarithmic opinion pool does not have the property of coherent marginalization which means assessments of elements comprising a variable produce two different results for the combined probability distribution (O'Hagan et al. 2006).

3.4.1.2.3 Generalized Opinion Pooling

Cooke (1991) presents a formula which generalizes opinion pooling, as shown in Equation 18 and Equation 19, where E is the number of experts, w_i to w_E are the non-negative weights which sum to one, p_{ij} represents expert i's probability distribution, $M_r(j)$ represents the weighted mean, $P_r(j)$ represents the combined probability distribution, and r is any real number.

Equation 18:

$$M_r(j) = \left(\sum_{i=1}^{E} w_i p_{ij}^{r} \right)^{1/r}$$

Equation 19:

$$P_r(j) = \frac{M_r(j)}{\sum_{k=1}^{n} M_r(k)}$$

When $r = 1$, the above becomes the linear opinion pool; when $r \to 0$, the above approaches the logarithmic opinion pool. For other values of r, the above gives other combinations of rules (Cooke 1991). Figure 3 compares combined probability distributions produced from the linear and logarithmic opinion pools with experts equally weighted.

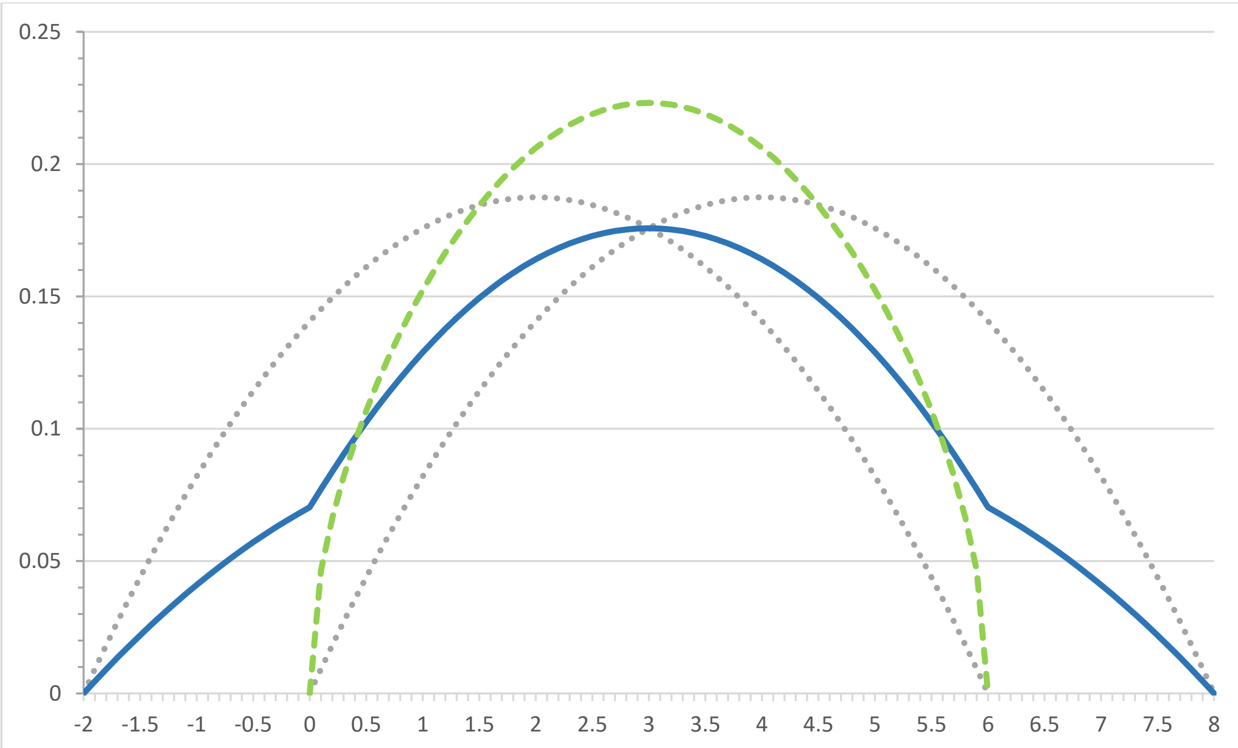


Figure 3: Comparison of opinion pooling approaches adapted from O'Hagan et al. 2006. The dotted lines are experts' individual probability distributions; the dashed green line is the combined distribution from logarithmic opinion pooling and the solid blue line is the combined distribution from linear opinion pooling.

Figure 3 shows that probability distributions produced from linear and logarithmic opinion pooling can vary significantly. This comparison illustrates that logarithmic opinion pooling typically provides a narrower distribution, while linear opinion pooling produces a broader distribution. According to O'Hagan et al. (2006), linear opinion pooling has been widely used in practice while logarithmic opinion pooling has been largely ignored due to perceptions that it produces unrealistically narrow distributions.

3.4.1.2.4 Establishing Expert Weights

The literature is extensive with respect to the weighted combination approaches described in Sections 3.4.1.2.1 to 3.4.1.2.3, however less attention has historically been given to methods of determining the weights themselves (Cooke 1991). Winkler (1968) presents four rules for determining weights:

1. Equal weighting for all experts.
2. Weights proportional to a decision makers' ranking of each experts' 'goodness'.
3. Weights proportional to an experts' self-rating.

4. Weights according to proper scoring rules (i.e., an assessment of each expert's calibration and entropy).

Applying equal weighting for all experts is the easiest and perhaps least contentious rule but it is generally considered more desirable to apply higher weighting to the experts with better performance (O'Hagan et al. 2006).

Cooke's method, also referred to as the Classical model (CM), is one of the most commonly cited and advanced methods of assigning weights using proper scoring rules (Cooke 1991). In the CM, experts are assessed individually against a set of seed questions within their field. Next, based on the accuracy and informativeness of their judgments, experts are assigned weightings which are then used in the mathematical aggregation of probability distributions for the variables of interest (also referred to as target variables).

Seed questions, also known as seed variables or calibration variables, are variables for which the facilitator knows the true value (or will know the true value during the timeframe of the study), but the experts do not. Between eight and twenty seed questions are typically used to assess the performance of experts in expressing their uncertainty using probabilities (Quigley et al. 2018). Seed questions can be classified as either domain or adjacent variables (European Food Safety Authority 2014). Domain seed questions concern variables that are either within the same area of expertise or use the same dimensions as the variables of interest, whereas adjacent seed questions are merely similar to the variables of interest with potentially different dimensions than the variables of interest. Seed questions can also be classified as either predictive or retrodictive (Cooke and Goossens 2000). Predictive seed questions concern future values which will become known within the timeframe of the study, whereas retrodictive questions concern past or present values that are already known to the facilitator. According to Quigley et al. (2018), practical experience indicates that the ideal seed questions are domain predictions, though these can be difficult to find in practice; domain retrodictions and adjacent predictions are considered the next best seed questions, and adjacent retrodictions are considered the weakest.

The CM, like many other elicitation methods and protocols, involves expert training prior to elicitation. As part of expert training in the CM, a facilitator should (Quigley et al. 2018):

- introduce the concept of structured expert judgment;

- describe the problem under investigation and justify why expert judgment is needed to analyze it;
- explain the use of seed questions to assess expert performance;
- warn participants about potential biases in elicitation and in particular overconfidence;
- indicate their preference for less informative, but statistically accurate judgments over informative, but inaccurate judgments; and,
- go through a few example questions with the experts and provide immediate feedback.

After experts have been trained, they are presented with a series of questions (a blend of seed and target variables) and asked to provide quantiles (typically 5th, 95th, and 50th percentiles in sequential order) based on their expert judgment for each variable (Quigley et al. 2018; O'Hagan et al. 2006). Once experts have provided quantiles for all variables, their responses for the seed variables are scored on two components: calibration (also called statistical accuracy) and information. Both components are calculated mathematically using the Kullback-Leibler (KL) distance measure which determines the distance between two probability distributions; for a detailed explanation of calibration and information scoring, refer to Cooke (1991).

The CM is generally well-supported in the literature, with over 200 expert elicitation panels having successfully used seed variables to provide consensus on important issues (Quigley et al. 2018). Cooke and Goossens (2008) provide evidence that distributions formed using CM usually outperform distributions formed using equal weighting (EW) in terms of calibration and informativeness. Some practioners and researchers of elicitation remain unconvinced of the CM. Morgan (2014) cautions that the use of the CM is potentially problematic in situations in which different experts make very different assumptions about underlying causal mechanisms; he also points to research by Clemen (2008) and Lin and Cheng (2009) indicating that the CM performs about as well as EW.

Correlation of experts in mathematical aggregation has received some attention in the literature. O'Hagan et al. (2006) explain the issue using a hypothetical situation wherein two experts have very similar information; the experts would be similarly weighted in the aggregation and their view would be given twice as much weight as it should receive. Cooke (1991) acknowledges correlation between experts as an issue but considers it to be "usually benign, and always unavoidable. Conversely, O'Hagan et al. (2006) write, "groups of similar experts will receive too much weight and minority views will be under-represented".

3.4.2 Behavioural Aggregation

Behavioural aggregation typically involves some form of interaction between experts followed by the development of a probability distribution by consensus. The intention with behavioural aggregation is that experts share their knowledge and interpretations to provide a better distribution than they would otherwise. The simplest behavioural aggregation technique is to convene a group of experts, instruct them to discuss the variable at hand, and ask them to generate a probability distribution for the group (Clemen and Winkler 1999). Most behavioural aggregation techniques are not a set of fixed procedures however there are more sophisticated techniques, such as Delphi or Nominal Group Technique, which can involve more restricted forms of interaction between experts. Depending on the chosen technique, selected experts, and nature of the problem, behavioural aggregation may not succeed in obtaining a single probability distribution for the group, in which case a mathematical aggregation may also be employed.

Since behavioural aggregation involves interaction between multiple experts, it is prone to numerous group biases as discussed in Section 3.3.4; for this reason the exercise is routinely facilitated. Behavioural aggregation avoids some of the challenges inherent in mathematical aggregation, namely deciding on a mathematical formula (e.g., linear versus logarithmic opinion pooling) or method of establishing expert weighting, and is considered a more attractive approach by some. For example, O'Hagan et al. (2006) writes, "group elicitation is advocated by a number of practioners of elicitation (see e.g., O'Hagan (2005), and particularly Phillips (1999))". This section explores some of the most widely cited, structured forms of behavioural aggregation, namely the Delphi method, the nominal group technique (NGT), and Sheffield elicitation framework (SHELF).

3.4.2.1 Delphi Method

The Delphi method, also known as the Delphi technique, is an indirect form of behavioural aggregation involving anonymous response, iterative and controlled feedback, and statistical group response (Dalkey 1969). The Delphi method is one of the oldest approaches to structuring group judgments; the method was developed in the 1950s by Helmer and Dalkey (Ungvarsky

2020). There are many variations of the Delphi method, but in the context of eliciting an aggregated probability distribution, O'Hagan et al. (2006) describe the process as follows:

1. Experts are asked to provide their individual assessment of uncertainty for each variable along with an explanation of their views.
2. All responses are anonymously supplied to the experts.
3. Experts are given the opportunity to adjust their individual responses in light of the other experts' responses.
4. Steps 2 and 3 are iterated until experts converge to a single distribution or a threshold for similarity is met (in which case mathematical aggregation is required to combine the different distributions).

By restricting interactions between experts, the Delphi method avoids some of the group biases that conventional behavioural aggregation methods are prone to. The Delphi method can either be conducted face-to-face or remotely which allows many experts to participate at a relatively low cost, however the iterative nature of the process can significantly increase the amount of time required to complete the study (Ungvarsky 2020). In an ideal scenario, experts would agree after several iterations but this is rarely the case and so mathematical aggregation is typically employed to the experts' distributions after the final round (Clemen and Winkler 1999) requiring consideration of pooling formulas and expert weighting. Criticisms of the Delphi method emerged soon after it gained popularity; Morgan (2014) refers to analyses by Sackman (1974) and Woudenberg (1991) that both questioned whether the consensus obtained by the method is genuine.

3.4.2.2 Nominal Group Technique

NGT is a behavioural aggregation approach similar to the Delphi method which was developed by Delbecq and Van de Ven in 1968 (Delbecq et al. 1975). According to O'Hagan et al. (2006), the NGT is "a variant of Delphi in which each expert presents her views to the group for group discussion". After experts present their views to the group, a facilitated group discussion occurs and then experts are given the opportunity to adjust their individual responses. Similar to the Delphi method, the process can be iterated until consensus is reached or mathematical aggregation is employed.

Like the Delphi method, NGT increases individual participation and avoids some of the group biases that conventional behavioural aggregation methods are prone to. As Delbecq et al. (1975) explain, an advantage of NGT is "the increased attention to each idea and increased opportunity for each individual to assure that his or her ideas are part of the group's frame of reference".

3.4.2.3 SHELF

SHELF is a behavioural aggregation framework involving group training and information sharing, individual judgments, distribution fitting and aggregation, and group feedback (Gosling 2018). Having been developed over several years and using many expert elicitation exercises, SHELF is supported by explanatory documentation, accompanying software, and templates that organize and standardize the process described below:

1. Exercise specification;
2. Expert selection;
3. Training;
4. Information sharing;
5. Individual judgments;
6. Distribution fitting;
7. Discussion and aggregation;
8. Feedback on distribution; and,
9. Iteration of steps 4 to 8 for each additional variable being elicited.

Since it became widely available in 2008, SHELF has been applied in a variety of fields including medicine, business planning, and environmental sciences. Gosling (2018) provides a detailed explanation of SHELF and O'Hagan (2019) presents a case study where an elicitation was conducted according to SHELF. An illustrative example of elicitation using SHELF in the context of probabilistic cost estimating is presented in Section 3.5.

3.4.3 Mixed Aggregation

Mixed aggregation is a combination of mathematical and behavioural aggregation. In cases where behavioural aggregation approaches, such as the Delphi method and NGT, resort to mathematical aggregation, they become forms of mixed aggregation. A relatively new mixed-aggregation approach called the IDEA protocol builds on recent developments in structured elicitation (Hanea et al. 2018b). The IDEA protocol combines classical elements, namely being a Delphi-like protocol and promoting performance-based mathematical aggregation, with modern elements such as remote elicitation. Hanea et al. (2018b) summarize the IDEA protocol as follows:

> "The acronym IDEA arises from the combination of the key features of the protocol that distinguish it from other structured elicitation procedures: it encourages experts to Investigate and estimate individual first round responses, Discuss, Estimate second round responses, following which judgements are combined using mathematical Aggregation".

Given the novelty of the IDEA protocol, real-world applications of the protocol have not been widely published, nor are there numerous studies currently available that compare it to other approaches. Nonetheless, Hanea et al. (2018a) conclude of the IDEA protocol and the issue of correlation between experts (discussed previously in Section 3.4.1.2.4):

> "[t]he data presented here suggest that some of the potential flaws of this new combined approach are not serious impediments to its deployment. In particular, the potential for generating unwanted correlation structures seems to be outweighed by the improvement in the quality of individual estimates, and subsequently (aggregated) in group judgements".

3.4.4 Performance of Aggregation Methods

Throughout Sections 3.4.1, 3.4.2, and 3.4.3, numerous elicitation methods and protocols have been described, along with their advantages and disadvantages, applications in the real world, and level of popularity; this section draws on work by Clemen and Winkler (1999) and O'Hagan et al. (2006) to summarize and compare their performance. Before comparing within and between formal methods, it is worth noting that evidence suggests that formal aggregation methods outperform intuitive aggregation; this is echoed by many in the literature that emphasize the importance of structured elicitation protocols, such as O'Hagan (2019) and Morgan (2014).

Among mathematical aggregation methods, simple averages have been shown to perform as well as or better than more complicated approaches (O'Hagan et al. 2006; Clemen and Winkler 1999). Conversely, more complicated approaches are thought to have the potential to extract more information however the conditions for which such benefits are realized are not clearly understood. Clemen and Winkler (1999) describe Bayesian approaches as an exercise in decomposition and tentatively conclude that the appropriate decomposition of subjective judgments into reasonable tasks may lead to better performance.

Among behavioural aggregation methods, conclusions have been mixed. Much of the literature on group judgment focuses on biases unique to groups discussed in Section 3.3.4 including groupthink and group polarization. The notion that groups are more accurate than the average group member is supported by a number of studies (Clemen and Winkler 1999) however some believe the opposite to generally be the case (O'Hagan et al. 2006). As previously mentioned, studies on the Delphi method have led to mixed results; the consensus is that the Delphi method has no clear advantage over other behavioural aggregation methods (Clemen and Winkler 1999).

Comparisons in the literature between mathematical and behavioural aggregation methods tend to use simple averaging for mathematical aggregation, rather than more complicated approaches; results vary but mathematical aggregation is thought by some to have a slight edge (Clemen and Winkler 1999). There are many real-world applications of both mathematical and behavioural aggregation methods in medicine, environmental sciences, economics, and more. O'Hagan et al (2006) note the following in regards to the debate between mathematical and behavioural aggregation:

- there is no evidence that conclusively favours one method over another;
- the simple average (equal-weighted linear opinion pool) is a simple and robust method for aggregating expert evaluations;
- the CM has the potential to perform better than EW however a well-structured elicitation process is essential to its success; and,
- behavioural aggregation methods have even greater potential however their success relies on skilled facilitation.

While the literature on mixed aggregation methods is sparse in comparison to mathematical and behavioural aggregation methods, novel mixed aggregation methods such as SHELF are promising as they are meant to take the best elements of mathematical and behavioural aggregation.

3.5 Elicitation Protocol Applied to Probabilistic Cost Estimating in Construction

This section presents an illustrative example to demonstrate the use of SHELF in elicitation for probabilistic cost estimating in construction. As previously discussed, SHELF is a behavioural aggregation framework that has been widely applied in health economics and medicine. Being a well-structured behavioural aggregation method, SHELF has the potential to yield some of the best possible elicitation results provided the protocol is appropriately facilitated. Additionally, SHELF is supported by supplementary materials including templates and guidance documents available online (Oakley and O'Hagan 2019). SHELF is divided into three phases: the pre-elicitation phase (described in Section 3.5.1 to 3.5.3), the elicitation phase (described in Section 3.5.4 to 3.5.8), and the post-elicitation phase (described in Section 3.5.9).

3.5.1 Exercise Specification

SHELF begins with identifying the uncertain quantities or variables requiring expert judgment; such variables are usually referred to as quantities of interest (QOI). Once QOIs have been identified, they need to be clearly defined, prioritized, and documented in an evidence dossier (Oakley and O'Hagan 2019).

3.5.1.1 Definitions

According to Oakley and O'Hagan (2019), QOIs need to be defined such that:

1. there is no ambiguity about the variable;
2. the variable will have a unique value; and,
3. the experts' judgments can be made as simple as possible.

Ambiguity in definitions can cause experts to interpret QOIs differently and thus provide assessments based on different variables. Consider the following QOI:

"The cost for track work on Project X assuming a two year construction schedule."

Some ambiguities for this QOI include:

- Should the cost be expressed in dollars or as a percentage relative to the estimated value? If dollars are desired, is it reasonable to assume that since the project is happening in Canada that the currency is Canadian dollars?

- What is the delineation of scope for "track work"? How do I know which components are considered to be track work?
- In which years will the two-year construction take place? Should escalation be accounted for in the response?

The following is a revised definition that removes those ambiguities:

"The cost (expressed as a percentage relative to the base estimate) for track work on Project X with construction beginning April 1, 2023 and ending October 31, 2024. Assume market escalation is accounted for separately. Refer to the 30% design drawing, specifications, and cost estimate included in the evidence dossier for a description of the track work scope."

The revised definition also satisfies the condition that QOIs have a unique value since there is only one cost for the "track work" package belonging to "Project X". If the QOI had been for a program rather than a project, or if the QOI referred to any cost item, then other qualifiers would have been necessary to achieve uniqueness.

To make assessments as simple as possible for evaluators, QOIs may need to be defined on a different scale (e.g. logarithmic) or in terms of two or more other variables for which it is easier for the experts to make judgments (Oakley and O'Hagan 2019). Elaboration is a technique for deriving a probability distribution from two or more quantities. For example, let X be a QOI defined as the cost of subgrade preparation for track work (in Canadian dollars). An expert may prefer to think of X in terms of a volume V (in cubic meters) and a unit cost UC (dollars per cubic meter). Using probability distributions elicited for V and UC and the equation $X = V * UC$, the probability distribution of X can be derived. Elaboration is typically employed where the ease of evaluation or quality of results is expected to improve enough to offset the greater number of variables to be elicited.

3.5.1.2 Many Quantities of Interest

For probabilistic cost estimating, there are likely to be several uncertain inputs at a minimum for which expert judgment is needed. Range estimating typically involves ranging no more than 20 critical items whereas other forms of probabilistic cost estimating involve ranging the full WBS. For more complex projects, the number of items to be ranged can reach the hundreds or thousands. When there are many QOIs, it is not feasible to establish probability distributions using a full SHELF elicitation for all QOIs. In this case, QOIs need to be prioritized and assessed accordingly.

The prioritization process begins by establishing the principal outcome (Oakley and O'Hagan 2019). In the context of cost uncertainty analysis, the principal outcome is typically the project cost. Next, the priority of each QOI is measured by a basic, one-way sensitivity analysis. To conduct the sensitivity analysis, the minimal assessment method is applied for all QOIs. Minimal assessment for a variable X involves two simple assessments: an estimate m_X intended to represent the best estimate, median, or most likely value and an uncertainty measure s_X such that X is approximately twice as likely to be in the range $m_X \pm s_X$ as to be outside that range. For each QOI, the total project cost is calculated once with that QOI set to $m_X - s_X$ and all other QOIs set to their m_X and again with that QOI set to $m_X + s_X$ and all other QOIs set to their m_X. The relative importance of each QOI is then the difference between the two total project costs when that QOI is varied. Table 1 shows the best estimate, uncertainty, and importance of sample QOIs.

Table 1: Assessments of importance for sample QOIs using minimal assessment and one-way sensitivity analysis.

QOI	m_X	s_X	Importance
Work Package A	1000	500	1000
Work Package B	100	50	100
Work Package C	500	250	500
Work Package D	300	150	300
Work Package E	900	300	600
Work Package F	400	100	200
Work Package G	700	350	700
Work Package H	650	200	400

For QOIs with relatively low importance, the uncertainty established using minimal assessment method is considered sufficient given their expected impact on the principal output (Oakley and O'Hagan 2019). Using the sample QOIs from Table 1, suppose Work Packages B and F are considered of low importance. If among the remaining QOIs there are too many variables to be elicited using the full SHELF elicitation, then only those with relatively high importance are earmarked. Continuing with the sample QOIs from Table 1, suppose Work Packages A, E, and G are considered of high importance. The remaining QOIs, if any, are typically elicited using a probabilistic Delphi method after the SHELF workshop(s). If among the remaining QOIs there are too many variables to be elicited using a probabilistic Delphi method, then they are subdivided into two groups: one group to be elicited using a probabilistic Delphi method and one group for which the minimal assessments are revised in light of the discussions and judgments made thus far. Continuing with the sample QOIs from Table 1 once again, suppose Work Packages C and H are considered important enough to be elicited using a probabilistic Delphi method, whereas Work Package D is considered of lower importance. Table 2 indicates the method for each sample QOI. The hierarchy of elicitation methods used in SHELF in order of precedence is (Oakley and O'Hagan 2019):

1. SHELF workshop for the most important QOIs;
2. Probabilistic Delphi for less important, but representative or related QOIs;
3. Revised minimal assessment for QOIs of non-negligible importance; and,
4. Minimal assessment for QOIs of negligible importance.

Table 2: Elicitation methods for sample QOIs based on their importance.

QOI	Importance	Method
Work Package A	1000	SHELF workshop
Work Package B	100	Minimal assessment
Work Package C	500	Probabilistic Delphi
Work Package D	300	Revised minimal assessment
Work Package E	600	SHELF workshop
Work Package F	200	Minimal assessment
Work Package G	700	SHELF workshop
Work Package H	400	Probabilistic Delphi

3.5.1.3 Evidence Dossier

An evidence dossier is a summary of available data for QOIs prepared in advance and provided to experts for their reference during the SHELF workshop(s) (Oakley and O'Hagan 2019). An evidence dossier typically includes an introduction which explains the purpose of the study as well as QOIs and their importance within the study. An evidence dossier also includes a summary of evidence where indirect evidence and lower quality evidence are flagged accordingly. When there are many QOIs, an evidence dossier is typically limited in scope to relatively important QOIs, i.e., QOIs for which a full SHELF elicitation or Probabilistic Delphi method will be used. While the use of an evidence dossier is not mandatory for SHELF, an evidence dossier is recommended as it:

1. reduces the likelihood that evidence is overlooked and documents what is known by the project team and experts;
2. aligns expert's understandings of the QOIs and ensures they all have access to the same body of evidence; and,
3. reduces bias caused by the availability heuristic (as discussed in Section 3.3.1.1).

For probabilistic cost estimating, an evidence dossier should include information pertaining to cost items and other uncertain model inputs. Examples of direct evidence include design drawings, specifications, and cost estimates. An evidence dossier should also include indirect evidence if available, such as variability experienced on similar projects and/or similar cost items. An evidence dossier adapted from Oakley and O'Hagan (2019) as an illustrative example is included in Appendix A.

3.5.2 Expert Selection

Given expert judgment is at the core of an elicitation exercise, recruiting the right experts is very important. In general, enlisting between four and eight experts is recommended for SHELF (Oakley and O'Hagan 2019). In an ideal scenario, potential experts would be thoroughly searched for and shortlisted by the elicitation team, however if there are few experts available, or if the project owner does not have the resources to extend beyond the most readily available experts, then there may not be much choice in the matter.

In the context of probabilistic cost estimating, one of the first places to search for experts is within the project team; suitable candidates include:

- designers (particularly those involved in preparing quantity take-offs);
- estimators;
- risk analysts or other individuals with a thorough understanding of project risks; and,
- project managers or other individuals with a thorough understanding of project characteristics and scope.

For projects with sufficient resources (typically complex or large-scale projects), potential experts can include external cost estimators, contractors, cost analysts, and advisors with experience conducting probabilistic cost estimating.

Once experts have been recruited, experts are sent a briefing package containing a draft evidence dossier (discussed in the previous section), an enquiry form, and an expert briefing. An expert enquiry form asks the expert for their name and title, background and expertise, and declarations of interest as well as additional evidence and clarifications or corrections. If any additional evidence or clarifications are submitted, the evidence dossier is revised and finalized for use in the SHELF workshop. The subject matter included in an expert enquiry form is expected to be fairly consistent regardless of the topic of elicitation. Similarly, the contents of an expert briefing broadly explain the principles of elicitation and uncertain judgments and do not necessarily need to be tailored to the study being conducted. Refer to the "Expert Enquiry" and "SHELF Expert Briefing" forms provided by Oakley and O'Hagan (2019) for templates.

3.5.3 Training

The last step in the pre-elicitation phase is training of experts in a group setting which occurs at the beginning of the SHELF workshop. Training in the SHELF workshop typically involves at least one practice elicitation including the use of forms and tools (Gosling 2018). The practice elicitation can involve either general knowledge quantities (e.g., population or size of a country, distance between cities, etc.) or quantities similar to QOIs. Consider the following practice QOI:

"The percent change in the Alberta non-residential construction deflator that occurred in 2019."

For longer term projects undertaking probabilistic cost estimating, future escalation rates are often relevant. A past escalation rate is a good candidate for practice elicitation since it is similar to potential QOIs (future escalation rates) and is knowable. For instance, the City of Edmonton (2020) found that the Alberta non-residential construction deflator increased by 0.7% in 2019.

The facilitator should carefully explain the probabilistic judgments required and any contradictions within and between experts' judgments. The facilitator should also use the practice elicitation as an opportunity to educate participants on the various biases that can affect elicitation outcomes, particularly forms of biases which can be reduced through training such as availability, overconfidence, and confirmation bias. Training on overconfidence can involve identifying and questioning experts with relatively narrow uncertainty ranges and then revealing the true value of the practice QOI to encourage a shift in opinion (Gosling 2018).

3.5.4 Information Sharing

Before making judgments about QOIs, experts are asked to share their responses to a series of questions which concern (Gosling 2018):

- participant interests related to the QOIs;
- participant expertise related to the QOIs;
- factors relevant to judgments of the QOIs;
- quantitative and/or qualitative evidence related to the QOIs; and,
- the structure of QOIs and whether adjustments can be made to improve the ease of evaluation.

Information sharing is meant to provide the owner with an understanding of the group members and their expertise, reiterate the purpose of the study and relevant sources of information, and improve the confidence of group members in each other. Once responses have been documented, they form part off the written record of the elicitation. Refer to the "SHELF 1" form provided by Oakley and O'Hagan (2019) for a template.

3.5.5 Individual Judgments

The first round of SHELF judgments is made by experts individually. SHELF supports the assessment of many different types of variables including continuous, discrete, and correlated quantities. In the context of probabilistic cost estimating, QOIs are generally continuous rather than discrete. Depending on the complexity of the project and type of probabilistic cost estimating methodology (i.e., range estimating versus ranging of entire WBS), QOIs may or may not be independent. If QOIs are dependent on one another, then either multivariate elicitation is necessary or correlation will need to be accounted for in the modelling approach (refer to Chapter 4 for detailed information on such an approach). For the illustrative example presented in this section, QOIs are continuous and independent.

For continuous QOIs, SHELF provides three methods of eliciting probability judgments individually: the tertile method, quartile method, and roulette method. Each method begins by asking the expert for their plausible limits for the QOI. The plausible limit consists of a lower plausible limit, L, and an upper plausible limit, U, where the value could theoretically reside outside the limits established but with such an occurrence being considered extremely unlikely. Plausible limits are established as the first judgment to avoid two common forms of bias: overconfidence and anchoring (as discussed in Section 3.3.1). Refer to the SHELF package provided by Oakley and O'Hagan (2019) for a description of each method and their advantages and disadvantages.

The quartile method is demonstrated as follows using a sample QOI, the cost (expressed as a percentage relative to the base estimate) of Work Package A, from the evidence dossier included in Appendix A. First, each expert is asked for their lower and upper plausible limits for the QOI. Next, each expert is asked to provide their median value, M, i.e., the value that they believe yields an equal probability of having the true value be either above or below. After providing a median value, each expert is asked to establish lower and upper quartiles, $Q1$ and $Q3$ respectively. $Q1$ is a value between L and M that the expert believes yields an equal probability of having the true value be either below $Q1$ or between $Q1$ and M. Similarly, $Q3$ is a value between M and U that the expert believes yields an equal probability of having the true value be either between M and $Q3$ or above $Q3$. Table 3 shows the plausible limits, median, and quartiles from various experts for Work Package A.

Table 3: Plausible limits, median, and quartiles from various experts for the cost of Work Package A (expressed as a percentage relative to the base estimate).

Expert	**Lower Plausible Limit (L)**	**Upper Plausible Limit (U)**	**Median (M)**	**Lower Quartile (Q1)**	**Upper Quartile (Q3)**
Expert A	50%	200%	100%	80%	150%
Expert B	70%	130%	100%	90%	110%
Expert C	70%	150%	110%	95%	125%
Expert D	50%	150%	100%	80%	130%
Expert E	40%	250%	120%	100%	170%

3.5.6 Distribution Fitting

After experts have made their individual judgments, distributions are fitted to their responses. Distributions can be fit using a variety of different procedures; the least-squares fitting procedure is commonly used (Gosling 2018). The least-squares fitting procedure involves comparing the elicited values against theoretical values from a candidate probability distribution with particular parameters. The parameter values corresponding to the lowest squared difference between elicited and theoretical values are chosen. With a finite number of elicited values, there are many distributions that will minimize the squared difference between elicited and theoretical values so the field of options is typically narrowed to a distribution or family of distributions. SHELF provides an R package which allows for the storing and fitting of expert judgments (Oakley 2021). Figure 4 shows fitted distributions for Work Package A corresponding to individual expert judgments shown in Table 3 as well as a linear opinion pool (discussed previously in Section 3.4.1.2.1). The full R package script for the illustrative example is included in Appendix A.

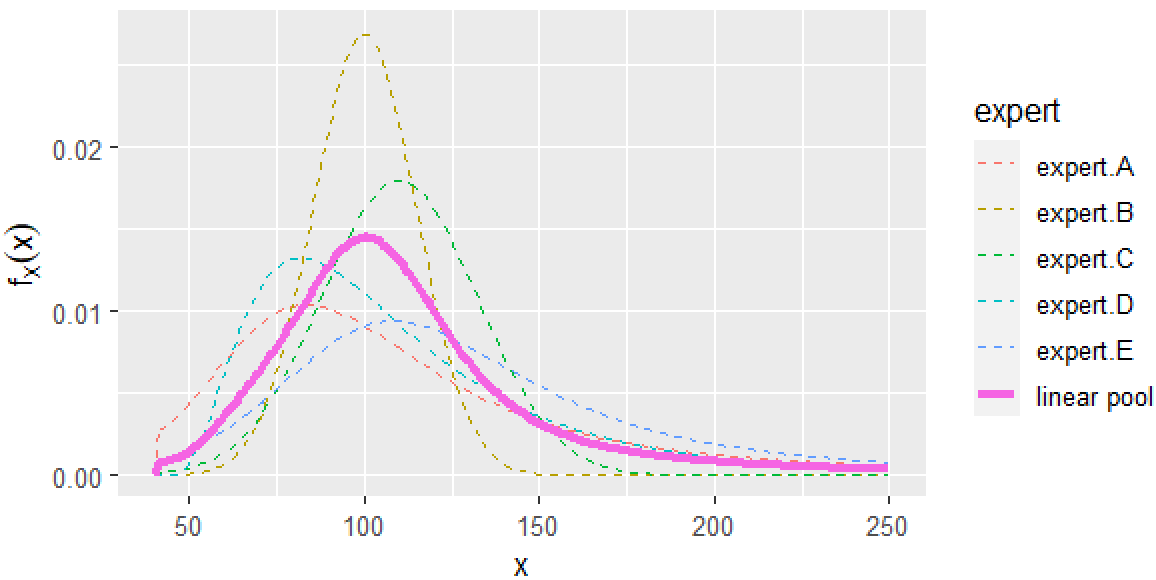


Figure 4: Fitted distributions corresponding to expert judgments, including a linear opinion pool, for the cost of Work Package A. Image generated using the R package "SHELF: Tools to Support the Sheffield Elicitation Framework" developed by Oakley (2021).

3.5.7 Discussion

After individual distributions have been fitted, they are shown to the experts as a group and followed by discussion focused on differences of opinion. For each area of the fitted distribution plot where there is a divergence of opinion, the group is encouraged to seek to understand each other's perspectives and understand the reasons for such divergence (Oakley and O'Hagan 2019). Using the distributions shown in Figure 4 as an example, Expert B's distribution is much more concentrated around a cost of 100% relative to the base cost estimate and so the facilitator could ask Expert B to explain their reason for such confidence and/or ask the other experts to explain why they think higher or lower values are more plausible.

Some practitioners of SHELF choose to share the pooled distribution with the group at this point, while others find that this can be distracting to experts and may be used as a fall-back by experts whom prefer more objective methods for combining opinions (Gosling 2018). Regardless, statistics for the pooled distribution, such as median and quartiles, can be collected and used during the next stage. For example, the pooled distribution shown in Figure 4 has a median, lower quartile, and upper quartile corresponding to 105.9%, 87.8%, and 129.2%, respectively.

3.5.8 Feedback

Two of the elicitation methods used in individual judgments, tertile and quartile, also apply to group judgments. Additionally, SHELF accommodates the probability method which involves asking the experts to make group judgments of three probabilities to encourage the experts to think directly about probabilities (Oakley and O'Hagan 2019). The facilitator, along with the other members of the elicitation team, determine which elicitation method will be used.

In keeping with the illustrative example, the quartile method is demonstrated for a sample QOI, the cost (expressed as a percentage relative to the base estimate) of Work Package A, from the evidence dossier included in Appendix A. First, the facilitator seeks to find a median value that the experts can agree on as being representative of the group's views. Oakley and O'Hagan (2019) emphasize the concept of a rational impartial observer (RIO). Using this concept, experts are not expected to reach complete agreement about an uncertainty quantity but rather they are asked to judge what a RIO might reasonably believe, having seen the experts' individual judgments and listened to the group discussion. After the group reaches consensus on a median value, the process is repeated for lower and upper quartiles. Once group judgments have been established, a distribution is fitted and shown to the experts along with statistics and suitable feedback. Suppose the group provides the following judgments for the cost of Work Package A:

- a median equal to 105%;
- a lower quartile equal to 85%; and,
- an upper quartile equal to 140%.

As previously mentioned, there are many distributions that will minimize the SSQ between elicited and theoretical values. Table 4 shows the sum of squared differences (SSQ) between elicited and fitted probabilities for the cost of Work Package A.

Table 4: Sum of squared difference between elicited and fitted probabilities for the cost of Work Package A (expressed as a percentage relative to the base estimate).

Normal	Student's-t	Gamma	Log-Normal	Log-Student's-t	Beta
0.00296	0.00288	0.00083	0.00025	0.00024	0.00170

Based solely on the SSQ, any of the distributions shown in Table 4 would likely be judged as adequate in this case, however further considerations affect the suitability of a distribution. For example, the log-Student's-t distribution, shown in Figure 5, has the potential to far exceed the upper plausible limit of 250% established by experts. The log-normal distribution has the same issue.

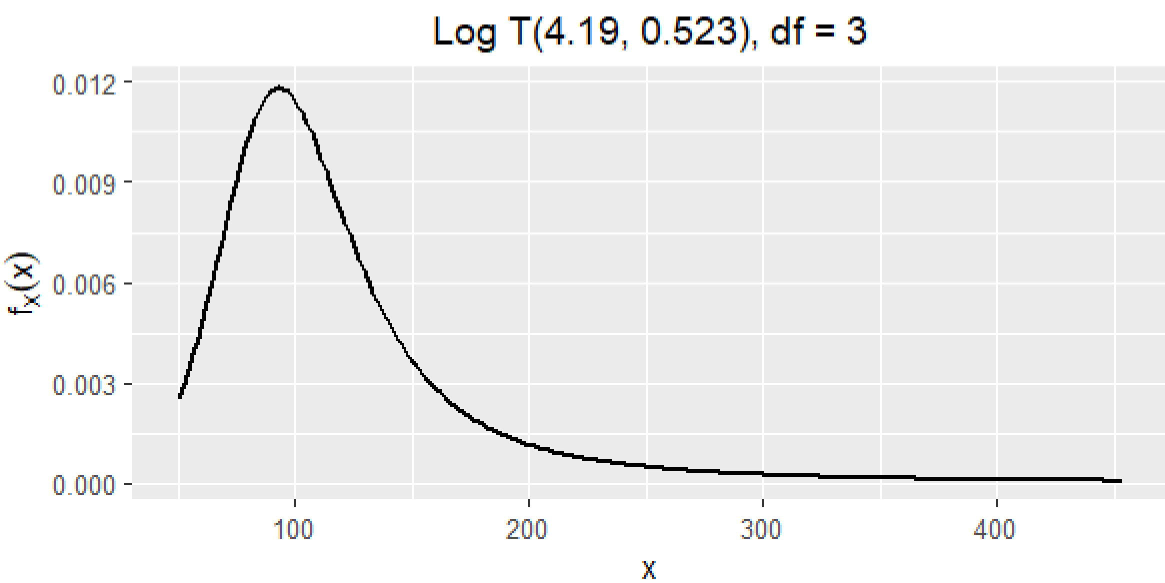


Figure 5: A fitted log-Student's-t distribution for the cost of Work Package A (expressed as a percentage relative to the base estimate). Image generated using the R package "SHELF: Tools to Support the Sheffield Elicitation Framework" developed by Oakley (2021).

In order of increasing SSQ, the next best distributions for this case are the gamma and beta distributions. The gamma distribution, shown in Figure 6, still has the potential to exceed the upper plausible limit established by experts, however this occurs to a lesser extent compared to the log-Student's-t and log-normal distributions. The beta distribution, shown in Figure 7, is the most promising in this case as it eliminates the issue of exceeding plausible limits altogether.

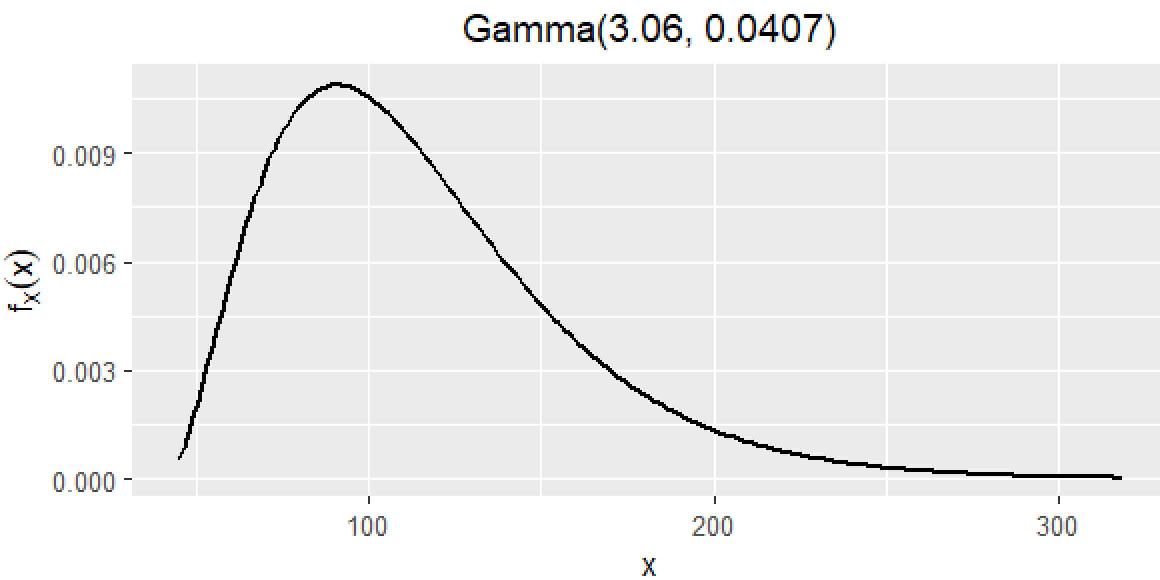


Figure 6: A fitted gamma distribution for the cost of Work Package A (expressed as a percentage relative to the base estimate). Image generated using the R package "SHELF: Tools to Support the Sheffield Elicitation Framework" developed by Oakley (2021)

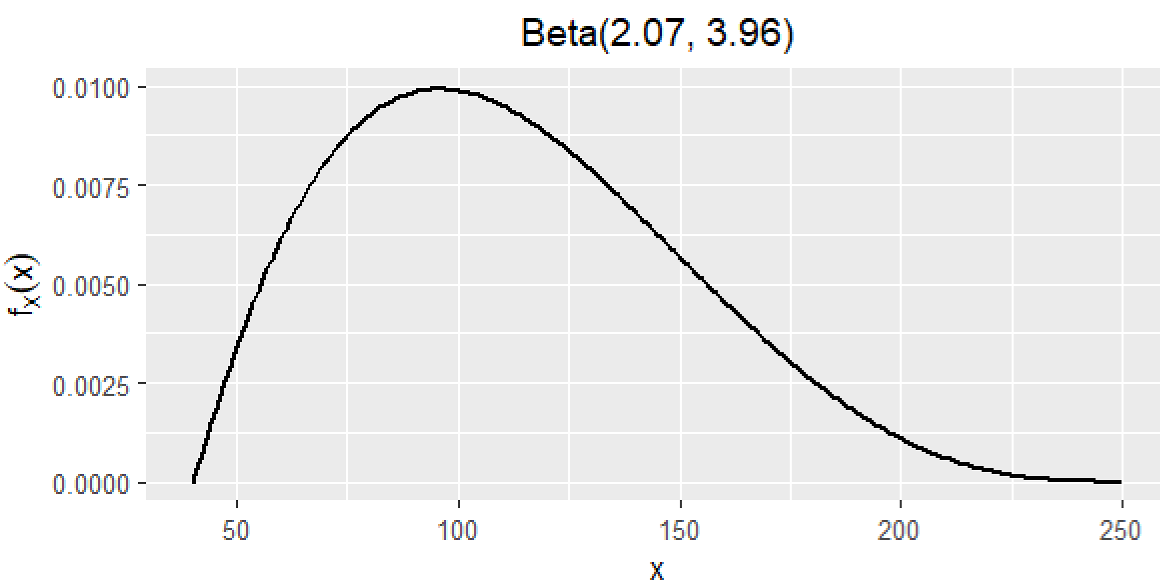


Figure 7: A fitted beta distribution for the cost of Work Package A (expressed as a percentage relative to the base estimate). Image generated using the R package "SHELF: Tools to Support the Sheffield Elicitation Framework" developed by Oakley (2021).

Using the 'feedback' function of the SHELF R package, fitted percentiles and probabilities can be obtained and used to further assess the suitability of fitted distributions. Fitted percentiles and fitted probabilities for the cost of Work Package A are summarized in Table 5 and Table 6, respectively.

Table 5: Cost for Work Package A (expressed as a percentage relative to the base estimate) from fitted distributions corresponding to 5th and 95 percentiles.

	Cost from Fitted Distributions					
Percentile	**Normal**	**Student's-t**	**Gamma**	**Log-Normal**	**Log-Student's-t**	**Beta**
5th	40.6%	22.5%	60.9%	64.9%	59.3%	57.3%
95th	178.0%	195.0%	197.0%	215.0%	266.0%	180.0%

Table 6: Probabilities from fitted distributions corresponding to elicited values for the cost of Work Package A (expressed as a percentage relative to the base estimate).

	Probability from Fitted Distributions						
Value	**Elicited**	**Normal**	**Student's-t**	**Gamma**	**Log-Normal**	**Log-Student's-t**	**Beta**
85%	0.25	0.280	0.281	0.265	0.258	0.258	0.272
105%	0.50	0.459	0.462	0.478	0.488	0.489	0.469
140%	0.75	0.769	0.762	0.762	0.757	0.757	0.766

The issue raised for the log-normal and log-Student's-t distributions during visual inspection is also evident in the 95th percentile costs shown in Table 5. The 95th percentile of 215% for the log-normal distribution is quite close to the upper plausible limit of 250%, while the 95th percentile of 266% for the log-Student's-t distribution exceeds it. As expected based on the relatively low SSQ values in Table 4, each of the fitted distributions shown in Table 6 produce probabilities that are relatively close to the elicited quartiles. For this illustrative example, the elicited values will be fitted to a beta distribution.

In addition to visual inspection of plots, experts often find it useful to be given information about the fitted distribution that is similar to the format of the briefing and/or elicitation questions (Gosling 2018). Examples of useful statements or summaries for this illustrative example include:

"According to your judgments, there is a 10% chance of Work Package A having costs outside the range (57%, 180%)."

"Your responses suggest that a cost of 205% for Work Package A is highly unlikely; is this reasonable?"

Several rounds of iteration as a group may be necessary to obtain a distribution that the group agrees appropriately represents their judgments, or at least what a RIO might reasonably believe. Once the group agrees on a fitted distribution for the QOI, the elicitation for that quantity is completed and the SHELF judgments are repeated for each other QOI.

3.5.9 Debrief

After the SHELF workshop is completed, templates are completed and distributed to the participating experts for their review. It is also considered good practice to send participants a feedback questionnaire to assess how the workshop was managed and how it may be improved (Oakley and O'Hagan 2019). The templates, along with their attachments, form the workshop record and are typically a portion of the final report to the client along with the actual outcomes of the probabilistic cost estimating analysis.

3.6 Conclusion

This chapter introduced inputs for probabilistic cost estimating, explained challenges associated with eliciting such inputs and how they can be overcome, and demonstrated the application of an elicitation protocol to probabilistic cost estimating in construction using an illustrative example. The illustrative example showcased a critical component of the cost uncertainty analysis process: establishing probability distributions to model cost uncertainty. Lastly, this chapter set the stage for the probabilistic cost estimating model in Chapter 4 which investigates the affect that model structure, inputs, and methodologies have on cost uncertainty outcomes.

Chapter 4 – Monte Carlo Simulation Experiment

4.1 Introduction

As previously noted, Monte Carlo simulation is a modelling technique involving the use of random numbers (Law 2015). Monte Carlo simulations rely on repetition of an experiment where random variates are sampled from probability distributions for use as variables within the model and the desired output parameters of a model are calculated. Cost uncertainty is often analyzed using Monte Carlo simulation, but Monte Carlo simulation can be used for a variety of other experiments and problems.

This chapter presents a Monte Carlo simulation experiment to test the hypothesis that since cost uncertainty (measured by standard deviation) is affected by correlation between model inputs, modelling correlated inputs independently results in underestimation of cost uncertainty. The simulation outputs a set of randomly generated cost items which are added together to yield a set of project costs. Each set of cost items represents the components of a sample construction project and the corresponding set of project costs represents the cost uncertainty experienced by that sample project. Various scenarios can be modelled by repeating the simulation with either random sampling or iteration of one or more target variables. The experiment is automated using code developed in RStudio (version 1.4.1103); an overview of the model is shown in Figure 8. Model outputs are used to assess the performance of different correlation modelling methods and specifically address the following:

- Which method produces the least error in terms of standard deviation or percentiles typically used to establish contingency (e.g., P80)?
- How does the number of work packages affect the extent of underestimation or overestimation for each method?
- Is there a number of work packages for which an independence assumption yields an acceptable result?

The following sections describe the process of generating cost item data, collecting statistics and generating samples, applying correlation methods, and evaluating each method relative to the generated data. A base model scenario is used to illustrate the execution of the model. The following sections also introduce various modelling scenarios, present scenario results, address model verification and validation, and discuss key findings.

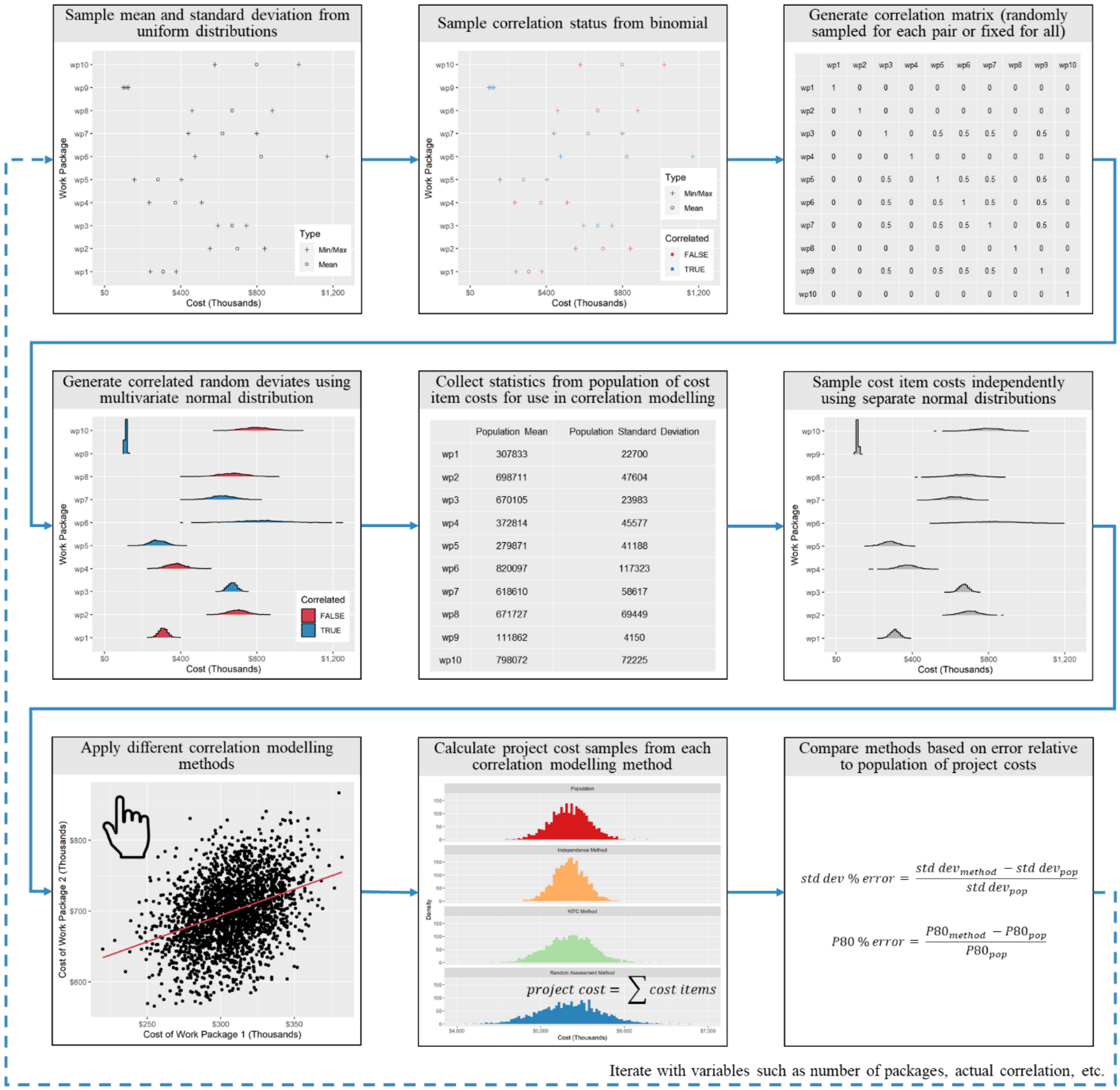


Figure 8: Overview of Monte Carlo simulation experiment.

4.2 Generating Correlated Cost Item Data

To generate correlated cost item data, the characteristics of a project and its cost items need to be determined. Each project is made up of a number of cost items each with their own locations (e.g., mean or median) and spread (e.g., standard deviation). Some or all of the cost items can be correlated with one another and the extent of correlation between cost items varies. This section describes the inputs involved in generating cost item data, how inputs are determined for a given scenario, and the actual process of cost item data generation.

4.2.1 Number of Cost Items

The number of cost items for a project varies based on many factors including project size, project complexity, the level of desired detail in the WBS, and the structure of the WBS itself (e.g., splitting cost items by geographic location, discipline, etc.). Very simple, small projects may only have a few cost items whereas very complex, large projects may have thousands of cost items.

Depending on the model scenario, the number of cost items in the model is either:

- fixed (e.g., set to 10 for all iterations);
- varied stepwise (e.g., set to 10 for a group of iterations, set to 30 for another group of iterations, and set to 100 for another group of iterations);
- varied iteratively (e.g., set from five to 100 in increments of one); or,
- randomly sampled (e.g., set using a random deviate of the uniform distribution on the interval [5, 100] rounded to the nearest integer).

Where the number of cost items is varied iteratively or randomly sampled, the values range from five to 100 cost items, representing a broad range of project sizes, complexities, etc. The number of cost items (in addition to the location and spread of costs) will determine the project cost for each iteration. For the base model scenario presented in this section, the number of cost items is fixed at 10.

4.2.2 Cost Item Values

Cost items vary significantly in value depending on the project and WBS. Ultimately, costs in the model are generated for each cost item using random deviates of a probability distribution (described in detail in Section 4.2.5). The parameters needed to generate a cost depend on the probability distribution selected for the model. For a normal distribution, two parameters are required: mean and standard deviation.

The mean for each cost item in the model is randomly sampled using a random deviate of the uniform distribution between \$100,000 and \$900,000, representing a broad range of cost item sizes. The standard deviation for each cost item in the model is also randomly sampled using a random deviate of the uniform distribution with the interval being a function of cost accuracy and the cost item mean. The minimum standard deviation is 10% of the mean divided by three and the maximum standard deviation is 50% of the mean divided by three; such an interval ensures that 99.7% of values fall within the boundaries of a cost accuracy between +/- 10% and +/- 50% (since 99.7% of values are within three standard deviations of the mean). The mean and standard deviation (in addition to the number of cost items) will determine the project cost for each iteration.

Figure 9 shows the mean and measure of spread (which is ultimately a function of standard deviation) of cost items for the base model scenario; Table 7 shows the mean and standard deviation of cost items for the same scenario.

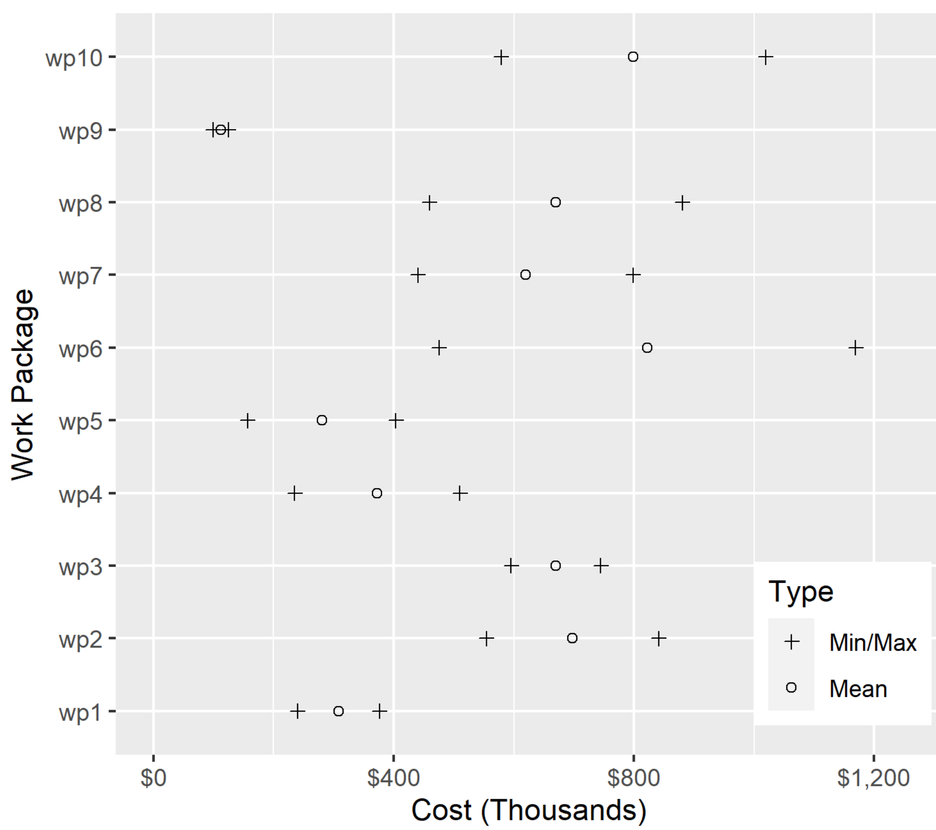


Figure 9: Mean, minimum, and maximum costs for cost items in the base model scenario.

Table 7: Mean and standard deviation of costs for cost items in the base model scenario.

Cost Item	Cost Mean	Cost Standard Deviation
Work Package 1	308,117	22,710
Work Package 2	697,931	47,772
Work Package 3	670,167	24,912
Work Package 4	372,510	45,922
Work Package 5	280,154	41,179
Work Package 6	822,434	115,636
Work Package 7	619,276	59,688
Work Package 8	669,998	70,206
Work Package 9	111,863	4,235
Work Package 10	799,120	73,370

4.2.3 Proportion of Correlated Cost Items

The proportion of cost items that are correlated for a project varies depending on many factors including the interdependency of project components, the level of desired detail in the WBS, and the type of cost uncertainty analysis being performed (recall that range estimating focuses on only critical items and requires related items to be linked or combined). Generally speaking, unless cost items have been expressly established to be independent, the proportion of correlated cost items for a project is likely non-zero. The proportion is a fraction defined from zero to one; for example a proportion of 0.5 indicates that 50% of cost items are correlated.

Depending on the model scenario, the proportion of correlated cost items is either:

- fixed (e.g., set to 1 for all iterations); or
- varied iteratively (e.g., set from zero to one in increments of 0.1).

Once the proportion of correlated cost items has been determined for an iteration, the correlation status for each cost item is randomly sampled using a random deviate of the binomial distribution where a value of zero (i.e., false) indicates the cost item is independent and a value of one (i.e., true) indicates the cost item is correlated. The correlation status (in addition to the correlation strength) will determine the correlation matrix for each iteration. In the base model scenario, the proportion of correlated cost items is fixed at 0.5. Figure 10 shows the correlation status of cost items for the base model scenario.

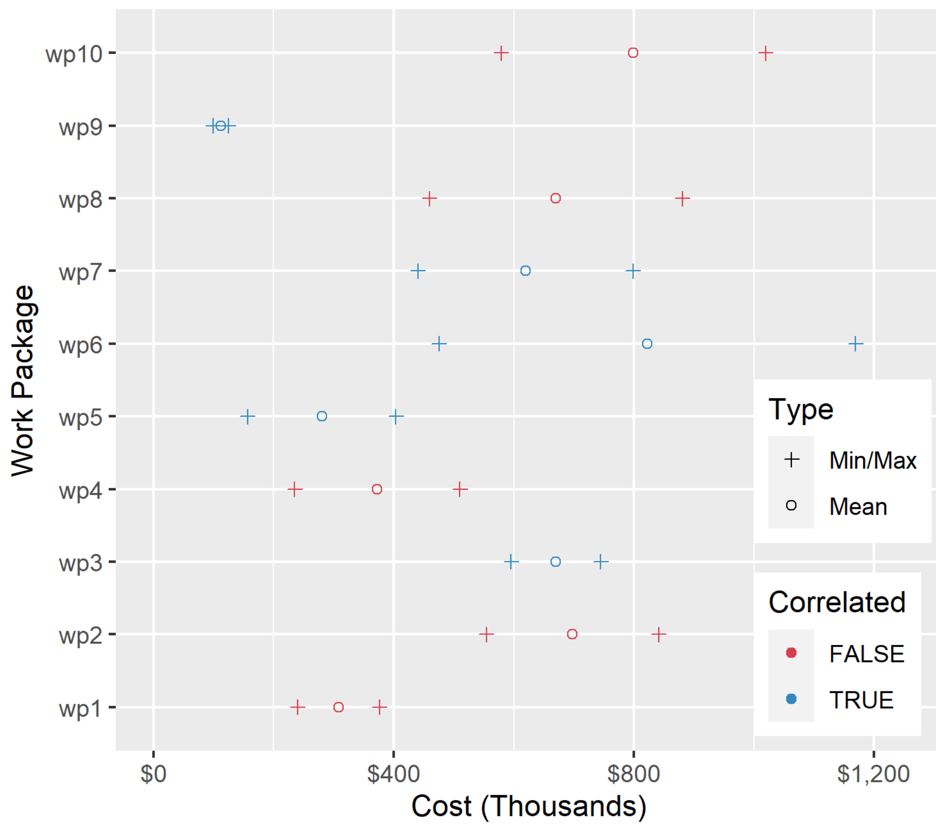


Figure 10: Correlation status for cost items in the base model scenario.

4.2.4 Strength of Correlation

The strength of correlation between cost items varies based on many factors including the interdependency of quantities within cost items and whether a pair of cost items shares materials or labour pools. Correlation strength is a continuous measure with values between negative one and one; variables represented using certain types of probability distributions, e.g. bivariate lognormal, can have bounds within this interval (Garvey et al. 2016). Positive correlation is on the interval (0,1] and indicates that a cost increase in one item is associated with a cost *increase* in another item. Conversely, negative correlation is on the interval [-1, 0) and indicates that a cost increase in one item is associated with a cost *decrease* in another item. Uncorrelated or independent cost items have a correlation equal to zero.

Correlation strength values in the model range from zero to one as positive correlations are more common and of particular concern from a cost underestimation perspective. Correlation strength is captured in a correlation matrix which has a number of important properties (Numpacharoen and Atsawarungruangkit 2012):

- the matrix must be symmetric;
- diagonal elements must be equal to one;
- non-diagonal elements must be in the interval [-1, 1]; and,
- the matrix must be positive semi-definite (i.e., the matrix must be internally consistent with all eigenvalues greater than or equal to zero).

The model provides two approaches for determining the correlation strength and creating the correlation matrix: the common correlation strength option and the random correlation strength option.

4.2.4.1 Common Correlation Strength

In the first approach, referred to as the common correlation strength option, the correlation strength is set common to all pairs of correlated cost items. Depending on the model scenario, the common correlation strength is either:

- fixed (e.g., set to 0.5 for all iterations);
- varied iteratively (e.g., set from zero to one in increments of 0.01); or,
- randomly sampled (e.g., set using a random deviate of the uniform distribution on the interval [0, 1]).

Once the correlation strength has been determined for an iteration, an n by n matrix is created (where n is the number of work packages for the iteration) with each matrix element set to the correlation strength. Next, each element of the matrix that involves an independent cost item (i.e., elements with a correlation status of zero) is set to zero since they have no correlation with any other cost item. Finally, diagonal elements of the matrix are set to one since each cost item is perfectly correlated with itself. In the base model scenario, the correlation strength is common for all correlated items and fixed at 0.5. Table 8 shows the correlation matrix of cost items for the base model scenario.

Table 8: Correlation matrix for cost items in the base model scenario.

	wp1	wp2	wp3	wp4	wp5	wp6	wp7	wp8	wp9	wp10
wp1	1	0	0	0	0	0	0	0	0	0
wp2	0	1	0	0	0	0	0	0	0	0
wp3	0	0	1	0	0.5	0.5	0.5	0	0.5	0
wp4	0	0	0	1	0	0	0	0	0	0
wp5	0	0	0.5	0	1	0.5	0.5	0	0.5	0
wp6	0	0	0.5	0	0.5	1	0.5	0	0.5	0
wp7	0	0	0.5	0	0.5	0.5	1	0	0.5	0
wp8	0	0	0	0	0	0	0	1	0	0
wp9	0	0	0.5	0	0.5	0.5	0.5	0	1	0
wp10	0	0	0	0	0	0	0	0	0	1

4.2.4.2 Random Correlation Strength

In the second approach, referred to as the random correlation strength option, the correlation strength is randomly determined for each pair of correlated cost items. The simplest method for creating a randomly generated correlation matrix is the rejection sampling method (Numpacharoen and Atsawarungruangkit 2012). The rejection sampling method uses random deviates of the uniform distribution on a specified interval to set each non-diagonal matrix element, checks whether the matrix is semi-definite, and, if the matrix is not semi-definite, generates a new matrix until a valid matrix is obtained. The rejection sampling method is sufficient for a low-dimensional matrix (i.e., three or fewer variables) but quickly becomes inefficient for large-dimensional problems. There are many different techniques for creating correlation matrices which can be classified by their objectives or constraints as follows (Numpacharoen and Atsawarungruangkit 2012):

- a matrix with predetermined eigenvalues and spectrum;
- a matrix with a given mean value;
- a matrix based on random Gram matrix; and,
- a matrix in which each element is distributed within its boundaries.

The model uses an R package developed by Makalic and Schmidt (2018) called randcorr for generating a random correlation matrix using Cholesky factorization and angles, sampling of angles from a distribution, and conversion to standard correlation matrix form. The package implements an algorithm by Pourahmadi and Wang (2015) but with angles being sampled using an efficient sampling algorithm (Makalic and Schmidt 2020). The sampling algorithm utilizes rejection sampling where the envelope distribution is a scaled symmetric Beta distribution. Correlation values within the randomly generated correlation matrix are generated on the interval [-1, 1] with a tendency towards central values.

Since the correlation coefficients produced by the *randcorr* function are on the interval [-1, 1], the model scales the correlation matrix to the determined interval for all iterations. For example, if the minimum bound, ρ_{min}, and maximum bound, ρ_{max}, of correlation strength are fixed at zero and one, respectively, for all iterations, then each matrix element would increase by one and be divided by two. A general form of the coefficient scaling is shown in Equation 20 where $\rho_{Y,X}$ is the correlation coefficient for two cost items, X and Y, generated by the *randcorr* function and $\rho_{Y,X}'$ is the scaled correlation coefficient for the same two cost items.

Equation 20:

$$\rho_{Y,X}' = \frac{(\rho_{Y,X} + 1) \times (\rho_{max} - \rho_{min})}{2} + \rho_{min}$$

Next, each element of the matrix that involves an independent cost item (i.e., elements with a correlation status of zero) is set to zero since they have no correlation with any other cost item. The *randcorr* function produces a correlation matrix where diagonal elements are already set to one, so no adjustment of these elements is necessary. Table 9 shows the correlation matrix in a scenario where the number of cost items is fixed at 10, the correlation strength is randomly generated on the interval [0, 1] for each cost item pair, and the proportion of correlated cost items is fixed at 0.5; this scenario is similar to the base model scenario except that the correlation strength is randomly determined for each pair of correlated cost items rather than set common to all pairs of correlated cost items and fixed.

Table 9: Correlation matrix in a scenario where the number of cost items is fixed at 10, the correlation strength is randomly generated on the interval [0, 1] for each cost item pair, and the proportion of correlated cost items is fixed at 0.5.

	wp1	wp2	wp3	wp4	wp5	wp6	wp7	wp8	wp9	wp10
wp1	1	0	0	0	0	0	0	0	0	0
wp2	0	1	0	0	0	0	0	0	0	0
wp3	0	0	1	0	0.592	0.581	0.611	0	0.402	0
wp4	0	0	0	1	0	0	0	0	0	0
wp5	0	0	0.592	0	1	0.600	0.440	0	0.510	0
wp6	0	0	0.581	0	0.600	1	0.303	0	0.411	0
wp7	0	0	0.611	0	0.440	0.303	1	0	0.419	0
wp8	0	0	0	0	0	0	0	1	0	0
wp9	0	0	0.402	0	0.510	0.411	0.419	0	1	0
wp10	0	0	0	0	0	0	0	0	0	1

4.2.5 Generating Correlated Random Deviates

Once the cost item inputs are determined for an iteration, generating correlated costs is relatively simple. In the model, random deviates are generated for cost items using a multivariate normal distribution. First, a vector of cost item means is created. Next, a standard deviation matrix is created where diagonal elements are the standard deviation of cost items. A covariance matrix is then created by pre-multiplying and post-multiplying the correlation matrix by the standard deviation matrix since the covariance between two cost items is equal to the corresponding correlation coefficient multiplied by the product of the standard deviations of the two cost items. Table 10 shows the covariance matrix of cost items for the base model scenario. Lastly, the *rmvnorm* function receives the mean vector and covariance matrix and generates m random deviates for each cost item per iteration where m is the size of the population of costs for the model.

Table 10: Covariance matrix for cost items in the base model scenario. Values are shown in millions.

	wp1	wp2	wp3	wp4	wp5	wp6	wp7	wp8	wp9	wp10
wp1	516	0	0	0	0	0	0	0	0	0
wp2	0	2282	0	0	0	0	0	0	0	0
wp3	0	0	621	0	513	1440	743	0	53	0
wp4	0	0	0	2109	0	0	0	0	0	0
wp5	0	0	513	0	1696	2381	1229	0	87	0
wp6	0	0	1440	0	2381	13372	3451	0	245	0
wp7	0	0	743	0	1229	3451	3563	0	126	0
wp8	0	0	0	0	0	0	0	4929	0	0
wp9	0	0	53	0	87	245	126	0	18	0
wp10	0	0	0	0	0	0	0	0	0	5383

Statistical literature provides guidelines for determining a sufficient number of deviates for a Monte Carlo simulation depending on the desired confidence interval. The confidence interval is an interval for which there is an associated confidence level (e.g., 95%) that the interval includes a population parameter (e.g., mean). In the context of generating correlated random deviates in this model, the population parameters for a given cost item are the randomly determined mean and standard deviation for that cost item.

Byrne (2013) provides a systematic approach to choosing the number of deviates based on confidence interval width. Assuming a 95% level of confidence, a confidence interval width of 0.01 requires 9604 iterations. The number of deviates required decreases to 2401, 385, and 97 for confidence interval widths of 0.02, 0.05, and 0.10 respectively. Assuming a 95% confidence level, a confidence interval width of 0.02 indicates that there is a 95% probability that the true median ($p = 0.50$) of a cost item falls between the 48^{th} and 52^{nd} percentile of random deviates.

In this model, the set of random deviates generated for a cost item represents the *population* of costs for that cost item rather than *samples* of costs. Statistics collected on the random deviates will be used as population parameters to generate independent samples for each cost item (described in detail in Section 4.3) and compare different correlation modelling methods (described in detail in Section 4.5). Since the mean and standard deviation for a cost item are determined randomly and the random deviates are the basis of comparison for subsequent calculations, similarity between the randomly determined parameters and statistics of the random deviates is not necessary and therefore the width of the confidence interval in this context is not important. Conversely, choosing an excessive number of deviates will increase the overall computational time of the model. Setting the number of random deviates for each cost item to 2401 provides a confidence interval width of 0.02 (which is more narrow than necessary but narrow nonetheless) and did not notably increase the computational time for the model; this value of m also provides a conservative degree of similarity between the randomly determined parameters and the statistics of the random deviates. Thus, 2401 random deviates are generated to produce the population of costs for each cost item in all model scenarios, including the base model scenario. Table 11 shows the first set of random deviates (i.e., costs) generated by the *rmvnorm* function for the base model scenario; Figure 11 shows the population of costs of the cost items for the same scenario.

Table 11: Correlated random deviates of cost for cost items in the base model scenario.

Cost Item	**First Random Deviate**
Work Package 1	341,770
Work Package 2	648,951
Work Package 3	658,024
Work Package 4	319,482
Work Package 5	315,814
Work Package 6	814,902
Work Package 7	566,430
Work Package 8	700,733
Work Package 9	104,431
Work Package 10	730,341

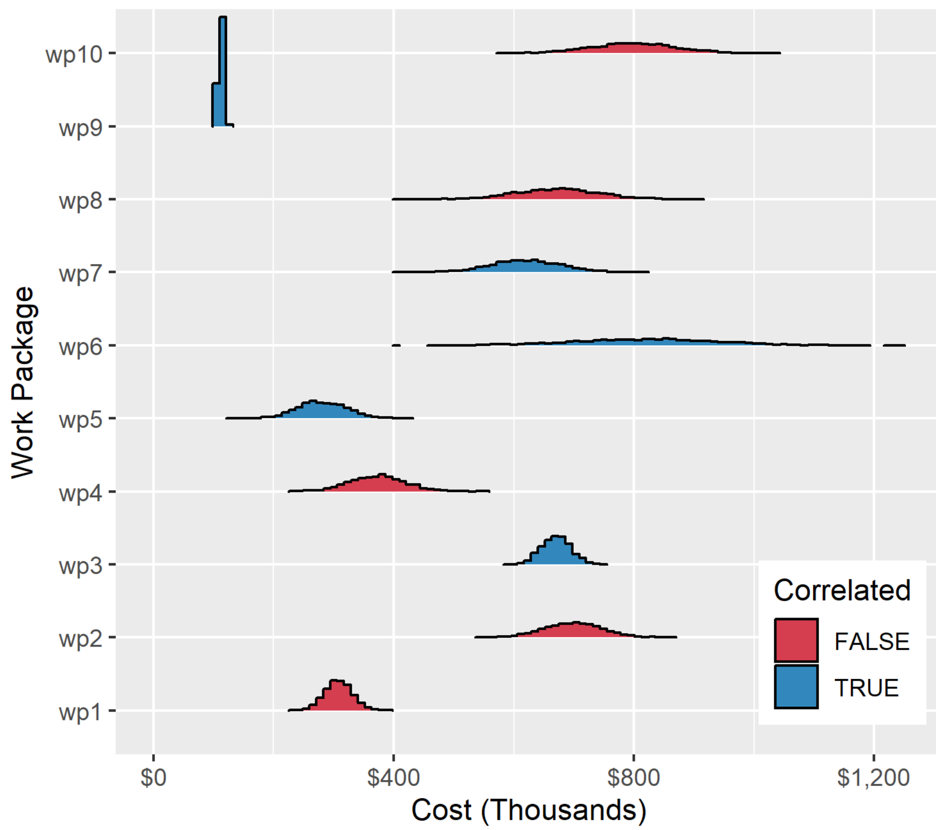


Figure 11: Population of costs for cost items in a scenario where the number of cost items is fixed at 10, the proportion of correlated cost items is fixed at 0.5, and the number of random deviates is fixed at 2401.

4.3 Collecting Population Statistics and Generating Independent Samples

After random deviates for each cost item have been generated, statistics are collected on the data and used to generate independent samples of each cost item. If the distribution(s) used to create the population of costs for each cost item vary or are unknown, then a distribution would need to be fitted for each cost item rather than simply collecting statistics and using the known distribution type. Since the model uses a multivariate normal distribution to generate correlated costs for each cost item, the statistics collected for each cost item can be directly fed into separate normal distributions that generate independent costs for each cost item. Table 12 shows the population mean and standard deviation for cost items in the base model scenario. The *rnorm* function receives the mean and standard deviation for a given cost item and generates n random deviates for the cost item per iteration where n is the number of cost samples generated for the model.

Table 12: Population mean and standard deviation for cost items in the base model scenario.

Cost Item	Population Mean	Population Standard Deviation
Work Package 1	307,833	22,700
Work Package 2	698,711	47,604
Work Package 3	670,105	23,983
Work Package 4	372,814	45,577
Work Package 5	279,871	41,188
Work Package 6	820,097	117,323
Work Package 7	618,610	58,617
Work Package 8	671,727	69,449
Work Package 9	111,862	4,150
Work Package 10	798,072	72,225

Considerations for determining the number of independent cost samples generated are similar to the considerations discussed in Section 4.2.5 for the number of correlated random deviates generated except that the population parameters are the mean and standard deviation of the correlated random deviates for each cost item. The independent cost samples produced by each normal distribution are samples of the population which will be used to apply different correlation modelling methods and which will ultimately be compared against population parameters. Unlike the circumstances described in Section 4.2.5, similarity between the population parameters and statistics of the samples is important. If the number of independent cost samples is too low then the mean and standard deviation of samples will vary significantly from the mean and standard deviation of the population of costs and the comparison of the correlation modelling methods will be unduly affected.

The number of random samples generated for each cost item in an iteration determines the standard error of the mean (SEM) for each cost item and affects the overall computational time required for the model. The SEM for a cost item is given by Equation 21 where σ is the standard deviation of the population of costs for a cost item, $\bar{x}$ is the sample of costs for a cost item, and n is the number of samples.

Equation 21:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

As indicated by the square root in the denominator of Equation 21, the relationship between SEM and the number of samples is nonlinear and eventually there are diminishing returns for error reductions. For example, a two-fold reduction in SEM requires four times as many samples and a ten-fold reduction in standard error requires one hundred times as many samples. As previously discussed, there is also a trade-off between error and computational time since an increase in the number of samples results in a decrease in SEM but an increase in computational time.

As explained in Section 4.2.2, the standard deviation of each cost item is initially sampled from a uniform distribution in the model, meaning that the standard deviation of each cost item is varied with respect to one another. Since the standard deviation varies for each cost item, the number of runs needed to achieve a desired value for SEM also varies for each cost item. Table 13 shows how the SEM for samples of cost items varies as the number of samples changes in the base model scenario.

Table 13: Standard error of the mean as a function of the number of samples in the base model scenario.

Cost Item	Population Mean	Population Standard Deviation	Standard Error of the Mean			
			n = 97	n = 385	n = 2401	n = 9604
Work Package 1	307,833	22,700	2,305	1,157	463	232
Work Package 2	698,711	47,604	4,833	2,426	972	486
Work Package 3	670,105	23,983	2,435	1,222	489	245
Work Package 4	372,814	45,577	4,628	2,323	930	465
Work Package 5	279,871	41,188	4,182	2,099	841	420
Work Package 6	820,097	117,323	11,912	5,979	2394	1197
Work Package 7	618,610	58,617	5,952	2,987	1196	598
Work Package 8	671,727	69,449	7,051	3,539	1417	709
Work Package 9	111,862	4,150	421	212	85	42
Work Package 10	798,072	72,225	7,333	3,681	1474	737

Rather than targeting a particular value for SEM, an alternative is to set a target relative to the standard deviation of a cost item. For example, the SEM associated with 97 samples is 10% relative to the standard deviation of a cost item, whereas the SEM associated with 1000 and 10000 samples are approximately 3.2% and 1% relative to the standard deviation of a cost item.

Table 14 shows how the SEM relative to the population mean for samples of cost items varies as the number of samples changes in the base model scenario.

Table 14: Standard error of the mean relative to the population mean as a function of the number of samples in the base model scenario.

Cost Item	**Standard Error of the Mean Relative to Population Mean**			
	n = 97	**n = 385**	**n = 2401**	**n = 9604**
Work Package 1	0.75%	0.38%	0.15%	0.08%
Work Package 2	0.69%	0.35%	0.14%	0.07%
Work Package 3	0.36%	0.18%	0.07%	0.04%
Work Package 4	1.24%	0.62%	0.25%	0.12%
Work Package 5	1.49%	0.75%	0.30%	0.15%
Work Package 6	1.45%	0.73%	0.29%	0.15%
Work Package 7	0.96%	0.48%	0.19%	0.10%
Work Package 8	1.05%	0.53%	0.21%	0.11%
Work Package 9	0.38%	0.19%	0.08%	0.04%
Work Package 10	0.92%	0.46%	0.18%	0.09%

For the base model scenario, the SEM relative to the population mean does not exceed 1% for any cost item if 385 samples are generated. As the number of samples is increased to 2401, the highest SEM relative to the population mean does not exceed 0.30%.

Regardless of the scenario under consideration, setting the number of samples for each cost item to 2401 provides a confidence interval width of 0.02; this value of n provides a good degree of similarity between the population parameters and the independent cost samples and did not notably increase the computational time for the model. Thus, 2401 random deviates are generated to produce the cost samples for each cost item in all model scenarios, including the base model scenario. Table 15 shows the first set of independent samples generated by separate normal distributions in the base model scenario; Figure 12 shows the full set of independent sample costs of the cost items for the same scenario.

Table 15: First set of independent cost samples for cost items in the base model scenario.

Cost Item	First Sample Cost
Work Package 1	337,168
Work Package 2	680,429
Work Package 3	660,793
Work Package 4	389,707
Work Package 5	319,349
Work Package 6	993,496
Work Package 7	562,886
Work Package 8	778,320
Work Package 9	111,922
Work Package 10	731,264

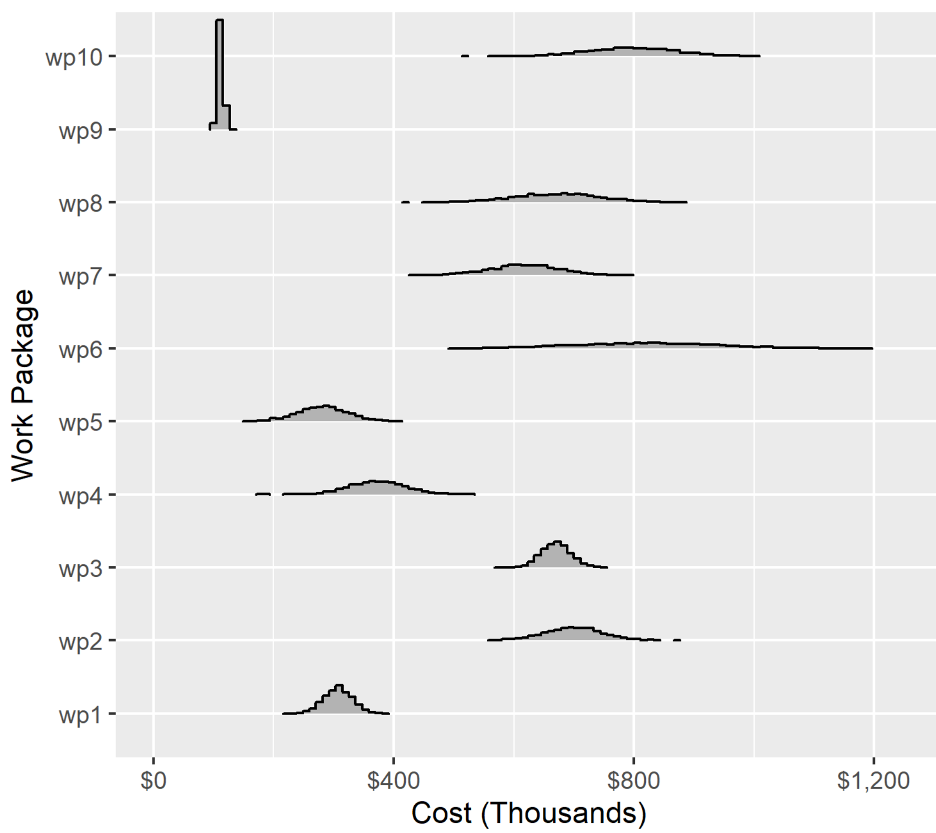


Figure 12: Full set of independent sample costs for cost items in the base model scenario.

Table 16 compares the mean and standard deviation of the sample costs for cost items with the mean and standard deviation of the population of cost items in the base model scenario. The percentage error for mean and standard deviation are calculated using conventional percentage error formulas as shown in Equation 22 and Equation 23, respectively, where μ is the mean of the population, $\bar{x}$ is the mean of the sample, σ is the standard deviation of the population, and s is the standard deviation of the sample.

Equation 22:

$$percentage\ error\ in\ mean = \frac{\mu - \bar{x}}{\mu} \times 100\%$$

Equation 23:

$$percentage\ error\ in\ standard\ devation = \frac{\sigma - s}{\sigma} \times 100\%$$

Table 16: Comparison of population and sample statistics for cost items in the base model scenario.

Cost Item	Population Mean	Sample Mean	Mean Percent Error	Population Standard Deviation	Sample Standard Deviation	Standard Deviation Percent Error
Work Package 1	307,833	306,861	0.32%	22,700	23,109	-1.80%
Work Package 2	698,711	698,899	-0.03%	47,604	47,210	0.83%
Work Package 3	670,105	670,245	-0.02%	23,983	23,885	0.41%
Work Package 4	372,814	372,733	0.02%	45,577	45,947	-0.81%
Work Package 5	279,871	281,169	-0.46%	41,188	41,441	-0.61%
Work Package 6	820,097	819,099	0.12%	117,323	115,751	1.34%
Work Package 7	618,610	615,772	0.46%	58,617	58,624	-0.01%
Work Package 8	671,727	671,480	0.04%	69,449	71,006	-2.24%
Work Package 9	111,862	111,746	0.10%	4,150	4,245	-2.28%
Work Package 10	798,072	797,212	0.11%	72,225	72,575	-0.49%

In the base model scenario, the percent error for mean is +/- 0.46% which is within expectations given the standard error of the mean relative to the population mean did not exceed 0.30% and 99.8% of sample means will fall within three times that value (i.e., +/- 0.90%). The percent error for standard deviation in the base model scenario is also quite low with the percent error in the interval [-2.28%, 1.34%]. The similarity of statistics between the population and samples provide confidence that the number of samples for the model is sufficient.

4.4 Applying Correlation Modelling Methods

As previously mentioned, correlation between cost items can be modelled in a Monte Carlo simulation in many different ways. Section 2.4.1.3.1 addressed several correlation modelling methods, namely:

- using model relationships to capture functional correlation;
- assigning correlations subjectively;
- deriving correlations mathematically from the structures of cost estimating relationships; or,
- determining correlations empirically using Monte Carlo simulation.

An alternative to modelling correlation is to combine related items and otherwise assume independence between cost items (i.e., similar to the range estimating approach described in Section 2.4.1.3). In terms of applying correlation to independently sampled costs, this approach is akin to “do nothing”.

Much like the data required to develop probability distributions for cost items, the data required to either develop mathematical relationships for modelling functional correlation or empirically determine correlation are often not available or of low quality. In the absence of quality data, subjective correlation assignments allow for correlation to be accounted for in the simulation.

The model includes three methods for applying (and not applying) correlation:

1. the independence method;
2. the “knee-in-the-curve” method; and,
3. the random assessment method.

4.4.1 Independence Method

The independence method simply assumes that there is no correlation between cost items; this assumption is seldom true and one of the main reasons for narrow probability distributions of project cost (Garvey et al. 2016). Nonetheless, the independence method is used by some in industry and for some applications may produce reasonably accurate results, as such it is a useful method to assess and to compare with correlation modelling methods that apply correlation. As previously described, sample costs for cost items are generated in the model by separate normal distributions, meaning that cost samples between cost items are already independent. To calculate a sample of the project cost, a sample cost for each cost item is added together as shown in Equation 24 where n is the number of cost items within the project for a given scenario; this calculation is repeated for each set of cost samples until 2401 samples of overall project cost have been generated. Figure 13 shows a histogram of project cost samples from the independence method in the base model scenario.

Equation 24:

$$project\ cost = \sum_{i=1}^{n} cost\ of\, cost\ item_i$$

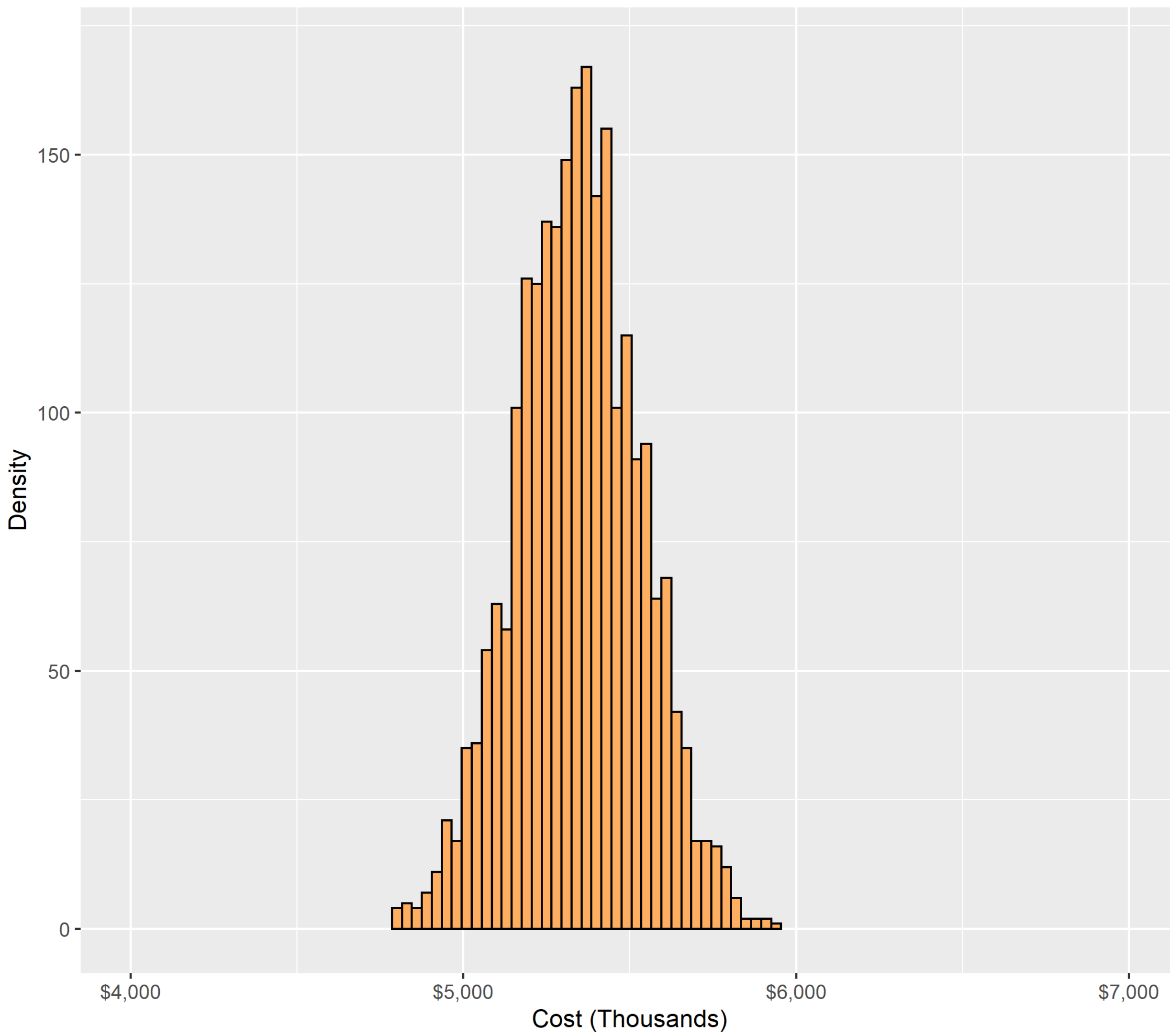


Figure 13: Samples of project cost from the independence method in the base model scenario.

The mean and standard deviation of sample project costs from the independence method are $5,345,217 and $184,188, respectively; the P80 is $5,502,511. Statistics collected about sample project costs from the independence method will be compared with project cost statistics calculated from population costs as well as from sample costs using the other correlation modelling methods (described in detail in Section 4.5).

4.4.2 Knee-in-the-Curve Method

One approach to subjectively assigning correlations involves qualitatively assessing each pair of cost items using linguistic terms such as weak, moderate, and strong (Touran 1993), or uncorrelated, correlated by a small amount, and correlated by a large amount (National

Aeronautics and Space Administration 2008). As previously discussed, time, information, and effort are required to assess correlation between cost items. The number of cost item pairs requiring assessment increases drastically as a function of the number of cost items since a project with n cost items has $n(n-1)/2$ cost item pairs (Garvey et al. 2016). Another challenge with developing a correlation matrix from subjective assessment is that the result is often an inconsistent correlation matrix, i.e., a matrix that is not positive semi-definite (Lurie and Goldberg 1998). Where subjective assessments yield an inconsistent correlation matrix, the matrix needs to be adjusted to satisfy the properties of a correlation matrix (including the positive semi-definite property) while keeping the assignments as close as possible to the original matrix.

Given the practical difficulties associated with assessing pairs of cost items individually, other options for subjective assessments need to be considered. The first such option is the knee-in-the-curve (KITC) method which involves choosing correlation for all cost items pairs from the interval [0.10, 0.30] where the "knee" (i.e. bend) in curves of uncertainty underestimation occur (Garvey et al. 2016). The KITC method is based on the notion that for correlation values greater than 0.30, there is little change in the percent that a project's cost uncertainty is underestimated by not capturing positive correlation (i.e., assuming independence) when it is present.

With an interval of correlation established for the KITC method, the next step is to generate a correlation matrix. The correlation matrix is randomly generated using the same process described in Section 4.2.4.2 except the model scales elements of the correlation matrix to the KITC interval [0.1, 0.3]. Table 17 shows the correlation matrix for cost items using the KITC method in the base case scenario.

Table 17: Correlation matrix for cost items using the knee-in-the-curve method in the base case scenario.

	wp1	wp2	wp3	wp4	wp5	wp6	wp7	wp8	wp9	wp10
wp1	1	0.150	0.212	0.172	0.169	0.230	0.204	0.218	0.265	0.224
wp2	0.150	1	0.222	0.228	0.222	0.157	0.177	0.226	0.148	0.164
wp3	0.212	0.222	1	0.181	0.150	0.181	0.166	0.250	0.192	0.165
wp4	0.172	0.228	0.181	1	0.201	0.204	0.250	0.163	0.160	0.191
wp5	0.169	0.222	0.150	0.201	1	0.217	0.205	0.170	0.212	0.165
wp6	0.230	0.157	0.181	0.204	0.217	1	0.245	0.175	0.236	0.200
wp7	0.204	0.177	0.166	0.250	0.205	0.245	1	0.164	0.231	0.211
wp8	0.218	0.226	0.250	0.163	0.170	0.175	0.164	1	0.214	0.218
wp9	0.265	0.148	0.192	0.160	0.212	0.236	0.231	0.214	1	0.208
wp10	0.224	0.164	0.165	0.191	0.165	0.200	0.211	0.218	0.208	1

Once the correlation matrix has been developed, the next step in the model is to induce correlation between independent cost samples. As previously mentioned, there are a variety of approaches for inducing correlation. The model uses an R package developed by Liu (2020) called SimJoint for inducing rank correlation among variables. The package implements an algorithm by Iman and Conover (1982) which is employed by commercially available products such as @Risk. The *SJpearson* function requires samples to be column sorted in a matrix. For the model, cost samples for each cost item are captured in their own column with samples sorted from lowest to highest cost. Table 18 shows the first five and last five independent cost samples for each cost item sorted from lowest to highest in the base case scenario.

Table 18: The first five and last five independent cost samples sorted from lowest to highest for each cost item in the base case scenario.

wp1	wp2	wp3	wp4	wp5	wp6	wp7	wp8	wp9	wp10
219,793	565,405	577,205	179,496	152,019	496,375	425,714	416,414	96,539	517,340
227,902	569,058	587,186	190,288	154,236	503,278	433,416	420,527	96,939	519,301
235,266	569,151	597,897	220,392	162,098	510,927	434,141	450,287	99,221	560,744
238,090	569,565	600,069	225,353	166,748	512,484	435,772	463,212	99,638	563,791
238,807	572,128	600,102	231,981	167,211	517,408	437,680	467,804	99,676	581,260
…	…	…	…	…	…	…	…	…	…
369,387	830,918	741,981	514,310	400,338	1,140,279	768,965	877,046	123,156	991,025
373,708	831,704	742,321	516,160	402,937	1,147,903	778,612	878,398	123,686	994,587
377,312	839,184	745,591	519,190	403,899	1,171,721	784,344	879,306	124,505	996,668
380,411	840,975	745,832	519,668	406,929	1,185,285	792,758	882,175	126,696	1,005,665
382,448	866,994	747,735	528,108	407,598	1,188,085	793,760	886,850	127,717	1,008,233

Next, the *SJpearson* function receives the KITC correlation matrix and the sorted independent cost samples and outputs reordered samples with correlation approximate to the values in the correlation matrix. The *cor* function, which computes the correlation between variables, can be used with the reordered samples to create a correlation matrix for comparison with the target correlation matrix. In the base model scenario, each element between the matrices were equal to three significant digits meaning the correlation of the reordered data matched that of the target correlation matrix. Finally, samples of project cost are calculated using the same process described in Section 4.4.1. Figure 14 shows a histogram of project cost samples from the KITC method in the base model scenario.

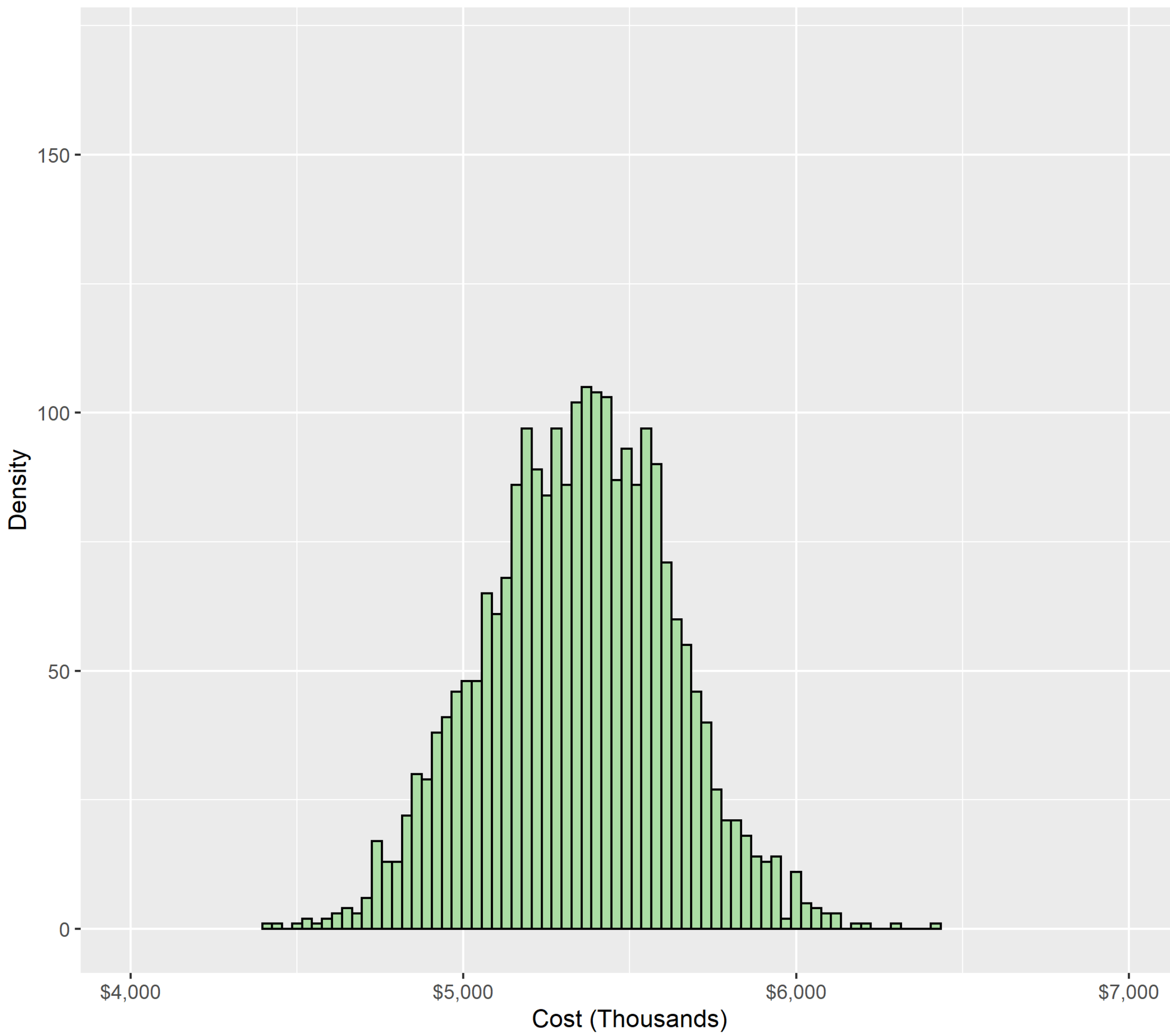


Figure 14: Samples of project cost from the knee-in-the-curve method in the base model scenario.

The mean and standard deviation of sample project costs from the KITC method are $5,345,217 and $278,631, respectively; the P80 is $5,576,823. The mean of sample project costs from the KITC method is the same as the mean of sample project costs from the independence method since correlation was induced through reordering sample costs and the mean is not affected by the order of the data. Conversely, the standard deviation of sample project costs from the KITC method ($278,631) is larger than the standard deviation of sample project costs from the independent method ($184,188) by a factor of approximately 1.51 in the base model scenario. Further comparisons of sample and population cost statistics are described in detail in Section 4.5.

4.4.3 Random Assessment Method

The random assessment method is presented as another option for subjectively assigning correlation. The random assessment method uses randomly generated values of correlation for all cost item pairs in the interval [0.1, 0.9] as a substitute for individual assessments of correlation between pairs of cost items. While subjective assessments of correlation between pairs of cost items are not random in reality, randomized values provide a lower performance bound of subjective assessment[2]. As opposed to the full interval for positive correlation (i.e., [0, 1]), the model uses the interval [0.1, 0.9] for random assessment on the basis that correlation between cost items is rarely entirely independent (i.e., correlation equal to zero) or entirely correlated (i.e., correlation equal to one).

Another way to conceptualize the random assessment method is to think of it as a more conservative[3] version of the KITC method where the correlation interval is expanded and the central tendency is increased from a correlation value of 0.2 (the midpoint in the interval [0.1, 0.3]) to 0.5 (the midpoint in the interval [0.1, 0.9]). Since it provides a greater positive correlation value on average than the KITC method, the random assessment method will yield a greater spread in project cost results.

Similar to the KITC method, the correlation matrix is randomly generated using the same process described in Section 4.2.4.2 except the model scales elements of the correlation matrix to the interval [0.1, 0.9]. Table 19 shows the correlation matrix for cost items using the random assessment method in the base case scenario.

[2] In practice, if subjective assessments were found to be too close to randomness, then their credibility would likely be questioned.

[3] Conservative in this context refers to a lesser underestimation of cost uncertainty when cost items are in fact positively correlated.

Table 19: Correlation matrix for cost items using the random assessment method in the base case scenario.

	wp1	wp2	wp3	wp4	wp5	wp6	wp7	wp8	wp9	wp10
wp1	1	0.364	0.631	0.692	0.557	0.339	0.452	0.638	0.668	0.423
wp2	0.364	1	0.393	0.613	0.433	0.503	0.381	0.614	0.504	0.568
wp3	0.631	0.393	1	0.612	0.547	0.265	0.395	0.393	0.708	0.463
wp4	0.692	0.613	0.612	1	0.332	0.388	0.323	0.606	0.721	0.521
wp5	0.557	0.433	0.547	0.332	1	0.445	0.741	0.480	0.280	0.318
wp6	0.339	0.503	0.265	0.388	0.445	1	0.670	0.403	0.361	0.413
wp7	0.452	0.381	0.395	0.323	0.741	0.670	1	0.473	0.258	0.323
wp8	0.638	0.614	0.393	0.606	0.480	0.403	0.473	1	0.548	0.603
wp9	0.668	0.504	0.708	0.721	0.280	0.361	0.258	0.548	1	0.627
wp10	0.423	0.568	0.463	0.521	0.318	0.413	0.323	0.603	0.627	1

After the correlation matrix has been developed, correlation is induced using the same process as for the KITC method, i.e., the sorted independent cost samples are reordered with correlation approximate to the values in the correlation matrix. Finally, samples of project cost are calculated using the same process described in Section 4.4.1. Figure 15 shows a histogram of project cost samples from the random assessment method in the base model scenario.

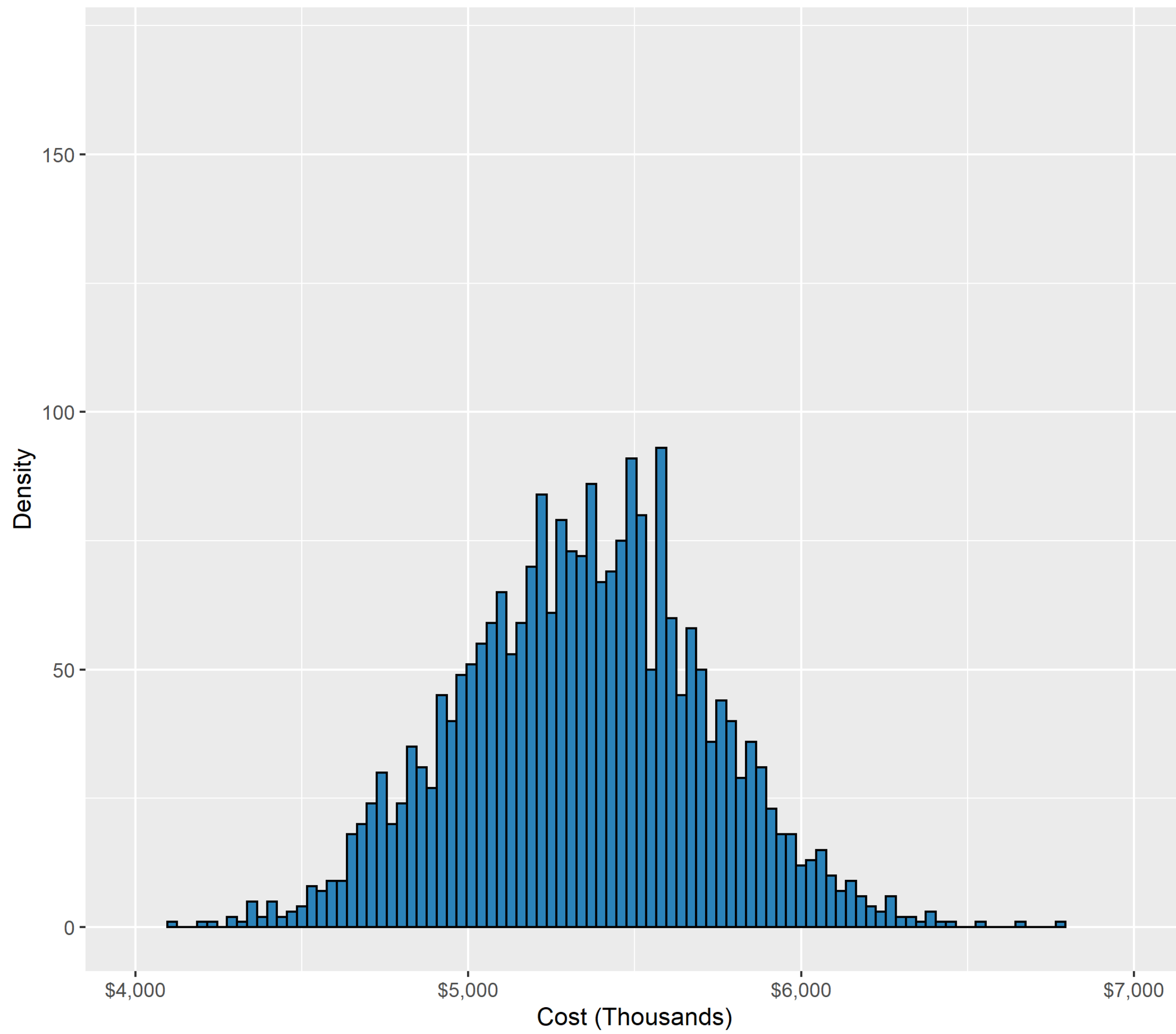


Figure 15: Samples of project cost from the random assessment method in the base model scenario.

The mean and standard deviation of sample project costs from the random assessment method are $5,345,217 and $372,120, respectively; the P80 is $5,655,646. The mean of sample costs from the random assessment method is the same as the means of sample project costs from the independent and KITC methods since correlation was induced through reordering sample costs and the mean is not affected by the order of the data. Conversely, the standard deviation of sample project costs from the random assessment method ($372,120) is larger than the standard deviation of sample project costs from the KITC method ($278, 631) by a factor of approximately 1.33 and is larger than the standard deviation of sample project costs from the independence method ($184,188) by a factor of approximately 2.02 in the base model scenario.

Further comparisons of sample and population cost statistics are described in detail in Section 4.5.

4.5 Evaluating Correlation Modelling Methods

The correlation modelling methods described in Section 4.4 each produce samples of project cost. Similarly, Equation 24 can be applied to the population of costs for each cost item to obtain a population of project costs for a given scenario. Figure 16 shows a histogram of the population of project costs in the base model scenario. The mean and standard deviation of the population of project costs are $5,349,702 and $230,752, respectively; the P80 is $5,540,653.

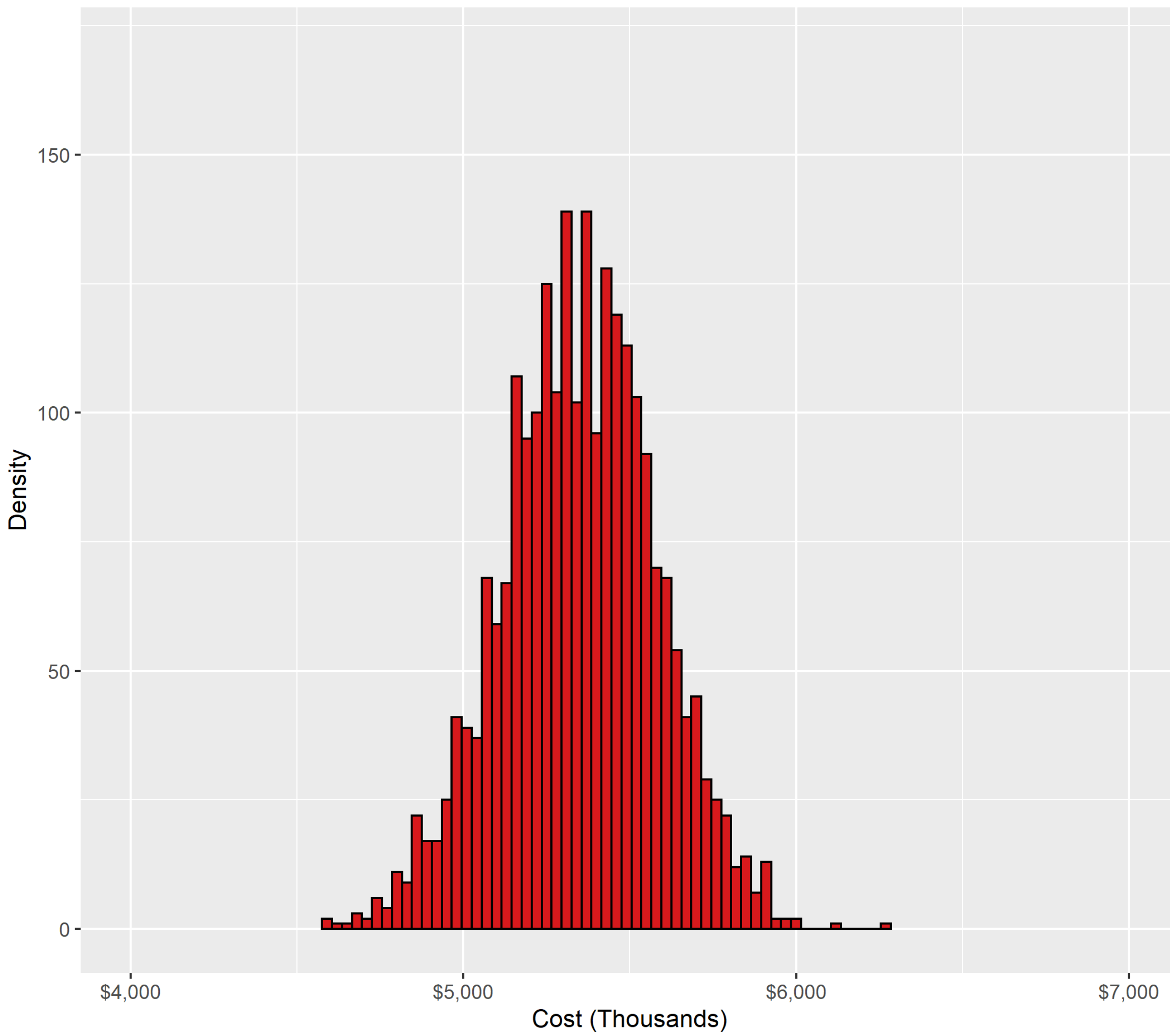


Figure 16: Population of project costs in the base scenario model.

Figure 17 shows histograms of project cost samples from each correlation modelling method as well as a histogram of the population of project costs in the base scenario model.

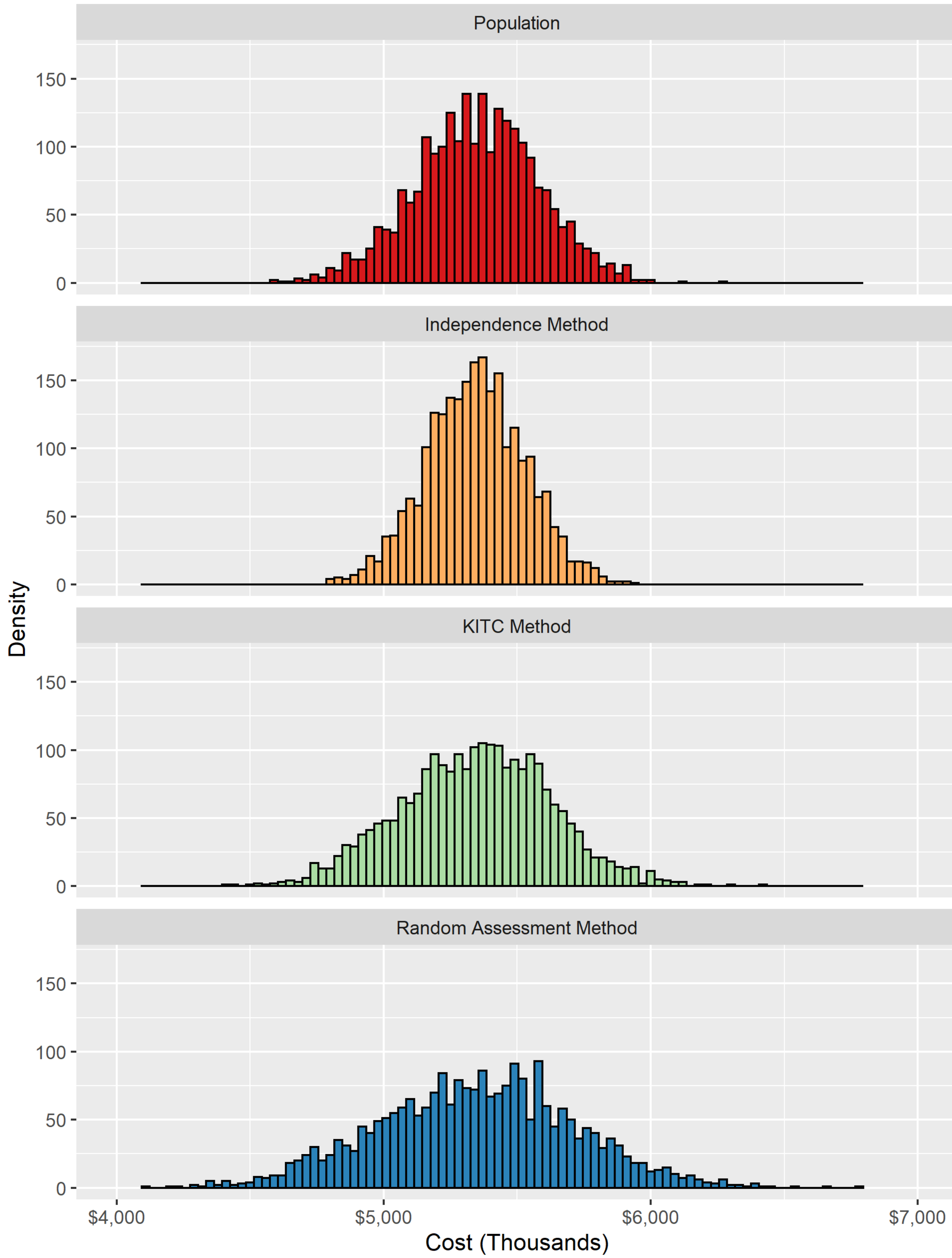


Figure 17: Samples of project cost from each correlation modelling method as well as the population of project costs in the base model scenario.

Upon visual inspection, samples produced by the independence method have less variance than the population (i.e., cost uncertainty is underestimated), whereas samples produced by the random assessment method have more variance than the population (i.e., cost uncertainty is overestimated). Table 20 summarizes the mean, standard deviation, and P80 of project costs corresponding to each correlation modelling method as well as to the population of project costs in the base scenario model.

Table 20: Mean, standard deviation, and P80 of project costs for correlation modelling methods and population in the base model scenario.

Data Source	**Project Cost Mean**	**Project Cost Standard Deviation**	**Project Cost P80**
Independence method	\$5,349,702	\$184,188	\$5,502,511
KITC method	\$5,345,217	\$278, 631	\$5,576,823
Random assessment method	\$5,345,217	\$372,120	\$5,655,646
Population	\$5,345,217	\$230,752	\$5,540,653

With project cost statistics calculated on both the samples from each correlation modelling method and the population, the correlation modelling methods can be evaluated for a given scenario. In particular, the difference between estimated statistics (i.e., statistics on the samples of project cost from each correlation modelling method) and the actual statistics (i.e., statistics on the population of project costs) is calculated in absolute or relative terms.

Absolute error, shown in Equation 25, refers to the difference between an actual statistic and an estimated statistic as an absolute value.

Equation 25:

$$absolute\ error = |statistic_{actual} - statistic_{estimated}|$$

In the context of correlation modelling method comparison, it is useful to understand the directionality of the difference between statistics. Depending on the scenario being modelled, a method could produce sample statistics that are above or below the population statistics, meaning that the method may have underestimated or overestimated cost uncertainty. The distinction between underestimation and overestimation of cost uncertainty is an important one for numerous reasons[4] previously described in this thesis. As such, the error, shown in Equation 26 is more useful than Equation 25 in the context of the model.

Equation 26:

$$error = statistic_{estimated} - statistic_{actual}$$

The order of the terms in Equation 26 is reversed relative to Equation 25 to produce results with a sign that is intuitive from the author's perspective. A *negative* value yielded by Equation 26 indicates that the estimated statistic is lower than the actual statistic, i.e., the estimated value produced by the correlation modelling method *underestimated* the actual value in the population.

A disadvantage of the absolute error and error calculations for comparing methods is that results are specific to the scenario under consideration. For example, in the base model scenario, the error in the standard deviation between samples from the independent method and the population is -$46,564. Consider another scenario called the large project scenario, identical to the base model scenario except that the number of cost items is 100, where the error in the standard deviation between samples from the independent method and the population is -$1,416,219. While the error in the large project scenario is larger than the error in base model scenario by a factor of approximately 30.4, the mean of project cost in the population also increased significantly (from $5,349,702 to $49,196,118). Similarly, the standard deviation of project cost in the population increased from $230,752 to $1,993,110. In light of the increase in population mean and standard deviation, the increase in standard deviation error seems less drastic. When the error is expressed relative to the actual value, i.e. as a percentage error, it becomes easier to interpret.

Percentage error, shown in Equation 27, refers to the difference between an actual statistic and an estimated statistic relative to the actual statistic expressed as a percentage. In the context of the model, the percentage error is calculated without an absolute value, similarly to Equation 26.

Equation 27:

$$percentage\ error = \frac{statistic_{estimated} - statistic_{actual}}{statistic_{actual}} \times 100\%$$

[4] The principal reason among them being that project costs have a demonstrated history of consistent underestimation by large margins (Hollmann 2012).

The percentage error in the standard deviation between samples from the independent method and the population are -20.2% and -71.1% for the base model scenario and large project scenario, respectively; these values for percentage error make it clear that despite the population mean and standard deviation increasing for the large project scenario, the difference in the standard deviation of samples from the independent method and population standard deviation has worsened; put simply, when the number of cost items increased from 10 to 100, the underestimation of project cost uncertainty worsened.

The standard deviation of project cost is a useful statistic for comparing samples and the population since it is a measure of the amount of variation in project cost and therefore a measure of project cost uncertainty. Thus, the primary metric for evaluating correlation modelling methods is the percentage error in project cost standard deviation as shown in Equation 28 where σ is the standard deviation of the population and s is the standard deviation of the sample. Note that Equation 28 is identical to Equation 23 except that the sign is reversed.

Equation 28:

$$standard\ deviation\ percentage\ error = \frac{s - \sigma}{\sigma} \times 100\%$$

Another useful statistic for evaluating correlation modelling methods is the cost corresponding to one or more percentiles, e.g., P70 or P80. Percentiles and standard deviation are related. For example, a low standard deviation indicates that values tend to be close to the mean and these are the same values that correspond to percentiles of project cost. Percentiles are a useful metric for evaluating correlation methods since owners typically establish their budgets based on their desired level of confidence, typically expressed as a percentile of possible results. Essentially, this metric can provide an indication of how different correlation modelling methods affect the values used to actually set budgets. There is no 'correct' percentile or rule set for determining which percentile to use, however some values are more commonly used than others. The P80 is commonly referenced in the literature (U. S. Government Accountability Office 2020; AACE International 2008; National Aeronautics and Space Administration 2008) as a risk averse value. Percentiles at or near P80 are also commonly used in industry. Thus, the secondary metric for evaluating correlation modelling methods in this thesis is the percentage error in P80 as shown in Equation 29.

Equation 29:

$$P80\ percentage\ error = \frac{P80_{sample} - P80_{population}}{P80_{population}} \times 100\%$$

Table 21 shows the standard deviation percentage error and P80 percentage error for each correlation modelling method in the base model scenario.

Table 21: Percentage error in standard deviation and P80 for each correlation modelling method in the base model scenario.

Data Source	Standard Deviation Percent Error	P80 Percent Error
Independence method	-20.2%	-0.7%
KITC method	20.7%	0.7%
Random assessment method	61.3%	2.1%

4.6 Model Scenarios

Throughout Sections 4.1 to 4.5, information and results have been provided for the base model scenario containing the following parameters:

- the number of cost items fixed at 10;
- the proportion of correlated cost items fixed at 0.5 (i.e., 50% of cost items correlated); and,
- correlation strength common to all pairs and fixed at 0.5.

The base model scenario is used to demonstrate the execution of the model and show how correlation modelling methods are evaluated in principal, but the results of the base model scenario apply to a narrowly defined set of parameters. To draw larger conclusions about the performance of different correlation modelling methods, these methods need to be applied to a wide range of situations. Model scenarios have been developed to allow for one or more key parameters to be manipulated while the performance of each correlation modelling method is monitored. The key parameters being adjusted in the model include:

- the number of cost items;
- the proportion of correlated cost items; and,
- correlation strength.

In each scenario, key parameters are manipulated to the largest extent possible (or in some cases feasible). For some parameters there are natural boundaries and for others, there are practical or logical reasons for establishing narrower boundaries. For the number of cost items, boundaries need to be defined. While a project can theoretically have as little as one cost item, that low of a number is extremely uncommon and such a project would probably not warrant cost uncertainty analysis. As such, a lower boundary of five is used as a more practical lower limit. In a theoretical sense, there is no upper limit to the number of cost items a project can have; large projects can be made up of hundreds or even thousands of cost items. Given the nature and extent of model processes, there are practical limitations to the number of cost items that can be modelled without requiring excessive amounts of computational time. In particular, the computational time required to generate random correlation matrices for the KITC and random assessment methods increases exponentially as the number of cost items increases. As such, an upper boundary of 100 is used as a practical upper limit which covers a healthy range of project sizes. For the proportion of correlated cost items, there are natural boundaries in the sense that the proportion of correlated variables can only be a value between zero and one. Correlation strength also has natural boundaries of negative one (completely negatively correlated) to one (completely positively correlated), however correlation strength in the model is limited to values between zero and one since positive correlations are more common in the context of cost and can have significant effects on the cost uncertainty of project costs.

As discussed throughout Sections 4.1 to 4.5, a number of model parameters are fixed for all scenarios, these include:

- cost mean for each cost item which is uniformly sampled between \$100,000 and \$900,000;
- cost standard deviation for each cost item which is uniformly sampled between 10% of the mean divided by three and 50% of the mean divided by three;
- the number of random deviates used to form the population of costs for each cost item which is fixed at 2401; and,
- the number of samples of cost generated for each cost item which is also fixed at 2401.

4.6.1 Varying Correlation Strength

The first model scenario variant (Scenario 1) involves varying the correlation strength between cost items in the population. In particular, correlation strength is common to all pairs of correlated cost items and varied iteratively from zero to one in increments of 0.01. Key model parameters for Scenario 1 are summarized in Table 22.

Table 22: Key model parameters for Scenario 1.

Model Parameter	Parameter Value(s)
Number of cost items	Fixed at 10
Proportion of correlated cost items	Fixed at 1
Correlation strength	Common to all pairs of correlated cost items, iterated from 0 to 1 in increments of 0.01

Given there are 101 values of correlation strength in Scenario 1, the model performs 101 iterations and produces 101 sets of population and sample statistics as well as 101 sets of error measurements. Table 23 shows the percentage error in standard deviation and P80 for each correlation modelling method found in the first iteration of Scenario 1 where correlation strength is set to zero.

Table 23: Percentage error in standard deviation and P80 for each correlation modelling method in the first iteration of Scenario 1 where the correlation strength of the population is zero.

Data Source	Standard Deviation Percent Error	P80 Percent Error
Independence method	0.5%	0%
KITC method	55.9%	1.5%
Random assessment method	110.0%	3.0%

The percentage error in standard deviation and P80 for the independence method shown in Table 23 are 0.5% and 0%, respectively, indicating that when a project has 10 cost items and there is no correlation between cost items, the independence method produces the best (i.e., least erroneous) results. Figure 18 shows the standard deviation percentage error for each correlation modelling method as a function of correlation strength of the population in Scenario 1.

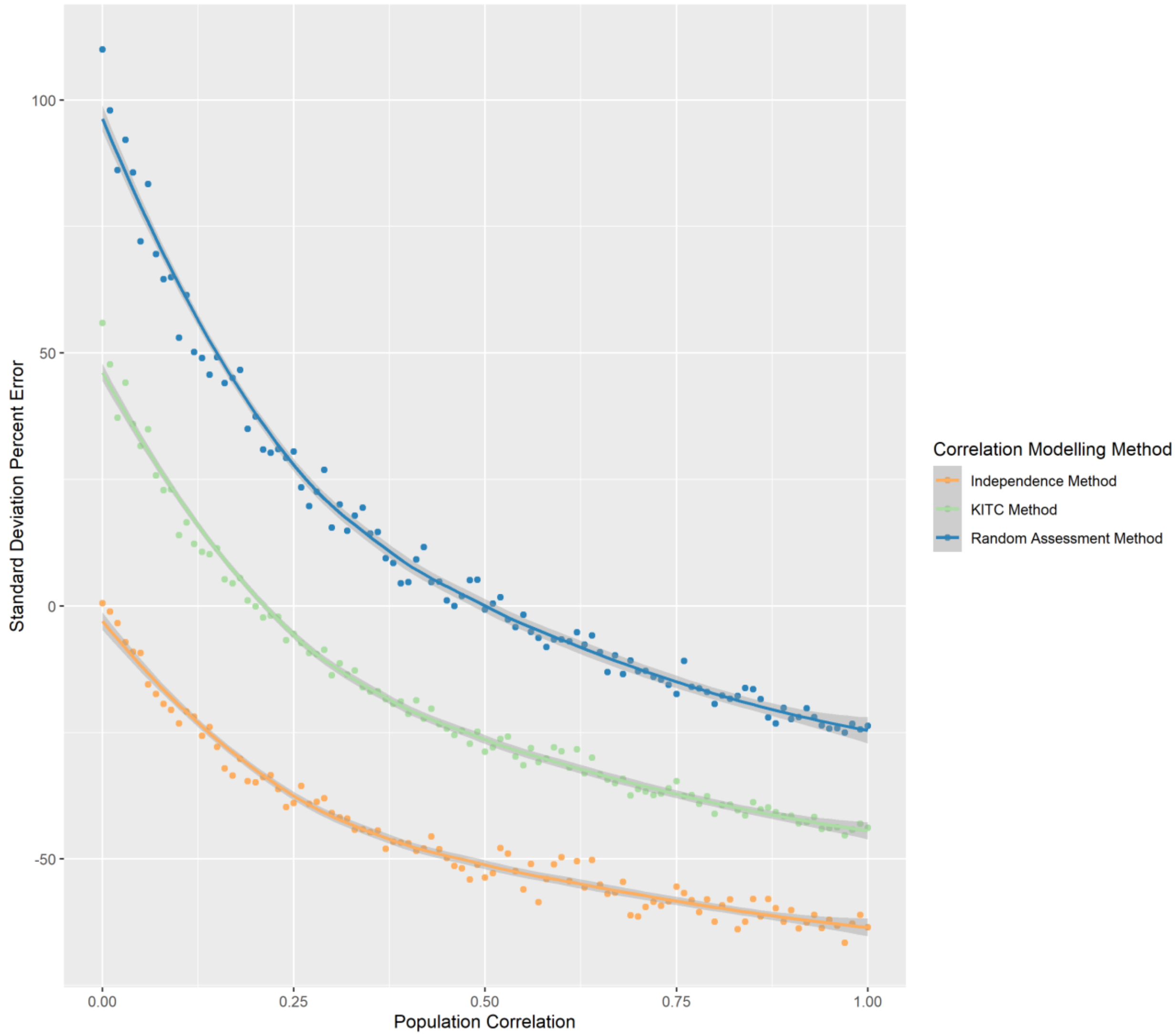


Figure 18: Standard deviation percentage error for each correlation modelling method as a function of correlation strength of the population in Scenario 1.

Results of Scenario 1 show that as the correlation in the population approaches zero, the percentage error in standard deviation of samples generated by the independence method also approaches zero, however correlation near zero is seldom the case. The random assessment method yields the most significant overestimation of cost uncertainty (with percentage error approaching 100%) whereas the independence method yields the most significant underestimation of cost uncertainty (with percentage error approaching -65%).

The KITC and random assessment methods had positive and negative values of percentage error, indicating that these methods are capable of underestimating or overestimating cost uncertainty depending on the actual correlation in the population. Conversely, the independence method resulted in only negative values of percentage error, indicating that this method can only underestimate, not overestimate, cost uncertainty.

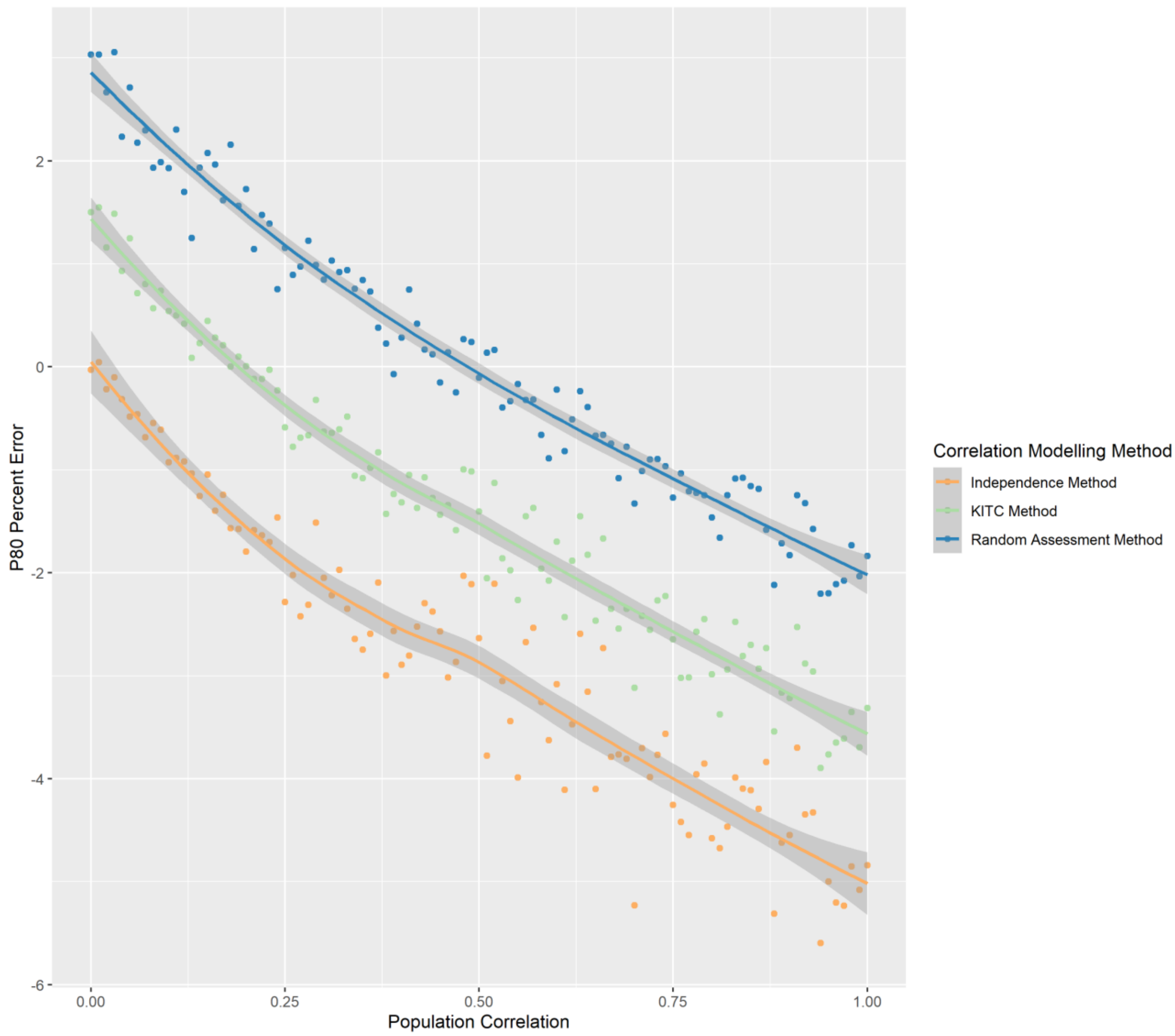


Figure 19: P80 percentage error for each correlation modelling method as a function of correlation strength of the population in Scenario 1.

Figure 19 shows the P80 percentage error for each correlation modelling method as a function of correlation strength of the population in Scenario 1. Patterns observed in the P80 percentage error are similar to those of the percentage error in standard deviation except the P80 percentage error is in the range of +3% and -5% for all methods, whereas the standard deviation percentage error is in the range of +100% and -65%. In Scenario 1, the independence method exposes an owner to a potential budget underestimation of about 5% with no potential for overestimation if the owner chooses the P80 to be their level of confidence. If the owner desires a higher level of confidence, e.g. P90, then the potential budget underestimation would increase. Likewise, if the owner is satisfied with a lower level of confidence, e.g. P70, then the potential budget underestimation would decrease[5].

4.6.2 Varying Correlation Strength and Number of Cost Items (Stepwise)

The next model scenario (Scenario 2) involves varying the correlation strength between cost items in the population (similarly to Scenario 1) along with stepwise variation in the number of cost items. Like Scenario 1, correlation strength is common to all pairs of correlated cost items and varied iteratively from zero to one in increments of 0.01 however the number of cost items will be first set at 10 (the same as Scenario 1), then set at 30, and then set at 100. The three discrete number of cost items represent small, medium, and large projects. Key model parameters for Scenario 2 are summarized in Table 24.

Table 24: Key model parameters for Scenario 2.

Model Parameter	Parameter Value(s)
Number of cost items	Varied stepwise from 10 to 30 to 100
Proportion of correlated cost items	Fixed at 1
Correlation strength	Common to all pairs of correlated cost items, iterated from 0 to 1 in increments of 0.01

Given there are 101 values of correlation strength and three values for the number of cost items, the model performs 303 iterations and produces 303 sets of population and sample statistics as well as 303 sets of error measurements. For the first 101 iterations, the number of cost items is 10 which means outputs are identical to Scenario 1. Figure 20 shows the standard deviation percentage error for each correlation modelling method as a function of correlation strength of the population for three discrete values for the number of cost items in Scenario 2.

[5] In fact, an owner taking a 'risk neutral approach' (i.e., P50) would gain no benefit from any correlation modelling methods as presented in this thesis since correlation affects the variance of project costs, not their mean.

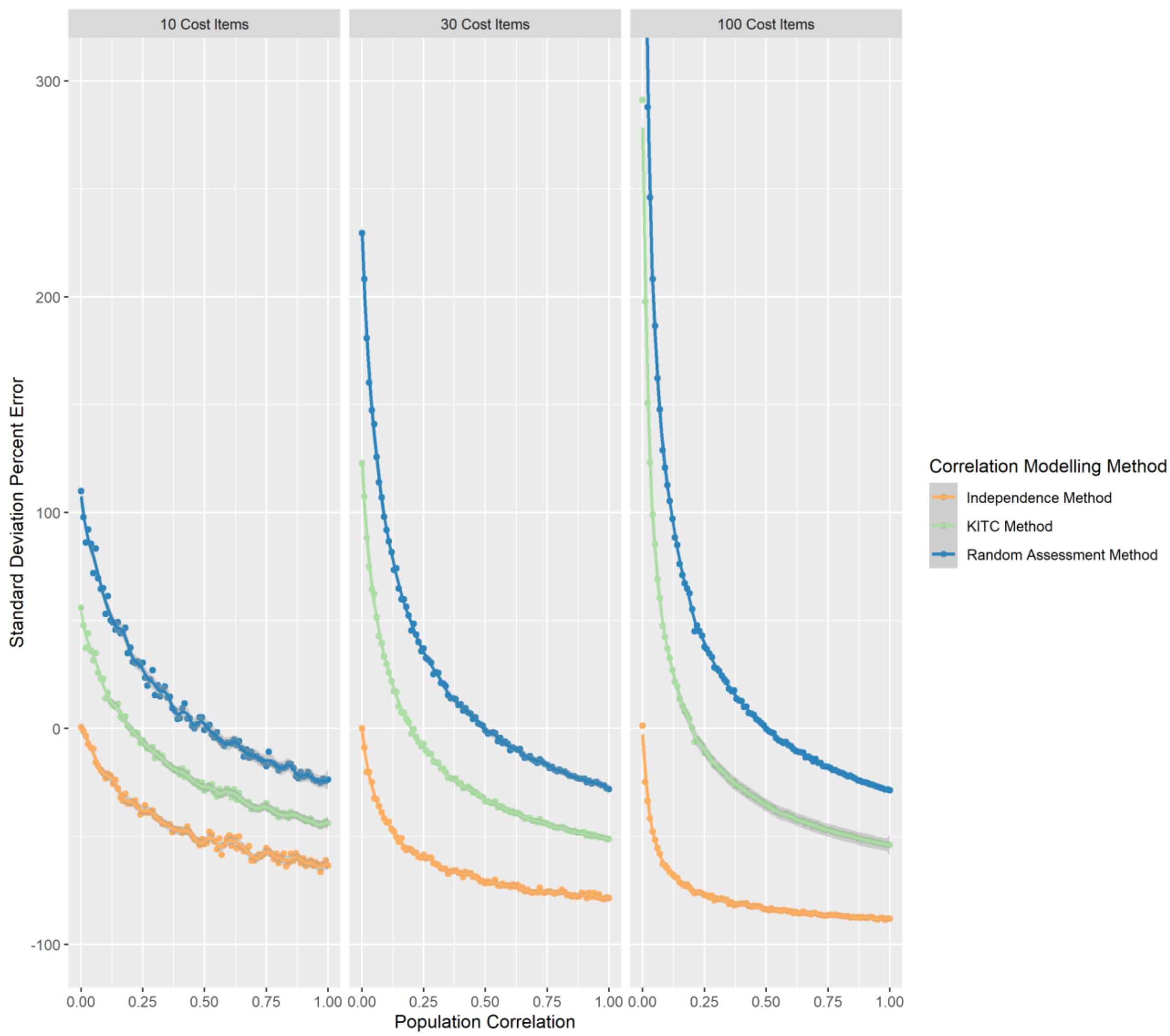


Figure 20: Standard deviation percentage error for each correlation modelling method as a function of correlation strength of the population for three discrete values for the number of cost items in Scenario 2.

Results of Scenario 2 illustrate how the number of cost items affects the underestimation and overestimation of cost uncertainty for the different correlation modelling methods under varying correlation strengths in the population. The percentage error in standard deviation of samples generated by the independence method still approaches zero where correlation in the population approaches zero, however as correlation is introduced, the extent of cost underestimation increases drastically as the number of cost items increases. Conversely, the percentage error in standard deviation of samples generated by the random assessment method increases drastically where correlation in the population approaches zero and the number of cost items increases. The random assessment method continues to yield the greatest overestimation of cost uncertainty and the independence method continues to yield the greatest underestimation of cost uncertainty.

The methods maintained their same capacity for underestimation and/or overestimation that they had for Scenario 1, however Scenario 2 highlights the volatile behaviour of each method as the number of cost items increases. Fortunately, the boundary conditions for correlation in the population are uncommon in reality as cost items tend to have at least some level of correlation but tend not to be perfectly correlated either. The independence method underestimates cost uncertainty for most possible values of correlation in the population and, for higher numbers of cost items, the percentage error in standard deviation increases rapidly beyond values capable with the other methods.

The correlation modelling methods respond differently to increases in the number of cost items in terms of their potential to underestimate cost uncertainty. For a highly correlated population, the random assessment method yields standard deviation percentage error of -23.7% and -28.6% where the number of cost items are 10 and 100, respectively (an increase in standard deviation percentage error of about -5%); the independence method yields -63.5% and -88.1% where the number of cost items are 10 and 100, respectively (an increase in standard deviation percentage error of about -24.6%).

Figure 21 shows the P80 percentage error for each correlation modelling method as a function of correlation strength of the population in Scenario 1.

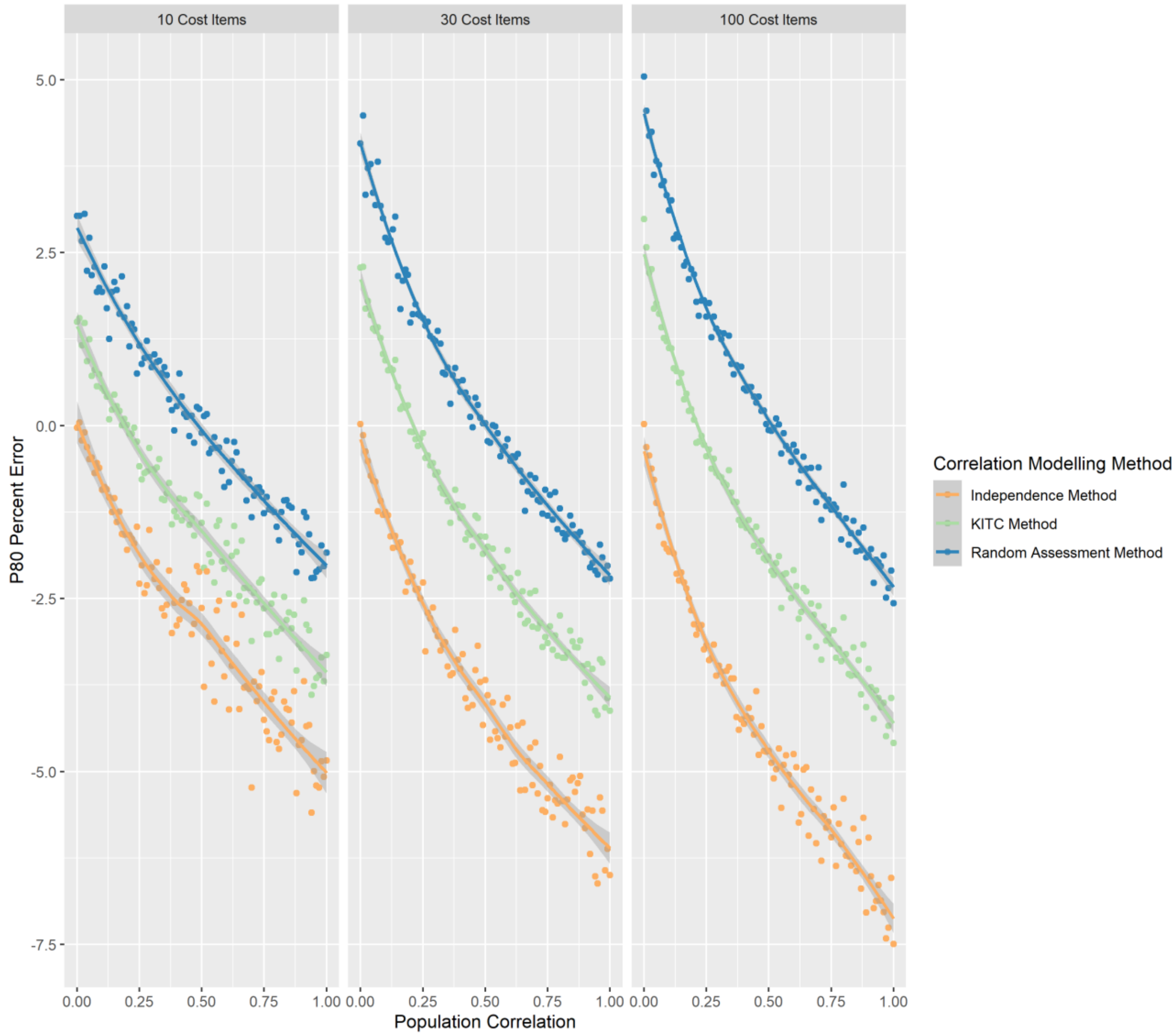


Figure 21: P80 percentage error for each correlation modelling method as a function of correlation strength of the population for three discrete values for the number of cost items in Scenario 2.

Similar to the standard deviation percentage error, the P80 percentage error increases in both directions as the number of cost items increases; for 10 cost items, the P80 percentage error spans about 5% for each method whereas for 100 cost items, the P80 percentage error spans about 7.5% for each method.

With 10 cost items, the independence method exposes an owner to a potential budget underestimation of about 5% if the owner chooses the P80 to be their level of confidence, whereas with 100 cost items, the independence method exposes an owner to a potential budget underestimation of about 7.5%. With 10 cost items, the random assessment method would expose the same owner to a mixed result where the potential budget underestimation is capped at about -2% and the potential budget overestimation is capped at about 3%; with 100 cost items, the random assessment method would expose that owner to a mixed result where the potential budget underestimation is capped at about -2.5% and the potential budget overestimation is capped at about 5%. Figure 21 supports the description of the random assessment method provided earlier, which was that the random assessment method is a more conservative version of the KITC method since it exposes an owner to less *underestimation* in cost uncertainty and ultimately budget.

The results from Scenario 2 for the independence method provide useful upper limits of potential budget underestimation resulting from a failure to account for correlation between cost items under different numbers of cost items and given a P80 level of confidence; namely potential underestimation limits of about 5%, 6%, and 7.5% for 10, 30, and 100 cost items, respectively.

4.6.3 Varying Correlation Strength and Number of Cost Items (Iterative)

The next model scenario (Scenario 3) involves varying the correlation strength between cost items in the population (similarly to Scenarios 1 and 2) along with iterative variation in the number of cost items. Like Scenarios 1 and 2, correlation strength is common to all pairs of correlated cost items and varied iteratively from zero to one in increments of 0.01 however the number of cost items is also varied iteratively except from five to 100 in increments of one. Key model parameters for Scenario 3 are summarized in Table 25.

Table 25: Key model parameters for Scenario 3.

Model Parameter	**Parameter Value(s)**
Number of cost items	Iterated from 5 to 100 in increments of 1
Proportion of correlated cost items	Fixed at 1
Correlation strength	Common to all pairs of correlated cost items, iterated from 0 to 1 in increments of 0.01

Given there are 101 values of correlation strength and 96 values for the number of cost items, the model performs 9696 iterations and produces 9696 sets of population and sample statistics as well as 9696 sets of error measurements. Figure 22 shows the standard deviation percentage error for the independence method as a function of correlation strength of the population and the number of cost items in Scenario 3; Figure 23 and Figure 24 show the same for the KITC and random assessment methods, respectively.

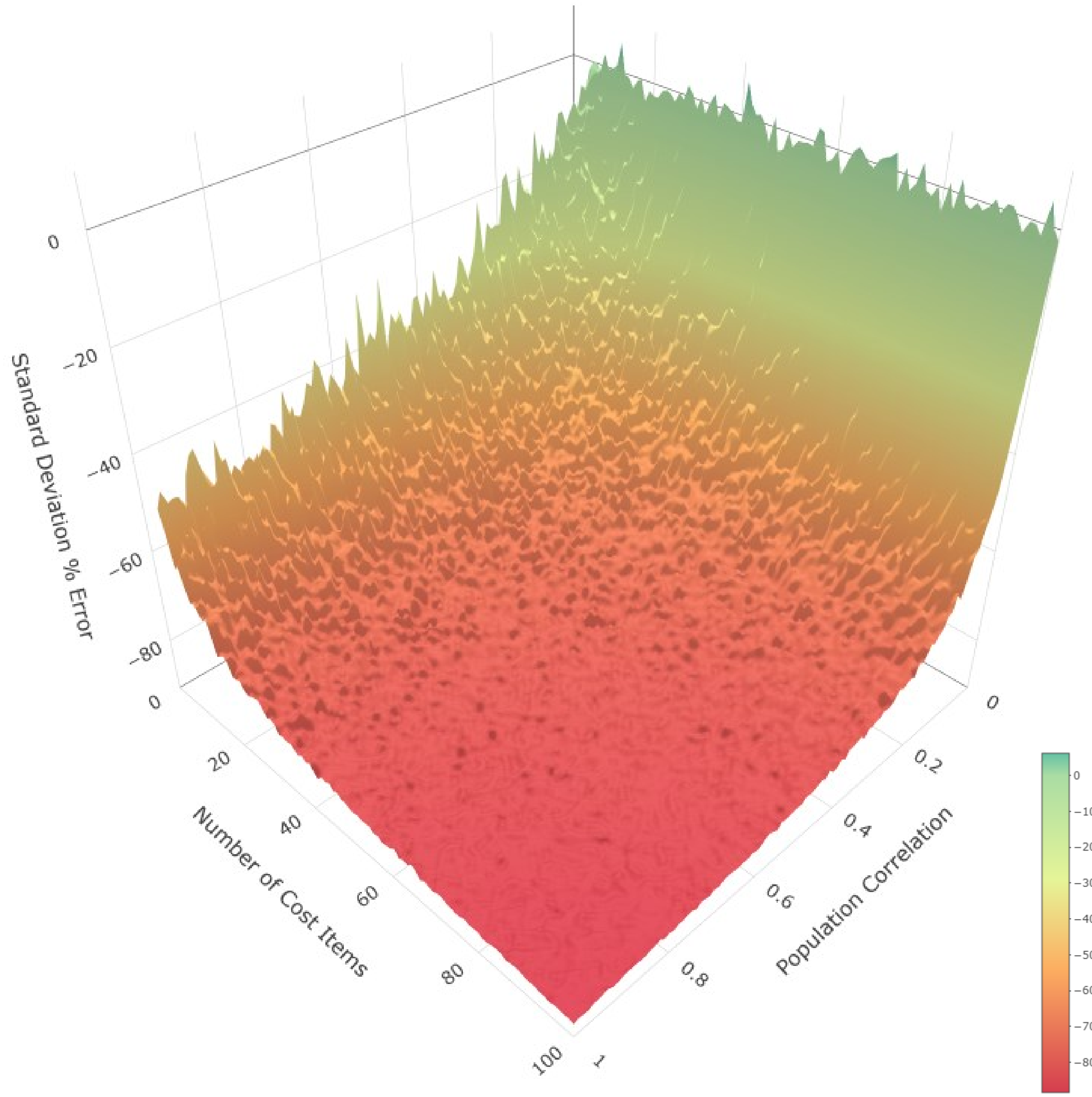


Figure 22: Standard deviation percentage error for the independence method as a function of correlation strength of the population and the number of cost items in Scenario 3.

Where Figure 20 shows the standard deviation percentage error as a function of correlation strength for three discrete numbers of cost items, Figure 22, Figure 23, and Figure 24 show the standard deviation percentage error as a function of correlation strength while the number of cost items is increased gradually. In Figure 22, nearly all combinations of population correlation and number of cost items (i.e., except where the population has no correlation) result in underestimation of cost uncertainty. For the portion of Figure 22 where the number of cost items is low (e.g., five), the standard deviation percentage error is less sensitive to the extent of correlation in the population which indicates the independence method is most suitable for low numbers of cost items.

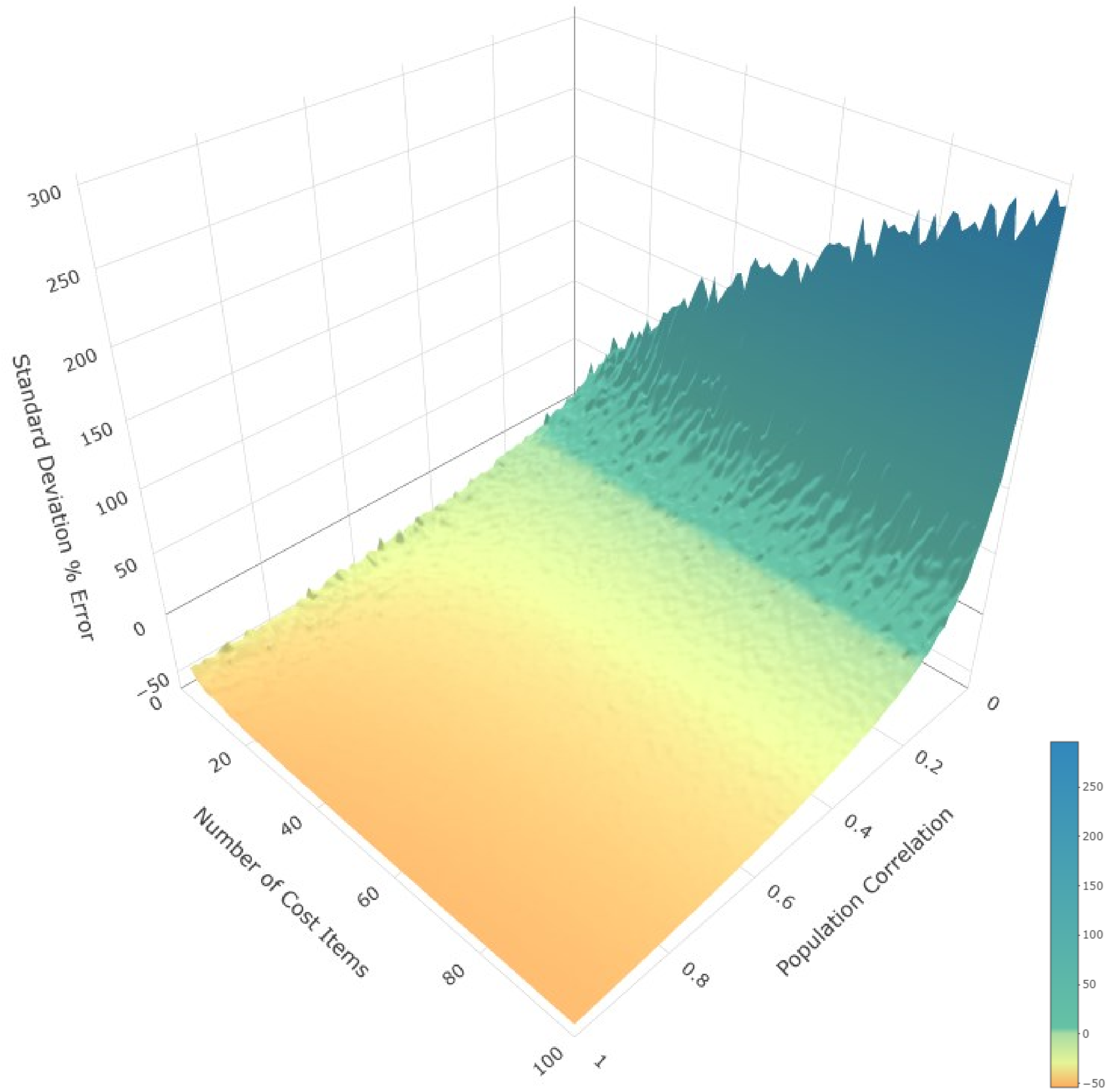


Figure 23: Standard deviation percentage error for the KITC method as a function of correlation strength of the population and the number of cost items in Scenario 3.

In Figure 23, a sizeable region of near-zero percentage error is observed where the population correlation is in the range of 0.1 to 0.3, illustrating that the KITC method performs as intended (recall that the knee in the curve is observable in this same interval) and yields low percentage error where at least some correlation is present regardless of the number of cost items. For the portion of Figure 23 where the population correlation approaches zero, the standard deviation percentage error is very sensitive to the number of cost items which indicates that when the number of cost items is high (e.g., 100), the KITC method should only be used if at least some correlation between cost items is known or expected to be present.

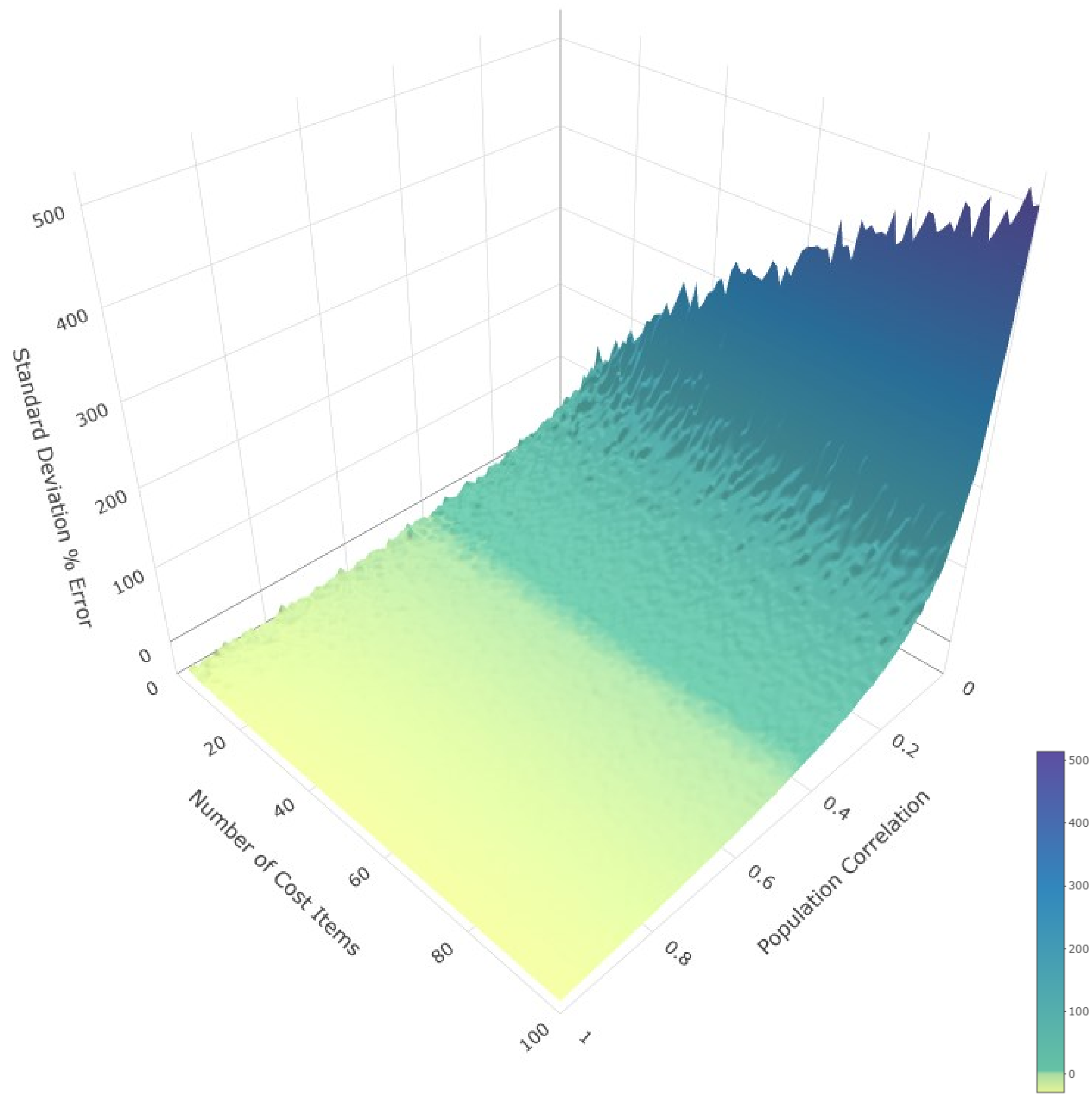


Figure 24: Standard deviation percentage error for the random assessment method as a function of correlation strength of the population and the number of cost items in Scenario 3.

A sizeable region of near-zero percentage error is also observed in Figure 24 except it is found where the population correlation is in the range of 0.4 to 0.6, illustrating that the random assessment method yields low percentage error where moderate correlation is present regardless of the number of cost items. Similarly to the KITC method, where the population correlation approaches zero, the standard deviation percentage error is extremely sensitive to the number of cost items which indicates that when the number of cost items is high (e.g., 100), the random assessment method should only be used if cost items are known to be correlated or if conservative results are desired (i.e., overestimation of cost uncertainty is considered acceptable).

4.6.4 Varying Correlation Strength and Proportion of Correlated Cost Items

The next model scenario (Scenario 4) involves varying the correlation strength between cost items in the population (similarly to Scenarios 1, 2, and 3) and the proportion of correlated cost items. Like Scenarios 1, 2, and 3, correlation strength is common to all pairs of correlated cost items and varied iteratively from zero to one in increments of 0.01 however the proportion of correlated cost items is also varied iteratively except from zero to one in increments of 0.1. The number of cost items is fixed at 100 for Scenario 4 (rather than being varied stepwise from 10 to 30 to 100) because the binomial distribution (used to determine whether cost items are correlated in each iteration) generates significant noise in simulation results when accompanied by low numbers of cost items. Key model parameters for Scenario 4 are summarized in Table 26.

Table 26: Key model parameters for Scenario 4.

Model Parameter	Parameter Value(s)
Number of cost items	Fixed at 100
Proportion of correlated cost items	Iterated from 0 to 1 in increments of 0.1
Correlation strength	Common to all pairs of correlated cost items, iterated from 0 to 1 in increments of 0.01

Given there are 101 values of correlation strength and 11 values for the proportion of correlated cost items, the model performs 1,111 iterations and produces 1,111 sets of population and sample statistics as well as 1,111 sets of error measurements. Figure 25 shows the standard deviation percentage error for the independence method as a function of correlation strength of the population and the proportion of correlated cost items in Scenario 4; Figure 26 and Figure 27 show the same for the KITC and random assessment methods, respectively.

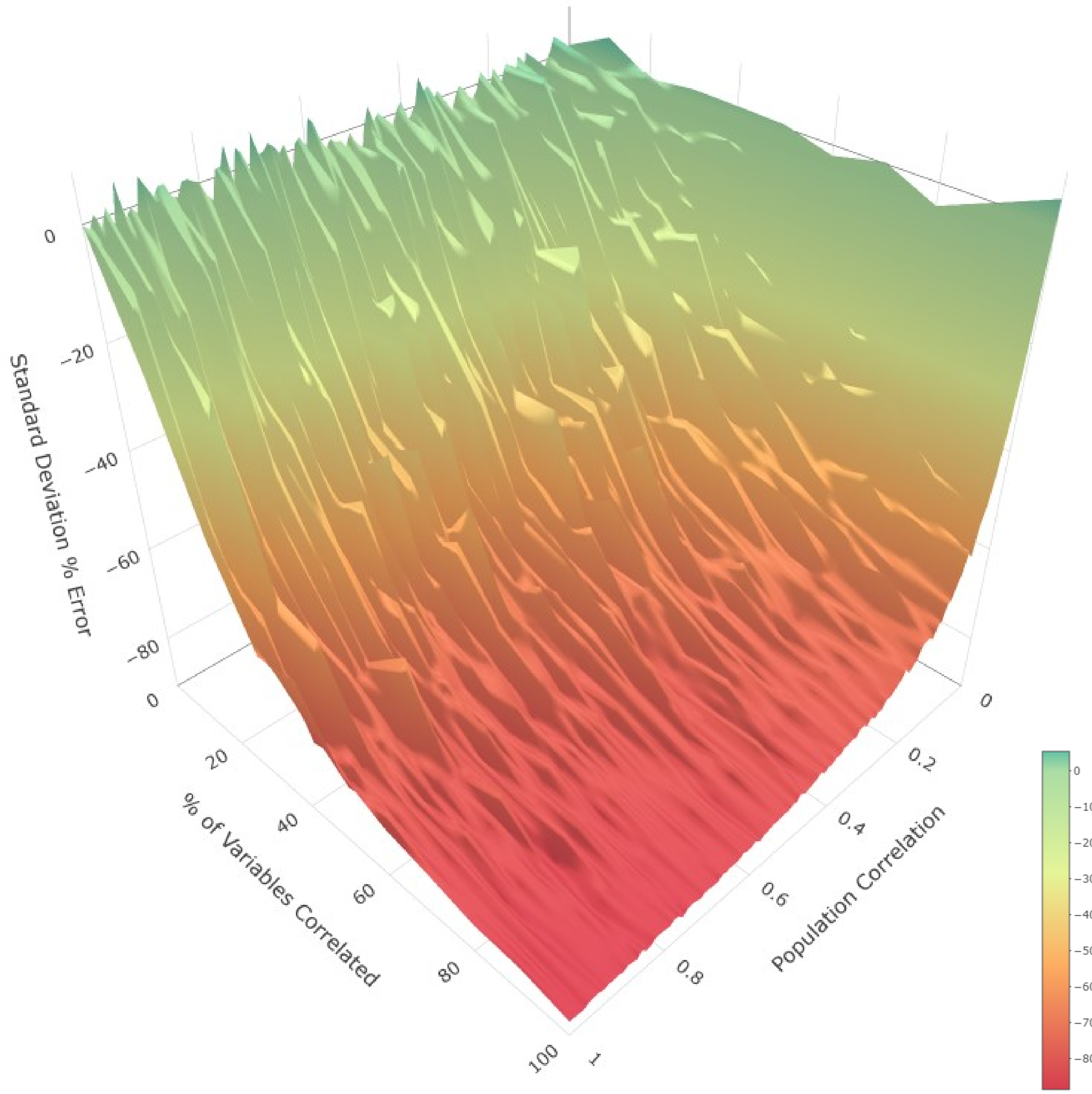


Figure 25: Standard deviation percentage error for the independence method as a function of correlation strength of the population and the proportion of correlated cost items in Scenario 4.

In Figure 25, the standard deviation percentage error for the independence method approaches zero where either the population correlation or the proportion of correlated cost items approaches zero. For combinations where the population is at least somewhat correlated and the proportion of correlated variables surpasses 20%, the extent of underestimation in cost uncertainty worsens exponentially.

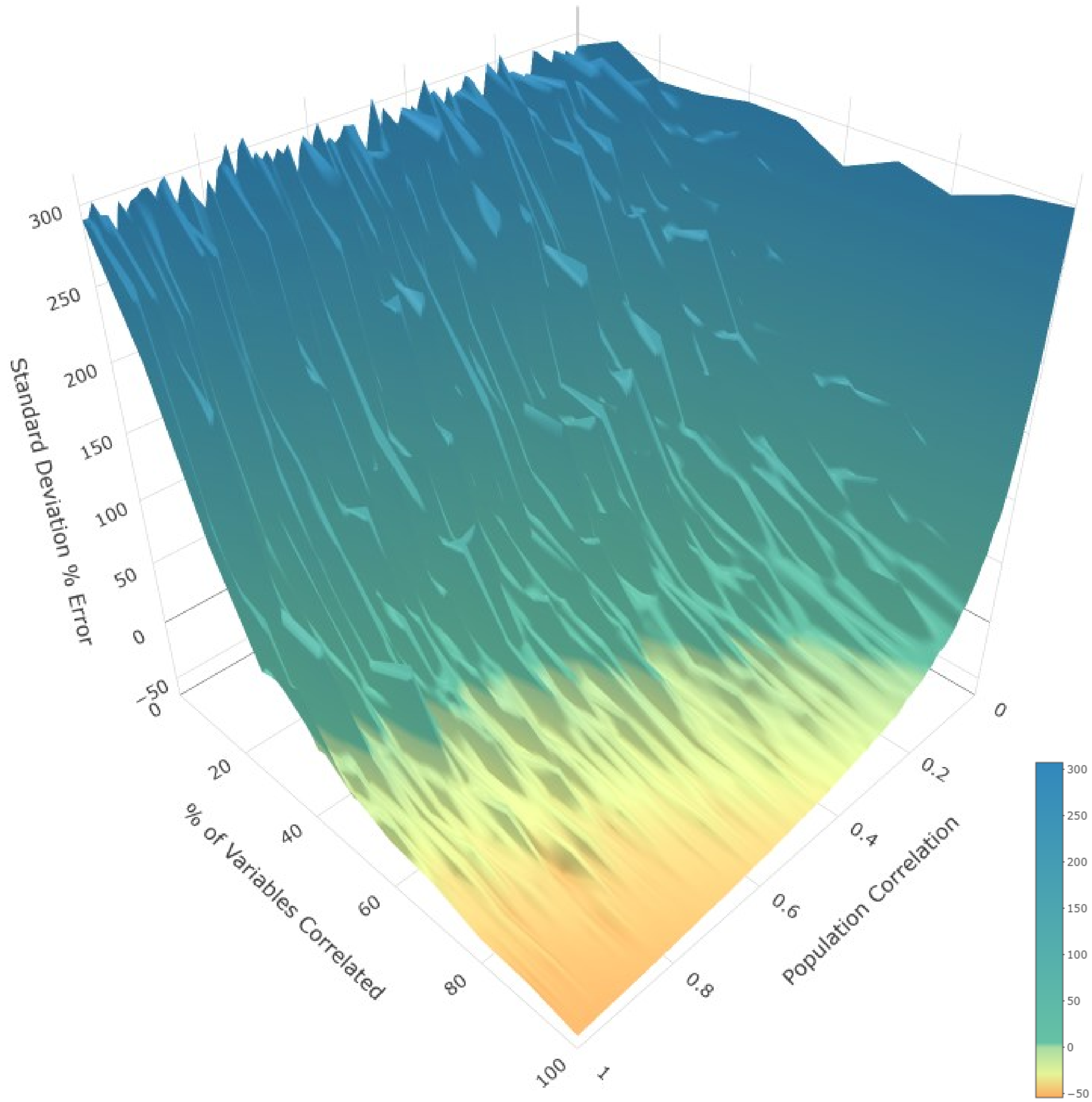


Figure 26: Standard deviation percentage error for the KITC method as a function of correlation strength of the population and the proportion of correlated cost items in Scenario 4.

In Figure 26, a band of near-zero percentage error is observed at the boundary of yellow and green surface, indicating that the KITC method yields low percentage error where low to moderate correlation is present in 50% or more cost items. Where the population correlation approaches one and a high proportion of cost items are known or expected to be correlated, the KITC method is shown to underestimate cost uncertainty (upwards of -50% percentage error in standard deviation). Conversely, where the population correlation or proportion of correlated cost items approaches zero, the KITC method is shown to significantly overestimate cost uncertainty (upwards of 300% percentage error in standard deviation).

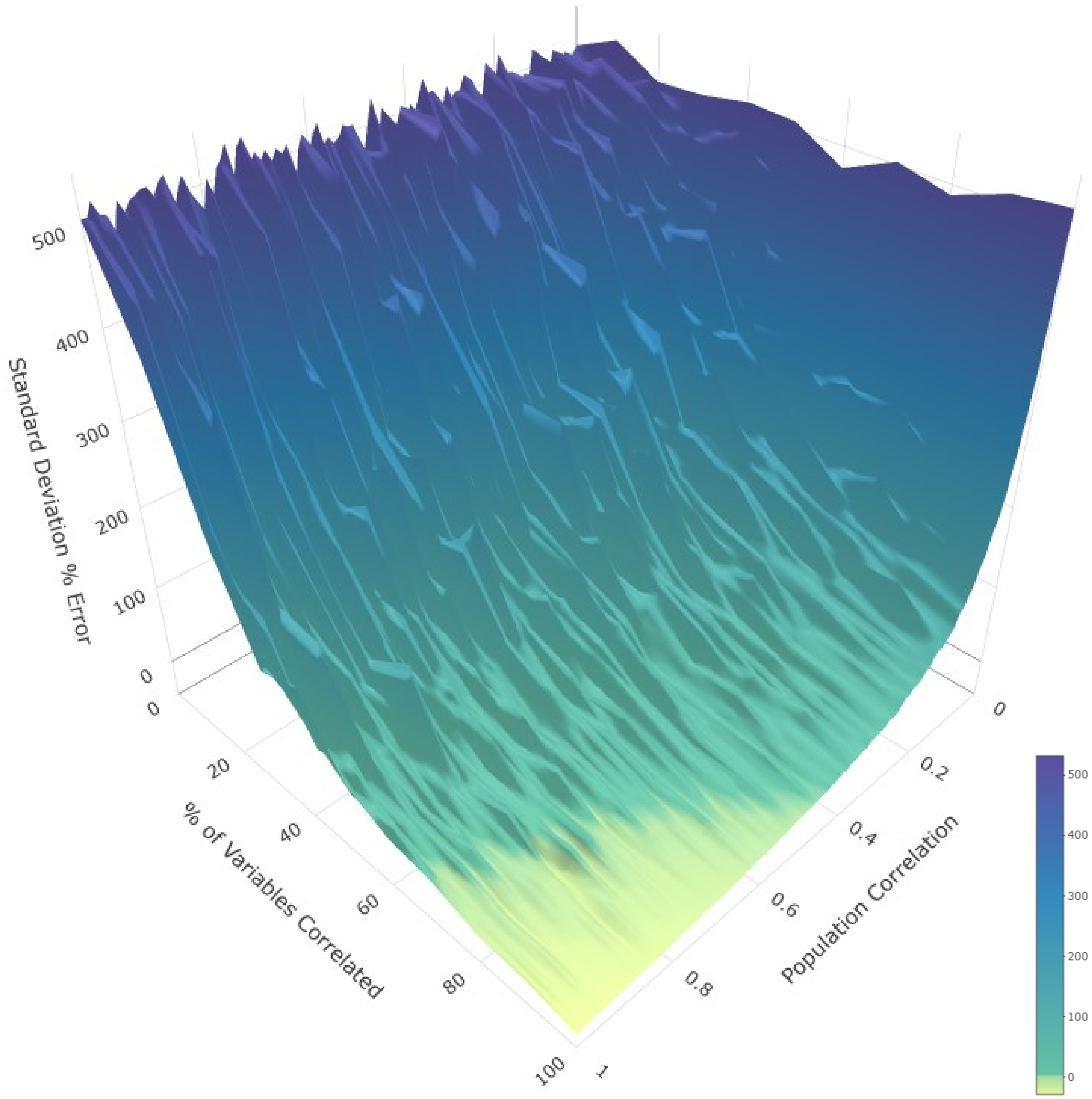


Figure 27: Standard deviation percentage error for the random assessment method as a function of correlation strength of the population and the proportion of correlated cost items in Scenario 4.

Figure 27 also shows a band of near-zero percentage error however it is situated towards higher values of correlation strength and proportion relative to Figure 26, indicating that the random assessment method yields low percentage error where moderate to high correlation is present in 60% to 70% or more cost items. The random assessment method underestimates cost uncertainty only for large proportions of highly correlated cost items. Where the population correlation or proportion of correlated cost items approaches zero, the random assessment method is shown to drastically overestimate cost uncertainty (upwards of 500% percentage error in standard deviation) even in comparison to the KITC method. The results of Scenario 4 align with the findings from Scenario 3 and indicate that the random assessment method is only suitable if a significant portion of cost items are known to be correlated or if conservative results are desired.

4.6.5 Multiple Sample Projects

The final model scenario (Scenario 5) involves random sampling of both correlation strength between cost items in the population and the number of cost items. In each iteration, the correlation strength is randomly sampled using a random deviate of the uniform distribution between zero and one and then set common to all pairs of correlated items. Meanwhile, the number of cost items is also randomly sampled using a random deviate of the uniform distribution between five and 100 and then rounded to the nearest integer for each iteration. Key model parameters for Scenario 5 are summarized in Table 27.

Table 27: Key model parameters for Scenario 5.

Model Parameter	**Parameter Value(s)**
Number of cost items	Randomly sampled from a uniform distribution with an interval [5, 100] and rounded to the nearest integer
Proportion of correlated cost items	Fixed at 1
Correlation strength	Randomly sampled from a uniform distribution with an interval [0, 1] and common to all pairs of correlated cost items

1000 iterations are performed to obtain a variety of combinations for correlation strength and number of cost items without requiring excessive computational time. Figure 28 shows a histogram of standard deviation percentage error for each correlation modelling method across all iterations in Scenario 5.

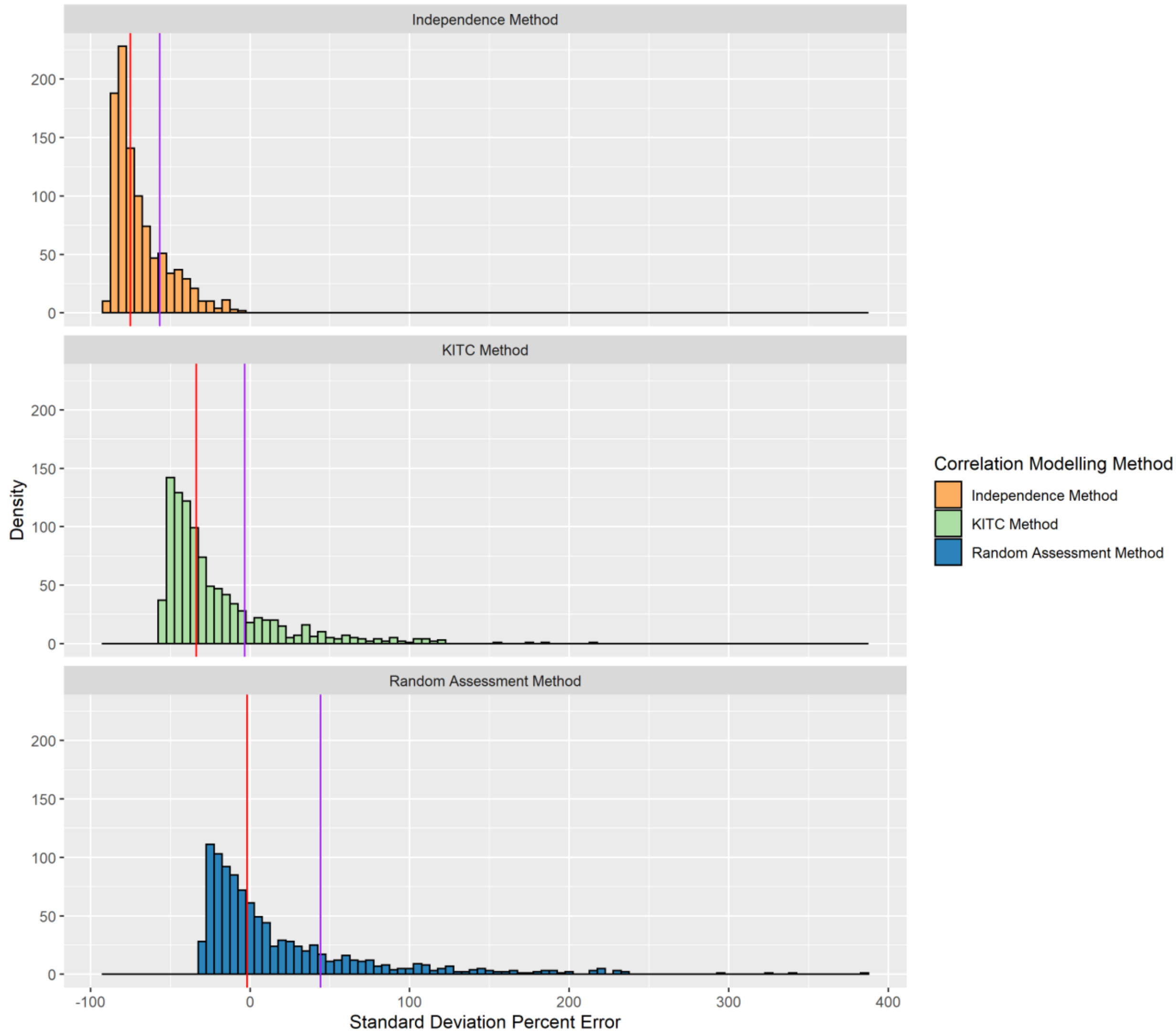


Figure 28: Standard deviation percentage error for each correlation modelling method across all iterations in Scenario 5. The red and purple vertical lines show the values for which 50% and 80% of results, respectively, fall below.

Each histogram shown in Figure 28 is skewed to the right indicating that for each correlation modelling method, standard deviation percentage error values are more concentrated on the low end of possible values in Scenario 5. As shown by the red vertical lines in Figure 28, 50% of values for the standard deviation percentage error fall below -75%, -33.7%, and -1.9% for the independence, KITC, and random assessment methods, respectively. Similarly, 80% of values for the standard deviation percentage error fall below -56.7%, -3.4%, and 44.3% for the independence, KITC, and random assessment methods, respectively, as shown by the purple vertical lines in Figure 28.

Results of Scenario 5 indicate that while the KITC and random assessment methods are capable of generating large percentage error in standard deviation (i.e., overestimation of cost uncertainty), such percentage error values represent a small subset of outcomes when the number of cost items and correlation strength are randomly sampled within large bounds. The independence method generated negative percentage error in standard deviation (i.e., underestimation of cost uncertainty) exclusively.

Average values for the standard deviation percentage error across all iterations are shown in Table 28. Based on the iterations in Scenario 5, the independence and KITC methods underestimate cost uncertainty by about 69% and 21% on average, whereas the random assessment method overestimates cost uncertainty by about 18% on average. In terms of absolute percentage error in standard deviation, the random assessment method performs best on average.

Table 28: Average values for the standard deviation percentage error across all iterations in Scenario 5.

Data Source	Average Standard Deviation Percent Error
Independence method	-69.3%
KITC method	-21.1%
Random assessment method	18.1%

Figure 29 shows a histogram of P80 percentage error for each correlation modelling method across all iterations in Scenario 5.

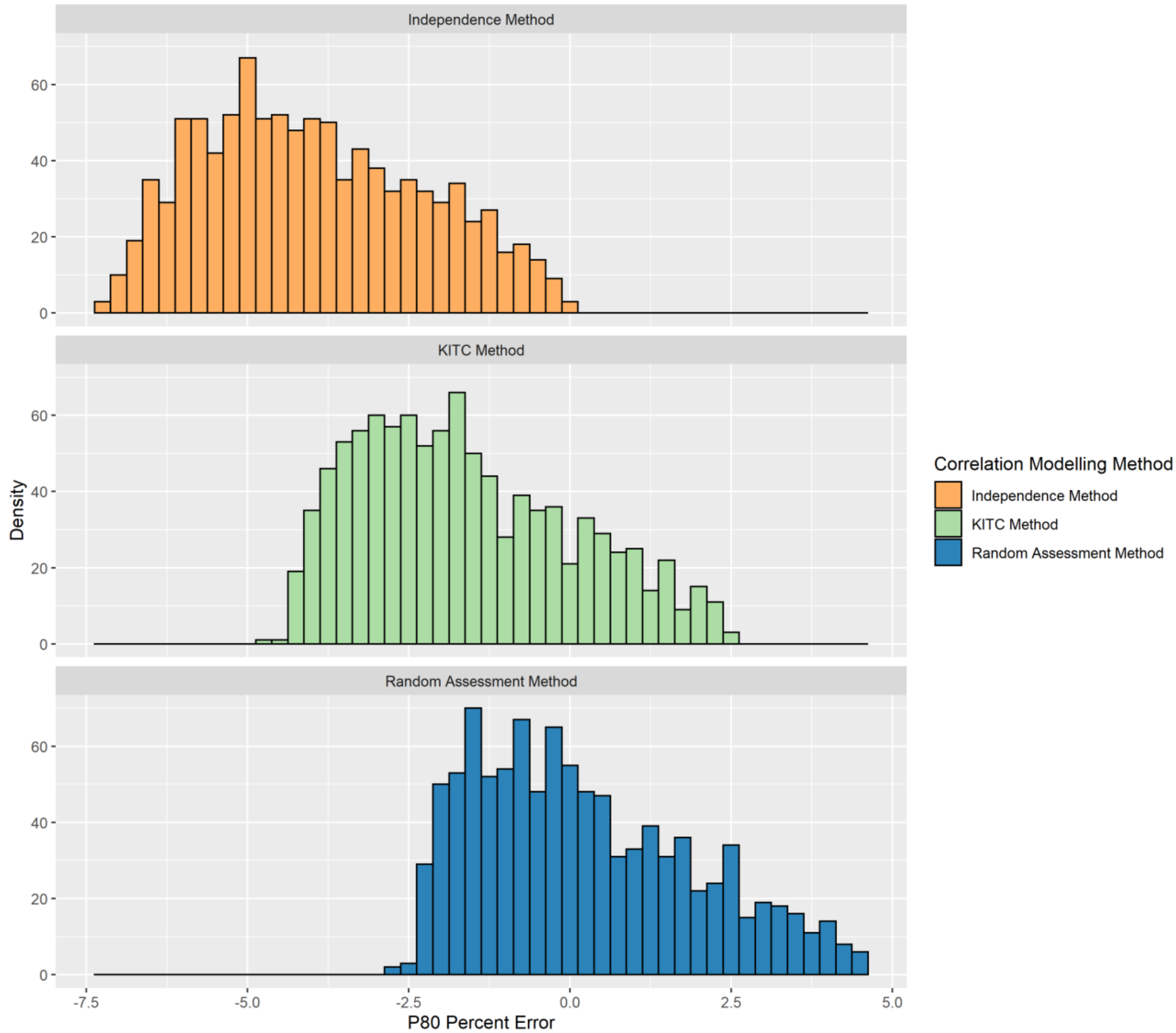


Figure 29: P80 percentage error for each correlation modelling method across all iterations in Scenario 5.

Similar to Figure 28, each histogram shown in Figure 29 is skewed to the right indicating that for each correlation modelling method, P80 percentage error values are more concentrated on the low end of possible values in Scenario 5. For the independence method, the observed skew implies that a high degree of budget underestimation is more common than a low degree of budget underestimation (provided the owner chooses the P80 to be their level of confidence).

Average values for the P80 percentage error across all iterations are shown in Table 29. Based on the iterations in Scenario 5, the independence and KITC methods result in budget underestimations of 4% and 2% on average, respectively, whereas the random assessment method accurately estimates the budget on average. In terms of absolute percentage error in P80, the random assessment method performs best on average.

Table 29: Average values for the P80 percentage error across all iterations in Scenario 5.

Data Source	Average P80 Percent Error
Independence method	-4.0%
KITC method	-1.6%
Random assessment method	0.2%

4.7 Verification and Validation

Since simulation-based cost uncertainty analysis is often used to understand potential costs and establish budgets for projects, model developers, users, and decision makers alike want to be confident that the simulation model and its results are correct. The same is true for the Monte Carlo simulation experiment presented in this section. Model verification and validation are the processes by which model competency can be tested and communicated. Verification ensures a simulation model is correctly programmed and implemented, while validation ensures a simulation model produces a satisfactory range of accuracy consistent with its intended purpose (i.e., the model accurately simulates the system it is representing).

Verification of the model focused on generating correlated cost item data, generating independent samples, and applying correlation modelling methods correctly. According to Sargent (2013), the major factor affecting the verification process is whether a simulation language or higher level programming language is used. Since the model described in this section was simulated using a higher level programming language, R, verification is primarily concerned with determining that the simulation functions and that programming has been performed correctly. The model has been tested using dynamic testing where the model is simulated under different conditions and outcomes are used to determine if the programming and implementation are correct. Dynamic testing techniques used to test the model included tracing and investigations of input-output relationships, for example:

- correlation between population costs (generated from the *rmvnorm* function) was computed and compared with the correlation matrix provided to the *rmvnorm* function;

- correlation between sample costs (generated from the KITC method) was computed and compared with the target correlation matrix; and,
- the mean and standard deviation of project costs simulated in the model were compared with corresponding values yielded by the method of moments (Appendix B shows this comparison for the base model scenario).

Validation of the experiment focused on establishing boundaries for key model parameters in various scenarios that represent a variety of project sizes, correlation strengths, and combinations thereof. Techniques used to validate the model included face validity, extreme condition testing, and parameter variability, for example:

- the logic in the conceptual model and boundaries for the number of cost items were discussed with experts in cost uncertainty analysis and construction management and deemed reasonable;
- random deviates generated from a multivariate normal distribution with a correlation strength of zero were compared with random samples generated from independent normal distributions; and,
- variations in correlation strength for the population affect the direction of cost uncertainty estimation for correlation modelling methods in alignment with the literature (Garvey et al. 2016).

4.8 Discussion

This chapter presented a Monte Carlo simulation experiment to apply different methods of modelling correlation and assess how well they represent cost uncertainty under a variety of conditions. Three correlation modelling methods were discussed: the independence method, the KITC method, and the random assessment method. In addition to a base case scenario, five model scenarios were explored where one or more key model parameters were manipulated. Each scenario included multiple iterations with the number of iterations dependent on the model parameters being manipulated. Key model parameters for the experiment included the number of cost items, the proportion of correlated cost items, and the correlation strength between correlated cost items.

In each iteration, the properties of each cost item were randomly generated based on model parameters and fed into a multivariate normal distribution to establish a population of costs for each cost item. Next, population statistics were calculated and fed into separate normal distributions to establish sample costs for each cost item. Then, sample costs were sorted and re-ordered to induce correlation (except in the case of the independence method where sample costs are already independent). Afterwards, project costs were established for the different sample sets and population and compared against population statistics to measure the percentage error. Finally, iterations were repeated until all possible combinations within a given scenario were simulated.

In every scenario modelled, the independence method produced sample sets ranging from 0% error in standard deviation to about -90% (indicating significant underestimation of cost uncertainty). The least erroneous results yielded by the independence method occurred where the correlation strength or proportion of correlated cost items approached zero, however this is considered uncommon in practice. Scenario 2 established boundaries of budget underestimation and overestimation for correlation modelling methods assuming an owner chooses the P80 to be their level of confidence. The independence method exposes an owner to potential budget underestimation of about 7.5% for a project with 100 cost items. Results from Scenario 3 showed that the independence method performs best for low numbers of cost items. When the number of cost items and correlation strength were generated randomly (Scenario 5), higher degrees of underestimation were more frequent than lower degrees of underestimation in both cost uncertainty and budget values.

The KITC method produced sample sets ranging from about 300% error in standard deviation (indicating significant overestimation of cost uncertainty) to about -55% error (indicating underestimation of cost uncertainty). The most erroneous results yielded by the KITC method occurred where the correlation strength or proportion of correlated cost items approached zero, however this is considered uncommon in practice. Assuming an owner chooses the P80 to be their level of confidence, the KITC method exposes an owner to a mixed result for a project with 100 cost items where the potential budget underestimation is capped at about -4.5% and the potential budget overestimation is capped at about 2.5%. Results from Scenario 3 showed that the KITC method performs well for any number of cost items provided the cost items are at least mildly correlated. Results from Scenario 5 indicated that extreme values of standard deviation percentage error produced from the KITC method form a small subset of possible outcomes.

The random assessment method produced sample sets ranging from about 500% error in standard deviation (indicating extreme overestimation of cost uncertainty) to about -30% error (indicating some underestimation of cost uncertainty). The most erroneous results yielded by the random assessment method occurred where the correlation strength or proportion of correlated cost items approached zero, however this is considered uncommon in practice. Assuming an owner chooses the P80 to be their level of confidence, the random assessment method exposes an owner to a mixed result for a project with 100 cost items where the potential budget underestimation is capped at about -2.5% and the potential budget overestimation is capped at about 5%. Results from Scenario 3 showed that the random method performs well for any number of cost items provided the cost items are moderately correlated. Results from Scenario 5 indicated that extreme values of standard deviation percentage error produced from the random assessment method form a small subset of possible outcomes. On average, the random assessment method yielded 0.2% error in the P80 compared to -1.6% and -4.0% for the KITC and independence methods, respectively.

Chapter 5 – Application of Correlated Modelling Methods in a Case Study

5.1 Introduction

This chapter provides a case study for the application of correlation modelling methods to cost uncertainty analysis of a light rail transit (LRT) project. The case study is based on an actual project however data has been modified to protect confidentiality. The cost uncertainty analysis is taking place following completion of preliminary design and on behalf of the owner. AAS Consulting (a fabricated name) is developing the model with participation from various team members to elicit model parameters.

The LRT project, for which cost uncertainty is being analyzed, involves an extension of the current line for a transit system within Western Canada. The LRT project also includes construction of a new operations, maintenance, and storage facility (OMSF) and upgrades to an existing OMSF. The project's budget was established at the beginning of the preliminary design phase and the owner wants to understand the current probability of delivering the project within the established budget before proceeding to the next phase. The current budget excludes escalation and risk and is set at $810 million. The project team consists of the owner's staff, design consultants, and specialty consultants such as AAS Consulting. Since the case study focuses specifically on cost uncertainty, discrete risks and time-based variables (e.g., cost escalation, cost increases associated with schedule delays, etc.) are excluded from the case study.

The cost uncertainty analysis process implemented on this project is similar to the process developed by the cost analysis community and described by Garvey et al. (2016) and is summarized as follows:

1. Review the point estimate and establish model inputs.
2. Establish probability distributions to model cost uncertainty.
3. Account for correlation between inputs.
4. Perform the cost uncertainty analysis using simulation.
5. Analyze results and identify the probability levels associated with the point estimate cost and current budget.

The remainder of this section elaborates on each step of the process as implemented for the project.

5.2 Cost Uncertainty Model Inputs

Before the cost uncertainty analysis started, a cost estimate was developed for the project by a cost consultant within the project team. The cost estimate is based on design drawings and reports provided by the design consultant at the conclusion of the preliminary design phase. The cost estimate was divided into the three main scopes of work, each with their own level of design and cost estimate accuracy:

- mainline extension (preliminary design, Class 3 estimate according to AACE);
- new OMSF (preliminary design, Class 3 estimate according to AACE); and,
- upgrades at existing OMSF (conceptual design, Class 5 estimate according to AACE).

In addition to the cost items included in the cost estimate, the project team identified and estimated additional scopes not reflected in the design drawings and reports, specifically:

- management and overhead for the project team;
- contractor management and overhead;
- public art;
- OMSF equipment;
- light rail vehicles (LRVs); and,
- off-corridor improvements.

Table 30 shows the list of cost items for the project along with their estimated costs (values are scaled to protect confidentiality). The total estimated cost of the project (excluding escalation and contingency) is $736,821,293.

Table 30: List of cost items and their estimated costs for the case study project.

Code	Task Name	Cost
0	Contract Award	-
1	Indirect Services	$184,714,983
1.1	Construction Management Services (Pre-Construction)	$2,644,270
1.1.1	Mainline	$1,359,123
1.1.2	New OMSF	$1,170,016
1.1.3	Upgrades at Existing OMSF	$115,132
1.1.4	Pre-Construction Services Milestone	-
1.2	Construction Management Services	$49,721,960
1.3	Owner Management	$24,842,150
1.4	Consultant	$105,448,753
1.4.1	Base Services	$82,905,800
1.4.1.1	Detailed Design	$44,627,100
1.4.1.2	Construction Administration & Engineering	$35,239,000
1.4.1.3	Post-Construction Services	$3,039,700
1.4.2	Additional Services	$22,542,953
1.4.2.1	Project Office	$4,384,700
1.4.2.2	Program Management & Specialized Services	$8,420,453
1.4.2.3	LRT Systems Assurance, Testing & Commissioning	$6,294,600
1.4.2.4	Partnering	$430,400
1.4.2.5	Provisional Items	$3,012,800
1.5	Public Art	$2,057,850
2	Direct Construction Costs	$455,266,310
2.1	Mainline	$255,999,631
2.1.1	Site Preparation (incl. clearing, demolitions, relocations, etc.)	$8,711,740
2.1.2	Environmental	$429,270
2.1.3	Civil Mechanical (incl. utilities)	$17,226,776
2.1.4	Civil Electrical (incl. utilities)	$163,014
2.1.5	Cycle track and sidewalk	$7,498,590
2.1.6	Avenue crossings	$4,496,470
2.1.7	Site Improvements (landscaping, misc. civil)	$16,370,192
2.1.8	Stops - Station 1	$23,568,292
2.1.9	Stops - Station 2	$23,568,292
2.1.10	Guideways / Track Work	$62,932,098
2.1.11	Traction Power	$29,478,365
2.1.12	Train Control/Communications/Security	$6,574,898

Code	Task Name	Cost
2.1.13	Lighting Systems/Traffic Control	$11,737,008
2.1.14	Sub-Stations - Station 1	$7,487,505
2.1.15	Sub-Stations - Station 2	$7,908,896
2.1.16	Handover to Owner	-
2.1.17	Owner Testing and Commissioning	-
2.1.18	Decommissioning Temporary Station	$ 2,037,675
2.1.19	Repurposing Temporary Station	$24,452,100
2.1.20	TPSS Relocation	$1,358,450
2.2	New OMSF	$152,681,027
2.2.1	A10- Foundations	$10,320,416
2.2.2	B10- Superstructure	$9,255,663
2.2.3	B20- Exterior Enclosure	$11,132,226
2.2.4	B30- Roofing	$5,360,172
2.2.5	C10- Interior Construction	$1,901,287
2.2.6	C20- Stairways	$186,379
2.2.7	C30- Interior Finishes	$1,141,913
2.2.8	D20- Plumbing Systems	$1,165,278
2.2.9	D30- HVAC Systems	$4,671,166
2.2.10	D40- Fire Protections	$1,801,305
2.2.11	D50- Electrical Systems- Building	$13,823,587
2.2.12	G10- Site Preparation	$6,778,666
2.2.13	G20- Site Improvements	$3,206,485
2.2.14	G30- Site Plumbing Utilities	$3,931,626
2.2.15	G50- Site Electrical Utilities	$27,793,887
2.2.16	G60- Track Work	$46,869,514
2.2.17	FF&E Equipment	$1,711,647
2.2.18	Sustainable Building Policy	$1,629,810
2.3	Upgrades at Existing OMSF	$46,585,652
2.3.1	Washbay Addition	$32,367,489
2.3.2	Maintenance Pits	$9,182,388
2.3.3	Heavy Maintenance Renovation	$5,035,774
3	LRV Procurement	$96,840,000
4	Construction End Milestone	-
	TOTAL	$736,821,293

5.3 Probability Distributions

In determining probability distributions to represent cost uncertainty, several contributing factors were identified by the project team. First, the preliminary engineering phase was undertaken using a risk-based approach, which sought to detail certain areas of the design more than others, based on a perceived level of risk. This means that the level of design detail (e.g., concept, preliminary, etc.) had to be evaluated for each cost item individually to determine a basis for the range of costs. Second, information on uncertainty in the unit rates and quantities required input from two groups within the project team: the cost consultant, for the unit rate uncertainty, and the design consultant, for the quantity uncertainty. Thus, two separate uncertainty factors were used to represent estimator confidence and conservativeness and designer confidence and conservativeness. This approach offered the additional benefit of a second perspective on the cost estimate for each cost item because the design consultant could comment on the cost estimate unit rates and the cost consultant could comment on the quantities and design components.

The set of cost uncertainty factors used for the project are described as follows:

- Level of design detail: the extent of design for a given cost item (e.g., concept, preliminary, detailed).
- Historic variability: dependent on the type of work (e.g., road work, track work, concrete work, etc.). This is based on historic records of projects in the area and available to AAS Consulting as well as information from literature.
- Estimator confidence and conservativeness: based on the contingency value selected by the cost consultant and assumptions used to develop the unit rates.
- Designer confidence and conservativeness: based on the level of confidence that the design consultant has with the quantities used in the estimate and how conservative the quantities are; also includes the design consultant and management gut feel on the level of uncertainty and conservativeness of the design and cost estimate.
- Flexibility for contractor changes to design: a contributor to the uncertainty with the estimate is the influence that the contractor will have during detailed design to make favourable modifications to the design to increase efficiency, standardization, etc.

- Design familiarity: relates to how established and proven the design specifications are historically and how familiar contractors will be with constructing them; a design that is very new to industry will have a higher degree of uncertainty with pricing versus a standard design.

Each factor was assigned a relative weighting of importance (shown in Table 31) and used to score the cost items qualitatively (i.e., assessed as either low, medium, and high for all factors except level of design detail). For example, if estimator (or designer) confidence and conservativeness was assessed by the project team to be low for a cost item, then that would tend to broaden the probability distribution beyond typical expectations given its level of design; conversely, if historic variability was assessed to be low for a cost item, then that would tend to tighten the probability distribution.

Table 31: Cost uncertainty factor weightings for the case study project.

Cost Uncertainty Factor	**Weighting**
Level of design detail	53%
Historic variability	4%
Estimator confidence and conservativeness	11%
Designer confidence and conservativeness	10%
Flexibility for contractor changes to design	12%
Design familiarity	10%

Scoring for each cost item was determined based on the following:

- level of design detail has been determined based on the risk-based approach for level of design as well as input from design leads captured in one-on-one meetings;
- historic variability has been set based on past cost uncertainty analyses and cost estimate studies over the past nine years in Western Canada;
- cost estimator confidence and conservativeness is based on input from the cost consultant;

- designer confidence and conservativeness and design familiarity were based on input from design leads captured in one-on-one meetings;
- flexibility for changes to design was determined based on past cost uncertainty analyses and cost estimate studies for LRT projects in Western Canada and input from design leads captured in one-on-one meetings; and,
- an overall review of the input and results was undertaken in meetings held with the design consultant and management to confirm the input assumptions.

An example of the raw data gained from interviews and past analyses for the “Stops – Station 1” cost item is as follows:

- level of design detail is preliminary based on current drawings and design report;
- historic variability is medium given past variability in estimating for LRT stations;
- estimator confidence and conservativeness is medium based on the applied unit rates;
- designer confidence and conservativeness is medium given the design is at a refinement stage (design changes are not expected) but there is potential for a key stakeholder to change requirements for column spacing;
- flexibility for contractor changes to design is low as the project team is open to adjustment but there is unlikely to be any highly impactful change due to the heavy review and vetting process; and,
- design familiarity is high as the design is fairly standard and while there are some complicated components, they are no more complicated than any other well-finished architectural structure.

After processing the raw data, the probability distribution for the cost uncertainty of the “Stops – Station 1” cost item is as shown in Figure 30.

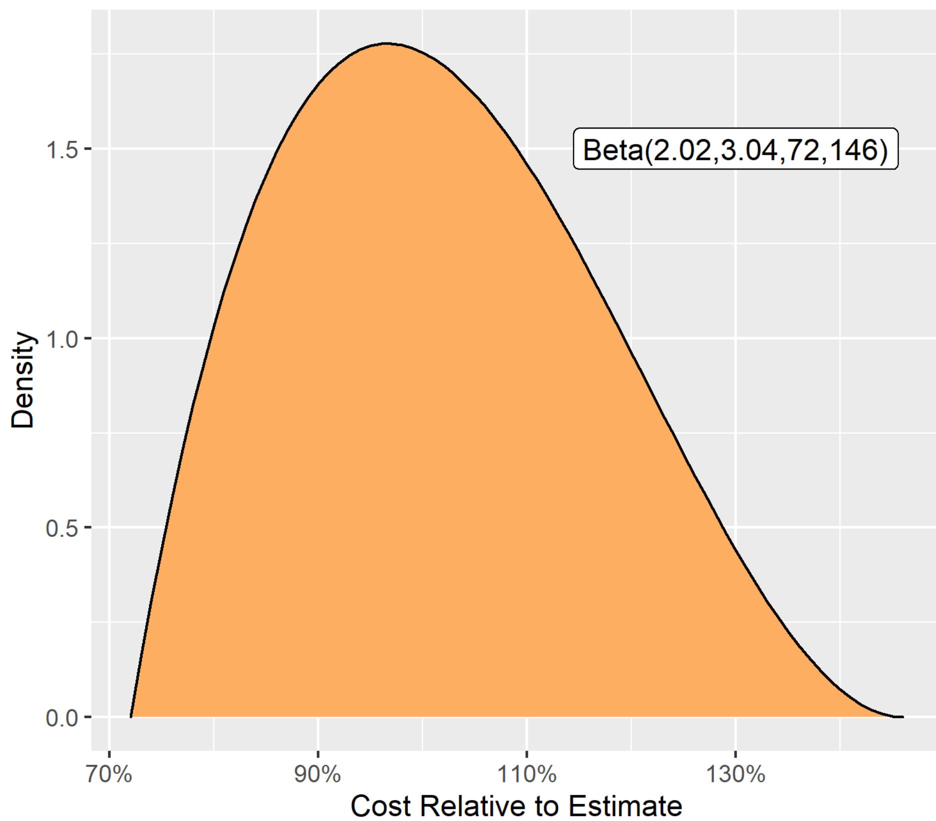


Figure 30: Beta distribution established for the cost uncertainty of the "Stops - Station 1" cost item. Cost values are relative to the cost included in the project estimate.

The minimum and maximum values for the "Stops – Station 1" cost item are 72% and 146%, respectively, resulting in a distribution that is slightly more narrow than the typical range used for a preliminary design level of design on the project. The distribution has been narrowed based on the raw data collected for the cost item, and in particular due to a high degree of design familiarity for this scope. The distribution shown in Figure 30 is skewed to the right which indicates that costs are considered more probable at the lower end of the range compared to the higher end of the range.

Beta distributions were chosen for the cost uncertainty analysis on this project as they are well-suited to representing skewed behaviour and can be shaped so that they return values that are spread across a wide or a narrow range, while avoiding the normal distribution's long-tail effects.

5.4 Correlation

Correlation between cost items was accounted for in one of two ways: correlation related to risk and correlation related to other sources (e.g., changes in labour rates or other market forces, shared labour or material costs, and changes in design with implications to multiple cost items).

To account for correlation related to risk, the analysis contemplated discrete risks which were each linked to one or more cost items. As risks with cost impacts are realized in the simulation, the costs of linked cost items increases accordingly based on anticipated proportions, thus capturing risk-related correlation using functional relationships within the model. As previously mentioned, the case study focuses specifically on cost uncertainty, so discrete risks and their correlation are excluded from the remaining analysis.

To account for correlation for sources other than risk, the analysis considered three correlation modelling methods: the independence method, the KITC method, and the random assessment method (refer to Section 4.4 for a detailed description of each method). These methods were applied in the case study since:

- data were not available to derive the correlation mathematically or empirically and,
- qualitative assessment of each cost item pair was considered infeasible since there are 1,326 assessments required for the 52 uncertain cost items within the project.

Where costs are modelled independently, the independence method requires no further action whereas the KITC and random assessment methods induce correlation by reordering of independent samples after the simulation (to be discussed in Section 5.5).

5.5 Simulation

After probability distributions were established for each cost item and an approach for modelling correlation was decided, costs for each cost item were sampled individually. AAS Consulting chose to generate 10,000 samples for each cost item since this yields a confidence interval width of less than 0.01 at a 95% level of confidence. Several cost items were determined to be constant (i.e., represented using a static value for each iteration) rather than represented using a probability distribution; these included:

- "Contract Award", "Pre-Construction Services Milestone", "Handover to Owner", "Owner Testing and Commissioning", and "Construction End Milestone" which were placeholder cost items intended for use in schedule analysis (excluded from the case study); and,
- "Public Art" and "Sustainable Building Policy" which were committed amounts with certainty on their value.

After samples were generated for each cost item, cost items were separated into two groups: those that were modelled with probability distributions and those that were static for each iteration. The cost items needed to be separated in preparation for correlation modelling methods since correlation cannot be induced in samples that are constant for all iterations. The remainder of this section describes how each correlation modelling method uses the independent cost item samples to generate samples of project cost.

5.5.1 Applying Independence Method

Since the sample costs for each cost item were generated using independent probability distributions, cost samples between cost items are already independent. Since no sorting of samples is required, the two groups of cost items are recombined. To apply the independence method, the sample costs for each cost item are simply added together to produce a sample project cost (as shown in Equation 24); this is repeated until 10,000 project cost samples have been calculated. Figure 31 shows a histogram of project cost samples from the independence method in the case study. The mean and standard deviation of sample project costs from the independence method are $742,988,200 and $22,442,732, respectively; the P80 is $761,875,985.

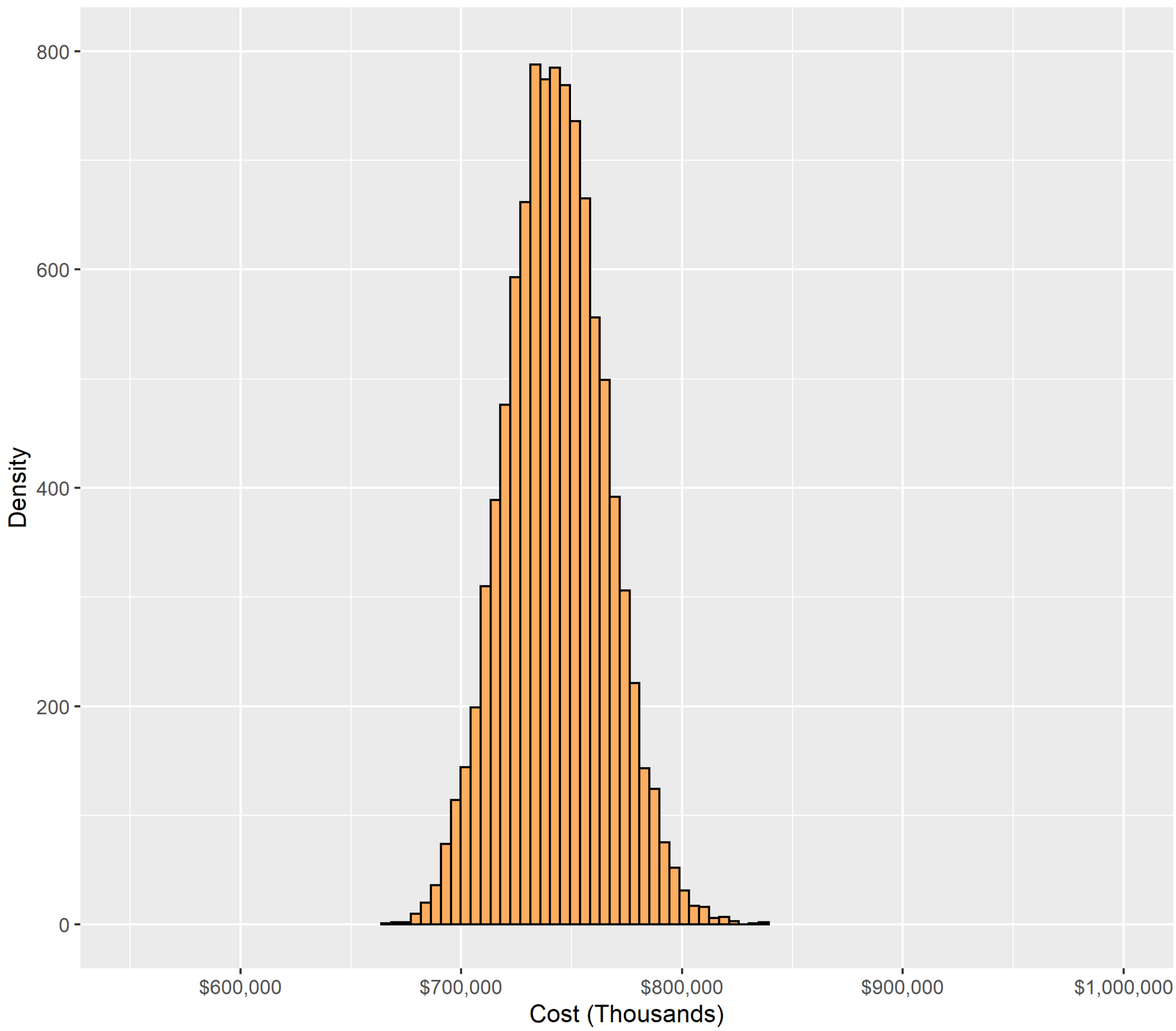


Figure 31: Samples of project cost from the independence method in the case study.

5.5.2 Applying Knee-in-the-Curve Method

To apply the KITC method, a correlation matrix is generated with correlation strength on the interval [0.10, 0.30] and with dimensions equal to the number of cost items modelled with probability distributions (refer to Section 4.4.2 for a detailed description of the process). Next, cost samples for each cost item are column sorted in a matrix and reordered using the *SJpearson* function to induce correlation between variables approximately equal to the target correlation matrix.

After the cost items modelled with probability distributions have been correlated, they are recombined with cost items that were static for each iteration. Finally, project cost samples are calculated by adding sample costs for each cost item together (as shown in Equation 24). Figure 32 shows a histogram of project cost samples from the KITC method in the case study. The mean and standard deviation of sample project costs from the KITC method are $742,988,200 and $49,932,817, respectively; the P80 is $785,645,832.

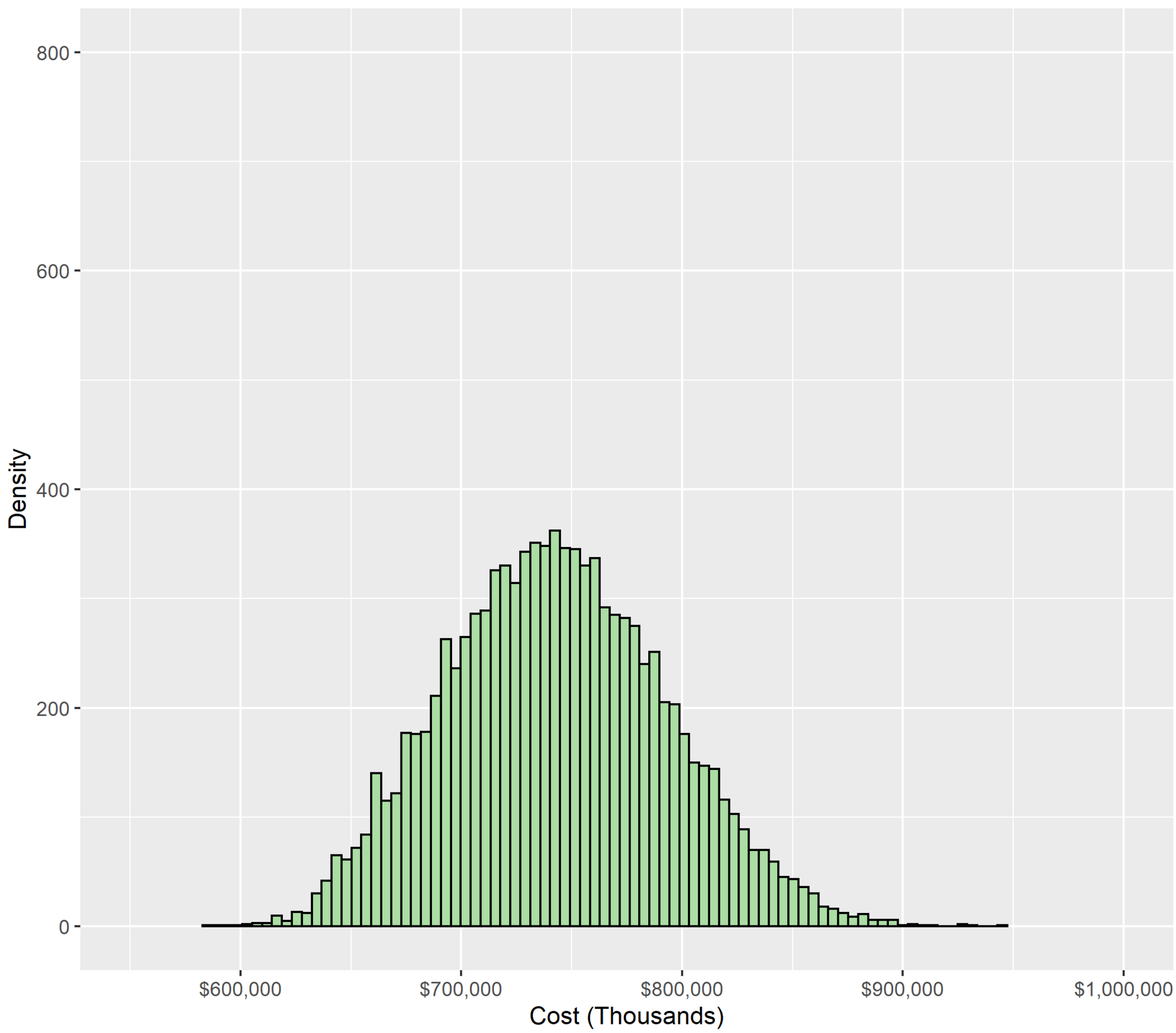


Figure 32: Samples of project cost from the knee-in-the-curve method in the case study.

5.5.3 Applying Random Assessment Method

The process for applying the random assessment method is the same as the process described for the KITC method in Section 5.5.2, except the correlation matrix is generated with correlation strength on the interval [0.10, 0.90]. Figure 33 shows a histogram of project cost samples from the random assessment method in the case study. The mean and standard deviation of sample project costs from the random assessment method are $742,988,200 and $73,706,897, respectively; the P80 is $808,095,518.

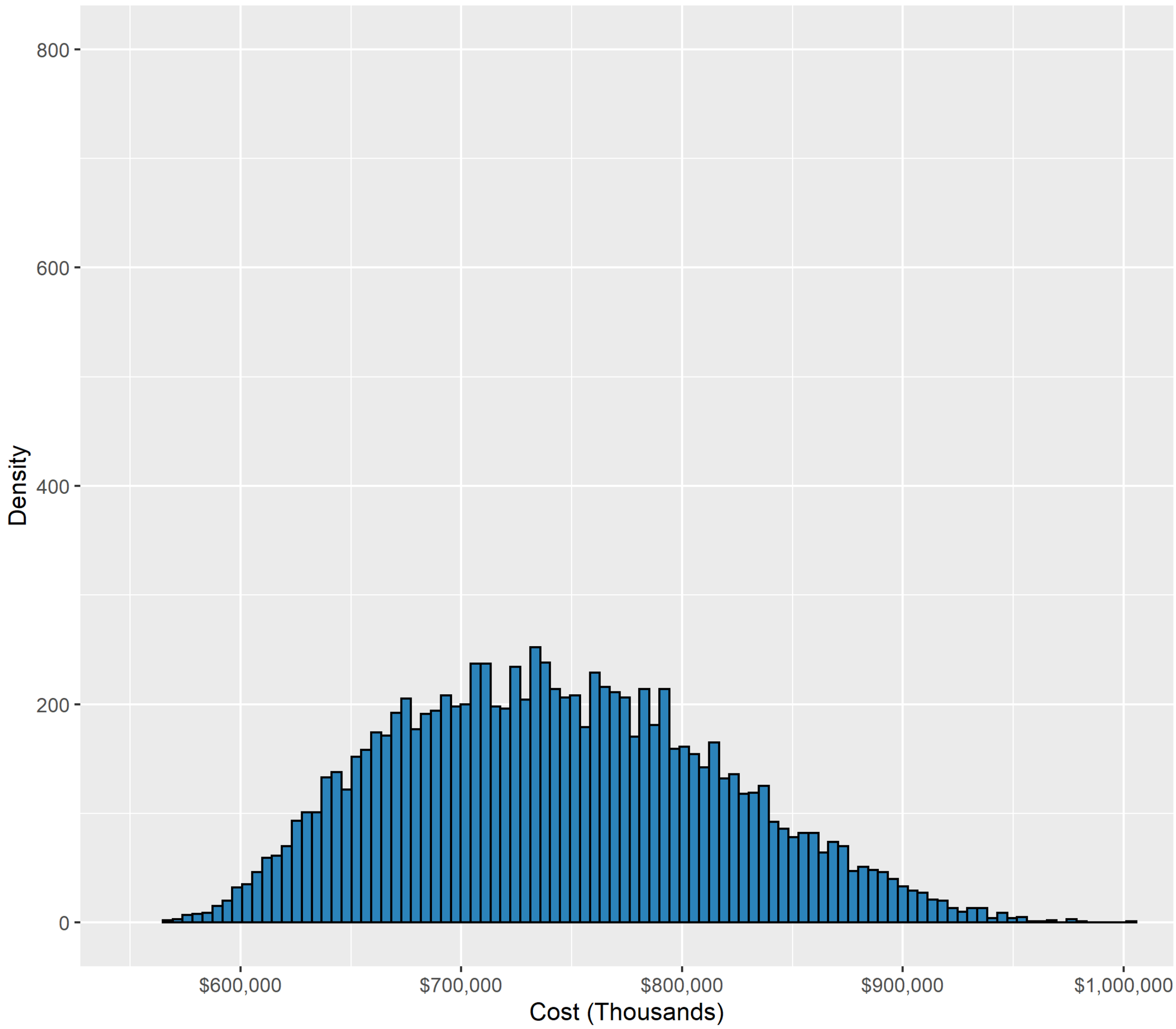


Figure 33: Samples of project cost from the random assessment method in the case study.

5.6 Results Analysis

After each correlation modelling method has been applied, the sample project costs produced by each method are compared. Figure 34 shows histograms of project cost samples from each correlation modelling method in the case study. Table 32 summarizes the mean, standard deviation, and P80 of project costs corresponding to each correlation modelling method in the case study.

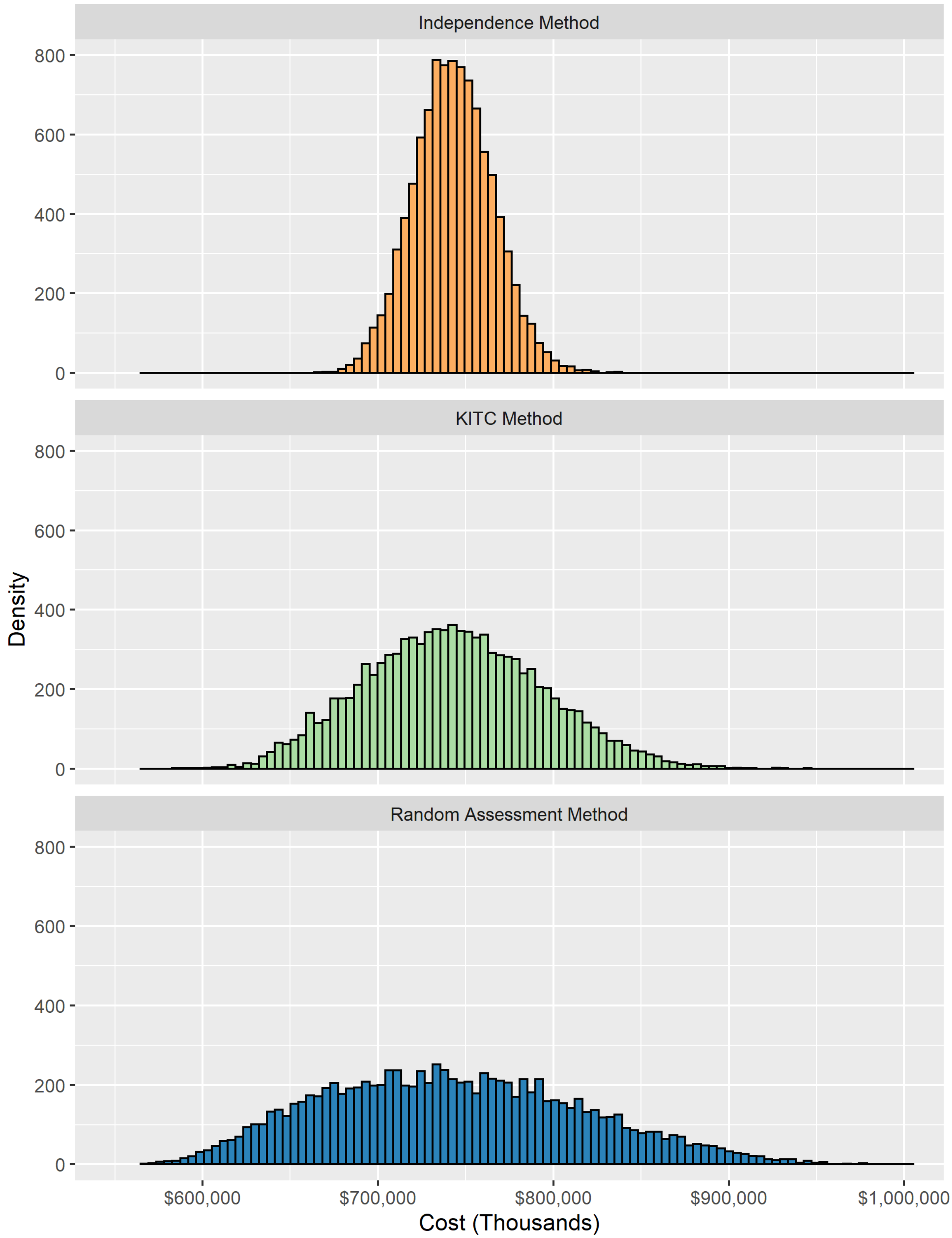


Figure 34: Samples of project cost from each correlation modelling method in the case study.

Table 32: Mean, standard deviation, and P80 of project costs for correlation modelling methods in the case study.

Correlation Modelling Method	Project Cost Mean	Project Cost Standard Deviation	Project Cost P80
Independence method	$742,988,200	$22,442,732	$761,875,985
KITC method	$742,988,200	$49,932,817	$785,645,832
Random assessment method	$742,988,200	$73,706,897	$808,095,518

As expected, the mean of sample project costs is consistent for all methods since the reordering that occurs for the KITC and random assessment methods does not affect the mean. The project cost standard deviation varies significantly between methods in the case study. For example, the standard deviation of project costs from the KITC and random assessment methods are larger than the same value from the independence method by factors of approximately 2.22 and 3.28, respectively. In terms of values that could be used to establish or validate budgetary amounts, the P80 project cost amounts vary from approximately $762 million to $808 million, depending on the correlation modelling method used. The spread between the most discrepant P80 values (independence method versus random assessment method) is approximately $46 million, or approximately 6% relative to the P80 yielded by the independent method.

When using the KITC method, Garvey et al. (2016) recommend conducting a sensitivity analysis to understand the model's sensitivity to the choice of correlation strength and determine a value for which increases in correlation strength are less consequential, particularly given the decision maker's desired level of confidence. As part of the case study, a sensitivity analysis scenario was implemented wherein the correlation matrix was scaled iteratively for various intervals of correlation strength. The correlation strength started with the interval [-0.1, 0.1] (centered on zero correlation) and increased in increments of 0.1 until the interval [0.6, 0.8] (centered on correlation of 0.7). Figure 35 shows project costs corresponding to different percentiles as correlation induced in the model varies.

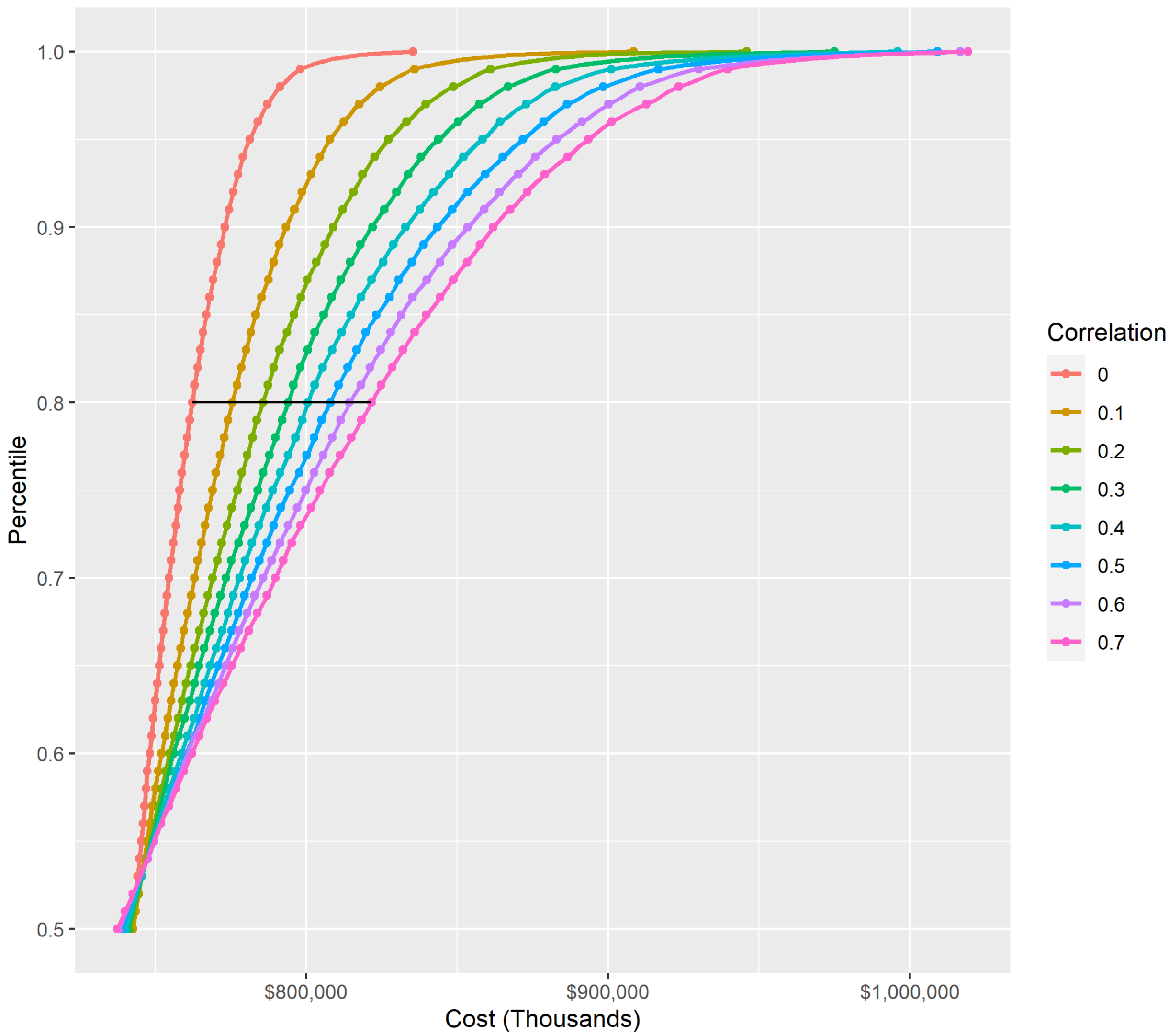


Figure 35: Project costs corresponding to different percentiles as correlation induced in the model varies. A horizontal black line is shown where the percentile is 0.8, i.e., P80.

The left-most curve (shown in red) has zero correlation and thus produces similar results to the independence method. The third curve from the left (shown in lime green) has correlation equal to 0.2 (the same midpoint as the KITC method) and thus produces similar results to the KITC method. The third curve from the right (shown in blue) has correlation equal to 0.5 (the same midpoint as the random assessment method) and thus produces similar results to the random assessment method.

Model sensitivity to correlation can be observed in Figure 35 by choosing a given percentile (e.g., P80) and examining how project cost (shown on the x axis) changes for different correlation strengths. In Figure 35, observe that as the correlation increases, the change in value at the 80th percentile decreases; beyond a correlation strength of 0.3, successive cost differences are progressively less consequential. Additionally, functional correlation related to risk has been captured separately for the project. Thus, results generated from the KITC method are deemed most appropriate and the remainder of the discussion is based on such results.

Finally, with the project cost samples generated from the KITC method, the probabilities associated with the point estimate and current budget can be calculated. The point estimate for the project was $736,821,293 which corresponds to approximately the 46th percentile of simulated project costs; this result indicates that 46% of simulated project costs are at or below the point estimate value. Conversely, 54% of simulated project costs exceed the point estimate value, representing a 54% chance of exceeding the point estimate. The current project budget is $810 million which corresponds to approximately the 90th percentile of simulated project costs. The owner's desired level of confidence is P85 which corresponds to a value of $795,935,906. The results of the cost uncertainty analysis indicate that the current project budget is in excess of the value yielded by the simulation by about $14 million. Depending on the owner and their assessment of the project's risk, surplus funds can be allocated to contingency or a separate management reserve (i.e., increase the level of confidence to P90 and maintain the budget as is) or released for use on other projects (i.e. decrease the budget to $796 million). For the case study, the project budget was maintained at $810 million and excess funds were held in contingency.

5.7 Discussion

This chapter demonstrated the application of correlation modelling methods to cost uncertainty analysis using a case study. The case study concerned cost uncertainty analysis completed for an LRT project in Western Canada with minor alterations made to the data to ensure confidentiality. The case study generally followed the process outlined in Section 2.4.1 for cost uncertainty analysis and applied correlation modelling methods introduced in Section 4.4. Results showed the sensitivity of the model to different values of correlation between cost items and confirmed that the current project budget slightly exceeds the amount required given a level of confidence set at P85.

Chapter 6 – Conclusions

6.1 Thesis Summary

This thesis intended to improve the accuracy of cost uncertainty analysis and in particular remedy deficiencies in current practice that contribute to the underestimation of project costs. The thesis began with an introduction of cost uncertainty analysis and the current state of the art for construction practice and other fields. Gaps were found in the construction engineering literature with respect to establishing subjective probability distributions and capturing correlation between model inputs.

To address the former gap, this thesis sought to improve the representation of cost uncertainty by understanding potential elicitation biases and how they can be reduced through structured elicitation protocols. An elicitation protocol called SHELF was applied to cost uncertainty analysis for a construction project as part of an illustrative example. To address the latter gap, a Monte Carlo simulation experiment was designed to examine the effects of correlation on project cost uncertainty and assess various methods for modelling correlation. The experiment involved generating cost item data, collecting statistics and generating samples, applying correlation methods, and evaluating each method relative to the generated data under a variety of scenarios.

Three correlation modelling methods were applied as part of the experiment: the independence method, the KITC method, and the random assessment method. In addition to a base case scenario, five model scenarios were explored where one or more of the following model parameters were manipulated: the number of cost items, the proportion of correlated cost items, and the correlation strength between correlated cost items.

In all scenarios, except where correlation between cost items was near-zero, the independence method consistently generated project costs which underestimated cost uncertainty. For greater numbers of cost items, the extent of underestimation produced by the independence method increased exponentially where at least some correlation was present. At near-zero correlation, the independence method estimated cost uncertainty accurately while the KITC and random assessment methods yielded significant overestimation, however near-zero correlation seldom occurs in practice.

In Scenario 2, the correlation strength between cost items in the population was varied for 10, 30, and 100 cost items. Using the results for the independence method in Scenario 2, Table 33 shows the upper bounds of potential budget underestimation resulting from a failure to account for correlation under different numbers of cost items and given a P80 level of confidence.

Table 33: Maximum potential budget underestimation for the independence method resulting from a failure to account for correlation given a P80 level of confidence and different numbers of cost items.

Number of Cost Items	Maximum Budget Underestimation
10	5%
30	6%
100	7.5%

Results from Scenario 3 indicated that the independence method performs at its best for low numbers of cost items[6]. Provided cost items are mildly correlated, the KITC method performs well for any number of cost items; similarly provided cost items are moderately correlated, the random assessment method performs well for any number of cost items.

Results from Scenario 5 indicated that for the independence method, higher degrees of underestimation were more frequent than lower degrees of underestimation in both cost uncertainty and budget values. Conversely, for both the KITC and random assessment methods, extreme values of standard deviation percentage error (i.e., significant overestimation) form a small subset of possible outcomes. Based on the results from Scenario 5 and assuming a P80 level of confidence, the independence and KITC methods underestimated the budget by 4% and 2% on average, respectively, whereas the random assessment method estimated the budget within 0.2% on average.

Finally, a case study was presented to illustrate the application of experimental modelling methods to an actual project. The case study applied the three correlation modelling methods included in the Monte Carlo simulation experiment to independently sampled costs. Before determining the most appropriate correlation modelling method, sensitivity analysis was performed to assess the model's sensitivity to the choice of correlation strength. Results generated from the KITC method were deemed most appropriate. The probabilities associated

[6] This finding aligns with a recommended practice from AACE International (2008) which cautions against independently ranging a high number of items.

with the point estimate and project budget values were determined based on project cost samples and the budget exceeded the desired level of confidence.

6.2 Overall Conclusion

Results of the Monte Carlo simulation experiment support the hypothesis that since cost uncertainty (measured by standard deviation) is affected by correlation between model inputs, modelling correlated inputs independently results in underestimation of cost uncertainty. In fact, independent modelling of inputs resulted in an underestimation of cost uncertainty except where all cost items are entirely uncorrelated which is seldom the case in the context of cost uncertainty analysis. Unlike the KITC and random assessment methods, the independent method did not result in an overestimation of cost uncertainty in any model scenario[7].

Based on outcomes from the Monte Carlo simulation experiment, recommendations for modelling correlation in cost uncertainty analysis are as follows:

- Where possible, capture functional relationships in the model to account for correlation;
- Where data are available, derive correlations mathematically or empirically;
- If data are not available; assign correlations subjectively;
    - For a low number of cost items, assess correlation between pairs qualitatively[8];
    - If the number of cost items is too high or if the time and/or resources required to qualitatively assess correlation are not available, then:
        - If cost items are believed to be mildly correlated, apply the KITC method or;
        - If cost items are believed to be moderately or strongly correlated, apply the random assessment method;
        - If the analyst and/or decision makers do not have the confidence to assess the extent of correlation more broadly, conduct analysis on the model's sensitivity to the choice of correlation strength, determine a value for which increases in correlation strength are less consequential, particularly

[7] Correlation strength in the model was limited to values between zero and one (i.e., positive correlation) for all scenarios. If negatively correlated cost items were present, the independence method could generate an overestimate of cost uncertainty.
[8] A subjectively determined correlation matrix may need adjustment to satisfy the properties of a correlation matrix (e.g. the matrix must be consistent).

given the decision maker's desired level of confidence, and induce correlation on the value identified from the analysis;

- Alternatively where there are relatively few cost items (e.g., in the range of 10 to 20), model cost items independently and increase the percentile value by the percentage of exposure in budget underestimation[9].

6.3 Contributions

The contributions from this thesis are as follows:

- an elicitation protocol previously applied in fields such as medicine and business planning was applied to cost uncertainty analysis in construction practice;
- the thesis provided a methodology for evaluating the performance of different correlation modelling methods in an unbiased and reproducible manner;
- the Monte Carlo simulation experiment demonstrated:
    - the range of costs produced by cost uncertainty analysis is very sensitive to the extent of correlation between cost items;
    - when correlation is present but unaccounted for in modelling methods, cost uncertainty can be significantly underestimated;
    - bounds for potential underestimation and overestimation can be empirically derived according to a desired level of confidence and number of cost items; and,
- the thesis provided analysts with practical recommendations for valuing and modelling correlation depending on access to data and resources, the number of cost items within a project, and the extent of correlation expected between cost items.

6.4 Recommendations for Further Research

Opportunities for further research related to this thesis are as follows:

- Schedule uncertainty analysis: In the same way a cost estimate is a single outcome in a probability distribution of possible cost outcomes, the schedule estimate is only a single outcome. Simulation-based schedule uncertainty analysis faces challenges which are

[9] For example, if an owner chooses the P80 to be their level of confidence and the project contains 100 cost items, then increase the 80^{th} percentile simulation result by 7.5% to account for potential underestimation of cost uncertainty (refer to Table 33 for the percentage exposure corresponding to other numbers of cost items).

similar to cost uncertainty analysis. For example, schedule activities may or may not be correlated which affects the potential range of project durations.

- Integrated cost-schedule uncertainty analysis: This thesis focused specifically on cost uncertainty analysis however there are models that integrate cost and schedule to assess uncertainty. In integrated models, additional inputs are required and may be correlated. For example, if an activity increases in duration, it is possible that the cost of that activity may also increase.
- Integration with quantitative risk analysis: The Monte Carlo simulation experiment and case study did not include discrete risks which would typically be modelling in cost uncertainty analysis using quantitative risk analysis. If discrete risks are linked to multiple cost items, this can introduce functional correlation between those cost items.
- Evaluation of the applied elicitation protocol: The thesis applied the SHELF protocol to an illustrative example rather than to one or more actual projects. Experiments could be designed to apply the SHELF method as well as other elicitation protocols and evaluate their performance.
- Qualitative assessments of correlation: For projects with fewer cost items, qualitative assessment is a viable alternative to the KITC and random assessment methods for valuing correlation. Data could be collected for such projects and used to evaluate the performance of qualitative assessments versus the correlation modelling methods presented in this thesis.
- Additional model scenarios: Limits had to be established within the Monte Carlo simulation experiment to manage the scope for this thesis and thus several model parameters were fixed for all scenarios, e.g., the intervals for sampling the mean and standard deviation of each cost item. Additional scenarios could be developed to examine the impact of alternative parameters.
- Performance of correlation modelling methods with other distribution types: The Monte Carlo simulation experiment used a multivariate normal distribution for generating correlated cost data and separate normal distributions for independent cost samples, however many other distribution types are used in cost uncertainty analysis. The experiment could be expanded for different types of distributions and for mixed distribution types.

Bibliography

AACE International. (2008). *Recommended Practice No. 41R-08, Risk Analysis and Contingency Determination Using Range Estimating*.

AACE International. (2012). *Recommended Practice No. 62R-11, Risk Assessment: Identification and Qualitative Analysis*.

AACE International. (2019). *Recommended Practice No. 57R-09, Integrated Cost and Schedule Risk Analysis using Risk Drivers and Monte Carlo Simulation of a CPM*.

AACE International. (2020). *Recommended Practice No. 17R-97, Cost Estimate Classification System*.

AbouRizk, H. (2013). "Understanding and Improving Input for Quantitative Risk Analysis in the Construction Industry." Master's Thesis, University of Alberta.

AbouRizk, S. M. (1990). "Input modeling for construction simulation." Doctoral Thesis, Purdue University.

AbouRizk, S. M., Halpin, D. W., and Wilson, J. R. (1991). "Visual Interactive Fitting of Beta Distributions." *Journal of Construction Engineering and Management*, 117(4), 589–605.

Ahuja, H. N., Dozzi, S. P., and AbouRizk, S. M. (1994). *Project Management: Techniques in Planning and Controlling Construction Projects*. Wiley, New York, New York.

Alhakami, A. S., and Slovic, P. (1994). "A Psychological Study of the Inverse Relationship Between Perceived Risk and Perceived Benefit." *Risk Analysis*, 14(6), 1085–1096.

Ayyub, B. M. (2003). *Risk Analysis in Engineering and Economics*. Chapman and Hall/CRC, Boca Raton, Florida.

Blanton, H., Axsom, D., McClive, K. P., and Price, S. (2001). "Pessimistic Bias in Comparative Evaluations: A Case of Perceived Vulnerability to the Effects of Negative Life Events." *Personality and Social Psychology Bulletin*, SAGE Publications Inc, 27(12), 1627–1636.

Byrne, M. D. (2013). "How Many Times Should a Stochastic Model Be Run? An Approach Based on Confidence Intervals." *Proceedings of the 12th International Conference on Cognitive Modeling*, Carleton University, Ottawa, Canada, 445–450.

Chapin, J. (2001). "Self-protective Pessimism: Optimistic Bias in Reverse." *North American Journal of Psychology*, North American Journal of Psychology, Winter Garden, 3(2), 253–261.

City of Edmonton. (2020). "Summary of Key Fall 2020 Forecast Indicators." City of Edmonton.

Clemen, R. T. (2008). “Comment on Cooke’s classical method.” *Expert Judgement*, 93(5), 760–765.

Clemen, R. T., and Winkler, R. L. (1999). “Combining Probability Distributions From Experts in Risk Analysis.” *Risk Analysis*, Kluwer Academic Publishers-Plenum Publishers, New York, 19(2), 187.

Cooke, R. (1991). *Experts in Uncertainty : Opinion and Subjective Probability in Science*. Environmental Ethics and Science Policy, Oxford University Press.

Cooke, R., and Goossens, L. H. J. (2000). “Procedures Guide for Structural Expert Judgement in Accident Consequence Modelling.” *Radiation Protection Dosimetry*, 90, 303–309.

Cooke, R. M., and Goossens, L. L. H. J. (2008). “TU Delft expert judgment data base.” *Expert Judgement*, 93(5), 657–674.

Curran, M. W. (1976). “Range Estimating: Coping with Uncertainty.” *AACE International Transactions*, I–6.

Curran, M. W. (1988). “Range Estimating: Reasoning with Risk.” *AACE International Transactions*, N.3.1-N.3.9.

Curran, M. W. (1989). “Range Estimating: Contingencies with Confidence.” *AACE International Transactions*, B.7.1-B.7.4.

Curran, M. W. (1990). “Range Estimating Reduces Iatrogenic Risk.” *AACE International Transactions*, K.3.1-K.3.3.

Dalkey, N. (1969). “An experimental study of group opinion: The Delphi method.” *Futures*, 1(5), 408–426.

Delbecq, A. L., Van de Ven, A. H., and Gustafson, D. H. (1975). *Group techniques for program planning: A guide to nominal group and Delphi processes.* Management applications series, Scott, Foresman.

Dias, L. C., Morton, A., and Quigley, J. (2018). “Elicitation: State of the Art and Science.” *Elicitation: The Science and Art of Structuring Judgement*, L. C. Dias, A. Morton, and J. Quigley, eds., Springer International Publishing, Cham, Switzerland.

Diekmann, J. E. (1983). “Probabilistic Estimating: Mathematics and Applications.” *Journal of Construction Engineering and Management*, 109(3), 297–308.

Dienemann, P. F. (1966). *Estimating Cost Uncertainty Using Monte Carlo Techniques*. RAND Corporation, Santa Monica, CA.

Dillon, R. L., John, R., and von Winterfeldt, D. (2002). "Assessment of Cost Uncertainties for Large Technology Projects: A Methodology and an Application." *INFORMS Journal on Applied Analytics*, 32(4), 52–66.

Dolinski, D., Gromski, W., and Zawisza, E. (1987). "Unrealistic pessimism." *The Journal of social psychology*, Heldref, Washington, DC, 127(5), 511–516.

European Food Safety Authority. (2014). "Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment." *EFSA Journal*, John Wiley & Sons, Ltd, 12(6), 3734.

Finucane, M. L., Slovic, P., Johnson, S. M., and Alhakami, A. (2000). "The Affect Heuristic in Judgments of Risks and Benefits." *Journal of Behavioral Decision Making*, John Wiley and Sons Ltd, 13(1), 1–17.

Fischhoff, B. (1982). "Debiasing." *Judgment under Uncertainty: Heuristics and Biases*, D. Kahneman, P. Slovic, and A. Tversky, eds., Cambridge University Press, 422–444.

Flyvbjerg, B., Holm, M. S., and Buhl, S. (2002). "Cost Underestimation in Public Works Projects: Error or Lie?" *Journal of the American Planning Association*, 68(3).

Furnham, A., and Boo, H. C. (2011). "A literature review of the anchoring effect." *The Journal of Socio-Economics*, 40(1), 35–42.

Garvey, P. R. (2008). "A Scenario-Based Method for Cost Risk Analysis." *Journal of Cost Analysis and Parametrics*, Taylor & Francis, 1(1), 65–76.

Garvey, P. R., Book, S. A., and Covert, R. P. (2016). *Probability Methods for Cost Uncertainty Analysis: A Systems Engineering Perspective*. Chapman and Hall/CRC, Boca Raton, Florida.

Garvey, P. R., Flynn, B., Braxton, P., and Lee, R. (2012). "Enhanced Scenario-Based Method for Cost Risk Analysis: Theory, Application, and Implementation." *Journal of Cost Analysis and Parametrics*, Taylor & Francis, 5(2), 98–142.

Gosling, J. P. (2018). "SHELF: The Sheffield Elicitation Framework." *Elicitation: The Science and Art of Structuring Judgement*, L. C. Dias, A. Morton, and J. Quigley, eds., Springer International Publishing, Cham, Switzerland.

Hanea, A. M., Burgman, M., and Hemming, V. (2018a). "IDEA for Uncertainty Quantification." *Elicitation: The Science and Art of Structuring Judgement*, L. C. Dias, A. Morton, and J. Quigley, eds., Springer International Publishing, Cham, Switzerland.

Hanea, A. M., McBride, M. F., Burgman, M. A., and Wintle, B. C. (2018b). "Classical meets modern in the IDEA protocol for structured expert judgement." *Journal of Risk Research*, Routledge, 21(4), 417–433.

Hartley, D., and French, S. (2018). "Elicitation and Calibration: A Bayesian Perspective." *Elicitation: The Science and Art of Structuring Judgement*, L. C. Dias, A. Morton, and J. Quigley, eds., Springer International Publishing, Cham, Switzerland.

Hatamleh, M. T., Hiyassat, M., Sweis, G. J., and Sweis, R. J. (2018). "Factors affecting the accuracy of cost estimate: case of Jordan." *Engineering, Construction and Architectural Management*, 25(1), 113–131.

Hollmann, J. K. (2012). "Estimate accuracy: Dealing with reality." *AACE International Transactions*, 3, 1753–1771.

Iman, R. L., and Conover, W. J. (1982). "A distribution-free approach to inducing rank correlation among input variables." *Communications in Statistics - Simulation and Computation*, Taylor & Francis, 11(3), 311–334.

Isidore, L. J., and Back, W. E. (2002). "Multiple Simulation Analysis for Probabilistic Cost and Schedule Integration." *Journal of Construction Engineering and Management*, 128(3), 211–219.

Janis, I. L. (1972). *Victims of Groupthink; A Psychological Study of Foreign Policy Decisions and Fiascoes*. Houghton, Mifflin.

Joyner, C. A. (1992). "Preventing overconfidence in decision-making: Debiasing with devil's advocacy." *ProQuest Dissertations and Theses*, Ph.D., Ohio University, Ann Arbor.

Kerr, N. L., MacCoun, R. J., and Kramer, G. P. (1996). "Bias in Judgment: Comparing Individuals and Groups." *Psychological Review*, 103(4), 687–719.

Kerr, N. L., and Tindale, R. S. (2004). "Group performance and decision making." *Annual review of psychology*, Annual Reviews, Palo Alto, CA, 55, 623–655.

Kerr, N. L., and Tindale, R. S. (2011). "Group-Based Forecasting?: A Social Psychological Analysis." *International Journal of Forecasting*, 27(1), 14–40.

Koriat, A., Lichtenstein, S., and Fischhoff, B. (1980). "Reasons For Confidence." *Journal of Experimental Psychology: Human Learning and Memory*, American Psychological Association, US, 6(2), 107–118.

Law, A. M. (2015). “Statistical Analysis of Simulation Output Data: The Practical State of the Art.” *Proceedings of the 2015 Winter Simulation Conference*.

Lichtenstein, S., Fischhoff, B., and Phillips, L. D. (1982). “Calibration of probabilities: the state of the art to 1980.” *Judgment Under Uncertainty: Heuristics and Biases*, D. Kahneman, P. Slovic, and A. Tversky, eds., Cambridge University Press, 306–334.

Linder, S. (2005). *Fifty years of Research on Accuracy of Capital Expenditure Project Estimates: A Review of the Findings and their Validity*. Finance, University Library of Munich, Germany.

Liu, C. W. (2020). “SimJoint: Simulate Joint Distribution.”

Lurie, P., and Goldberg, M. (1998). “An Approximate Method for Sampling Correlated Random Variables from Partially-Specified Distributions.” *Management Science*, 44, 203–218.

Mahamid, I. (2015). “Factors affecting cost estimate accuracy: Evidence from Palestinian construction projects.” *International Journal of Management Science and Engineering Management*.

Makalic, E., and Schmidt, D. F. (2018). “An efficient algorithm for sampling from sin^k(x) for generating random correlation matrices.”

Makalic, E., and Schmidt, D. F. (2020). “An efficient algorithm for sampling from sink (x) for generating random correlation matrices.” *Communications in Statistics - Simulation and Computation*, Taylor & Francis, 1–5.

Montibeller, G., and von Winterfeldt, D. (2015). “Cognitive and Motivational Biases in Decision and Risk Analysis.” *Risk Analysis: An International Journal*, 35(7), 1230–1251.

Montibeller, G., and von Winterfeldt, D. (2018). “Individual and Group Biases in Value and Uncertainty Judgments.” *Elicitation: The Science and Art of Structuring Judgement*, L. C. Dias, A. Morton, and J. Quigley, eds., Springer International Publishing, Cham, Switzerland.

Morgan, M. G. (2014). “Use (and abuse) of expert elicitation in support of decision making for public policy.” *Proceedings of the National Academy of Sciences*, National Academy of Sciences, 111(20), 7176–7184.

Morgan, M. G., and Henrion, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press, New York, New York.

Morris, P. A. (1974). “Decision Analysis Expert Use.” *Management Science*, Institute of Management Sciences, 20(9), 1233–1241.

Morris, P. A. (1977). “Combining Expert Judgments: A Bayesian Approach.” *Management Science*, Institute of Management Sciences, 23(7), 679–693.

Moussa, M. A. E.-F. M. (2013). “Unified Simulation Methodology and Project Risk Assessment Framework.” PhD Thesis, University of Calgary.

Murphy, A. H., and Winkler, R. L. (1974). “Credible Interval Temperature Forecasting: Some Experimental Results.” *Monthly Weather Review*, American Meteorological Society, Boston MA, USA, 102(11), 784–794.

Mussweiler, T., Strack, F., and Pfeiffer, T. (2000). “Overcoming the Inevitable Anchoring Effect: Considering the Opposite Compensates for Selective Accessibility.” *Personality and Social Psychology Bulletin*, 26(9), 1142–1150.

Myers, D. G., and Lamm, H. (1976). “The Group Polarization Phenomenon.” *Psychological Bulletin*, 83(4), 602–627.

National Aeronautics and Space Administration. (2008). *2008 NASA Cost Estimating Handbook*. National Aeronautics and Space Administration, Washington, DC.

National Research Council. (1990). *Improving the Accuracy of Early Cost Estimates for Federal Construction Projects*. The National Academies Press, Washington, DC.

Nickerson, R. S. (1998). “Confirmation Bias: A Ubiquitous Phenomenon in Many Guises.” *Review of General Psychology*, 2(2), 175–220.

Numpacharoen, K., and Atsawarungruangkit, A. (2012). “Generating Correlation Matrices Based on the Boundaries of Their Coefficients.” *PLOS ONE*, Public Library of Science, 7(11), e48902.

Oakley, J. (2021). *SHELF: Tools to Support the Sheffield Elicitation Framework*. R package version 1.8.0. Accessed November 8, 2021. https://CRAN.R-project.org/package=SHELF.

Oakley, J. E., and O’Hagan, A. (2019). “SHELF: the Sheffield Elicitation Framework (version 4).” *School of Mathematics and Statistics, University of Sheffield, UK*, <http://tonyohagan.co.uk/shelf/> (Oct. 20, 2021).

O’Hagan, A. (2019). “Expert Knowledge Elicitation: Subjective but Scientific.” *The American Statistician*, Taylor & Francis, 73(sup1), 69–81.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. John Wiley & Sons, Ltd, West Sussex, England.

O'Hagan, T. (2005). "Elicitation." *Significance*, 2(2), 84.

Phillips, L. D. (1999). "Group elicitation of probability distributions: Are many heads better than one?" *Decision Science and Technology: Reflections on the Contributions of Ward Edwards*, J. Shanteau, B. A. Mellers, and D. A. Schum, eds., Springer Science+Business Media, New York, New York, 313–330.

Plous, S. (1995). "A Comparison of Strategies for Reducing Interval Overconfidence in Group Judgments." *Journal of applied psychology*, American Psychological Association, Washington, DC, 80(4), 443–454.

Pourahmadi, M., and Wang, X. (2015). "Distribution of random correlation matrices: Hyperspherical parameterization of the Cholesky factor." *Statistics & Probability Letters*, 106, 5–12.

Project Management Institute. (2017). *A guide to the project management body of knowledge (PMBOK guide)*. Project Management Institute, Newtown Square, Pennsylvania.

Quigley, J., Colson, A., Aspinall, W., and Cooke, R. (2018). "Elicitation in the Classical Model." *Elicitation: The Science and Art of Structuring Judgement*, L. C. Dias, A. Morton, and J. Quigley, eds., Springer International Publishing, Cham, Switzerland.

Ramsey, F. P. (1926). "Truth and Probability." *Studies in Subjective Probability*, H. E. Kyburg and H. E. Smokler, eds., Wiley, New York, New York.

Sackman, H. (1974). *Delphi Assessment: Expert Opinion, Forecasting, and Group Process*. RAND Corporation, Santa Monica, CA.

Sadeghi, N. (2009). "Combined Fuzzy and Probabilistic Simulation for Construction Management." Master's Thesis, University of Alberta.

Sargent, R. G. (2013). "Verification and validation of simulation models." *Journal of Simulation*, Taylor & Francis, 7(1), 12–24.

Seaver, D. A., Winterfeldt, D. von, and Edwards, W. (1978). "Eliciting subjective probability distributions on continuous variables." *Organizational Behavior and Human Performance*, 21(3), 379–391.

Shaheen, A. A., Fayek, A. R., and AbouRizk, S. M. (2007). "Fuzzy Numbers in Cost Range Estimating." *Journal of Construction Engineering and Management*, 133(4), 325–334.

Shi-Woei Lin and Chih-Hsing Cheng. (2009). "The reliability of aggregated probability judgments obtained through Cooke's classical model." *Journal of Modelling in Management*, Emerald Group Publishing Limited, Bradford, 4(2), 149–161.

Skagerlund, K., Forsblad, M., Västfjäll, D., and Slovic, P. (2020). "The Affect Heuristic and Risk Perception – Stability Across Elicitation Methods and Individual Cognitive Abilities." *Frontiers in Psychology*, Frontiers Media S.A., 11.

Slovic, P. (1972). "From Shakespeare to Simon: speculations–and some evidence– about man's ability to process information." *Oregon Research Institute*, 12(2).

Slovic, P., Finucane, M. L., Peters, E., and MacGregor, D. G. (2007). "The affect heuristic." *European Journal of Operational Research*, Elsevier, Amsterdam, 177(3), 1333–1352.

Stoner, J. A. F. (1961). "A Comparison of Individual and Group Decisions Involving Risk." Master's Thesis, Massachusetts Institute of Technology.

Touran, A. (1993). "Probabilistic Cost Estimating with Subjective Correlations." *Journal of Construction Engineering and Management*, 119(1), 58–71.

Touran, A., and Wiser, E. P. (1992). "Monte Carlo Technique with Correlated Random Variables." *Journal of Construction Engineering and Management*, 118(2), 258–272.

Tversky, A., and Kahneman, D. (1974). "Judgment under Uncertainty: Heuristics and Biases." *Science*, 185(4157), 1124–1131.

U. S. Air Force. (2007). *Cost Risk and Uncertainty Analysis Handbook*. U.S. Air Force, Hanscom Air Force Base, Massachusetts.

U. S. Government Accountability Office. (2020). *Cost Estimating and Assessment Guide: Best Practices for Developing and Managing Program Costs*. U.S. Government Accountability Office, Washington, DC.

Ungvarsky, J. (2020). "Delphi method." *Salem Press Encyclopedia*, Salem Press.

Weinstein, N. D. (1980). "Unrealistic Optimism About Future Life Events." *Journal of Personality & Social Psychology*, 39(5), 806–820.

Welsh, M. B., and Begg, S. H. (2018). "More-or-less elicitation (MOLE): reducing bias in range estimation and forecasting." *EURO Journal on Decision Processes*, Springer Berlin Heidelberg, Berlin/Heidelberg, 6(1–2), 171.

Welsh, M. B., Begg, S. H., Bratvold, R. B., and Lee, M. D. (2004). “Problems with the Elicitation of Uncertainty.”

Winkler, R. L. (1968). “The Consensus of Subjective Probability Distributions.” *Management Science*, Institute of Management Sciences, 15(2), B61–B75.

von Winterfeldt, D., and Edwards, W. (1986). *Decision Analysis and Behavioral Research*. Cambridge University Press.

Woudenberg, F. (1991). “An evaluation of Delphi.” *Technological Forecasting and Social Change*, 40(2), 131–150.

Yaniv, I., and Foster, D. P. (1995). “Graininess of judgment under uncertainty: An accuracy-informativeness trade-off.” *Journal of Experimental Psychology: General*, American Psychological Association, US, 124(4), 424–432.

Zadeh, L. A. (1965). “Fuzzy sets.” *Information and Control*, 8(3), 338–353.

Appendix A

Sample Materials from Illustrative Example using SHELF

Evidence Dossier[10] – Cost Variability on Project X

1. Context

Project X is a light rail transit project located in Edmonton, Alberta with a budget of $100 million. The project recently completed the preliminary design stage (i.e., 30% design documentation) and has had an estimate developed based on the current design with the understanding that the project will be delivered using the design-build delivery method. Before proceeding to the next phase, Project X must seek approval from decision makers. Decision makers require an assessment of the project's cost uncertainty and confirmation that the requested budget provides adequate contingency. In particular, decision makers are interested in the following question: "What is the level of confidence (i.e., percentile) associated with the requested budget for Project X?" This elicitation exercise is being conducted to obtain probability distributions for the various cost items (i.e., work packages) comprising Project X based on project documents as well as data from past light rail transit projects in Alberta; these probability distributions will form the inputs of a model that simulates the range of project costs and provides the level of confidence associated with budget values within the range of costs.

2. Quantities of Interest

There are eight major cost items comprising Project X, as shown in Table A-1 below, along with their estimated values and expected start and end dates for construction.

Table A-1: Cost items for Project X and their estimated costs and construction dates.

QOI	Base Estimate (Thousands of Dollars)	Start Date	End Date
Work Package A*	1000	April 1, 2022	October 31, 2023
Work Package B	100	June 1, 2022	September 30, 2022
Work Package C*	500	October 1, 2022	March 31, 2023
Work Package D	300	April 1, 2023	August 31, 2023
Work Package E*	900	April 1, 2023	October 31, 2024
Work Package F	400	September 1, 2023	May 31, 2024
Work Package G*	700	June 1, 2024	September 30, 2024

[10] This evidence dossier is adapted from Oakley and O'Hagan (2019) as an illustrative example. References are included in the Bibliography rather than shown at the end of the dossier.

Work Package H*	650	November 1, 2024	April 2025

There are five quantities of interest (QOIs) corresponding to the most important cost items listed in Table A-1 denoted with an asterisk and defined as follows:

"The cost (expressed as a percentage relative to the base estimate) for Work Packages A, C, E, G, and H on Project X given construction is anticipated according to the timelines shown in Table 1. Assume market escalation is accounted for separately. Refer to the 30% design drawing, specifications, and cost estimate attached to this evidence dossier for a description of the work package scopes."

3. Scientific Opinion

Difficulties arise in representing the cost uncertainty of cost items because:

1. numerous sources contribute to cost uncertainty including risk, cost estimate accuracy, and bias;
2. cost estimate accuracy in turn is driven by design definition as well as a number of secondary factors including project complexity and estimating techniques;
3. the delineation of scope between cost items can be complex and the contents of a cost item type may not be consistent between projects;
4. actual cost data from other projects is typically not readily available in part due to the sensitive nature of such information; and,
5. where actual cost data is available, it may not be relevant to other projects as each project is unique and the circumstances surrounding a particular cost outcome may not be representative of other projects.

In addition to the difficulties above, cost estimates have historically been underestimated by very large margins (Hollmann 2012). The causes of cost underestimation are debated; a number of explanations are presented as follows.

Strategic Misrepresentation and Willful Biases

The authors of the first statistically significant study of cost underestimation in transportation infrastructure provide the following opinion on cost underestimation (Flyvbjerg et al. 2002):

> "Based on a sample of 258 transportation infrastructure projects worth US$90 billion and representing different project types, geographical regions, and historical periods, it is found with overwhelming statistical significance that the cost estimates used to decide whether such projects should be built are highly and systematically misleading. Underestimation cannot be explained by error and is best explained by strategic misrepresentation, that is, lying. The policy implications are clear: legislators, administrators, investors, media representatives, and members of the public who value honest numbers should not trust cost estimates and cost-benefit analyses produced by project promoters and their analysts."

Cognitive and Motivational Biases

Hollmann (2012) attributes cost underestimation to willful biases as well as flawed analysis:

> "Tragically, many cost engineers are facilitating management's collective and sometimes willful biases regarding accuracy by using flawed, unreliable risk analysis methods…"

Garvey et al. (2016) offer a similar, albeit less cynical, explanation:

> "While optimistic assumptions continue to be a factor in cost growth above initial estimates, a key methodology deficiency was inadequate attention to correlations between a program's work breakdown structure (WBS) element costs."

Dillion et al. (2012) suggest that underestimation is caused by biases in human judgment and as well as a number of other reasons:

> "(1) Cost estimates are often made in an environment of advocacy with pressures to produce low, can-do estimates for projects.
>
> (2) Cost estimates typically reflect what reasonable costs should be, not what they could be considering all types of uncertainties.

> (3) Uncertainties are accounted for by adding contingencies (typically 20 to 50 percent of the capital costs). However, these contingencies rarely cover extraordinary events, such as major technical or regulatory problems.
>
> (4) Attempts to incorporate uncertainties in cost estimates using probability distributions are often influenced by biases in human judgment (Tversky and Kahneman 1974)."

Dillion et al. (2012) also reference optimism in cost estimates which they observed in the application of their method for estimating cost uncertainties:

> "…base-case cost estimates are usually at the very low end of the probability distributions, indicating that these estimates are typically much too optimistic…"

Conversely, Flyvbjerg et al. (2002) reject the notion that genuine optimism is a primary cause of cost estimation stating:

> "As observed elsewhere, the incentive to publish and justify optimistic estimates is very strong, and the penalties for having been overoptimistic are generally insignificant (Davidson & Huot, 1989, p. 137; Flyvbjerg et al., in press). This is a better explanation of the pervasive existence of optimistic estimates than an inherent bias for optimism in the psyche of promoters and forecasters. And 'optimism' calculated on the basis of incentives is not optimism, of course; it is deliberate deception."

4. Data

Evidence provided in this section is divided into three categories based on its quality and relationship to the QOIs: project cost variances, project cost estimate accuracy, and work package cost variances.

Project Cost Variance from Comparative Study

Evidence in this category is considered high-quality as it was produced from a statistically significant and widely-cited study (Flyvbjerg et al. 2002), however it is also considered to be indirect evidence since it concerns cost variance at the *project level*. Flyvbjerg et al. describe cost variance observed in a sample of 258 transportation projects using a cost escalation metric explained as follows:

> "We follow international convention and measure the inaccuracy of cost estimates as so-called "cost escalation" (often also called "cost overrun"; i.e., actual costs minus estimated costs in percent of estimated costs). Actual costs are defined as real, accounted construction costs determined at the time of project completion. Estimated costs are defined as budgeted, or forecasted, construction costs at the time of decision to build."

Key observations from Flyvbjerg et al. (2002) are repeated in Table A-2 for convenience:

Table A-2: Inaccuracy of transportation project cost estimates by type of project (fixed prices). Source: Flyvbjerg et al. 2012.

Project Type	**Number of cases (N)**	**Average cost escalation (%)**	**Standard deviation**	**Level of significance (p)**
Rail	58	44.7	338.4	<0.001
Fixed-link	33	33.8	62.4	<0.004
Road	167	20.4	29.9	<0.001
All projects	258	27.6	38.7	<0.001

Participants are also encouraged to familiarize themselves with Figure 1 from Flyvbjerg et al. (2002) which shows cost escalation ranging from -80% to 300% at the project level.

Project Cost Estimate Accuracy from Cost Estimate Classification

Evidence in this category is considered high-quality as it is an amalgamation of the most common cost estimate classification systems (all of which have been made by reputable organizations), however it is considered to be indirect evidence since it concerns *cost estimate accuracy* (which is only one component of cost variability) at the *project level*. Nonetheless, it is useful to understand how cost estimate accuracy varies as a function of design definition. In addition to illustrating this, Figure A-1 compares cost estimate classification systems developed by the Association for the Advancement of Cost Engineering (AACE), the Department of National Defence (DND), the Royal Architectural Institute of Canada (RAIC), and the Government of Canada (GOC).

COST ESTIMATE CLASSIFICATION SYSTEMS					
AACE	**Class 4**	**Class 3**		**Class 1**	**Class 1**
DND		Indicative		Substantive	
RAIC	Sketch Design	Design Development		Construction Documents	Tender Documents
GOC	D	C	B		A
Design Documentation % Complete	**12.5%**	**25.0%**		**95.0%**	**100.0%**
Cost Estimate Accuracy (+/- %)	**30.0%**	**25.0%**		**15.0%**	**10.0%**

Figure A-1: Cost Estimate Classification Chart adapted from the Joint Federal Government / Industry Cost Predictability Taskforce (Guide to Cost Predictability in Construction: An Analysis of Issues Affecting the Accuracy of Construction Cost Estimates, unpublished report, 2012).

Work Package Cost Variance from Similar Projects

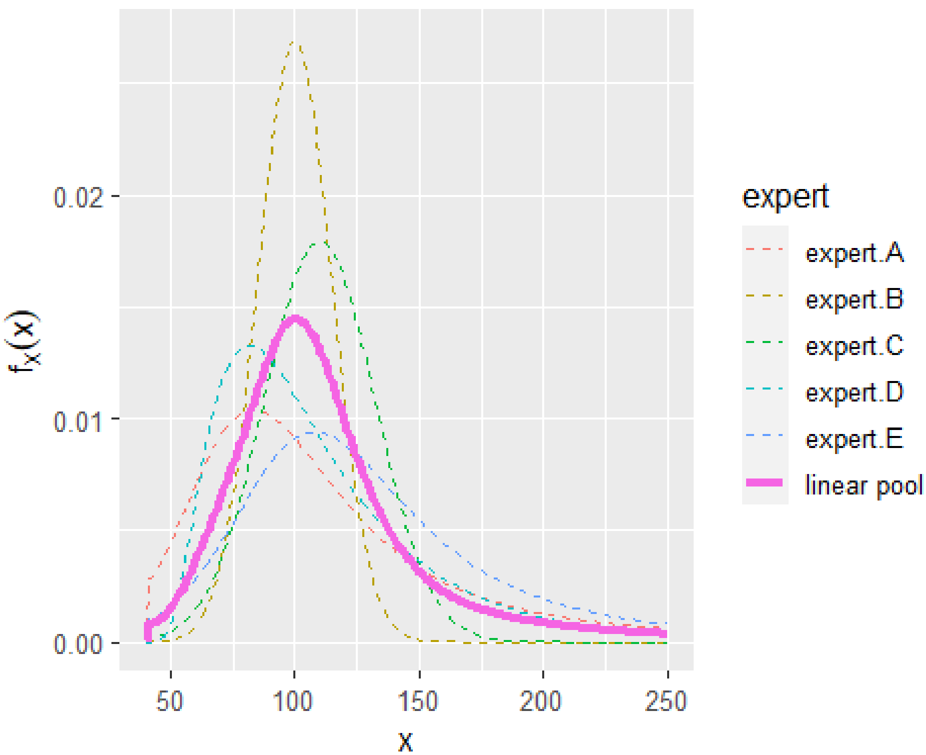
Evidence in this category is considered lower quality as the sample size is limited, however it is considered to be the most direct evidence currently available since it concerns the cost variance of *work packages* (rather than projects) completed in the same region. Unfortunately, this evidence is still somewhat indirect as several project characteristics vary (e.g., different timelines and sizes). Table A-3 shows the cost escalation, calculated in the same manner as Flyvbjerg et al. (2002), encountered for each work package on similar projects to Project X.

Table A-3: Cost escalation for various work packages on projects similar to Project X.

	Cost escalation (%)					
Work Package	**Project L**	**Project M**	**Project N**	**Project O**	**Project P**	**Project Q**
Work Package A	107.5	37.3	-48.6	57.1	16.6	-31.1
Work Package C	-30.3	-42.5	119.0	52.2	-2.0	59.7
Work Package E	-39.2	10.2	11.0	24.8	-44.4	17.7
Work Package G	48.0	-61.9	27.1	21.5	-44.6	35.6
Work Package H	20.3	20.2	19.1	-4.2	9.1	-18.5

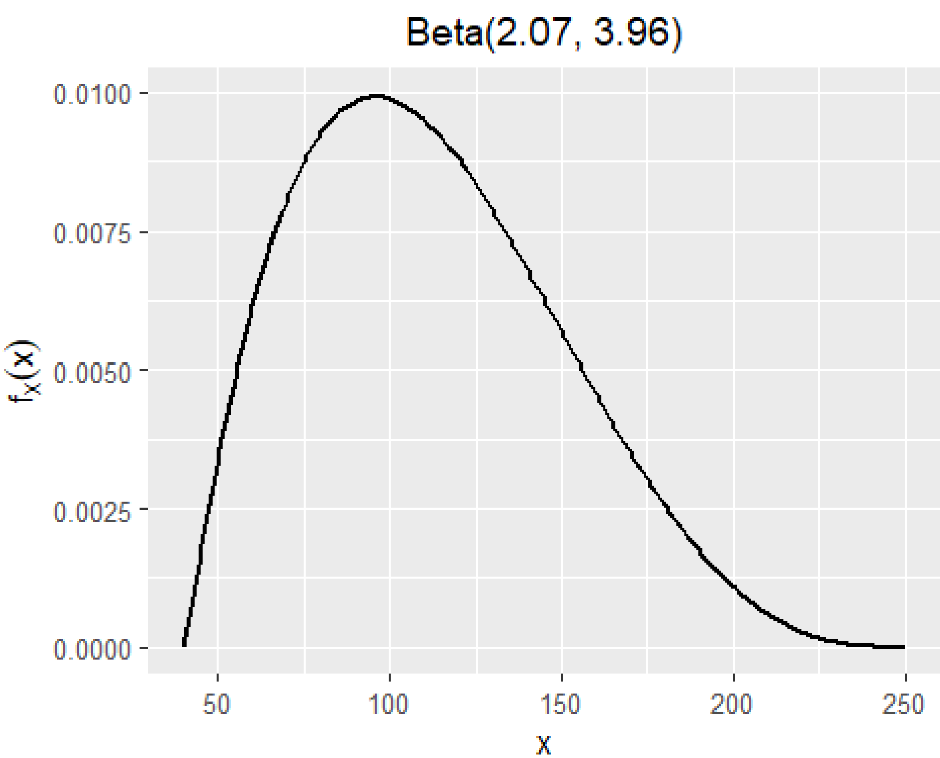
Script for Distribution Fitting of Work Package A using R Package "SHELF: Tools to Support the Sheffield Elicitation Framework"

```r
library("SHELF")

#clears all variables
rm(list=ls())

#captures quantiles for each expert into a matrix
v <- matrix(c(80,100,150,90,100,110,95,110,125,80,100,130,100,120,170),nrow = 3,ncol = 5)

#the probabilities are stored in a vector
p <- c(0.25, 0.5, 0.75)

#distributions are fitted for each expert according to their quantiles as well as the plausible limits
myfit <- fitdist(vals = v, probs = p, lower = 40, upper = 250)

#the best-fitting distributions for each expert are plotted along with a linear opinion pool
plotfit(myfit, lp = TRUE)
```




```
#the quartiles of the linear pool distribution are calculated
qlinearpool(myfit, q=c(0.25,0.5,0.75))

## [1]  87.86646 105.88243 129.23834

#a consensus distribution is fitted according to group judgments
consensus <- fitdist(vals = c(85,105,140), probs = p, lower = 40, upper = 250)

#the sum of squared differences between elicited and fitted probabilities are shown
consensus$ssq

##        normal           t        gamma    lognormal         logt        beta
## 1 0.002956741 0.002876766 0.0008269468 0.0002521479 0.0002438239 0.001700706
```

```
##   mirrorgamma mirrorlognormal  mirrorlogt
## 1 0.004374776     0.005152279 0.005035667

#a fitted beta distribution is plotted
plotfit(consensus, d = "beta")
```




```
#fitted percentiles and probabilities for various fitted distributions are shown
feedback(consensus, quantiles = c(0.01,0.05,0.95,0.99))

## $fitted.quantiles
##      normal     t gamma lognormal  logt  beta  hist mirrorgamma mirrorlognormal
## 0.01   12.2 -57.8  51.2      56.6  46.1  47.6  41.8       -14.9           -38.1
## 0.05   40.6  22.5  60.9      64.9  59.3  57.3  49.0        27.2            17.3
```

```
## 0.95  178.0 195.0 197.0     215.0 266.0 180.0 228.0       170.0           167.0
## 0.99  206.0 275.0 249.0     303.0 749.0 205.0 246.0       188.0           183.0
##      mirrorlogt
## 0.01     -237.0
## 0.05      -16.6
## 0.95      177.0
## 0.99      210.0
##
## $fitted.probabilities
##     elicited normal     t gamma lognormal  logt  beta hist mirrorgamma
## 85      0.25  0.280 0.281 0.265     0.258 0.258 0.272 0.25       0.288
## 105     0.50  0.459 0.462 0.478     0.488 0.489 0.469 0.50       0.451
## 140     0.75  0.769 0.771 0.762     0.757 0.757 0.766 0.75       0.772
##     mirrorlognormal mirrorlogt
## 85            0.292      0.293
## 105           0.446      0.449
## 140           0.772      0.775
```

Appendix B

Comparison of Monte Carlo Simulation Experiment Results to Method of Moments

As part of the simulation model verification process, outputs from the model for the mean and standard deviation of project costs are compared against results produced from classical statistics, i.e., the method of moments. The base model scenario is used for the comparison; therefore there are 10 cost items, 50% of cost items are correlated, and the correlation strength between correlated cost items is 0.5. The mean, standard deviation, and variance of costs for each cost item are summarized in Table B-1 and the covariance matrix is shown in Table B-2; these are the inputs provided to the *rmvnorm* function to generate the population of costs for each cost item which are subsequently combined to generate the population of project costs.

Table B-1: Mean, standard deviation, and cost variance of costs for cost items in the base model scenario.

Cost Item	**Cost Mean**	**Cost Standard Deviation**	**Cost Variance**
Work Package 1	308,117	22,710	515,755,269
Work Package 2	697,931	47,772	2,282,202,213
Work Package 3	670,167	24,912	620,601,589
Work Package 4	372,510	45,922	2,108,873,186
Work Package 5	280,154	41,179	1,695,683,554
Work Package 6	822,434	115,636	13,371,788,384
Work Package 7	619,276	59,688	3,562,714,185
Work Package 8	669,998	70,206	4,928,837,022
Work Package 9	111,863	4,235	17,931,014
Work Package 10	799,120	73,370	5,383,135,429

Table B-2: Covariance matrix for cost items in the base model scenario. Values are shown in millions.

	wp1	wp2	wp3	wp4	wp5	wp6	wp7	wp8	wp9	wp10
wp1	516	0	0	0	0	0	0	0	0	0
wp2	0	2282	0	0	0	0	0	0	0	0
wp3	0	0	621	0	513	1440	743	0	53	0
wp4	0	0	0	2109	0	0	0	0	0	0
wp5	0	0	513	0	1696	2381	1229	0	87	0
wp6	0	0	1440	0	2381	13372	3451	0	245	0
wp7	0	0	743	0	1229	3451	3563	0	126	0
wp8	0	0	0	0	0	0	0	4929	0	0
wp9	0	0	53	0	87	245	126	0	18	0
wp10	0	0	0	0	0	0	0	0	0	5383

According to the method of moments, the mean of project costs is the sum of the mean of the cost items within the project as shown in Equation B-1. Applying Equation B-1 to the cost item means in Table B-1, the mean project cost is as shown in Equation B-2.

Equation B-1:

$$Mean\left(cost_{proj}\right) = Mean(X_1) + Mean(X_2) + Mean(X_3) + \cdots + Mean(X_n)$$

Equation B-2:

$$\begin{aligned} &Mean\left(cost_{proj}\right) \\ &\qquad = 308{,}117 + 697{,}931 + 670{,}167 + 372{,}510 + 280{,}154 + 822{,}434 + 619{,}276 \\ &\qquad + 669{,}998 + 111{,}863 + 799{,}120 = 5{,}351{,}570 \end{aligned}$$

In the base case scenario, the project cost mean yielded from the method of moments is \$5,351,570 compared to a project cost mean in the population of \$5,349,702 yielded from the simulation. The percent difference between the two values is 0.03% which can be accounted for by the random nature of Monte Carlo simulation and finite number of random deviates generated.

According to the method of moments, the variance of project costs is the sum of the variance of the cost items plus the covariance between cost items, as shown in Equation B-3. Applying

Equation B-3 to the cost item variance and covariance in Table B-2, the variance in project cost is as shown in Equation B-4.

Equation B-3:

$$Var(cost_{project}) = Var(X_1) + Var(X_2) + Var(X_3) + \cdots + Var(X_n) + Covar(X_1, X_2, X_3, \ldots, X_n)$$

Equation B-4:

$$\begin{aligned} Var(cost_{project}) &= 515{,}744{,}100 + 2{,}282{,}163{,}984 + 620{,}607{,}744 + 2{,}108{,}830{,}084 \\ &+ 1{,}695{,}710{,}041 + 13{,}371{,}684{,}496 + 3{,}562{,}657{,}344 + 4{,}928{,}882{,}436 \\ &+ 17{,}935{,}225 + 5{,}383{,}156{,}900 \\ &+ 2 \times [513 + 1440 + 743 + 53 + 2381 + 1229 + 87 + 3451 + 245 + 126] \\ &= 55{,}025{,}036{,}179 \end{aligned}$$

In the base case scenario, the variance yielded from the method of moments is 55,025,036,179 compared to a project cost variance in the population of 53,246,485,504 yielded from the simulation. The percent difference between the two values is 3.2% which can be accounted for by the random nature of Monte Carlo simulation and finite number of random deviates generated. The comparison results in the base case scenario verify that the simulation model produces empirically derived statistics that are very close to those produced analytically.

As part of model validation process, an extreme condition test can be performed where the scenario is similar to the base case scenario except that the cost items are entirely independent (referred to as the independence scenario). The variance yielded from the methods of moments and independence method in the simulation model can be compared. First, for the method of moments, the covariance term becomes zero in Equation B-4 which produces Equation B-5.

Equation B-5:

$$\begin{aligned} Var(cost_{project}) &= 515{,}744{,}100 + 2{,}282{,}163{,}984 + 620{,}607{,}744 + 2{,}108{,}830{,}084 \\ &+ 1{,}695{,}710{,}041 + 13{,}371{,}684{,}496 + 3{,}562{,}657{,}344 + 4{,}928{,}882{,}436 \\ &+ 17{,}935{,}225 + 5{,}383{,}156{,}900 = 34{,}487{,}372{,}354 \end{aligned}$$

In the independence scenario, the variance yielded from the method of moments is 34,487,372,354 compared to a project cost variance in the population of 33,925,071,994 yielded from the simulation. The percent difference between the two values is 1.6% which can be accounted for by the random nature of Monte Carlo simulation and finite number of random deviates generated. The comparison result in the independence scenario validates that the model is plausible for a correlation value of zero (i.e., a boundary case).