**Aalto University
School of Business**

BAYESIAN-ADJUSTED ESTIMATES IN PROJECT SELECTION

A comparison study of Bayesian and non-Bayesian decision makers with empirical evidence from the pharmaceutical industry

Master's Thesis
Tri Tran
26 May 2017
Information and Service Economy

Approved in the Department of Information and Service Economy

__ / __ / 20__ and awarded the grade

_____

Author: Tri Tran	
Title of thesis: Bayesian-adjusted Estimates in Project Selection – A comparison study of Bayesian and non-Bayesian decision makers with empirical evidence from the pharmaceutical industry	
Degree: Master of Science in Economics and Business Administration	
Degree program: Information and Service Management	
Thesis advisors: Eeva Vilkkumaa and Juuso Liesiö	

Year of approval: 2017	**Number of pages:** 114	**Language:** English (US)

Abstract

Companies select projects to invest in based on uncertain estimates of their performance. Theories and empirical evidence suggest that if the uncertain estimates are taken at face value, the true performance of the selected projects tends to be lower than estimated, causing the decision makers (DMs) to experience post-decision disappointment. Taking prior information into account through Bayesian adjustment can result in more realistic estimates of the project performances and thus higher expected performance among the selected projects. However, Bayesian adjustment makes it less likely to predict extreme outcomes and, consequently, may lead to missing out on big wins.

This thesis studies the differences in the investment strategies of Bayesian and non-Bayesian DMs and outcomes of these strategies. This is done by employing a combined approach of both qualitative and qualitative research methodology. The quantitative approach of this thesis is in the form of a mathematical model that is used both to derive analytic results and for Monte Carlo simulation. The qualitative approach of this thesis is utilized to test theoretical findings empirically.

The key results reveal that when fewer than 50% of project proposals would truly perform well (e.g., have truly positive NPV), a Bayesian DM invests in too few and a non-Bayesian DM to too many projects. Moreover, the average *ex post* performance of the projects funded by a Bayesian DM is higher than that of a non-Bayesian DM. However, a non-Bayesian DM will have a higher proportion of funded projects that result in big wins, but also a higher proportion of projects that result in losses. If, on the other hand, more than 50% of project proposals would truly perform well, the roles of a Bayesian and non-Bayesian DM are reversed. The less accurate the performance estimates, the more pronounced the differences between the investment outcomes of a Bayesian and a non-Bayesian DM.

These analytic results are testified empirically in the R&D portfolio selection decisions in the pharmaceutical industry. Accordingly, the decision-making environment of the pharmaceutical industry displays characteristics of an environment with high estimate errors that amplify the differences in outcomes of a Bayesian DM's versus a non-Bayesian DM's investment decisions. As the DMs in this industry show quintessential characteristics of non-Bayesian DMs and the observed empirical outcomes perfectly coincide with the theoretical outcomes for non-Bayesian investment decisions, our theoretical findings are well-reflected empirically.

From a theoretical perspective, this thesis contributes novel analytic results on the differences between the investment strategies adopted by Bayesian and non-Bayesian DMs and validates these results with empirical evidence. From a practitioner's point of view, this thesis gives insights into how estimation uncertainties affect investment decisions and outcomes. Understanding such effects can help managers make better-informed decisions.

Keywords: Bayesian modeling of estimation uncertainties, portfolio selection, pharmaceutical industry	

Acknowledgements

Firstly, I would like to dedicate this thesis to Dr. Eeva Vilkkumaa. Words cannot explain how grateful I am to have her as my thesis instructor. She introduced me to the topic, patiently taught me Bayesian statistics from scratch, meticulously read every single line of my 114-page thesis, and gave me incredibly constructive, insightful, and detailed comments. Her knowledge and dedication inspire me to overcome the challenges and finish this thesis. Her high standards encourage me to push my limits and move beyond mediocrity. Thank you so much Eeva! I could not have wished for a better instructor.

Secondly, I would like to thank Dr. Juuso Liesiö for providing detailed feedback and for grading my thesis, and Dr. Merja Halme for organizing the thesis seminars and for providing constructive comments on my thesis presentations. I would also like to thank doctoral candidate Peng Xu for sharing with me his academic experience and for referring me to Dr. Eeva Vilkkumaa.

My thesis journey would have been more difficult without the great friends I fortunately have. My warmest thanks go to My Nguyen, Trinh Huynh, Que-Anh Nguyen, Saana Kallioinen, Ida Mauko, Noora Kumpulainen, Mariana Pacheco, Mikko Tokola, and many other wonderful friends.

Lastly, I would like to thank my family for the unconditional love and support. I am blessed to be born in such a loving family. I would like to thank my mom and dad, my brother and sister-in-law, and my baby niece for being amazingly supportive of every decision I have made, including my biggest (and best) decision to move to Finland; but my biggest thanks go to my Mom, for setting a great example of resilience, extraordinary work ethic, and discipline.

This thesis means a lot to me. It marks both the end of my journey as a student and the beginning of another challenging yet exciting journey.

Helsinki, 26 May 2017

Tri Tran

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Mathematical notations

Abbreviation/ Notation	Meaning
Var	Variance
Std	Standard deviation
Cov	Covariance
$Corr$	Correlation
V	True value
V^E	Value estimate (non-Bayesian)
V^B	Bayesian-adjusted estimate
$N\sim(\mu, \sigma^2)$	Normal distribution
	μ: mean
	σ: standard deviation
	σ^2: variance
$\mathbb{E}$	Expected value
$Pr(A|B)$	Conditional probability of A given B
$\mathbb{E}(A|B)$	Conditional expected value of A given B
(E)NPV	(Expected) Net Present Value
R&D	Research and Development
DM	Decision maker
pp.	Page

1 Introduction

"The only certainty is that nothing is certain"

- Pliny the Elder

Uncertainty, as a matter of fact, is the only certainty that exists. Uncertainty creates flaws and inaccuracies in forecasts and predictions. Especially in this VUCA[1] world, uncertainty becomes an undeniable part of life that people must cope with when making decisions. Consequently, every forecast or estimate, no matter how informed and unbiased it is, is still treated with skepticism. While such skepticism is perceived by many DMs as an effective counteract to uncertainty, shielding them from post-decision disappointment, is it statistically justified?

1.1 Motivation

Companies select projects to invest in based on uncertain estimates of their performance. DMs usually seek to maximize values by choosing projects with the highest estimated performances. Theories and empirical evidence suggest that if the uncertain estimates are taken at face value, the true performance of the selected projects tends to be lower than estimated, causing the DMs to experience post-decision disappointment. Taking prior information into account through Bayesian adjustment can result in more realistic estimates of the project performances and thus higher expected values among the selected projects. However, Bayesian adjustment makes it less likely to predict extreme outcomes and, as a consequence, may lead to missing out on big wins.

Empirical studies show that there are differences in how much DMs assign weight to prior information. At one extreme, there are so-called Bayesian DMs who weigh prior information optimally in accordance with Bayes' theorem. At the other extreme, there are

---

[1] VUCA: Volatility, Uncertainty, Complexity, and Ambiguity

non-Bayesian DMs who neglect prior information and take the uncertain estimates at face value. The extent to which a DM adopts a Bayesian state of mind is called *Bayesianess*.

This thesis studies the differences in the investment strategies of Bayesian and non-Bayesian DMs and outcomes of these strategies. From a theoretical perspective, this thesis will contribute novel analytic results on these differences and the impacts of different problem parameters on these differences by building a mathematical model. Subsequently, this thesis will validate theoretical findings with empirical evidence of the differences between Bayesian and non-Bayesian estimations in terms of decision-making behavior and financial gains and losses from the selected investments. From a practitioner's point of view, this thesis gives insights into how estimation uncertainties affect investment decisions and outcomes. Understanding such effects can help managers make better-informed decisions.

1.2   Research problems and research questions

Companies often select investment projects based on some threshold value of the project performance, measured by, e.g. net present value, return on investment, expected multi-attribute utility, etc. Evidently, the threshold value varies in practice depending on the nature of the investments.

With that assumption, this thesis will study the differences and the impacts of the differences between Bayesian and non-Bayesian project performance estimation on investment strategies, their respective theoretical and empirical outcomes, and the variables that affect these differences. The investment strategies are studied in terms of the proportion of funded projects out of project proposals. The strategies' impacts will be studied through (i) the average performance of funded projects (i.e. expected value of funded projects), (ii) the proportion of projects that result in losses (i.e. the true values turn out to be less than the threshold), and (iii) the proportion of funded projects with very high *ex post* performance (i.e., higher than 95- or 99-percentile). Besides, this thesis also examines and observes the factors that affect these differences. The factors include (i) the accuracy of the value estimates, as represented by the ratio of the variance in the projects' true values and the

variance of the estimation error and (ii) the proportion of project proposals with truly positive *ex post* values. Additionally, this thesis also testifies such differences with empirical data.

Based on the research problems, the research questions are formed as the three overarching questions as follows:

1. **What is the difference between a Bayesian and a non-Bayesian DM in terms of:**
   a. The proportion of funded projects,
   b. The average *ex post* value of funded project,
   c. The proportion of funded projects that result in losses, and
   d. The proportion of funded projects that result in very high gains (top 5% or 1%)?

2. **How are these differences affected by:**
   a. The accuracy of the value estimates and
   b. The proportion of project proposals with *ex post* performance above the threshold?

3. **How well are these differences reflected in empirical data?**

The answers to these three questions will provide us with a more thorough grasp of the differences between Bayesian and non-Bayesian investment strategies. In addition, this thesis will demonstrate the differences between Bayesian and non-Bayesian DMs in terms of expected post-decision disappointment and expected portfolio value. Based on this understanding, this thesis will develop a framework for better-informed and more accurate decision-making processes.

1.3   Methods and data

To study the differences between Bayesian and non-Bayesian DMs, a mathematical model for the projects' true and estimated values is built. In particular, the projects' true and estimated values are modeled as random variables that follow a bivariate normal distribution. Based on this model, the Bayesian-adjusted estimates (defined as the projects' expected true values given the value estimates) can be computed. The model is used to obtain analytic

results about the differences between Bayesian and non-Bayesian DMs. Specifically, the analytic model is expected to answer the questions about the differences between a Bayesian and a non-Bayesian DM in questions 1a and 1b and observe the effects of the elements listed in question 2. The expected results of the analytic model are illustrated in the following hypotheses:

Hypothesis 1: *If the proportion of project proposals with truly positive performance is lower than 50%, a Bayesian DM is likely to underinvest, whereas a non-Bayesian DM is likely to overinvest. The less accurate the value estimates are, the smaller the share of projects funded by the Bayesian DM and the larger the share of projects funded by a non-Bayesian DM. On the other hand, if the proportion of truly good proposals is higher than 50%, the roles of Bayesian and non-Bayesian DMs are reversed.*

Hypothesis 2: *If the proportion of project proposals with truly positive performance is lower than 50%, the projects funded by a Bayesian DM are on average more valuable than those funded by a non-Bayesian DM. The less accurate the value estimates are, the more valuable the projects funded by a Bayesian DM are compared to those funded by a non-Bayesian DM. On the other hand, if the proportion of truly good proposals is higher than 50%, the roles of Bayesian and non-Bayesian DMs are reversed.*

In addition to validating these hypotheses, the analytic model is also used to simulate decision-making processes based on Bayesian-adjusted and unadjusted estimates. Subsequently, this thesis will compare and analyze the differences between a Bayesian and non-Bayesian investment strategies based on the results of those simulations. Like the results from the analytic model, the simulation results are also expected to show distinctive differences between Bayesian-adjusted and unadjusted estimates. More specifically, the simulation results are expected to validate the hypotheses 1 and 2. In addition, the simulation results are also expected to answer the remaining of the research questions that are too complex for the analytic model (i.e. questions 1c and 1d with regards to question 2). The hypotheses for the simulation part are as follows:

Hypothesis 3: *If the proportion of project proposals with truly positive performance is lower than 50%, the funded projects of a Bayesian DM have a lower probability of incurring loss than those of a non-Bayesian DM. The less accurate the value estimates are, the lower the probability of incurring loss of the Bayesian DM's funded projects compared to that of the non-Bayesian DM's funded projects. On the other hand, if the proportion of truly good proposals is higher than 50%, the roles of Bayesian and non-Bayesian DMs are reversed.*

Hypothesis 4: *If the proportion of project proposals with truly positive performance is lower than 50%, the funded projects of a Bayesian DM have a lower probability of resulting in high gains than those of a non-Bayesian DM. The less accurate the value estimates are, the lower the probability of resulting in high gains of the Bayesian DM's funded projects compared to that of the non-Bayesian DM's funded projects.*

In the empirical part, to seek empirical evidence to endorse the theoretical findings, this thesis will study ***R&D investment decisions in the pharmaceutical industry***. This empirical part aims to explain the natures of R&D investment decisions in the pharmaceutical industry by using the theoretical findings and to examine how well the theoretical findings are reflected in empirical data. The pharmaceutical industry is selected because of its poor estimation accuracy. Such poor accuracy is the result of the lack of quantitative support in decision making, the decentralization of decision making causing information asymmetry, and the highly disruptive nature of the industry. Consequently, the author expects this industry to reflect well the differences between a Bayesian DM and a non-Bayesian DM.

The empirical part will be conducted through secondary research based on existing literature (i.e. academic publications and industry reports) on R&D investment decision in the pharmaceutical industry. This thesis will consolidate relevant characteristics of the pharmaceutical industry and reflect them to the theoretical findings from the analytic model and simulation model.

1.4   Structure of the thesis

This thesis has three primary parts as follows: literature review, mathematical models and empirical evidence. These parts are further divided into seven chapters in total. The thesis embarks with Chapter One, where the general background and research motivation are discussed. This introductory chapter will also position and frame this study, describing the significance of the topic and defining the research objectives. Chapter Two will review the literature on (i) project selection under uncertainty, (ii) modeling uncertainties in portfolio selection, and (iii) Bayesian modeling of estimation uncertainties. The main purpose of this chapter is to examine related issues categorically through earlier research. Chapter Three will discuss the methodology of the study, which is a combined methodology of quantitative and qualitative methods. This chapter aims to provide a thorough understanding of the thesis' methodological approaches and the link between those two seemingly separate methodologies. More specifically, the quantitative approach of this thesis is in the form of a mathematical model that is used both to derive analytic results and for Monte Carlo simulation. The qualitative approach of this thesis is utilized to test theoretical findings empirically in the pharmaceutical industry. Chapter Four will be dedicated to building a mathematical model to answer the research questions. After Chapter Four, all the research questions will be answered. The answers for the research questions will serve as a basis for the Empirical part in Chapter Five.

In Chapter Five, we will scrutinize the findings in the pharmaceutical industry through industry reports and existing publications on Research and Development (R&D) Portfolio selection in this industry. This chapter aims to validate theoretical findings and how relevant they are in a real-life environment. Subsequently, Chapter Six will consolidate and combine the findings and their empirical implications to propose actionable strategic recommendations for companies in terms of decision-making strategies in portfolio selection. In Chapter Seven, the conclusions will be presented. Theoretical contributions, managerial recommendations, research limitations, and suggestions for future research will also be discussed in this chapter.

2   Literature Review

The literature review part of this thesis is thematically structured, starting from a more general topic to a more specific topic. Firstly, it will examine the backdrop where the phenomenon takes place, i.e., project portfolio selection under uncertainty. Subsequently, it will review literature on the modeling of uncertainties in project portfolio selection. Lastly, it will review literature on Bayesian modeling of estimation uncertainties. The ultimate purpose of this literature review part is to better understand the environment where Bayesian and non-Bayesian DMs make decisions, the externalities that influence their decision-making processes, and the existing models that assist DMs in making better decisions. Reviewing existing literature also helps build the analytic model and subsequently build the simulation model.

2.1   Project portfolio selection under uncertainty

In practice, nearly all organizations aim to create value by choosing and executing portfolio of actions. The task of selecting project portfolios is a vital and frequent activity in many organizations. Those portfolios of actions typically consume resources (e.g., financial, human, time, etc.) and are therefore critically constrained by the availability of such resources (Kavadias *et al.*, 2004).

Project selection is a strategic decision problem which is often characterized by multiple, conflicting, and incommensurate criteria (Liesiö *et al.*, 2007), while the DM has to decide a portfolio of the most attractive alternative by taking into account different aspects of the projects' values (Mavrotas *et al.*, 2003). In other words, in a project selection problem, a DM allocates limited resource to a set of competing projects, considering one or more corporate goals or objectives (Medaglia *et al.*, 2007). It is very typical in a project selection process that analysts identify a set of project proposals, estimate the value of each proposal and accordingly recommend the DMs to choose the combination or portfolio of projects that has the highest estimated value overall, subject to resource and other constraints. The choice of the *"right"* projects must also consider consequences over multiple periods and uncertain

outcomes (Kavadias *et al.*, 2004). Consequently, project selection is a very complex decision-making process since it is affected by many critical factors, e.g. raw materials availability, probability of technical success, market condition, and government regulations (Bard *et al.*, 1988). Because of such complexity, Archer & Ghasemzadeh (1999) develop an integrated project portfolio selection framework that separates the selection process into different stage with consequential objectives.

Besides being highly complex, project portfolio selection processes also faces with great uncertainties. As the realizations are uncertain at the time of decision making, organizations must make decision based merely on *ex ante* estimates of the future value (Vilkkumaa *et al.*, 2014). According to Lindley *et al.* (1979) and Lindley (1986), value estimates are always uncertain. Estimated values are only *"true"* if the business executive could devote unlimited time, money, and computational resources to make the expected value calculation. Due to uncertainty and limited resources, there are no such things as *"true"* value estimates and, consequently, there are no perfect analyses. These value estimates are thus subject to error. Therefore, there exists a high level of risk for the uncertainty or incompleteness of project information that will consequently make it harder for the DM to choose the correct alternatives (Wang *et al.*, 2009).

A *"wrong"* decision in portfolio selection has a two-fold effect: (i) *direct costs:* resources are spent on unsuitable projects, and (ii) *opportunity costs:* the organization loses the benefits it could have gained if the resources had been spent on more suitable projects (Martino, 1995). Vilkkumaa *et al.* (2014) present two types of outcomes of project selection under uncertainty: (i) suboptimal portfolio and (ii) post-decision disappointment.

2.1.1   *Suboptimal portfolios*

Suboptimal portfolios are portfolio that contains projects that do not belong to the optimal portfolios (Vilkkumaa *et al.*, 2014). Due to the inherent uncertainty, some non-optimal projects have higher value estimates than do optimal projects, causing DMs to choose *"wrong"* projects to invest in. According to Loch & Kavadias (2002), optimal portfolios are difficult to define also because of the combinatorial complexity of project combinations. As

uncertainty is unavoidable in practice, the probability of selecting *"wrong"* projects, resulting in a suboptimal portfolio, is always larger than zero.

2.1.2 Post-decision disappointment

In a decision-making process, the *ex-ante* value estimate of a selected alternative serves as a reference point, to which the true value is compared when it is realized. Due to estimate errors, the *ex-ante* estimates of costs and benefits are often very different from actual *ex post* cots and benefits (see e.g. Flyvbjerg *et al.*, 2002; Flyvbjerg, 2009; Odek 2004; Jørgensen, 2013). Yet, in practical applications, it has been noted that the *ex post* observed values of selected projects are lower than estimated *ex ante* values and/or *ex post* costs are higher than *ex ante* estimated costs, causing the DM to experience post-decision disappointment, defined by Bell (1985, pp. 1) as *"a psychological reaction caused by comparing the actual outcome of a lottery to one's prior expectations."* Typically, post-decision disappointment manifests itself in terms of cost overruns and benefit underruns.

2.1.2.1   <u>Empirical evidence of cost overruns and benefit underruns</u>

Cost overruns in project selection under uncertainty are very pronounced in practice. There have been several empirical studies on the issue of cost overruns across different industries. According to Flyvbjerg *et al.* (2002), 90% of transportation infrastructure projects experience cost overruns. Odeck (2004) reveals a noticeable discrepancy between actual and estimated costs of road construction using data from Norwegian road construction over the years 1992-1995. Jørgensen (2013) seeks the reason for effort overruns in software projects. The results from the statistical model and the experiment demonstrate that selection bias can explain cost overruns (e.g. effort overruns).

In terms of geographic prevalence, cost overrun is a global phenomenon. It is observed across 20 nations and 5 continents (Flyvbjerg *et al.*, 2002). Cost overrun has not decreased over the past 70 years, proving that development in computation capacity, information availability, and data accessibility does not improve cost estimation accuracy. Hence, cost underestimation cannot be explained by errors and seems to be best explained by strategic misinterpretation (i.e. deliberate lies). Moreover, cost overrun is a ubiquitous phenomenon

across all types of projects. Transportation infrastructure projects are not prone to cost underestimation than are other types of large projects (Flyvbjerg *et al.*, 2002).

Benefit underruns can also be observed empirically. An empirical study by Pruitt & Gitman (1987) concludes that capital budgeting forecasts are optimistically biased. In their study, Pruitt & Gitman (1987) survey executives who confirm that *ex post* values of large projects fell far below their expected values. Consequently, these executives tend to adjust project profitability estimates downward in an effort to compensate for the bias. A more recent empirical study by Brous *et al.* (2009) compares the forecasted and actual performance of each investment made by one firm from 1996 to 2001 and concludes that in its sample, optimism bias leads to significant discrepancies between expected and actual NPVs. Accordingly, 80.41% of investments made between 1996 and 2001 realize an actual NPV that fell short of the expected NPV. The actual NPV, on average, is less than 1/3 of the expected NPV.

2.1.2.2   Causes of post-decision disappointment

Very uncertain value estimates have two effects on project selection. Firstly, uncertain value estimates make it more difficult to identify the projects with highest values. Secondly, they make it more likely that the values of the selected projects will be systematically overestimated. Earlier literature has identified two main sources for post-decision disappointment: (i) systematic bias in the projects' value/cost estimates, and (ii) the optimizer's curse (i.e. the consequences of the optimization-based selection process).

a.   Systematic bias

Systematic bias refers to the estimation errors that cannot be explained by random noise and thus these estimation errors do not average out. According to Flyvbjerg (2009), the difference between *ex ante* estimates and actual *ex post* values of ventures' costs and benefits are very pronounced for large infrastructure projects, where substantial cost underestimates are often combined with equally significant benefit overestimates, causing cost-benefit analyses of projects not only inaccurate but also systematically biased. By definition, the magnitude of such difference is positively correlated with the magnitude of post-decision disappointment.

Flyvbjerg (2009) identifies three main types of reasons that cause systematic bias in selecting infrastructure projects: (i) technical, (ii) psychological, and (iii) political-economic. The causes of systematic bias are illustrated in *Figure 1*.
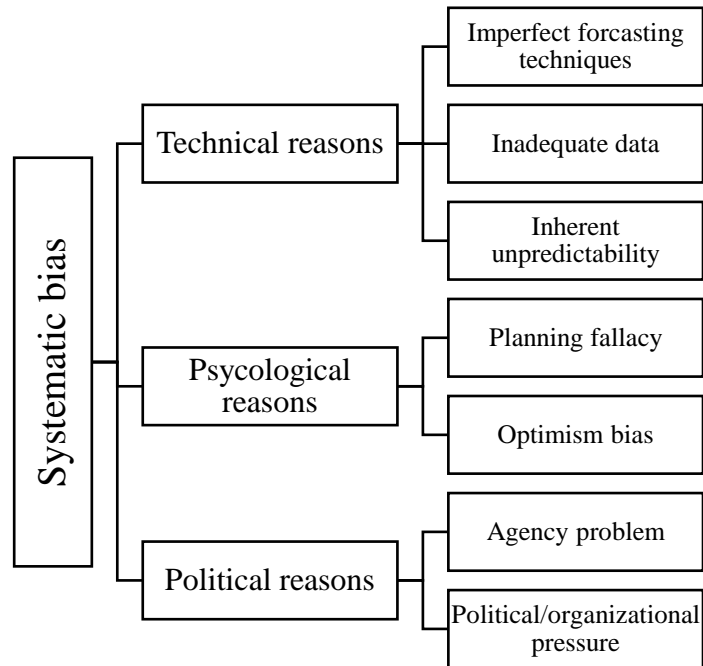
Figure 1: Causes of systematic bias (developed from Flyvbjerg, 2009)

Technical causes of systematic bias encompass of imperfect forecasting techniques (e.g. honest mistakes, forecasters' lack of experience, etc.), inadequate data, and the inherent unpredictability of the future values (Ascher, 1978; Flyvbjerg et al, 2002, 2005; Flyvbjerg, 2009). Thus, technological causes can be overcome by developing better forecast mechanism, improving data availability, and hiring more experienced forecasters.

Psychological causes include planning fallacy and optimism bias. Planning fallacy is defined by Kahneman & Tversky (1979) as a phenomenon where predictions about the time needed to complete a future task display an optimism bias and underestimate the time needed. In decision making, planning fallacy occurs when managers make decisions based on delusional optimism rather than on a rational weighting of gains, losses, and probabilities. Those managers involuntarily spin scenarios of success and oversee the possibility for errors and inaccuracies (Flyvbjerg *et al.*, 2005; Flyvbjerg, 2009). Over-optimism is rooted in

cognitive biases, i.e. errors in the way the mind process information. Optimism bias is a cognitive bias that causes a person to believe that they are at a lesser risk of experiencing a negative event compared to others (O'Sullivan, 2015). An empirical study of 97 investments by Brous *et al.* (2009) shows that optimism bias can results in substantial discrepancies between expected and actual NPVs.

Lastly, political-economic causes include project planners' and promoters' deliberate and strategic overestimation of benefits and underestimation of costs to increase the likelihood that the projects (and not the competitors' projects) are approved and funded (Wachs, 1989, 1990; Flyvbjerg *et al.*, 2002, 2005; Flyvbjerg, 2009). This strategic misinterpretation can be traced back to agency problems and political and organizational pressure (i.e. competition for scarce funds or jockeying for position). This deliberate misinterpretation thus can be considered as lies by Bok's (1979, pp. 14) and Cliffe *et al.*'s (2000, pp.3) definition.

b. Optimizer's curse

Even if there is no systematic bias in the value/cost estimates, when projects are selected based on the highest value estimates, the values of selected projects are thus more likely to have been overestimated than underestimated. Likewise, if projects are selected based on the lowest estimated costs, the selected projects are very likely to incur cost overruns. (Brown, 1974, 1978). This phenomenon is named *optimizer's curse*. The optimizer's curse therefore is a consequence of the optimization-based selection process and is one of the two main causes of the post-decision disappointment.

Without behavioral adjustments, higher expected benefits will be associated with greater expected disappointments; as an alternative with a relatively high estimated value will have, on average, a relatively large positive error associated with it (Pearson, 1897; Snedecor, 1946; Harrison & March, 1984; Bell, 1985). Thus, a DM who chooses alternatives based only on the face value of her estimates, on average, should expect to be disappointed, even if the individual value estimates are conditionally unbiased. In such cases, the inherent uncertainty in these value estimates coupled with the optimization-based selection process

makes the value estimates for the suggested actions to be biased high (e.g., Miller, 1978; Harrison & March, 1984; Miller, 1986; Smith & Winkler, 2006; Vilkkumaa *et al.*, 2014). In many aspects, this optimizer's curse is similar to the winner's curse (Capen *et al.*, 1971; Thaler, 1992). However, the optimizer's curse affects all kinds of intelligent decision making that involves attempts to optimize based on imperfect estimates.

Research has shown that there are three parameters that influence the magnitude of the optimizer's curse: (i) the signal-to-noise ratio, (ii) the number of alternative considered, (iii) the correlation among value estimates, and (iv) the correlation among the true values (Harrison & March, 1984; Smith & Winkler, 2006; Vilkkumaa *et al.*, 2014) (See ***Figure 2***).
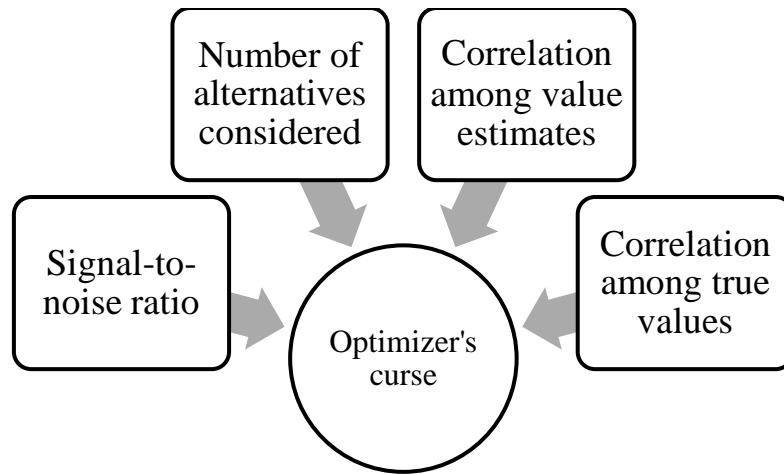
Figure 2: Elements that affect the magnitude of the optimizer's curse

(1) The signal-to-noise effect

Any selection decision based on *noisy* estimates (i.e. estimates with relatively large estimate errors) will exhibit a structural tendency toward disappointment. Studies have shown that the more uncertain the estimates, the more disappointment is to be expected (Harrison & March, 1984; Vilkkumaa *et al.*, 2014).

(2) The number of alternatives effect

The expected level of disappointment is also affected by the number of alternatives considered. In selection of a single alternative out of many, the expected disappointment is positively correlated with the number of alternatives considered (Hastings *et al.*, 1947;

Harter, 1969; Harrison & March, 1984; Smith & Winkler, 2006). Consequently, in portfolio selection, the smaller the share of selected projects out of project proposals, the larger the disappointment.

(3) Correlation between value estimates effect

In some cases, the projects' value estimates are correlated because the projects share common elements, such as common probability for technical success among R&D projects, or common probability distribution of the amount of oil available among oil development projects). Smith & Winkler (2006) show that if the correlation between the value estimates is positive, the expected post-decision disappointment is decreased. On the contrary, if the correlation between the value estimates is negative, the expected post-decision disappointment is increased.

(4) Correlation among true values effect

As the true values are uncertain in practice, one can expect the true values to be positively correlated for the same reasons as those for value estimates (e.g. true values for alternatives that depend on the same probability for technical success). Harrison & March (1984) and Smith & Winkler (2006) show that positive correlation among the true values increases the magnitude of expected disappointment. Smith & Winkler (2006) also observe the combined effect of the correlation among value estimates and the correlation among true values and conclude that increasing both correlations of value estimates and true values at the same time leads to a net decrease in the magnitude of post-decision disappointment.

For the sake of simplicity, the following parts of this thesis will focus on studying the case where there is no systematic bias in the projects' value or cost estimates; thus, post-decision disappointment is assumed to be entirely caused by the optimizer's curse.

2.2   Modeling uncertainties in portfolio selection

Earlier literature has presented models that incorporate uncertainties into portfolio selection process. Among publications on the topic of project selection under uncertainty, Charnes & Stedry (1966) present a technique called chance-constrained programming model. In such

model, random variables are defined to consider uncertainty for availability of facilities that are needed to perform R&D projects. However, this model fails to consider other kinds of uncertainties. Li & Sinha (2004) propose a method for highway investment decision making under uncertainty that utilizes Shackle's model for uncertainty-based project benefit analysis and system optimization for project selection. With the use of Shackle's model, the limitations of the risk-based life-cycle cost analysis approach are overcome by using degree of surprise as a measure of uncertainty associated with possible outcomes of performance measures used in project benefit analysis (Shakhsi-Niaei *et al.*, 2011). Li & Sinha (2004) also present a case study that reveals substantial differences in project selection results from using the proposed method and from using the current risk-based approach.

Loch & Kavadias (2002) develop a dynamic model of resource allocation, characterize optimal policies in closed form, and derive qualitative decision rules for managers. Medaglia *et al.* (2007) scrutinize project selection as a stochastic multi-objective linearly-constrained optimization. Consequently, Medaglia *et al.* (2007) propose an evolutionary method with partially funded projects, multiple (stochastic) objectives, project interdependencies (in the objectives), and a linear structure for resource constraints. The basis for this method is posterior articulation of preferences. This method is therefore capable of approximating the efficient frontier from stochastically non-dominated solutions. Additionally, Medaglia *et al.* (2007) compare the method with the stochastic parameter space investigation method and demonstrate it with a R&D portfolio problem under uncertainty based on Monte Carlo simulation.

Wey (2008) examines uncertainty of available budget, the chance of success and the efficient allocation of the project team in the selection of urban renewal projects with three techniques: integer-constrained multi-objective optimization, Monte Carlo simulation, and Analytic Network Process where costs are described with a probability distribution. Li (2009) examines budget uncertainty in highway investment decision making with a stochastic optimization model that addresses explicitly the inherent budget uncertainty. Li & Madanu (2009) introduce an uncertainty-based method for highway project level lifecycle benefit-cost analysis and project assessment, and study project benefits in three approaches:

deterministic, risk-based and uncertainty-based. The three sets of estimated project benefits are then executed on a stochastic optimization model for project selection. Consequently, it is concluded that there are substantial differences when uncertainty is considered versus when it is not considered.

Toppila *et al.* (2011) present a decision model for resources allocation to a portfolio of R&D projects that captures the dynamic structure of the decision problem with decision trees, recognizes uncertainties about sales parameters by employing interval estimates, and considers possible project interaction. The consideration of project interaction (e.g. mutual enabling, incompatibility, competition for the same scarce resources. In such context) is emphasized by Kavadias *et al.* (2004) by promoting the portfolio view in project selection. Subsequently, Abbassi *et al.* (2014) present a binary non-linear mathematical programing model for balancing portfolio values and their associated risks and develop a cross-entropy algorithm to solve the model.

Preference programming is another approach to modeling uncertainties in portfolio selection. *Preference programming* an umbrella term for multi-criteria weighting model under incomplete information that accommodate incomplete preference information, provide dominance concepts and decision rules to generate decision recommendations, and support the iterative exploration of the decision maker's preferences (Salo & Hämäläinen, 2010). Liesiö *et al.* (2007, 2008) present preference programming for robust portfolio modeling and project selection that extends preference programming methods into portfolio problems and supports project portfolio selection in the presence of multiple evaluation criteria and incomplete information (i.e. uncertainty). Subsequently, Liesiö & Salo (2012) develop a framework for scenario-based portfolio selection of investment projects that accommodates interactive decision support processes where the implications of additional probability and utility information or further risk constraints are demonstrated with regard to corresponding decision recommendations.

2.3   Bayesian modeling of uncertainties

Bayesian modeling of uncertainties is another type of modeling uncertainty in portfolio selection. The concept of Bayesian process essentially is to treat the results of an analysis (i.e. value estimates) as uncertain and before choosing an alternative by combining the value estimates with prior estimates of the value using Bayes' rule. More specifically, instead of making decision by taking uncertain estimates at face value, DMs explicitly model estimation uncertainties by associating a prior distribution with the projects' values and a conditional distribution on the value estimates. As a result, posterior distribution of the projects' true values given the observed value estimates can be obtained from the prior distribution and value estimates. Subsequently, this posterior distribution can be used to compute the projects' expected true values given the value estimates. This process therefore recognizes the inherent uncertainty and corrects for the upward bias that is associated with the process of optimization by adjusting the overestimated values downward. In order to adjust those values properly, DMs need to know the degree of uncertainty of both the estimates and the true values (i.e. noise and signal, respectively). (Gelman *et al.*, 2014).

It may be challenging to formulate and assess sophisticated models that illustrate the uncertainty in true values and value estimates given by a complex analysis in practice. In essence, it is recommended to view the result of an analysis as being analogous to an *expert report* and handle it in the manner way as Morris (1974) suggests. In such context, the Bayesian methods for adjusting value estimates can be considered as a disciplined method for discounting the results of an analysis to avoid post-decision disappointment (e.g. Gelman *et al.*, 2004). Accordingly, the Bayesian methods require the DM to consider her prior value estimates and the accuracy of such estimates and subsequently integrate her prior opinions into the analysis (Smith & Winkler, 2006).

Earlier literature in decision analysis has also examined the various levels at which different DMs take into account prior information and the effects of these differences in decision-making such as post-decision disappointment. According to earlier behavioral studies, people tend to underestimate prior information about the *base rate* of the event when

facing such prediction tasks. They instead base their decisions on the most recent evidence. Such behavior can result in errors in predicting rare events and extreme realizations (Kahneman & Tversky, 1973, 1979).

Theories suggest that if the true values and estimation errors are normally distributed, the Bayesian model results in the shrinkage of value estimates toward a prior mean, toward the mean of all value estimates, or toward some type of combination of those two (e.g. Gelman *et al.*, 1995; Carlin & Louis, 2000; Smith & Winkler, 2006). Thus, Bayesian-adjusted estimates have lower variance than unadjusted estimates. In other words, Bayesian adjusted estimates are less likely to be extremely high or low. However, the expected value of a project portfolio selected based on Bayesian-adjusted estimates is higher than that of a portfolio selected based on unadjusted estimates (Vilkkumaa *et al.*, 2014). To some extent, the Bayesian process of interpreting value estimates exhibits, similar characteristics to ambiguity aversion or uncertainty aversion (Ellsberg, 1961), in which good alternatives are penalized for uncertainty in their value estimates. The magnitude of the penalty is positively correlated with the uncertainty. Earlier research has also studied the implication of Bayesian or non-Bayesian estimation in the context of predicting *"the next big thing"* (i.e. rare events such as breakthrough technologies, radical innovations, disruptive technologies, etc.). According to an empirical study by Denrell & Fang (2010), forecasters that have been able to predict some unusual event tend to neglect the low *base rate* of such events, whereby their overall accuracy rates are very low. On the other hand, forecasters with good overall accuracy rates tend to fail in predicting rare events.

Bayesian modeling of uncertainties has long been present in financial portfolio optimization. Winkler & Barry (1975) introduce general models for portfolio selection and revision and utilize Bayesian inferential procedure to formally update probability distribution as ancillary information is received. Barry & Winkler (1976) present a Bayesian inferential model for forecasting future security prices under nonstationary parameters and compare it with a corresponding stationary model. Aguilar & West (2000) develop and explore Bayesian inference and computation in an empirical study of the dynamic factor structure of daily spot exchange rates for a selection of international currencies. Polson & Tew (2000) conduct an

empirical analysis of the S&P 500 index 1970-1996 that describes several statistical issues involved in quantitative approaches to portfolio, including predictive Bayesian approach using hierarchical models that incorporate parameter uncertainty and nonstationarity, and propose a technique for implementing large-scale optimal portfolio by using high-frequency data to obtain valuable statistical information in asset returns. Brandt *et al.* (2005) adopt a simulation approach to dynamic portfolio choice with an application to learning about return predictability that solve discrete-time portfolio choice problems involving non-standard preferences, numerous assets with arbitrary return distribution, and numerous state variables with potentially path-dependent or non-stationary dynamics. Soyer & Tanyeri (2006) present a simulation-based method to solve the multi-period portfolio selection problems with a Bayesian method, assuming the security returns follow multivariate random multivariate stochastic variance models.

Despite the prevalence of Bayesian modeling in financial portfolio optimization, its implications in project portfolio optimization have not yet been thoroughly explored. Project portfolio optimization and financial portfolio optimization are very similar in many aspects, but they are also fundamentally distinctive. Project portfolio selection, to some extent, is more complex than financial portfolio selection. According to Vilkkumaa *et al.* (2014), project portfolio selection differs from financial portfolio optimization in three major aspects: (i) the value of a project cannot be observed at the time of decision making as in the case of financial portfolio optimization, where market prices can be easily observed from the market; (ii) project portfolio selection decisions are either selected or rejected, while in financial portfolio optimization, an investor can invest any fractional amount of resources in any security; and (iii) there is more interdependence in project portfolio selection due to the sequel nature of some projects, in financial portfolio selection, although there is some correlation in security prices, decisions to invest in securities are logically independent to each other. As a consequence, Bayesian modeling of uncertainties in project selection is worth studying in more detail.

In the context of project portfolio selection under uncertainty, Bayesian modeling of uncertainty help DMs (i) increase the expected value of future realizations of the selected

projects, (ii) increase the number of selected projects that belong to the *ex post* optimal portfolio, and (iii) decrease the expected post-decision disappointment (Vilkkumaa *et al.*, 2014) (See *Figure 3*). Thus, a Bayesian DM will have a higher number of selected projects that belong to the optimal portfolio and a more realistic expectation about the performance of her selected projects, whereas a non-Bayesian DM is more likely to select a suboptimal portfolio with overestimated value.
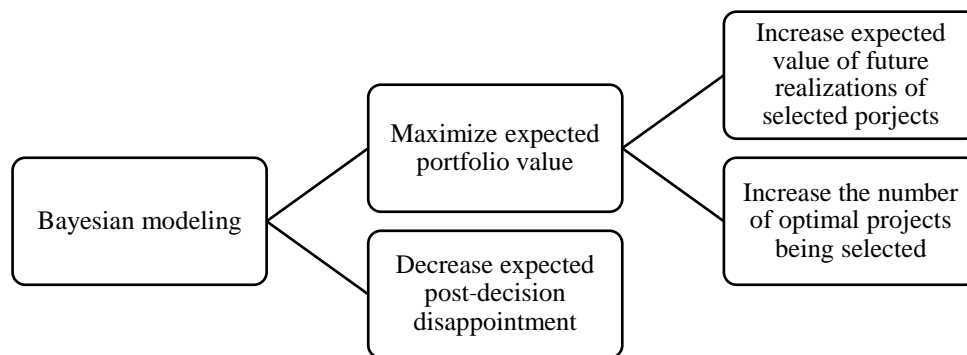
Figure 3: Benefits of Bayesian modeling of estimation uncertainty in project portfolio selection

2.3.1   Maximization of the expected portfolio value

The Bayesian modeling of estimate uncertainty are widely used in optimization problems in portfolio selection. The aims of such optimization problems include: (i) to maximize the expected performance of selected projects or (ii) to maximize the expected number of selected projects that belong to the truly optimal portfolio. Vilkkumaa *et al.* (2014) conclude that under very general assumptions, selecting projects based on these Bayesian estimates will yield at least as much value as the portfolio which maximizes the sum of value estimates. Nonetheless, if the use of Bayesian estimates leads to the selection of a different portfolio other than that of the use of value estimates, the expected value of the selected projects based on Bayesian estimates is strictly better (Vilkkumaa *et al.*, 2014).

2.3.2   Mitigation of post-decision disappointment

The Bayesian methods are widely recognized as an effective way to overcome the post-decision disappointment (e.g. Nickerson & Boyd, 1980; Harrison & March, 1984; Lindley, 1986; Gelman *et al.*, 1995; Carlin & Louis, 2000; Smith & Winkler, 2006; Vilkkumaa *et al.*, 2014, 2015). As the Bayesian-adjusted estimates will shrink, both upwards and downwards, the expectations from the initial estimates toward the prior mean, the Bayesian-adjusted estimates tend to yield more conservative information (i.e. less overestimated information) about the value of the selected portfolio. Vilkkumaa *et al.* (2014) state that, even without specific assumptions about the distribution of the true values and the value estimates given a true value, or about the problem constraints, using Bayesian-adjusted estimates will eliminate the expected discrepancy between the true values and estimated values (i.e. post-decision disappointment) of the chosen portfolio.

Accordingly, in the selected portfolio resulting from using Bayesian-adjusted estimates, the expected difference between the true value (i.e. realized *ex post* value) and the Bayesian-adjusted estimates is zero. In individual cases of each chosen project, the Bayesian estimates help alleviate post-decision disappointment, as it is less probable to have extreme Bayesian estimates than extreme non-Bayesian estimates (see e.g. Smith and Winkler, 2006; Vilkkumaa *et al.*, 2014).

3   Research Methodology

This chapter will present the methods of the research by explaining what type of data are analyzed and how, where, and when they are collected for the study. Furthermore, this chapter will also discuss briefly data analysis methods.

3.1   Overview of methodology

Amaratunga *et al.* (2002) classify research methodologies into two categories: quantitative and qualitative. Accordingly, the quantitative method relies primarily on numerical data, utilizes standardized measurements to test hypotheses and discover distinguishing characteristics, or empirical barriers. On the other hand, the qualitative method aims to describe, examine, and construct an understanding of culture, social behavior, or other similar phenomena. The qualitative method also takes into account the differences between individuals (Amaratunga *et al.*, 2002).

Consequently, a mixed method is appropriate for this thesis as the phenomenon being studied in this thesis includes both quantitative and qualitative aspects. On the one hand, it concerns about the statistical and analytical differences of a Bayesian and a non-Bayesian DM. On the other hand, it also concerns about the empirical relevance of such differences in a real-world environment, which involves more social and human factors that would be easier to analyze with a qualitative approach. Therefore, a mixed method of both quantitative and qualitative would be the best approach to such complex phenomenon.

The quantitative approach of this thesis will be carried out with a mathematical model. The mathematical model encompasses of an analytic model and simulation. The results from the analytic models will be used as the basis for Monte Carlo simulations, and the Monte Carlo simulations serve as a confirmation and validation for the results derived from the analytic model. In addition, the Monte Carlo simulation also allows the author to examine aspects of the model for which analytic results could not be derived.

The qualitative approach of this thesis will be employed in the empirical evidence part. As the portfolio selection decisions are influenced by many sociological factors, such as industry characteristics, corporate cultures, organizational structures (centralized decision making vs. decentralized decision making), DMs' preferences, etc., it would be appropriate to adopt a qualitative approach. Accordingly, this thesis will analyze the characteristics of the decision-making environment of the pharmaceutical industry, hypothesize the different outcomes of Bayesian and non-Bayesian investment decisions based on theoretical findings. Consequently, this thesis will study the characteristics of DMs in this industry and compare their observed investment outcomes with the theoretical outcomes.

3.2   Quantitative method

The quantitative method in this thesis is executed in the form of a mathematical model. The mathematical research methodology is by far the most popular methodology (Wacker, 1998). According to Dym (2004, pp. 6), *"mathematical modeling is a principled activity that has both principles behind it and methods that can be successfully applied"*. The principles are either overarching or meta-principles phrased as questions about the objectives, intentions and purposes of mathematical modeling. The principles are outlined and visually illustrated in *Figure 4*, which lists the questions asked in a principled approach to building a model relate to the development of that model.
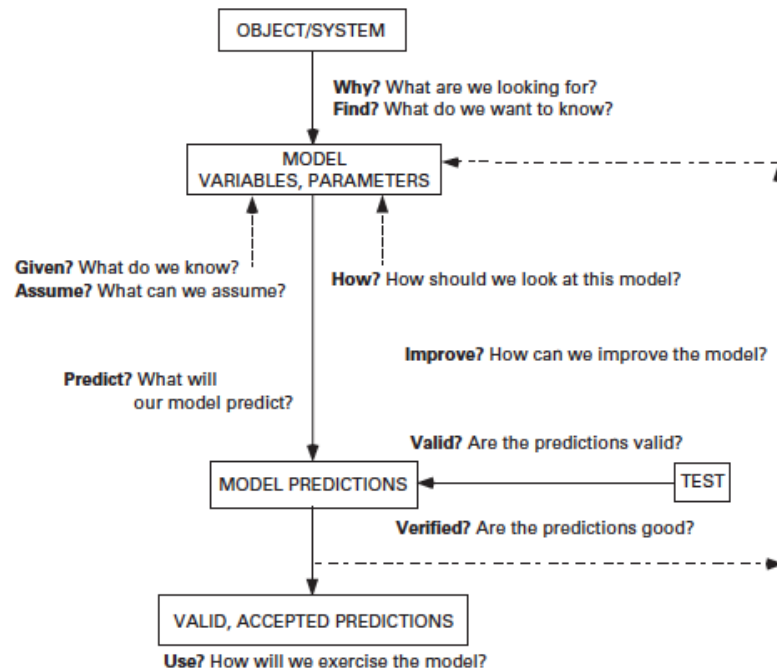
Figure 4: A first-order view of mathematical modeling that shows how the questions asked in a principled approach to building a model relate to the development of that model (developed from Carson and Cobelli, 2013).

3.2.1   *Analytic model*

Analytic models are a subset of mathematical models. By definition, analytic models are mathematical models that have a closed form solution, meaning that the solution to the equations used to describe changes in a system can be expressed as a mathematical analytic function. Gershenfeld (1999, pp. 7) defines analytic models as those that *"you can at least in theory write down with nothing more than a pencil and a piece of paper, hopefully arriving at an explicit closed-form solution"*. Analytic models are of great significance because of their power: where they are appropriate, it can be possible to deduce everything there is to know about a system (Gershenfeld, 1999).

In the beginning of the mathematical part of this thesis, we will define the variables, parameters and assumptions. Most of these variables, parameters and assumptions are based on those in existing literature. Subsequently, this thesis will utilize widely known

mathematical and statistical formulae and transformations to convert those formulae into a final form where conclusions that will answer the research questions can be made. The general formulae used in this thesis include, but not limited to, calculations of variance, covariance, correlation, the inverse Mill's ratio, etc.

In a nutshell, the analytic model in this thesis aims to provide answers to the research questions. In addition, the analytic model also provides an analytic framework based on which the Monte Carlo simulation model is built.

3.2.2   *Monte Carlo simulation model*

When some certain regression assumptions need to be validated, Monte Carlo simulation can be a way out. Monte Carlo simulation is a commonly used technique in the probabilistic analysis. In an essence, it is a numerical experimentation technique to obtain the statistics of the output variables of a system computational model, given the statistics of the input (Mahadevan, 1997). Mooney (1997) explains that when the population of interest is simulated, from the so-called *pseudo population*, repeated random samples are drawn. The statistic under study is computed under each *pseudo-sample,* and its sampling distribution is examined for insights into its behavior (Mooney, 1997).

Monte Carlo simulation method has been used for two primary purposes: (i) validation of the analytic results, and (ii) solution to the large complex systems when analytic approximations are not easy to make (Mahadevan, 1997). Mooney (1997) states that Monte Carlo simulation method can offer an alternative to analytic mathematics for understanding a statistic's sampling distribution and assessing its behavior in random samples. In terms of mechanism, Monte Carlo simulation operates by using random samples from known populations of simulated data to observe a statistic's behavior (Mooney, 1997). The basic concept of Monte Carlo simulation is relatively straightforward: if a statistic's sampling distribution is the density function of the value it could take on in a given population, then its estimate is the relative frequency distribution of the values of that statistic that were actually observed in many samples drawn from that population (Mooney, 1997). As it usually is unfeasible for social scientists to sample actual data multiple times, they use artificially

generated data that resemble the real thing in applicable ways. The recent availability of high-speed computers makes this approach now widely practical.

In this thesis, the Monte Carlo simulation model is built on the basis of findings from the analytic model. It aims to simulate the result of selecting an unlimited number of projects out of 1000 alternatives to invest in based on some predefined threshold. Subsequently, it is run 1000 times. More specifically, the Monte Carlo simulation in this thesis is conducted on Microsoft Excel software, using random number function to draw values from normal distribution and using one-way data table to run it 1000 times.

3.3   Qualitative method

Liamputtong & Ezzy (2005) argue that in addition to quantitative research method of numbers and statistics, there need to be other research methods that are able to explore the complexity of human behavior. In such context, qualitative research method is the solution. It is the embodiment of the means of eliciting evidence from diverse individuals, population groups, and contexts. Owning to its flexibility and fluidity, qualitative research is appropriate when researchers seek to understand the meanings, interpretation, and subjective experiences of individuals. While the modeling part of this thesis is heavily quantitative, the empirical part highly involves numerous human and social factors. Thus, the choice of adopting a qualitative method is but appropriate.

Specifically, this thesis examines the characteristics of the decision-making environment of the pharmaceutical industry using earlier literature and industry reports. From such understanding, this thesis will devise possible outcomes with regard to whether the DM is a Bayesian or a non-Bayesian DM. Subsequently, we will observe empirically the outcomes of R&D investments in the pharmaceutical industry, reflect them to the theoretical outcomes, and categorize them in terms of Bayesianess. We will then analyze the decision-making behavior of DMs in this industry also in terms of Bayesianess and, finally, conclude whether or not the Bayesianess of the DMs is reflected in their investment outcomes.

The pharmaceutical industry is chosen to study in this thesis due to its complex decision-making environment. Drug R&D investment decisions are affected by many social and political variables and thus is expected to display characteristics of decision-making environment with low estimation accuracy. According to our hypotheses, the low estimation accuracy will amplify the differences between a Bayesian and a non-Bayesian DM, therefore making it easier for the author to categorize DMs in this industry based on their Bayesianess.

4 Mathematical Model

In this section, the author will build a mathematical model to investigate the fundamental differences in terms of decision-making behaviors between a Bayesian and a non-Bayesian DM, and the outcomes with respect to different parameters values of the model. The mathematical model will provide both analytic results and simulated results.

4.1 Model description

Consider a DM who wants to select a subset or portfolio out of m project proposals $i = 1, \dots, m$. The selected portfolio is represented by the binary decision variable $z = [z_1, \dots z_m]$ such that $z_i = 1$ if and only if project i is selected. The set of portfolios satisfying relevant feasibility constraints (related to, e.g., budget or mutual exclusiveness) is denoted by Z. The true values of the selected projects are denoted by $v = [v_1, \dots, v_m]'$. These values, which may represent projects' net present value, return on investment or multivariate utility, are modelled as a realization of a vector-valued random variable $V = [V_1, \dots, V_m]' \sim f(v)$ where the joint distribution $f(v)$ is assumed to be known.

If the values v were known to the DM, she would choose the optimal portfolio $z(v)$ by solving the optimization problem[2] (Vilkkumaa *et al.*, 2014)

$$z(v) = \arg \max_{z \in \mathbb{Z}} zv. \tag{1}$$

However, when making decision in reality, the DM does not know the true values $v = [v_1, \dots, v_m]'$ but rather the value estimates $v^E = [v_1^E, \dots v_m^E]'$ of the true value v. These value estimates are assumed to be a realization of random variable $(V^E | v) \sim f(v^E | v)$ with a known distribution function $f(v^E | v)$. Here, we assume that there is no systematic bias in these estimates. Technically, this is done by assuming that the estimates are conditionally unbiased so that

---

[2] If there are multiple solutions to the maximization problem, $z(\cdot)$ is selected at random among these solutions.

$$\mathbb{E}[V_i^E|V = v] = \int_{-\infty}^{\infty} v_i^E f(v_i^E|v)dv_i^E = v_i. \tag{2}$$

If the DM were to choose projects based on these value estimates v^E, she would subsequently choose portfolio $z(v^E)$ which satisfies the optimization problem

$$z(v^E) = \arg\max_{z \in \mathbb{Z}} zv^E. \tag{3}$$

To study the average performance of the selection rule in (3) above, we define estimator $V^E = [V_1^E, \dots, V_m^E]$ (random variable) without conditioning on the true value v. Then, $V_i^E \sim f(v^E) = \int_{-\infty}^{\infty} f(v^E|v)f(v)dv$.

It can be shown that the true value of portfolio $z(v^E)$ selected based on value estimates v^E is expected to be lower than its estimated value. This result is formalized in Proposition 1, which was originally presented and proved by Vilkkumaa *et al.* (2014).

Proposition 1 (Vilkkumaa *et al.*, 2014, pp. 774). *Let V^E be a conditionally unbiased estimator of V. Then*

$$\mathbb{E}[z(V^E)V - z(V^E)V^E] \leq 0,$$

where $z(V^E)$ fulfills (2). Moreover, if $\mathbb{P}(z(V) \neq z(V^E)) > 0$, where $z(V)$ fulfills (1), then $\mathbb{E}[z(V^E)V - z(V^E)V^E] < 0$.

Proposition 1 indicates that the value of the chosen portfolio $z(v^E)$ will not be more than its estimated value on average. Moreover, if there exists a probability of selecting *non-optimal* projects, which is highly likely as the estimates are not perfectly accurate, the expected discrepancy between the realized and estimated portfolio values is strictly less than zero. Even if the value estimates are unbiased, the portfolio value will be systematically overestimated so that the DM will experience *post-decision disappointment*. The magnitude of this post-decision disappointment is positively correlated with the uncertainty in value estimates. (Smith & Winkler, 2006; Vilkkumaa *et al.*, 2014).

To mitigate post-decision disappointment, we can revise value estimates with Bayesian methods, i.e. by forming the posterior distribution $f(v|v^E)$ for project values given the estimates. The posterior distribution can be obtained from the prior distribution $f(v)$ and the likelihood distribution $f(v^E|v)$ by using the Bayes' rule: $f(v|v^E) \propto f(v)f(v^E|v)$. The posterior distribution can then be used to determine the expected value $\mathbb{E}[V_i|V^E = v^E]$ for project i given the value estimates, or the probability $\mathbb{P}(z_i(V) = 1|V^E = v^E)$ with which the project i belongs to the optimal portfolio. These expected values are the projects' *Bayesian estimates* $v^B = [v_1^B, \dots, v_m^B]'$, which can be computed from the observed estimates v^E through

$$v_i^B = \mathbb{E}[V_i|V^E = v^E] = \int_{-\infty}^{\infty} v_i f(v_i|v^E) dv_i. \tag{4}$$

The Bayesian estimate for project i is thus the mean of the posterior distribution $f(v_i|v^E)$. The portfolio that maximizes the expected value is now determined by the optimization problem

$$z(v^B) = \arg\max_{z \in \mathbb{Z}} \mathbb{E}[V|V^E = v^E] = \arg\max_{z \in \mathbb{Z}} z v^B. \tag{5}$$

Bayesian estimates shift expectations from the initial estimates v_i^E toward prior value information $f(v)$, more specifically, toward the prior mean μ_i. Consequently, these estimates will yield more conservative and less overestimated forecasts about the true value of the selected portfolio. To study the average performance of the selection rule in (5) above, we define Bayesian estimator $V^B = [V_1^B, \dots, V_m^B]$ (random variable), which is obtained from (4) by replacing the observed estimates v^E with the random variable V^E as follows:

$$V_i^B = \mathbb{E}[V_i|V^E] = \int_{-\infty}^{\infty} v_i f(v_i|V^E) dv_i. \tag{6}$$

Proposition 2 below shows that Bayesian estimates *eliminate* the expected post-decision disappointment. This result is also presented and proved by Vilkkumaa *et al.* (2014). The result is formalized in Proposition 2.

Proportion 2 (Originally as Proposition 3 in Vilkkumaa at al., 2014, pp. 775). *Let V,*
V^E, V^B *and* $z(.)$ *be as in Proposition 1, then*

$$\mathbb{E}[z(v^B)V - z(v^B)v^B | V^E = v^E] = 0$$

for all v^E, *and thus* $\mathbb{E}[z(V^B)V - z(V^B)V^B] = 0$.

Additionally, Vilkkumaa *et al.* (2014) also prove that selected portfolios resulting from
Bayesian estimates will yield at least as much value as the portfolio that maximizes the sum
of value estimates v^E. This result is formalized in Proposition 3.

Proposition 3 (Originally as Proposition 2 in Vilkkumaa *et al.*, 2014, pp. 775*). Let* V^E,
V and $z(V^E)$ *be as in Proposition 1, then*

$$\mathbb{E}[z(V^E)V - z(V^B)V] \leq 0,$$

where V^B *is given by (6) and* $z(V^B)$ *satisfies (5). Moreover, if there exists a possibility that*
$z(v^E) \neq z(v^B)$, *then* $\mathbb{E}[z(v^E)V - z(v^B)V] < 0$ *and therefore* $\mathbb{E}[z(V^E)V - z(V^B)V] < 0$.

As from (5), $z(v^B)$ maximizes the expected portfolio performance, this is an intuitive
outcome. However, if there exists a possibility that the use of Bayesian estimates results in
the selection of a different portfolio than that of the use of the value estimate v^E, the expected
value of the portfolio based on Bayesian estimates is strictly higher.

4.2   Analytic results for normally distributed values and estimate errors

From this point forward, we will assume that the projects' true and estimated values are
normally distributed in order to obtain analytic results. The assumptions are as follows:

Values: $V_i = \mu_i + E_i$, with $E_i \sim N(0, \sigma_i^2)$. Thus, for all V: $V \sim N(\mu, \sigma^2)$.

Estimates: $(V_i^E | V_i = v_i) = v_i + \Delta_i$, with $\Delta_i \sim N(0, \tau_i^2)$. Thus, for all V^E,
              $V^E \sim N(\mu, \tau^2 + \sigma^2)$.

This setting is illustrated in *Figure 5*.
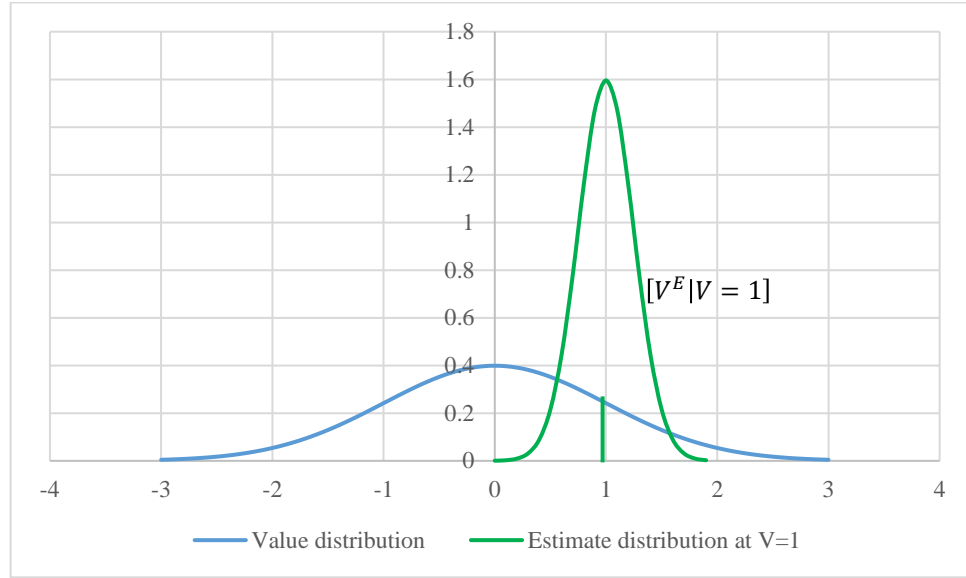
Figure 5: Illustration of value and estimate distributions

Under these assumptions, it can be proven that the Bayesian estimates also follow a normal distribution with the same mean but smaller variance than that of true values. All proofs can be found in Appendix A.

Proposition 4. *Let V, V^E, and V^B be as in Proposition 2 and assume that project values and value estimates follow the normal distribution as above, then we have*

- *Bayesian estimates:* $\quad V^B \sim N(\mu, \frac{\sigma^4}{\sigma^2 + \tau^2})$, *and*

- *The closed form expression for Bayesian estimates:*

$$v_i^B = \alpha_i v_i^E + (1 - \alpha_i)\mu_i, \qquad where\ \alpha_i = \left(1 + \frac{\tau_i^2}{\sigma_i^2}\right)^{-1}.$$

Figure 6 illustrates the normal distribution of true values, value estimates and Bayesian-adjusted estimates. The parameters are set as follows: $\mu = -0.5$, $\sigma = 0.3$, $\tau = 0.4$.
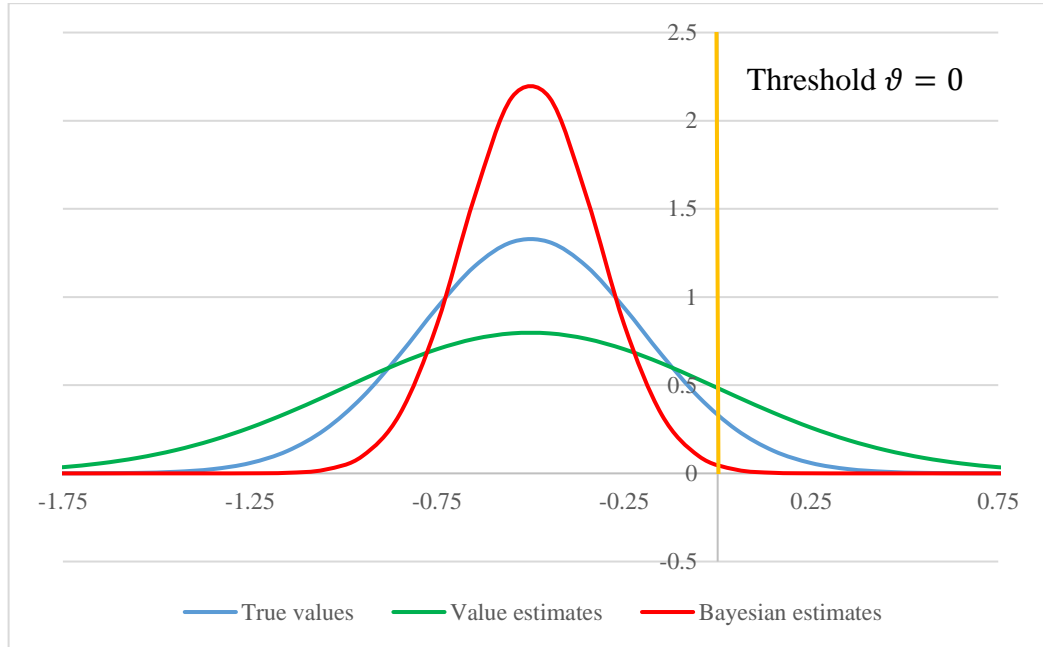
Figure 6: Illustration of value distribution, estimate distribution and Bayesian-adjusted estimate distribution.

By definition, if the values and estimation errors are normally distributed, the Bayesian estimates shift expectations from the initial estimates towards prior mean μ_i. This shift is clearly demonstrated in *Figure 6*. Proposition 4 is consistent with the findings from previous studies on Bayesian estimate, as it proves that the variance of Bayesian-adjusted estimates is lower than that of unadjusted estimates. Alternatively speaking, Bayesian-adjusted estimates are less likely to be extremely high or low.

In the following parts, we will examine the case in which all projects with estimated values above some predetermined threshold ϑ are selected, i.e. $z^E = 1$ if and only if $v_i^E > \vartheta$ where ϑ is a predetermined threshold. Without loss of generality, we will assume that the threshold value is zero ($\vartheta = 0$) if not specified otherwise. The threshold value of zero can be easily observed in practice, as most investors will invest in portfolios that yield positive returns. In addition, to get results that do not depend on the number of proposals, we will assume that this number approaches infinity. In that case, the shares of funded projects based

on true values, value estimates, and Bayesian estimates are $Pr(V \geq \vartheta)$, $Pr(V^E \geq \vartheta)$, and $Pr(V^B \geq \vartheta)$ respectively.

When there are fewer than 50% of the project proposals with truly positive *ex post* performance (e.g, truly positive NPVs), the mean of the three normal distributions is smaller than zero ($\mu < 0$). In this case, the proportion of funded projects is larger than optimal when estimates are used, and smaller than optimal when Bayesian estimates are used. On the contrary, if there are more than 50% of the project proposals with truly positive performance, then the mean $\mu > 0$. In this case, the proportion of funded projects is larger than optimal when Bayesian estimates are used, and smaller than optimal when estimates are used. This result is formalized in Proposition 5. All proofs can be found in Appendix A.

Proposition 5. *Let V, V^E, and V^B be as in Proposition 2 and $\vartheta = 0$.*

- *If $\mu < 0$, then $Pr(V^B > 0) < Pr(V > 0) < Pr(V^E > 0)$, and*
- *if $\mu > 0$, then $Pr(V^B > 0) > Pr(V > 0) > Pr(V^E > 0)$,*

where V^B is given by (6). Moreover, when the threshold is equal to the mean (i.e., $\mu = 0$), then $Pr(V^B > 0) = Pr(V > 0) = Pr(V^E > 0) = 0.5$.

Proposition 5 states that if $\mu < 0$, the proportion of funded projects of a Bayesian DM (i.e. $Pr(V^B > 0)$) is lower than the proportion of funded projects of a non-Bayesian DM (i.e. $Pr(V^E > 0)$). This relation can also be observed from *Figure 6*. As the variance of the Bayesian distribution is lower than that of estimate distribution, if the threshold is to the right of the mean, the proportion of funded projects in Bayesian distribution will be lower than that of estimate distribution because of the fat-tailed distribution of the estimates. On the contrary, if $\mu > 0$, the proportion of funded projects of a Bayesian DM is higher than the proportion of funded projects of a non-Bayesian DM. This relation can be observed from *Figure 7* with the following set of parameters: $\mu = 0.5$, $\sigma = 0.3$, $\tau = 0.4$.
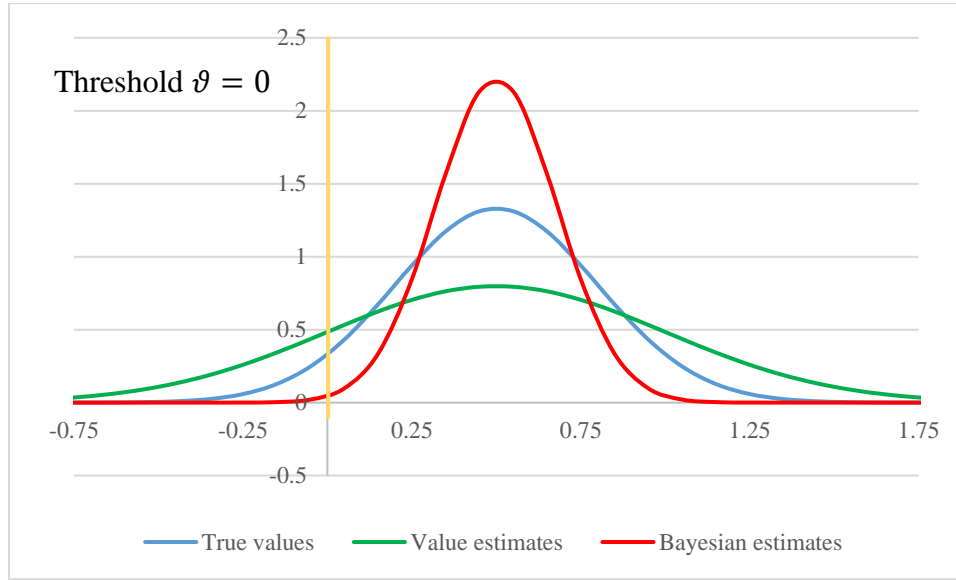
Figure 7: Proportions of funded projects when $\mu > 0$

The standard deviation σ of the normal distribution of true values is called *"signal"* and the standard deviation of the estimate error τ of the true values is called *"noise"*. The signal-to-noise ratio is thus defined as $\frac{\sigma}{\tau}$. With the same distribution of true values (i.e. *"signal"* is fixed), the absolute values of the differences among the proportions of funded projects using true values, value estimates, and Bayesian estimates decrease in $\frac{\sigma}{\tau}$. Simply put, the higher the signal-to-noise ratio (i.e., the more accurate the estimates), the less significant the differences between the proportions of funded projects are. This result is independent of the value of prior mean μ. This result is formalized in Proposition 6 and visualized in ***Figure 8*** (with $\mu < 0$) and ***Figure 9*** (with $\mu > 0$). All proofs can be found in Appendix A.

Proposition 6. *Let V, V^E, and V^B be as in Proposition 2, then*

$$|Pr(V > 0) - Pr(V^B > 0)|, |Pr(V^E > 0) - Pr(V > 0)|, \text{ and } |Pr(V^E > 0) - Pr(V^B > 0)| \text{ will decrease in } \frac{\sigma}{\tau} \text{ with all } \mu.$$

The differences among the proportions of funded projects resulting from true values, value estimates and Bayesian estimates decreases in the signal-to-noise ratio with all values of prior mean μ. In other words, the differences are more pronounced in *noisy* environments.
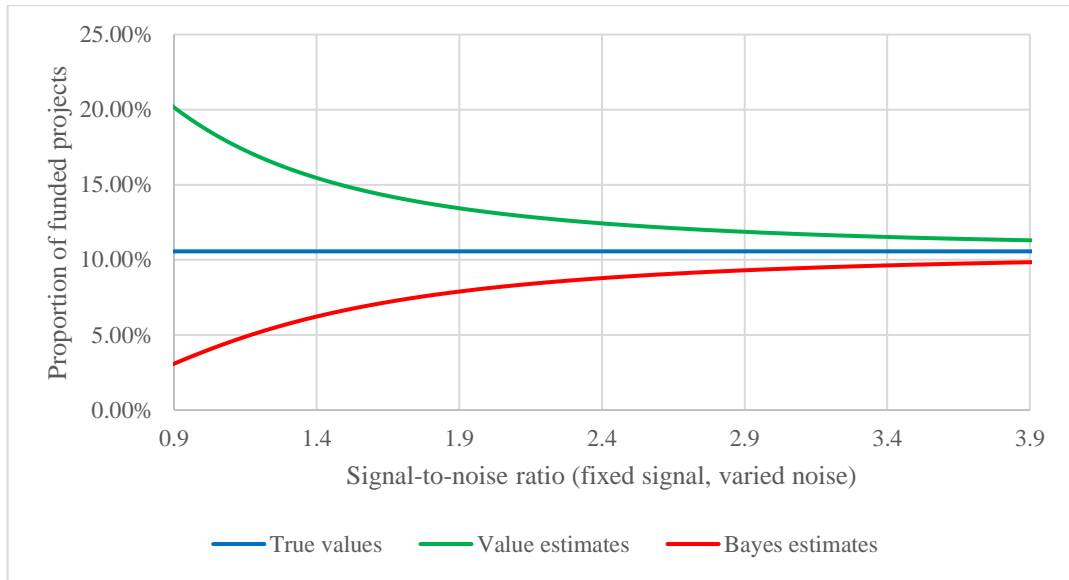
Figure 8: Proportion of funded projects at different signal-to-noise ratios
$(\mu = -5, \sigma = 4, \vartheta = 0)$
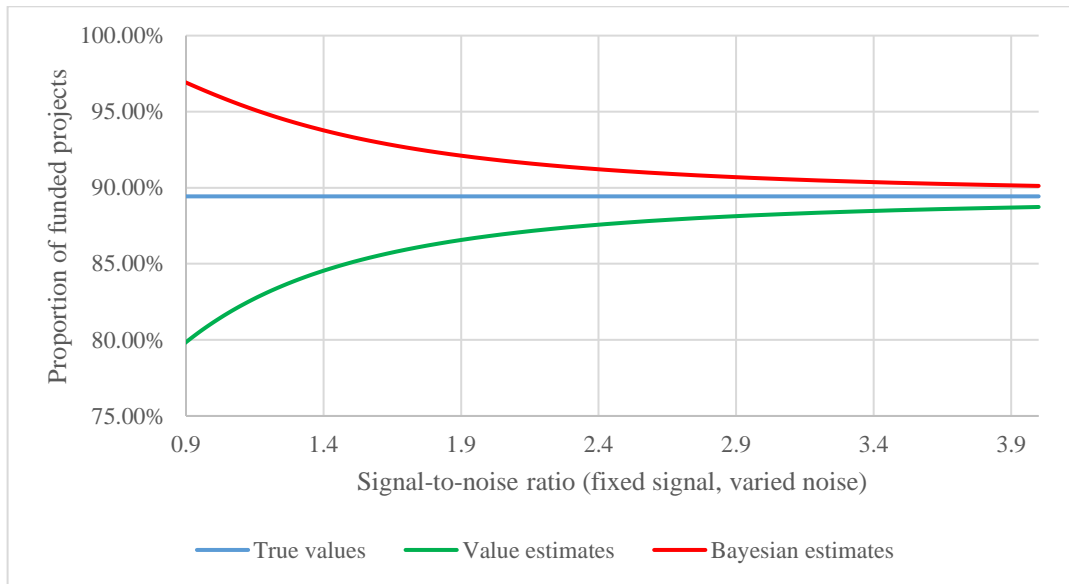
Figure 9: Proportion of funded projects at different signal-to-noise ratios
$(\mu = 5, \sigma = 4, \vartheta = 0)$

Furthermore, when other parameters remain constant, the differences among the Bayesian, optimal, and non-Bayesian proportions of funded projects also vary based on the value of prior mean. ***Figure 10*** illustrates the proportion of funded projects in relation to prior mean.
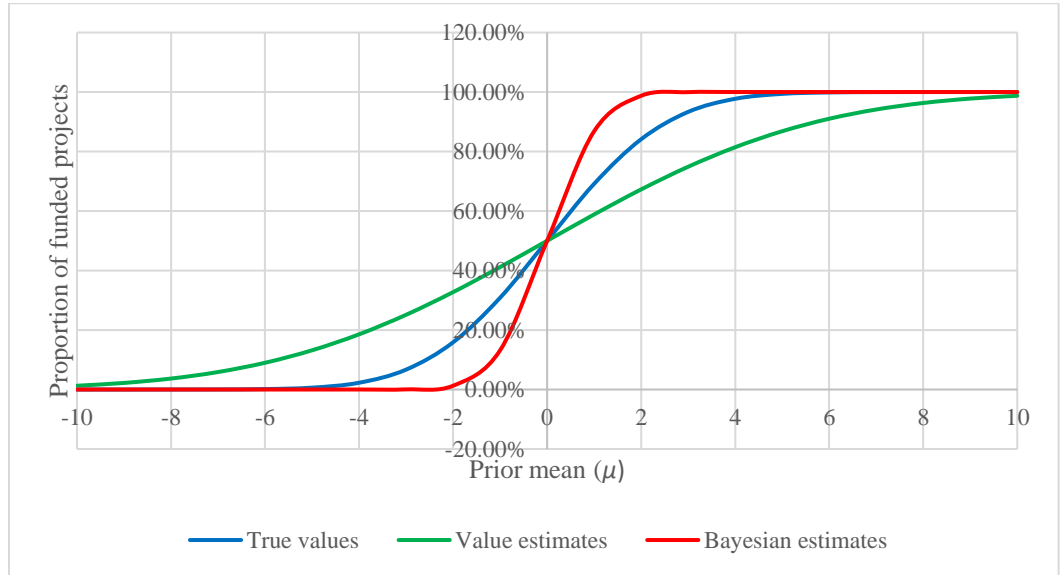
Figure 10: Proportion of funded projects at different values of prior means μ
$(\sigma = 2, \tau = 4, \vartheta = 0)$

Given that the variances and the threshold remain constant, it can be observed that the proportions of funded projects are essentially the cumulative normal distribution functions of V, V^E, and V^B with the moving prior mean being the variable and the threshold being the mean. When prior mean is smaller than the threshold, the proportion of funded projects resulting from value estimates is higher than optimal and the proportion of funded projects resulting from Bayesian estimates is lower than optimal. However, in cases when the proportion of projects with truly positive performance is higher than 50%, the proportion of funded projects resulting from Bayesian estimates is higher than optimal and the proportion of funded project resulting from value estimates is lower than optimal. For instance, with the parameter set as in *Figure 10*, if the prior mean $\mu = -2$, it would be optimal to fund 15.87% of the projects. A Bayesian DM in such situation would only fund approximately 1.27% of the projects, while a non-Bayesian DM would fund 32.74% of the projects.

With the same assumptions as in Proposition 5, the expected value of the portfolio resulting from Bayesian estimates is higher than the expected value of the portfolio resulting from value estimates. All proofs can be found in Appendix A.

Proposition 7. *Let V^E, V, and V^B as in Proposition 2 and $\vartheta = 0$.*

- *If $\mu < 0$, then $\mathbb{E}[V|V^B > 0] \geq \mathbb{E}[V|V^E > 0]$, and*
- *if $\mu > 0$, then $\mathbb{E}[V|V^B > 0] \leq \mathbb{E}[V|V^E > 0]$.*

The equality holds when $\mu = 0$ or $\tau = 0$.

Proposition 7 indicates that with the assumptions above, the expected *ex post* performance of funded projects of a Bayesian DM is higher than that of a non-Bayesian DM when fewer than 50% of the project proposals have truly positive performance. On the other hand, the expected *ex post* performance of funded projects of a Bayesian DM is lower than that of a non-Bayesian DM when more than 50% of the project proposals have truly positive performance. These relations can be observed in *Figure 11* and *Figure 12*.
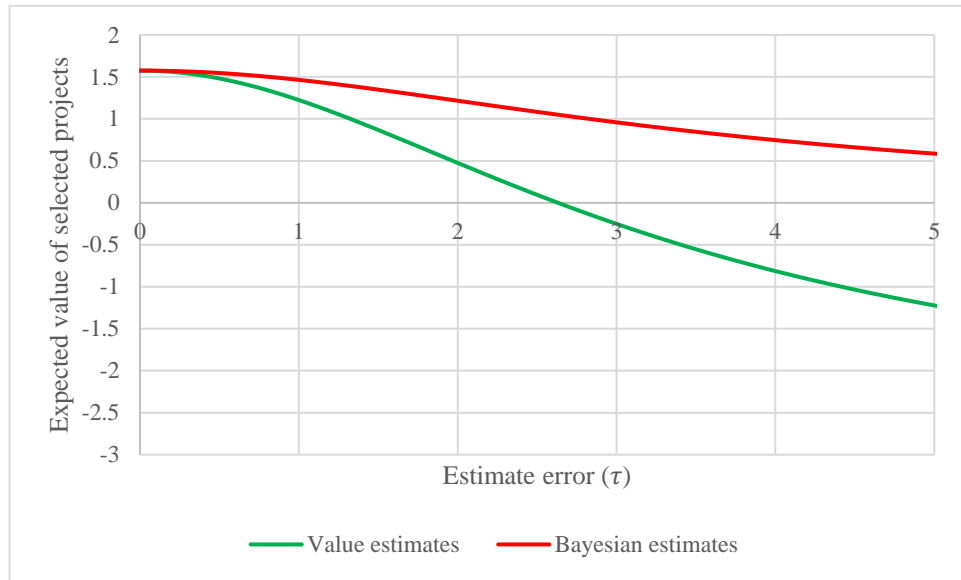
Figure 11: Expected value of selected projects at different estimate errors τ
$(\mu = -3, \vartheta = 0, \sigma = 3)$
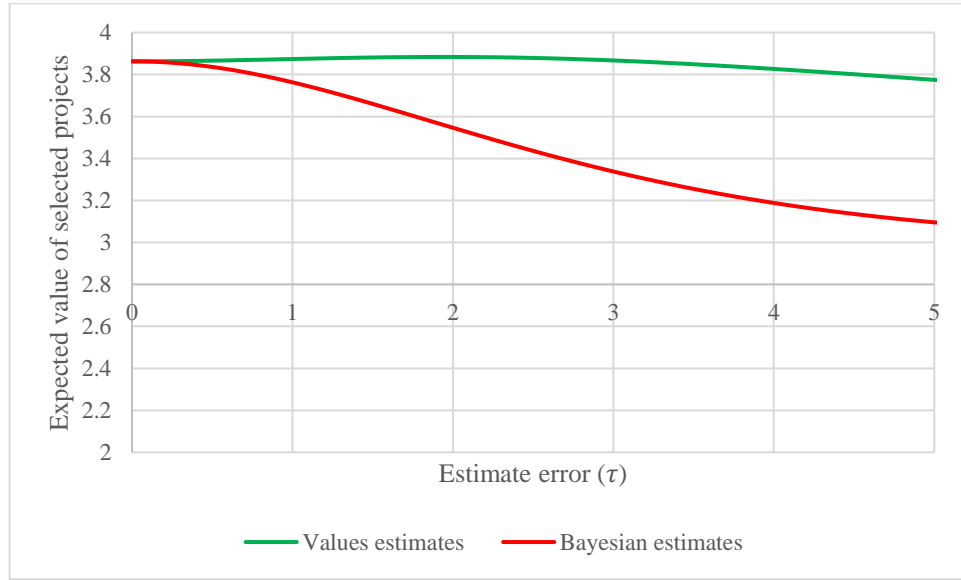
Figure 12: Expected value of selected projects at different estimate errors τ
$(\mu = 3, \vartheta = 0, \sigma = 3)$

It can be observed in both *Figure 11* (with $\mu < 0$) and *Figure 12* (with $\mu > 0$) that the higher the estimate error τ is, the more pronounced the difference between the expected values of selected projects resulting from value estimates and Bayesian estimates is. In other words, the difference increases in noise (i.e. the noisier the environment, the higher the difference), and thus, decreases in signal-to-noise ratio. This phenomenon is formalized in Proposition 8 and can also be observed in *Figure 11* and *Figure 12*. All proofs can be found in Appendix A.

Proposition 8. *Let V, V^E, and V^B be as in Proposition 2, then*

$$|\mathbb{E}[V|V^B > 0] - \mathbb{E}[V|V^E > 0]| \text{ decreases in } \frac{\sigma}{\tau} \text{ with all } \mu.$$

Accordingly, the higher the estimate errors, the more pronounced the difference between the expected value. Conversely, when the estimate error is low (i.e. the signal-to-noise ratio is high), there is a smaller difference between the expected values of Bayesian and non-Bayesian funded projects. This is quite intuitive, as the higher the estimate errors for a given value distribution, the higher the chance of selecting *"wrong"* projects. It is also

consistent with previous finding that the higher the signal-to-noise ratio, the lower the expected post-decision disappointment (see e.g. Harrison & March, 1984; Vilkkumaa *et al.*, 2014), as it can be observed from *Figure 11* and *Figure 12* that both of the expected values decrease in noise (i.e. increase in signal-to-noise).

4.3 Summary of analytic results

Proposition 4 shows that Bayesian-adjusted estimates have a lower variance than unadjusted estimates, indicating that Bayesian-adjusted estimates are less likely to be extremely high or low. Propositions 5 and 7 show that, although the probability of funded projects of a Bayesian DM is lower than that of a non-Bayesian DM, these funded projects of the Bayesian DM yield higher *ex post* performance. Proposition 6 indicates that the differences among the proportions of funded projects based on Bayesian estimates/ true values/ non-Bayesian estimates decrease in signal-to-noise. Proposition 8 indicates that the difference between the expected values of the selected portfolios based on Bayesian and non-Bayesian estimates decreases in signal-to-noise. Proposition 6 and 8 collectively indicate that the differences between Bayesian and non-Bayesian estimates (and their differences compared to the optimal one) are more pronounced in *noisy* environments.

4.4 Simulation Results

In this section, the analytic results from the previous sections are demonstrated and validated with Monte Carlo simulation. The simulation also allows the author to obtain some new results that are too complex to derive from the mathematical model analytically. The simulation is done on Microsoft Excel for portfolio selection from 1000 alternatives. The result of such selection is subsequently simulated 1000 times.

4.4.1 Distributions of true values, value estimates and Bayesian estimates

The Monte Carlo simulation results for portfolio selection from 1000 alternatives are summarized in *Table 1* and illustrated in *Figure 13*. The set of parameters used is ($\mu = -5, \sigma = 5, \tau = 10$).

Table 1: Summary of simulated results of value distribution of 1000 alternatives

		True values	**Estimates**	**Bayesian estimates**
Analytic results	*Average*	*-5.0*	*-5.0*	*-5.0*
	Standard deviation	*5.0*	*11.180340*	*2.236068*
Simulated results	Average	-4.80345	-5.06209	-5.01242
	Standard deviation	4.974985	11.22323	2.244646
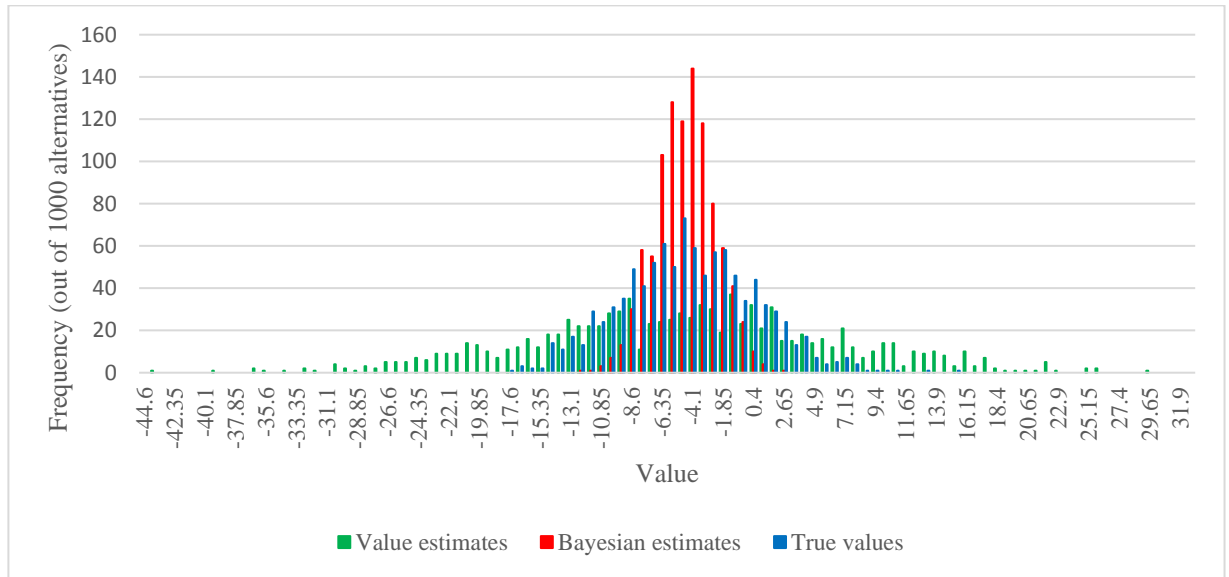
Figure 13: Simulated true value, value estimate, and Bayesian estimate distributions

As we can observe from ***Figure 13***, the variance of Bayesian-adjusted estimates is smaller than the variance of unadjusted estimates. The simulated distributions resemble those of the analytic model. Therefore, it again proves that Bayesian-adjusted estimates are less likely to be extremely high or low.

4.4.2   Proportion of funded projects

The simulation results for the proportion of funded projects resulting from Bayesian estimates, true values, and value estimates are summarized in ***Table 2***. The set of parameters used for this simulation is: $(\mu = -5, \sigma = 5, \tau = 3, \vartheta = 0)$. The difference among the

proportions of funded projects when funding decisions are based on Bayesian-adjusted estimates/ true values/ unadjusted estimates can be visually observed from *Figure 14*.

Table 2: Summary of simulated results of proportions of funded projects

	True values	**Estimates**	**Bayesian estimates**
Analytic results	*15.87%*	*19.56%*	*12.18%*
Simulated average	15.8%	19.6%	12.1%
Standard deviation	1.15%	1.22%	1.02%
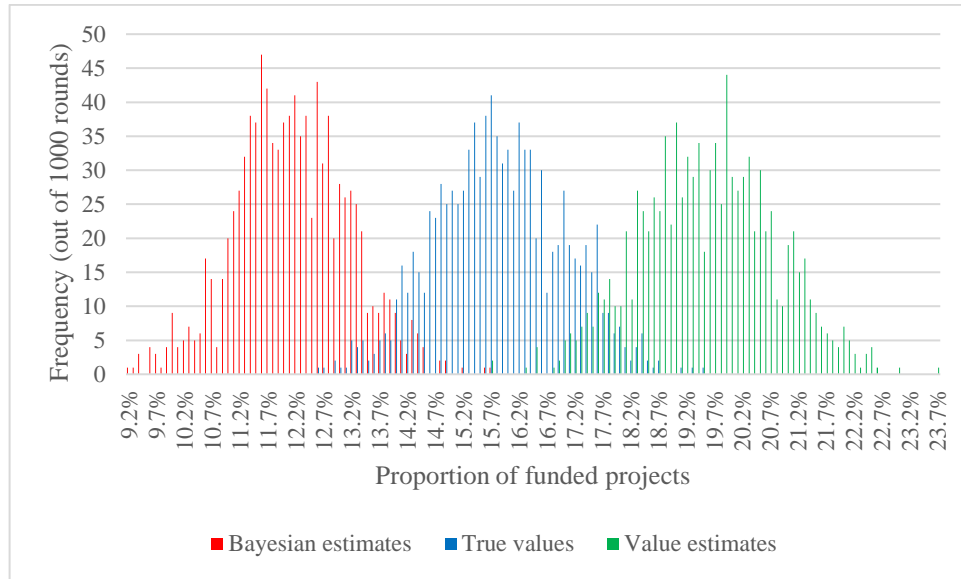
Figure 14: Number of funded projects out of 1000 proposals (Summary from 1000 rounds of Monte Carlo simulation)

Accordingly, the number of funded projects when funding decisions are made based on Bayesian estimates are smaller than optimal, while the number of funded projects based on unadjusted estimates are higher than optimal. These results coincide with the analytic results from Proposition 5. *Figure 15* and *Figure 16* summarize the simulated results for proportions of funded projects at different μ and τ. The results coincide with our analytic results, that if the prior mean $\mu > 0$, the roles of Bayesian DMs and non-Bayesian DMs are reversed, and the noisier the estimates are, the more significant the difference between Bayesian and non-Bayesian DMs, and between them and the optimal one.
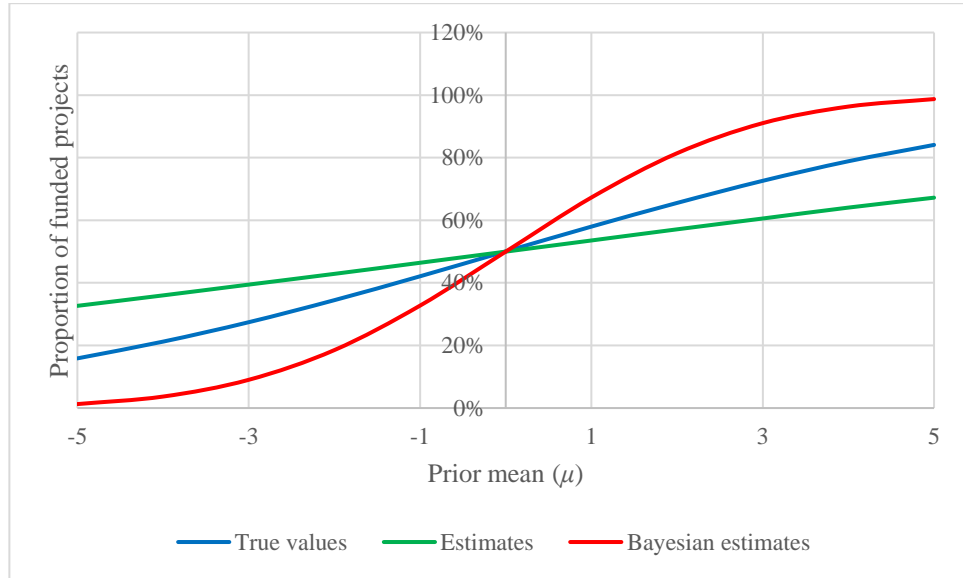
Figure 15: Simulated proportion of funded projects at different prior means μ

$$(\sigma = 5, \tau = 10, \vartheta = 0)$$
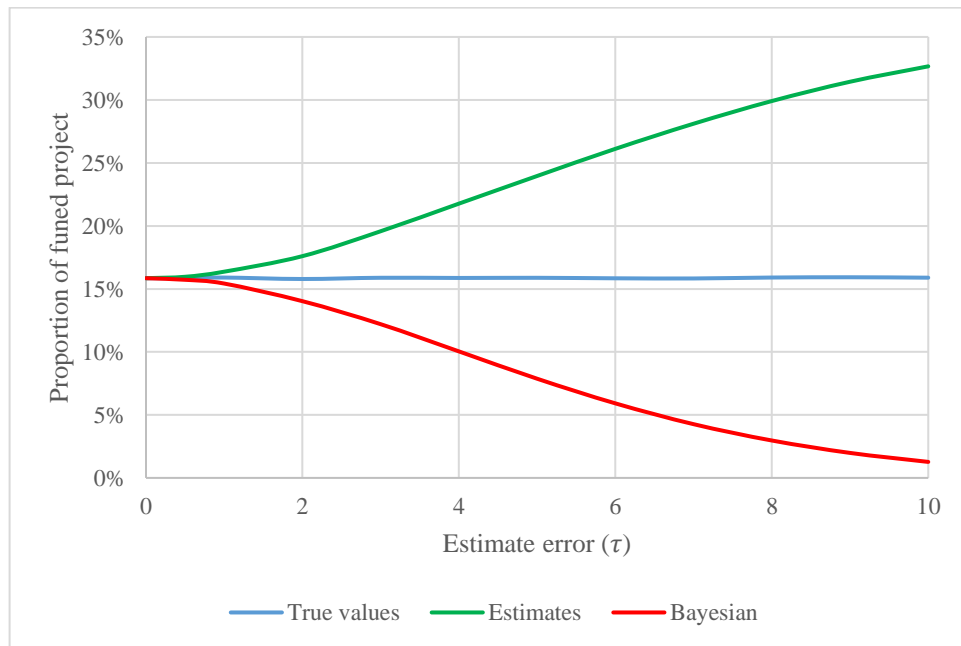
Figure 16: Simulated proportion of funded projects at different estimate errors τ

$$(\mu = -5, \sigma = 5, \vartheta = 0)$$

4.4.3 *Average NPV of funded projects*

The average *ex post* performance of funded projects when using different types of estimates to make funding decisions can also be simulated with Monte Carlo simulation. The set of parameters used is as follows: $(\mu = -5, \sigma = 5, \tau = 3, \vartheta = 0)$. The results are summarized in *Table 3* and visualized in *Figure 17*.

Table 3: Summary of simulated results of expected performance of funded projects

	True values	Estimates	Bayesian estimates
Analytic results	*2.625676*	*1.054875*	*2.116305*
Average	2.622747	1.048881	2.109116
Standard deviation	0.178999	0.235639	0.291269
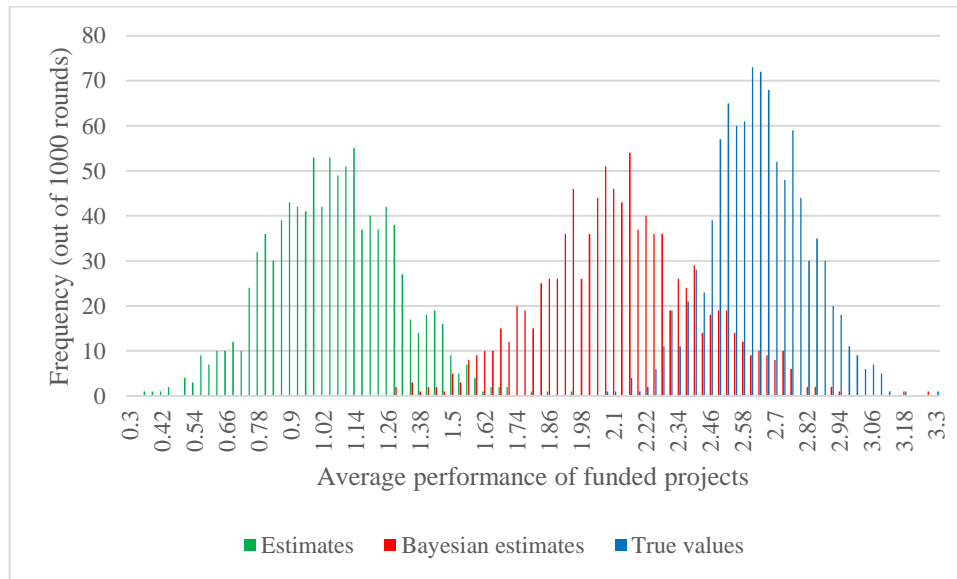
Figure 17: Simulated average performance of funded projects (Summary from 1000 rounds of Monte Carlo Simulation)

It can be seen from the simulated results that although using Bayesian-adjusted estimates will result in a lower proportion of funded projects, Bayesian funded projects will yield higher average *ex post* performance than that of unadjusted estimates. The average *ex post* performance of funded projects resulting from Bayesian decisions is also closer to the optimal *ex post* performance. This perfectly reflects the analytic findings in Proposition 6.

4.4.4   Funded projects that result in loss

This part presents the simulated results on funded projects that result in loss. There are two main aspects that are examined: (i) proportion of funded projects that result in loss and (ii) average loss among funded projects that result in loss.

4.4.4.1   Proportion of funded projects that result in loss

Firstly, to maintain the generalizability, *"loss"* is defined as occasions when the *ex post* performance is lower than the threshold. Thus, if the threshold is zero, "*loss*" would mean having a negative return; this is the most popular definition of loss. The simulated average proportion of funded projects that result in loss (selection out of 1000 proposals, simulated 1000 rounds) is summary in *Table 4* and visualized in *Figure 18*. The set of parameters used in this simulation is ($\mu = -5, \sigma = 5, \tau = 3, \vartheta = 0$).

Table 4: Summary of simulated results of proportion of funded projects that result in loss

	Estimates	Bayesian estimates
Average	7.62%	3.11%
Standard deviation	0.84%	0.54%
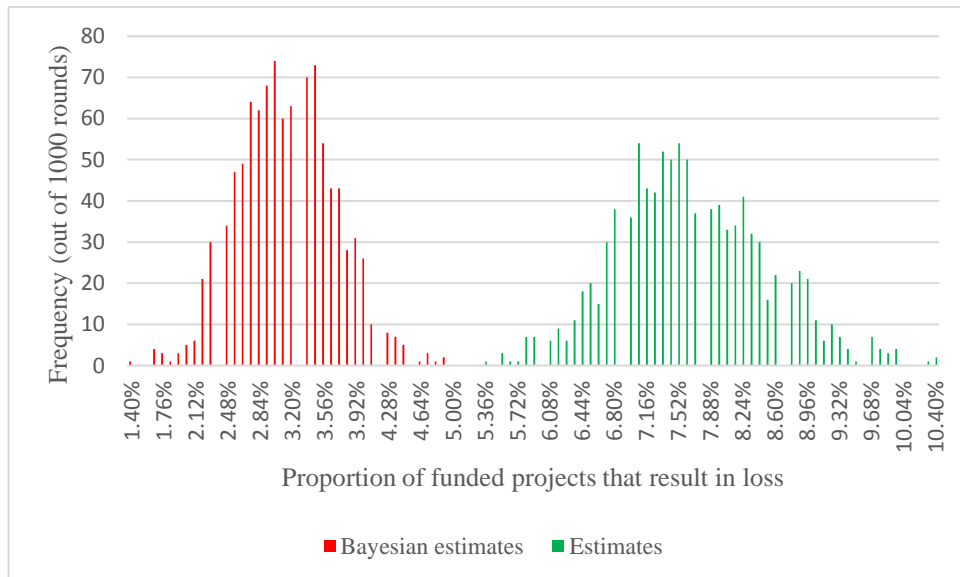
Figure 18: Simulated proportion of funded projects that result in loss

As can be easily observed from the proportion of funded projects that result in loss in the case when $\mu < 0$, which usually is case in practice, the Bayesian-adjusted estimates not only improve the average performance of funded projects, they also alleviate the possibility of having funded projects with lower *ex post* values than the threshold.

This result is unsurprising, as Bayesian model of estimate uncertainties shrinks the value estimates toward prior mean, and in this case prior mean is larger than the threshold, the probability of having a project funded with Bayesian estimates that yields a lower *ex post* performance than the threshold is alleviated. In other words, assuming the threshold is larger than the prior mean, if a Bayesian estimate is larger than the threshold, the probability of its corresponding true value being less than the threshold is lessened.

Figure 19 and *Figure 20* summarize the simulated results for proportion of funded projects at different prior means μ and estimate errors τ. When the prior mean μ is larger than zero, the proportion of funded projects that results in loss of a Bayesian DM is higher than that of a non-Bayesian DM. And the higher the estimate error τ, the more significant the difference between the results of a Bayesian and a non-Bayesian DM.
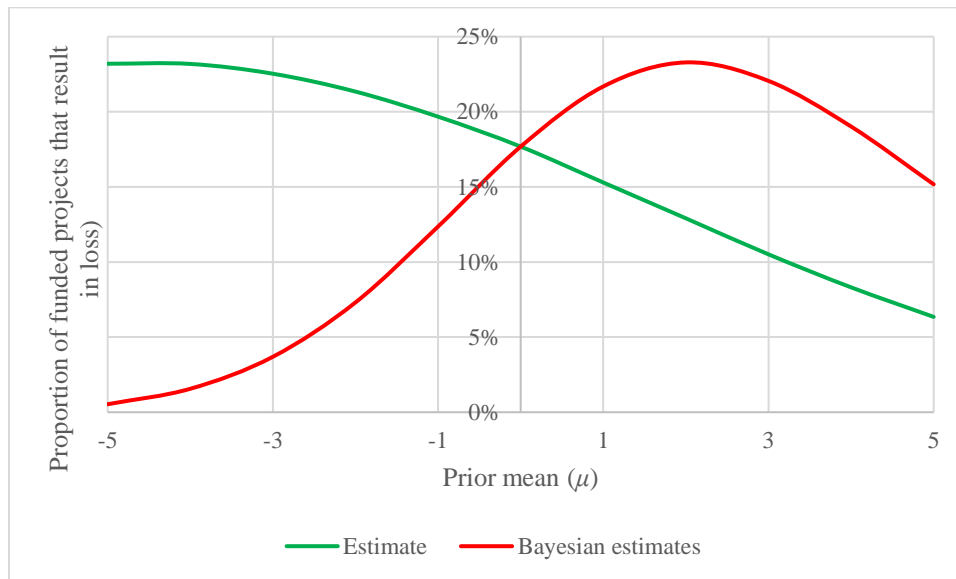
Figure 19: Simulated results for proportion of funded projects that result in loss at different prior means μ ($\sigma = 10, \tau = 5, \vartheta = 0$)
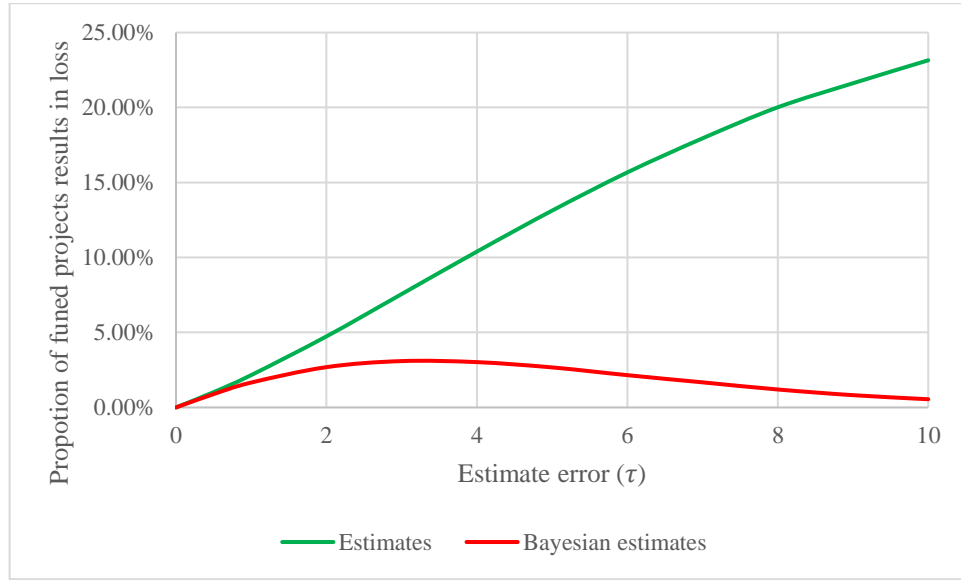
Figure 20: Simulated results for proportion of funded projects results in loss at different estimate error τ $(\mu = -5, \sigma = 5, \vartheta = 0)$

4.4.4.2   Average loss among funded projects that result in loss

Using the same method as the method Smith & Winkler (2006) use to calculate the average post-decision disappointment, we can calculate the simulated results for average loss among n funded projects that result in loss as follows:

$$Average\ loss = \frac{\sum_{j=1}^{n} \dfrac{v_j - \vartheta}{\sigma}}{n} = \frac{\sum_{j=1}^{n} v_j - \vartheta}{n\sigma}$$

Accordingly, the magnitude of loss for each of the selected projects that results in loss is presented as percentage of true values' standard deviation. The simulated results are summarized in **Table 5** and visualized in **Figure 21**. The set of parameters used in this simulation is: $(\mu = -5, \sigma = 5, \tau = 3, \vartheta = 0)$

Table 5: Summary of simulated results of average loss among funded projects that result in loss

	Estimates	Bayesian estimates
Average	-42.01%	-34.76%
Standard deviation	3.92%	5.17%
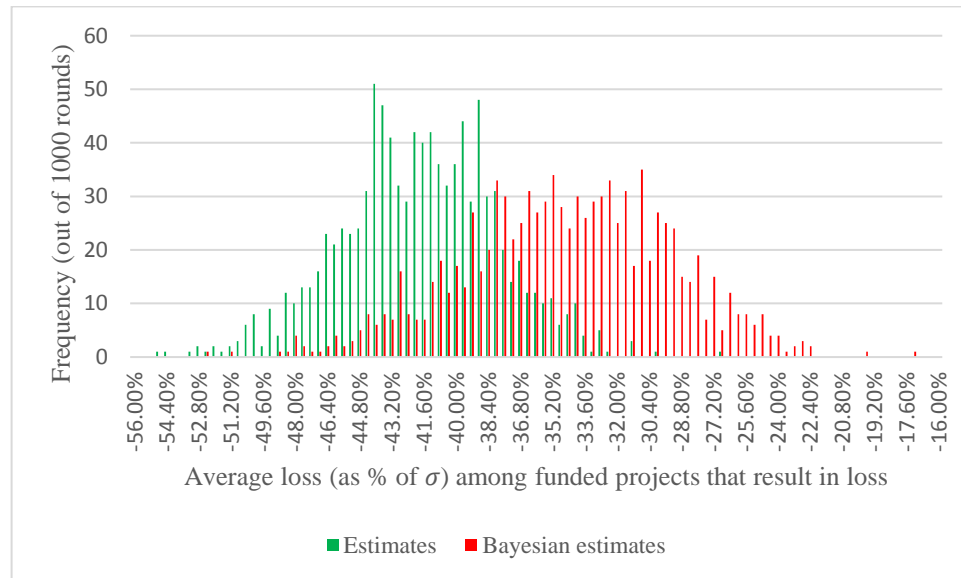
Figure 21: Simulated results for average loss as percentage of σ among funded projects that result in loss

Figure 21 illustrates the simulated average loss among funded projects that result in loss as percentage of the true values' standard deviation from a Bayesian portfolio and a non-Bayesian portfolio. The simulated result shows that when $\mu < 0$, not only the Bayesian portfolio has smaller proportion of funded projects that result in loss, those lost projects have lower average loss than that of a non-Bayesian portfolio. Bayesian estimates therefore decrease the proportion of funded projects that result in loss and simultaneously decrease the average loss among those projects when $\mu < 0$.
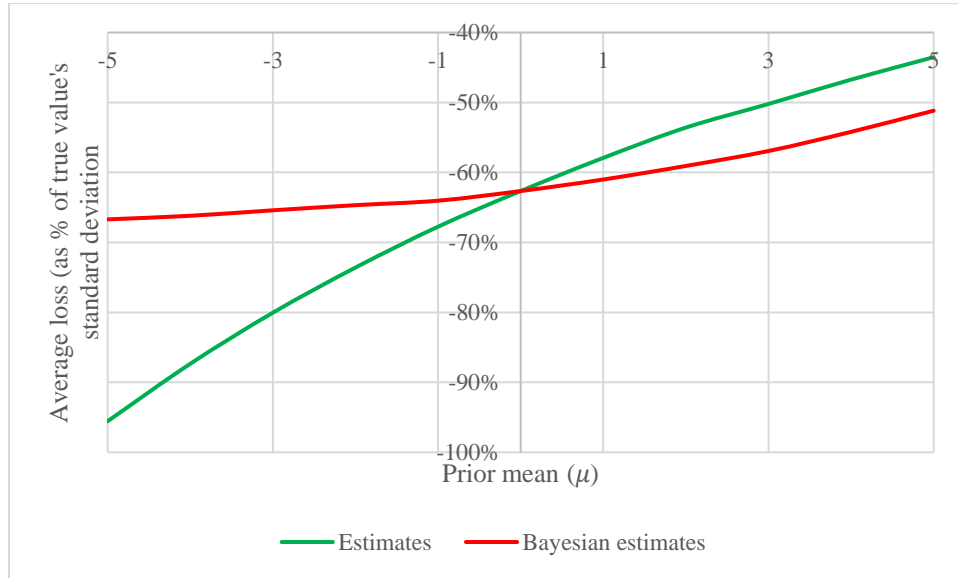
Figure 22: Simulated results for average loss at different prior means μ

$(\sigma = 5, \tau = 10, \vartheta = 0)$
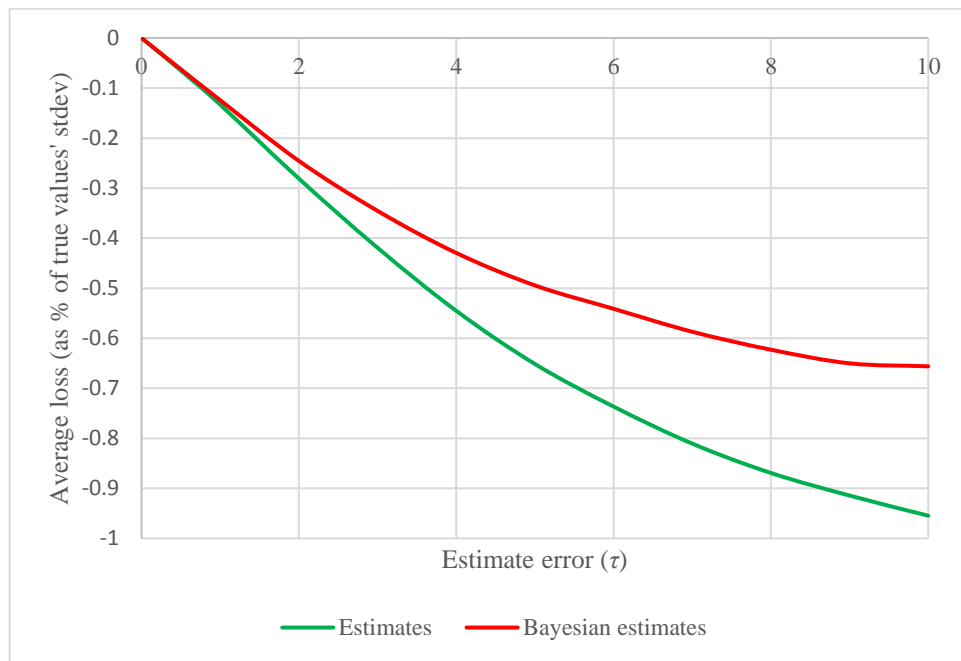
Figure 23: Simulated results for average loss at different estimate errors τ

$(\mu = -5, \sigma = 5, \vartheta = 0)$

Figure 22 and *Figure 23* summarize the simulated results for average loss at different prior means μ and estimate errors τ. Accordingly, when $\mu > 0$, the relation between the average loss of a Bayesian DM and that of a non-Bayesian DM is reversed. Accordingly, when $\mu > 0$, the projects funded by a Bayesian DM will have higher probability of resulting in loss and have higher average loss. And the higher the estimate error τ is (i.e. the lower the estimation accuracy is), the higher the difference between the losses is. These simulated results summarized in *Figure 18* to *Figure 23* validate our Hypothesis 3.

4.4.5   Post-decision surprise

Post-decision surprise indicates the difference between (Bayesian/non-Bayesian) value estimates and their corresponding true values. Negative post-decision surprise is called post-decision disappointment. Post-decision disappointment occurs when an *ex post* value is smaller than its estimate. To calculate post-decision surprise, we use Smith & Winkler (2006) method and calculate the difference between (Bayesian/non-Bayesian) estimates and their corresponding true values among the funded projects as percentage of the true values' standard deviation. The method aims to maintain the comparability between different value distributions thus it calculates post-decision surprise in a relative term. More specifically:

$$Post\text{-}decision\ surprise = \frac{v_i - v_i^E}{\sigma}$$

If the result is negative, post-decision surprise is post-decision disappointment. On the other hand, if the result is positive, it would be a positive surprise. The results for post-decision surprise for both Bayesian and non-Bayesian estimates are summarized in *Table 6*. *Figure 24* visualizes the post-decision surprise for Bayesian estimates and non-Bayesian estimates. The set of parameters used in this simulation are ($\mu = -5, \sigma = 5, \tau = 1, \vartheta = 0$).

Table 6: Summary of simulation results of post-decision surprise

	Estimates	**Bayesian estimates**
Average	-5.89%	0.03%
Standard deviation	1.59%	1.62%
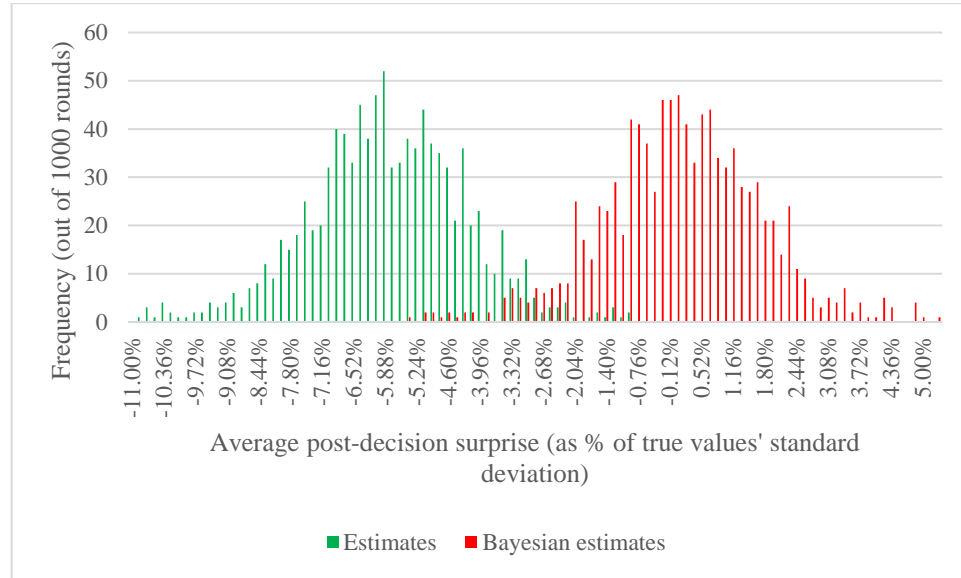
Figure 24: Simulated post-decision surprise (as percentage of true values' standard deviation) (Summary of 1000 rounds of Monte Carlo simulation)

Funding decisions based on non-Bayesian estimates, on average, result in a post-decision disappointment. This result coincides with earlier literature (e.g. Harrison & March, 1984; Vilkkumaa *et al.*, 2014). On the contrary, funding decisions based on Bayesian estimates result in, on average, zero surprise, as reflected in the simulated results. Such results can be explained by the definition of Bayesian modeling of estimate uncertainties. Accordingly, Bayesian modeling shrinks value estimates toward the prior mean thus eliminates the effects of estimate errors when averaged over a large number of decision processes.

4.4.6   *Funded projects that result in high gain*

As from Proposition 4, Bayesian-adjusted estimates have smaller variance than unadjusted estimates. They are thus less likely to be extremely high or low. Consequently, while Bayesian-adjusted estimates help eliminate post-decision disappointment, they also have lower the probability of identifying projects with extremely high gain. *Table 7* summarizes the simulation results of the percentage of the 50 projects among the truly best 5% get funded. Accordingly, the average percentage of Bayesian projects that yield *ex post* values higher than 95-percentile value is 84.16%, while the number of non-Bayesian projects that yield *ex*

post values higher than 95-percentile value is 93.84%. ***Figure 25*** only visualizes the frequency distribution of the percentages of the 50 truly best projects get funded based on Bayesian estimates from 1000 rounds of simulation. The set of parameters used in this simulation are $(\mu = -5, \sigma = 5, \tau = 5, \vartheta = 0)$.

Table 7: Summary of simulation results of percentage of the 50 projects among the truly best 5% get funded

	Estimates	Bayesian estimates
Average	93.84%	84.16%
Standard deviation	3.54%	5.73%
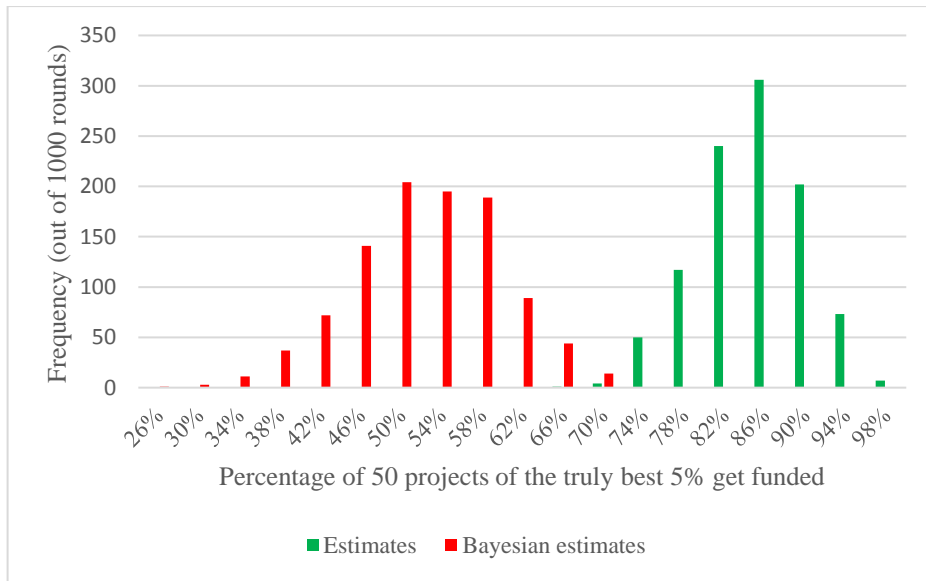
Figure 25: Percentage of the 50 projects of the truly best 5% get funded

Similarly, ***Table 8*** summarizes and ***Figure 26*** illustrates the simulated results of the percentages of the 10 projects among the truly best 1% get funded based on Bayesian and non-Bayesian estimates. The average percentage of the 10 projects among the truly best 1% get funded based on Bayesian estimates is 62.62%, while that of non-Bayesian estimate is 93.94%. The set of parameters used in this simulation is $(\mu = -5, \sigma = 5, \tau = 5, \vartheta = 0)$

Table 8: Summary of simulated results for percenatge of the 10 projects among the truly best 1% get funded

	Estimates	Bayesian estimates
Average	93.94%	72.40%
Standard deviation	8.05%	14.44%
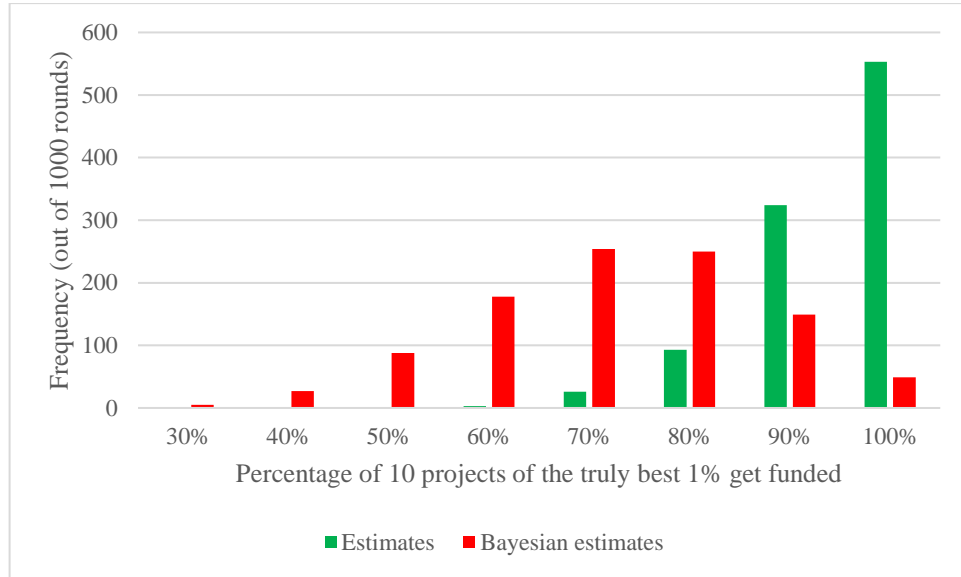
Figure 26: Simulated results for the percentage of the 10 projects of the truly best 1% get funded

The simulated results show that, although using Bayesian-adjusted estimates alleviates post-decision disappointment, it also decreases the probability of selecting projects with high gains. The simulated results validate our Hypothesis 4.

Figure 27 summarizes the simulated results for proportions of funded projects that result in high gains (truly best 5%) when more than 50% of the project proposals have truly positive performance at different estimate errors. According to the simulated results, when the estimate error τ is small in relation to the distance between the threshold and the prior mean, the proportions of funded projects that result in high gains are equal between a Bayesian DM and a non-Bayesian DM at approximately the optimal level (i.e. 100% of the

truly best projects get funded). However, when the estimate error τ is large in relation to the distance between the threshold and the prior mean, a Bayesian DM has higher proportion of funded projects that results in high gains.

Figure 27: Simulated results for proportions of funded projects resulting in high gains when $\mu > 0$ at different estimate errors τ ($\mu = 5, \sigma = 5, \vartheta = 0$)

4.4.7 *Expected portfolio values*

Thus far, we have found that when there are fewer than 50% of project proposals with truly positive performance, a Bayesian DM funds fewer but on average better projects. The results are reversed when there are more than 50% of project proposals with truly positive performance. Therefore, it is interesting to study whether the benefit of funding better projects on average offsets the harm of funded fewer projects by comparing the overall portfolio values. Portfolio values of a Bayesian and a non-Bayesian DM are the products of the proportions of funded projects and the expected values of funded projects: $\Pr(V^B > 0) \cdot \mathbb{E}[V|V^B > 0]$ and $\Pr(V^E > 0) \cdot \mathbb{E}[V|V^E > 0]$. Based on our simulations, the expected portfolio value of a Bayesian DM is higher than that of a non-Bayesian DM in both scenario both when $\mu > 0$ and when $\mu < 0$. The simulated results are demonstrated in *Figure 28* and *Figure 29*.
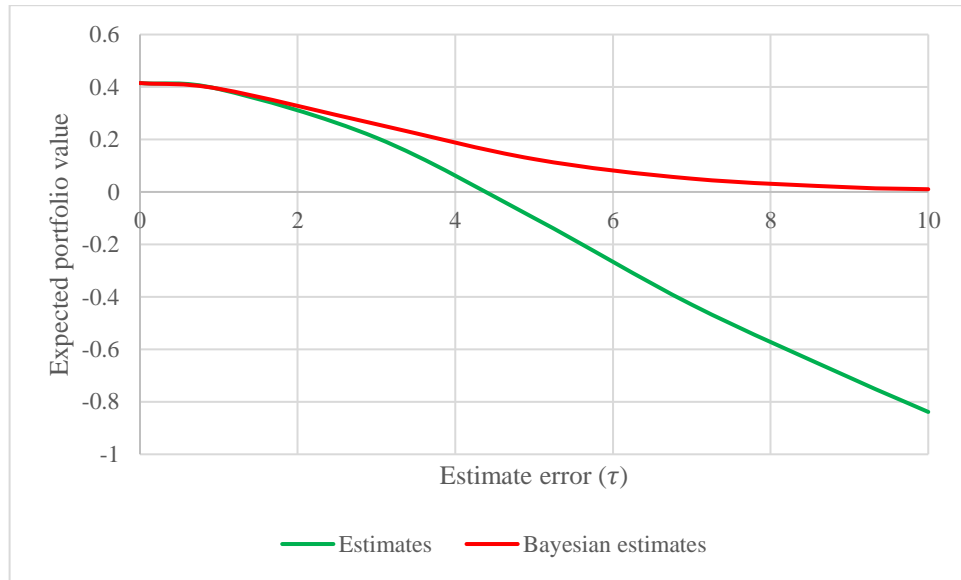
Figure 28: Simulated results for expected portfolio values when $\mu < 0$ at different estimate errors τ ($\mu = -5, \sigma = 5, \vartheta = 0$)
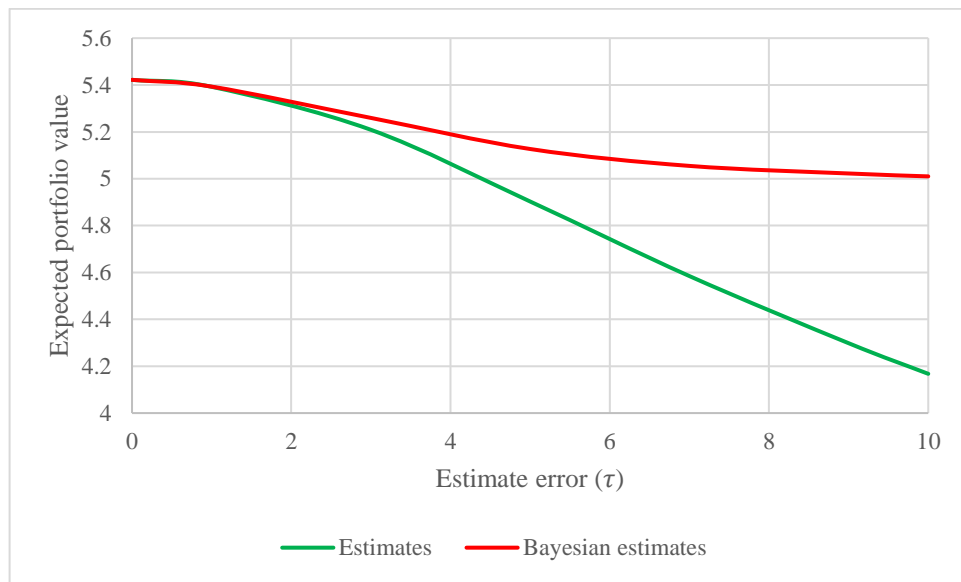
Figure 29: Simulated results for expected portfolio values when $\mu > 0$ at different estimate errors τ ($\mu = 5, \sigma = 5, \vartheta = 0$)

4.4.8   Summary of simulated results

The simulated results testify our previous analytic results. Moreover, it provides further insights on the differences between a Bayesian and a non-Bayesian DM in terms of the proportion of projects results in loss, the proportion of projects that results in high gains, and how these proportions change at different levels of estimation accuracy and the proportion of projects with truly positive performance.

More specifically, when fewer than 50% of the project proposals have truly positive performance, projects funded by a Bayesian DM have lower probability of incurring loss than those funded by a non-Bayesian DM. The average loss incurred among Bayesian projects is also lower than that among non-Bayesian projects. Subsequently, it is also observed that the expected post-decision disappointment among non-Bayesian projects is higher than that among Bayesian projects, which is approximately zero. When more than 50% of the project proposals have truly positive performance, the relation between a Bayesian DM and a non-Bayesian DM is reversed.

Furthermore, when fewer than 50% of the project proposals have truly positive performance, projects funded by a Bayesian DM have lower probability of incurring loss than those funded by a non-Bayesian DM. When more than 50% of the project proposals have truly positive performance, the proportions of funded projects that results in high gains for a Bayesian and a non-Bayesian DM have two possible outcomes: (i) they can be equal at approximately the optimal amount (i.e. 100% of the truly best 5% and 1% projects get funded) if the estimate error is small in relation to the distance between the threshold and the prior mean, or (ii) a Bayesian DM will hav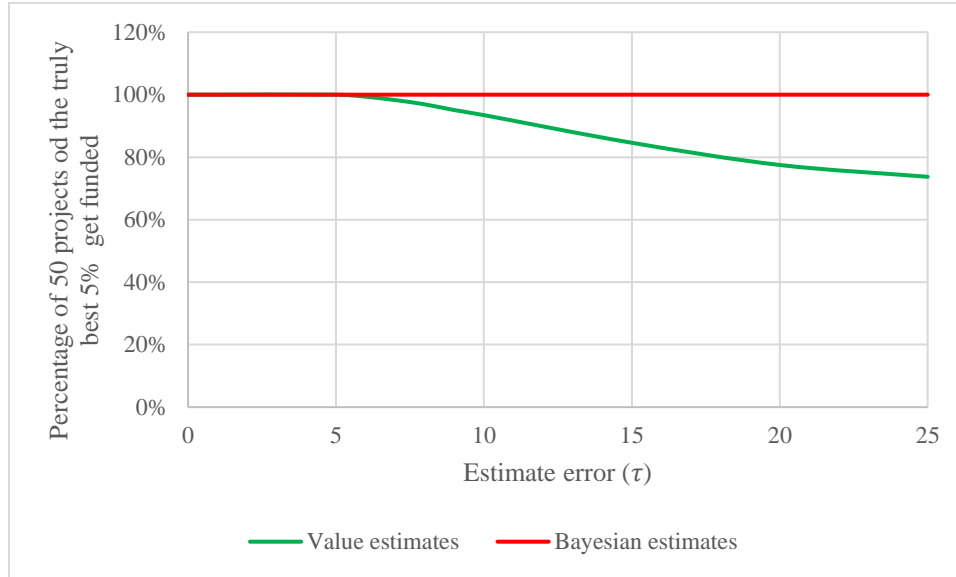e a higher proportion of funded projects that results in high gains if the estimate error is high in relation to the distance between the threshold and the mean. The expected portfolio value of a Bayesian DM is higher than that of a non-Bayesian DM in both scenarios. The analytic and simulated results are summarized in *Table 9*.

Table 9: Summary of analytic and simulated results

	$\mu < 0$	$\mu > 0$
Bayesian	- Underinvested - Higher average performance - Lower probability of loss - Lower probability of high gains - Higher expected portfolio value	- Overinvested - Lower average performance - Higher probability of loss - Equal/higher probability of high gain - Higher expected portfolio value
Non-Bayesian	- Overinvested - Lower average performance - Higher probability of loss - Higher probability of high gains - Lower expected portfolio value	- Underinvested - Higher average performance - Lower probably of loss - Equal/lower probability of high gains - Lower expected portfolio value

The analytic and simulated results in *Table 9* outline the differences between a Bayesian and a non-Bayesian DM in both scenarios when there are more and fewer than 50% of the project proposals with truly positive performance. However, in practice, it is highly unlikely to have more than 50% of the projects proposals with truly positive performance. Even in such cases, due to the scarcity of resources, the DMs will increase the threshold ϑ above the prior mean μ. In such cases, the results for $\mu < 0$ still holds true. Owing to its broader practical applications, the following parts will only examine the case when $\mu < 0$.

5   Empirical Evidence

This chapter serves to explore the characteristics of the decision-making environment of the pharmaceutical industry using earlier literature and industry reports. From such understanding, we will devise possible theoretical outcomes with regard to whether the DM is a Bayesian or a non-Bayesian DM. Subsequently, will observe empirically the outcomes of R&D investments in the pharmaceutical industry, reflect them to the theoretical outcomes, and categorize them in terms of Bayesianess. We will then analyze the decision-making behavior of DMs in this industry also in terms of Bayesianess and, finally, conclude whether or not the Bayesianess of the DMs is reflected in their investment outcomes.

5.1   Description of project portfolio selection in the pharmaceutical industry

Project portfolio selection criteria in the pharmaceutical industry are usually financial measurements, such as expected net present value (ENPV). The constraints in optimizing pharmaceutical portfolio are great in number, ranging from the portfolio level (e.g. budgets), to the individual component level (e.g. regulatory, payers) (Antonijevic, 2014). Furthermore, as the development of new drugs is long, costly, and risky, and R&D investment decisions are of high strategic significance (Miller, 2005).

The pressure on the pharmaceutical industry to improve the cost effectiveness, productivity, and quality of their product development has been increasing in the past few years. Such increasing pressure stems from the increasing cost and diminishing returns. The cause of such issue is the drug developers' false focus on the cost and speed of development. The most important parameter, the Probability of Success (PoS), has been overlooked (Antoijevic, 2014). With such behavior, drugs were developed as if the success was pending. Companies not only assume that their drug will be succeed, but they also assume their competitors will succeed with certainty (Antoijevic, 2014).

Besides, there are numerous challenges that pharmaceutical companies need to face with, ranging from regulatory challenges (e.g. new safety and efficacy requirement that will result in additional clinical trials), delay of submission, increased difficulty to get regulatory approval, to challenges caused by increasing costs of drug development. In addition, new drugs face more competition and possess less flexibility for extending patent times (Antoijevic, 2014).

The pharmaceutical industry counts on innovations to overcome such challenges. There are two main types of innovation: radical innovations and incremental innovation. A radical innovation generates major changes in technology, including the discovery of new knowledge, substantial technical risk, time and cost, while an incremental innovation generates minor changes to existing technology involving small advances based on an established foundation of knowledge (Roussel *et al.*, 1991). In the case of the pharmaceutical industry, new drugs would be considered as radical innovations. More precisely, these new drugs are defined as *"new Chemical Entities (NCEs), […] which, in most cases, represent significant therapeutic advances"* (Cool, 1985, pp. 250). They are also defined by the Food and Drug Administration (FDA) as *"those products representing new chemical structure never previously available to treat a particular disease"* (Pharmaceutical Manufacturers Association, 1989, pp. 22). On the other hand, incremental innovations, according to FDA criteria, are drug enhancements which involve combinations of existing drugs, new dosage forms, new indication, and formula changes. Similar definitions of radical and incremental definition in pharmaceutical industry are also coined by Abernathy & Clark (1985), Banbury & Mitchell (1995), Freeman (1982), and Roussel *et al.* (1991).

The degree of radicalness in the output of R&D activities, in general, can vary significantly. Unlike mechanical assembled product industries (e.g., disk drives, mainframe computers, automobiles, etc.) which have complex systems and numerous components, the pharmaceutical industry's core product concentrates on a molecule (Henderson, 1994). Consequently, the degree of radicalness of drug innovation is a function of new technological and scientific embedded in the drug (Abernathy & Clark, 1985).

5.2   Decision-making environment in the pharmaceutical industry

In this part, the characteristics of the decision-making environment in the pharmaceutical industry will be analyzed and reflected through the lens of project portfolio selection theory. According to our analyses, decision-making environment in the pharmaceutical industry has low estimation accuracy and high threshold.

5.2.1   Low estimation accuracy

Based on existing literature and industry reports, the pharmaceutical industry displays characteristics of a decision-making environment with low estimate accuracy. The characteristics include: (i) lack of quantitative support, (ii) decentralization of decision-making processes, (iii) information asymmetry, and (iv) high degree of disruption.

5.2.1.1   Lack of quantitative support for decision making

Although the need for a more quantitative decision making based on measurable parameters has been recognized in this increasingly challenging environment, the current process of decision making in most pharmaceutical companies seriously lacks quantitative support. According to Antoijevic (2014), it is due to three main reasons: (i) decision analysis is not sufficiently utilized, (ii) inadequate use of statistical resources, and (iii) lack of utilization of modeling and simulation. The pharmaceutical industry is, in fact, far behind other major industries in terms of utilization of quantitative methods as a basis of decision-making process. One possible explanation for a lack of scientific approach in decision making in this industry is that in the past, getting marketing approvals was less challenging while revenues for approved drugs were enormous. Consequently, the profits in those pharmaceutical companies were large, and DMs (e.g. executives) did not have the incentive to change in their decision-making process. Another frequently mentioned explanation is that pharmaceutical industry executives are not quantitative people, hence they are unfamiliar with and reluctant to base decision on quantitative method or simulation outputs. (Antoijevic, 2014). ***Figure 30*** summarizes the causes of such lack of quantitative decision-making method.
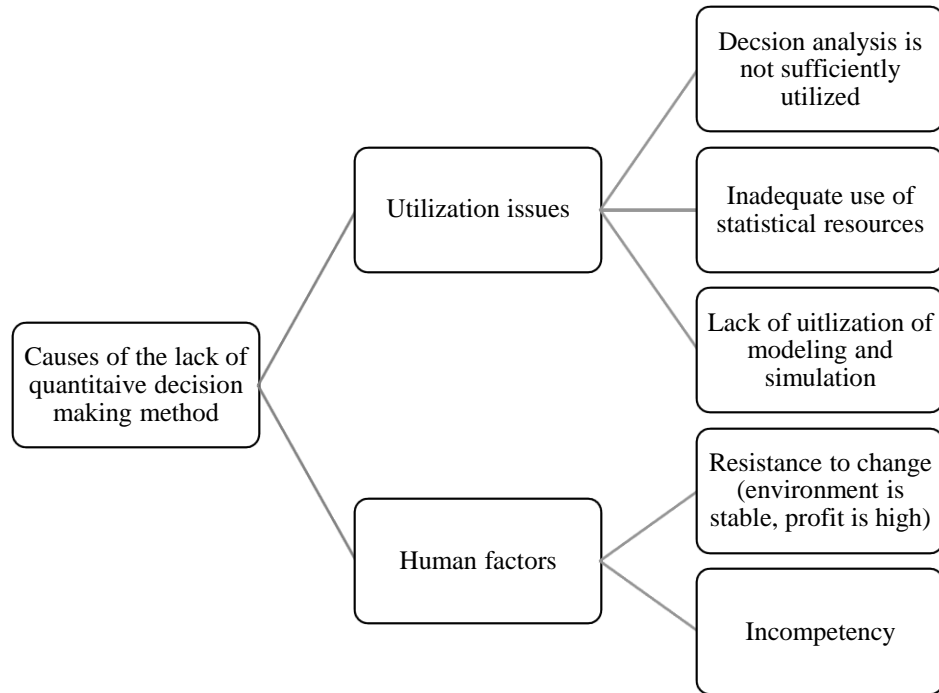
Figure 30: Causes of the lack of quantitative project portfolio selection decision making in the pharmaceutical industry (Antoijevic, 2014)

As statistical resources are not sufficiently utilized, the estimate errors in this environment is expected to be high. This observation is validated by Kola & Landis (2004) in their empirical study.

5.2.1.2 Decentralization of decision-making processes

The process of decision making is *"siloed"* within individual departments and is handled by only executive members of these departments (Antoijevic, 2014). Thomas (2011) reveals that such decentralization creates a misalignment between the firms' strategies and the managers' preference, or even conflicts of interest, as the decentralization distances the decision-making points from the strategic headquarters of the organizations. Such strategic misalignment increases the estimate errors, or even potential systematic bias caused by strategic misinterpretation, in the decision-making environment and thus decreases the estimation accuracy.

5.2.1.3   Information asymmetry

Schlapp *et al.* (2015) suggest that there exist information asymmetry and the lack/absence of information sharing during product evaluation across different departments. The decentralized organizational structure also creates a favorable environment for information asymmetry, as it extends the information flow back and forth. This information asymmetry reflects a low signal-to-noise ratio, in which the estimate errors are significant compared to the standard deviation of the true values (Antoijevic, 2014).

5.2.1.4   High degree of disruption

According to Christensen *et al.* (2009), the pharmaceutical industry is highly disruptive. In such disruptive environment, the ability to accurately forecast future performance of a given project is highly limited. As a consequence, estimate errors are expected to be high.

5.2.2   High threshold

Besides having a *noisy* decision-making environment, the pharmaceutical industry also has a relatively low approval rate in R&D portfolio investment decision (Kola & Landis, 2004, Antoijevic, 2014). This means that very few projects are approved. Hence, one can assume the threshold is high compared to the prior mean.

The overall characteristics of the decision-making environment and their corresponding reflections to portfolio selection theory are illustrated below. *Figure 31* also shows the corresponding theoretical investment outcomes of Bayesian and non-Bayesian DMs according to our theoretical findings.
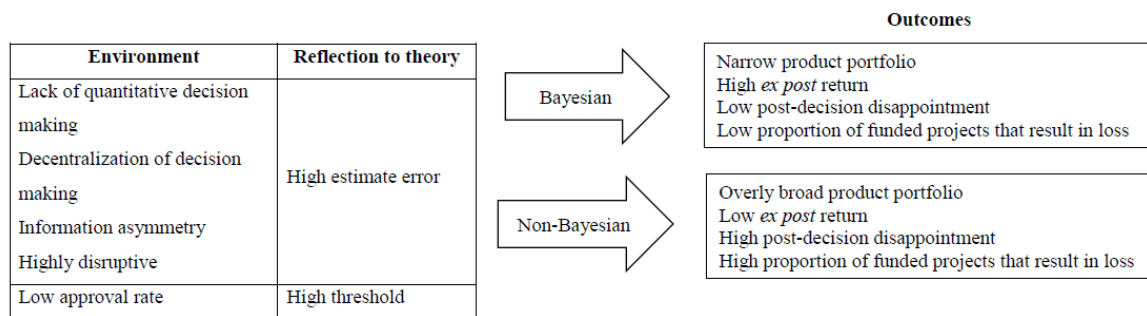
Figure 31: Reflection of decision-making environment characteristics to project portfolio selection theory and possible investment outcomes of the pharmaceutical industry

According to our findings in Chapter Four, having a decision-making environment with low estimation accuracy and a high threshold, the pharmaceutical industry will amplify the differences in outcomes between a Bayesian and a non-Bayesian DM. Accordingly, if a DM is Bayesian, she is expected to fund a narrow portfolio which results in a relatively high *ex post* return, low post-decision disappointment, and low probability of incurring loss. On the other hand, if a DM is a non-Bayesian DM, she is expected to fund a broad portfolio which results in a relatively low *ex post* return, high post-decision disappointment, and high probability of incurring loss. In order to validate these theoretical findings empirically, we will examine the characteristics of DMs in pharmaceutical industry, observe the outcomes of their investment decisions, and reflect these empirical outcomes to our theoretical outcomes in Chapter Four. Then, we will analyze the decision-making behavior of DMs in this industry and conclude whether or not their Bayesianess is reflected in their investment outcomes.

5.2.3 *Observed outcomes of the R&D investments in the pharmaceutical industry*

The outcomes of R&D investments in the pharmaceutical industry have been well documented in earlier literature and industry reports. Accordingly, R&D investments in the pharmaceutical industry result in (i) overly broad product selection, (ii) high attrition rate and low success rate, and (iii) low productivity and significant inefficiency.

5.2.3.1 Overly broad product selection

Schlapp *et al.* (2015) state that pharmaceutical companies often have overly broad product portfolios. Thomas (2011) suggests that one of the causes for such overly broad project selection is the decentralized decision-making process pharmaceutical firms. The overly broad product portfolio represents the high proportion of funded projects.

5.2.3.2 High attrition rate and low success rate

Although R&D investment decisions usually result in overly broad product selection, the rate of projects being terminated before hitting the market is noticeably high (Kola & Landis, 2004). According to Kola & Landis (2004), only one in nine compounds makes it through developments and gets approved by the European and/or the US regulatory authority. Roughly one in four compounds fails at the registration stage. Those failed compounds incur a significant loss in terms of both monetary (discovery and development costs) and opportunity costs (averagely 12 years 10 months). Some of the causes that are identified include costs overruns and commercial failure (Kola & Landis, 2004).

The alarmingly high attrition rate and low success rate indicates a high proportion funded projects that are terminated midway due to their unprofitable prospects. Among the projects that make it to the end, there is still a significantly low success rate. These phenomena prove that the proportion of funded projects that result in loss in this industry is extremely high. The low success rate and low productivity also project the high post-decision disappointment.

5.2.3.3 Low productivity and significant inefficiency

Grabowski (1997) presents an empirical study proving that for most marketed drugs, NPV revenue is less than average development costs. Kola & Landis (2004) also discuss about the low productivity and the acute inefficiency of R&D investment in the pharmaceutical industry; specifically, after the high attrition rate, there are only three out of ten drugs that made it to the market recover the original investment made in them. The low productivity and high inefficiency from R&D investment characterize the low *ex post* returns.

5.2.4   Comparing the observed outcomes with the theoretical outcomes

These observed outcomes perfectly coincide with the expected outcomes based on our theoretical findings in Chapter Four for a non-Bayesian DM in ***Figure 31***. In other words, the analytic results of the outcomes of investment decisions based on non-Bayesian estimates with a low signal-to-noise ratio is well-reflected in the outcomes of R&D investment decisions in pharmaceutical industry.

5.2.5   Decision-making behavior of DMs in the pharmaceutical industry

DMs in the pharmaceutical industry display quintessential non-Bayesian characteristics. Antoijevic (2014) reveals that DMs in the pharmaceutical industry make their decisions merely based on *"gut feel"* and many DMs in the pharmaceutical industry are against adopting a more quantitative approach to decision making. Besides implying high estimate errors, such random decision-making process ignore analytic data, specifically the *base rate* (i.e. prior mean). This characteristic perfectly corresponds with the outcomes observed from the R&D investments in the pharmaceutical industry.

5.3   Recommended strategies to improve the performance of R&D investments in the pharmaceutical industry

Earlier research recommends numerous strategies to improve the performance of R&D investments in the pharmaceutical industry. Those strategies can be classified into two categories: (i) creating a more favorable decision-making environment (e.g. increasing the signal-to-noise ratio, increasing success rate, etc.) and (ii) adopting a more Bayesian approach when making R&D investment decisions.

5.3.1   Create a more favorable decision-making environment

Our analytic results show that in an environment with lower estimate error (i.e. higher estimate accuracy/ higher signal-to-noise ratio), the differences a Bayesian and a non-Bayesian strategy, and the differences between each of them and the optimal strategy are smaller. Thus, in order to alleviate the gaps between both strategies with the optimal one, we

recommend creating a more favorable environment for decision-making processes (i.e. increase the signal-to-noise ratio).

Increasing the signal-to-noise ratio can significantly increase the performance of project portfolio selection decisions. Increasing the signal-to-noise ratio can be done by increasing the transparency and democracy of the decision-making process. A recent study of Held *et al.* (2009) of eight big pharma firms shows that those companies have attempted to increase the transparency and democracy in the product emulation process and consequently they outperform their competitors across multiple metrics of R&D productivity. Flyvbjerg (2009) also states that increasing transparency and democracy in decision making will improve average performance of the investment decisions. Besides, such transparency and democracy also help avoid political-economic issues (i.e. agency cost and organizational pressure) with can further bias the estimates.

Increasing the signal-to-noise ratio can also be done by increasing information availability for DMs. Thomas (2011) suggests that increasing information sharing will lower information asymmetry and consequently have a positive impact on the outcomes of R&D investments in the pharmaceutical industry. Kola & Landis (2004) and Thomas (2011) demonstrate that pharmaceutical companies with better information sharing and lighter organizational structures will yield better R&D performances than those with redundant information flows. More specifically to the pharmaceutical industry, Kola & Landis (2004) suggest improving pre-clinical testing. Such practice would improve the signal-to-noise ratio as it allows companies to observe and collect more information about the products before making decisions.

Miller (2005) proposes five techniques to improve R&D performance in the pharmaceutical industry. Among these five, there are four techniques that will increase the signal-to-noise ratio: (i) clinical trial simulation: estimate efficacy and tolerability profiles before clinical data are available, (ii) option pricing: can show the value of different clinical program designs, sequencing of studies and stop decisions, (iii) investment appraisal: compare ENVP of different product profiles and study design, (iv) threshold analysis:

understand development drug profile requirements given partial data. Such analyses will continually provide more information about the process and consequently increase the signal-to-noise ratio. Additionally, it also increases the availability of information and analytic data to the DMs.

5.3.2 *Adopting a Bayesian approach*

The analytic results in Chapter Four prove that investments made based on Bayesian estimates, on average, yield higher *ex post* returns than those of non-Bayesian estimates. These results coincide with many other literature on R&D investment decisions in the pharmaceutical industry. More specifically, this section will focus on the two main types of recommendations regarding the use of Bayesian modeling: (i) increase the use of pharmacoeconomic analysis and (ii) evaluate more conservatively (as a Bayesian DM would).

5.3.2.1   Use pharmacoeconomic analysis

Owing to the significance of R&D investment decisions in the pharmaceutical industry, pharmacoeconomics has an essential role in decision-making during drug research and development. Pharmacoeconomics can enhance the efficiency of R&D resource use and consequently increase commercial success significantly (Miller, 2005). Pharmacoeconomics analysis reflects the use of Bayesian modeling of estimate uncertainty that is enabled by information availability and estimate accuracy. The Bayesian analytic framework is proven to be well-suited to pharmacoeconomics in R&D since it explicitly acknowledges the purpose of data from clinical trials and experiments is to update knowledge (O'Hagan & Luce, 2003; Miller, 2005). As the Bayesian modeling of estimate uncertainty enables the synthesis of different information, it is of great value in the drug development process as pieces of information about the drug come from various sources, at various points of time, and in various forms; such synthesis of information is highly beneficial for DMs to stay informed about their value estimates. Thus, Bayesian methods offer DMs in pharmaceutical industry the efficiency of information and allow them to use such information in a systematic manner (O'Hagan & Stevens, 2001, 2002; Briggs, 2003; O'Hagan & Luce, 2003; Shih, 2003; Miller, 2005). Generally, the Bayesian analytic model of estimate uncertainty can be utilized in

practice to aid drug R&D decision-making processes.

Besides Bayesian modeling of estimate uncertainties, Miller (2005) suggests the use of other forms of pharmacoeconomic analysis in earlier stage of the R&D process, such as value of information analysis (i.e. assisting risk management by quantifying uncertainty and assessing the economic viability of gathering further info on the development drug). Such analysis also reflects the Bayesian approach in taking into account the *base rate* and prior information with a more quantitative method than the *gut feel* method that most pharmaceutical companies currently employ. In one way or another, it is proven that adopting a more Bayesian approach is an effective way for DMs in the pharmaceutical industry to overcome the current unproductivity of R&D activities.

5.3.2.2   Evaluate more conservatively

Earlier literature also encourages a more conservative evaluation process. Kola and Landis (2004) states that the pressure to improve R&D productivity calls for a stricter evaluation of all elements that influence R&D process; thus, companies that invest in a smaller product portfolio yield higher results than those with expansive product portfolios. Thomas (2011) claims that increase product range standardization would also reduce cost overruns and increase profits. These recommendations and their corresponding outcomes reflect a more Bayesian approach to making project portfolio investment decisions; as it is proven in Chapter Four that a Bayesian DM would invest in a smaller portion of project proposals yet yield a higher average return.

In a nutshell, adopting a more Bayesian approach is proven to be an effective way to solve the current inefficiency, unproductivity, and unprofitability of R&D investment in the pharmaceutical industry. It is observed that companies who adopt a Bayesian decision-making strategy will increase their average *ex post* return and outperform those who do not. The recommended strategies from earlier literature and their observed outcomes perfectly coincide with and well reflect the analytic results from Chapter Four.

6   Managerial Recommendations

Based on our findings, in practice, using Bayesian estimates as a basis for project portfolio selection results in higher average project value and eliminates expected post-decision disappointment. On the other hand, using non-Bayesian estimates would help ensure that truly best projects get funded. The less accurate the projects' value estimates, the greater the differences between Bayesian and non-Bayesian investment decision outcomes. However, if the estimates are highly accurate, Bayesian and non-Bayesian decisions make no difference. Thus, the extent to which a DM should adopt a Bayesian state of mind on depends on estimation accuracy and the DM's preferences, i.e., whether (s)he finds it important to have better projects on average or to seek big wins. Thus, this chapter of the thesis will propose managerial recommendations based on only two parameters: (i) the accuracy of the estimate, and (ii) the preferences of the DMs.

6.1   Preferences of the decision makers

In cases where fewer than 50% of the total project proposals have truly positive performance, or more generally, the threshold is higher than the prior mean, it is proven that adopting a Bayesian approach will improve the average performance of the selected projects. However, adopting a Bayesian approach will lower than chance of successfully predicting "the next big thing" (e.g. breakthroughs in technology). This characteristic is demonstrated with the simulation results in Chapter Four and in an empirical study by Denrell & Fang (2010). Denrell & Fang (2010) state that *"success [in predicting the next big thing] is a sign of poor judgement"* as most successful forecasters of the next breakthroughs tend to ignore the *base rate* and make decisions arbitrarily.

The Propositions in Chapter Four also prove in cases where more than 50% of the total project proposals have truly positive performance (i.e. threshold is lower than mean), the results mentioned in the previous paragraph will be reversed. In such cases where the expected performance of the alternatives is higher than the threshold, adopting a non-Bayesian approach results in a smaller proportion of funded projects and improves the

average performance of those funded projects. However, the probability of having such favorable project proposals is relatively low in practice. Additionally, due to the scarcity of resources (e.g. financial resources, human resources, etc.), it is highly unlikely that DMs will set the threshold lower than the prior mean.

Thus, in practice, when a DM aims to maximize average *ex post* performance, it is recommended to adopt a Bayesian approach, as it is proven that investment decisions based on Bayesian estimates will yield higher expected performance and lower post-decision disappointment. However, in cases where it is crucial to capture big wins, a non-Bayesian approach is recommended. Vilkkumaa *et al.* (2015) suggest a hybrid approach that allows DMs to utilize the strengths of both Bayesian and non-Bayesian approach. Accordingly, DMs are recommended to start with an overly broad portfolio (as a non-Bayesian DM would) and subsequently, when more information becomes available, abandon underperformed projects. The latter part of this approach displays Bayesian characteristics, as it takes into account prior information. Abandoning underperformed projects also results in a narrower portfolio, which is also a Bayesian characteristic.

The pharmaceutical industry is a quintessential example of environments where the non-Bayesian approach showcases its strengths. As the drug development process is long and highly unpredictable, and the threshold is extremely high, it is crucial for DMs in this industry to capture big wins. In such case, the choice of adopting a non-Bayesian approach can be justified.

6.2   Estimation accuracy

The estimation accuracy in our portfolio selection model is represented by the signal-to-noise ratio which reflects the magnitude of the standard deviation of the true values in relation to the estimate errors. It is proven that estimation accuracy is an element that amplifies the differences between a Bayesian and a non-Bayesian investment strategy. The lower the estimation accuracy is (i.e. the lower the signal-to-noise ratio is, the noisier the estimates are), the more pronounced the differences are. Thus, in environments with high estimation

accuracy, the differences between a Bayesian and a non-Bayesian investment strategies are subtler.

As a consequence, the Bayesianess of the optimal strategy depends on the decision-making environment and the preferences of the DMs. In *noisy* environments, the Bayesianess versus non-Bayesianess are more distinct in nature. More explicitly, if the DMs prefer to maximize average performance in a *noisy* environment, they should adopt a comprehensively Bayesian approach. However, if they want to have a higher probability of capturing the next big things, they should adopt an approach with a high degree of non-Bayesianess.

On the contrary, the distinction of Bayesianess versus non-Bayesianess is less pronounced in environments with higher estimation accuracy. As estimates are accurate, the estimate error τ is insignificant. Thus, the posterior distribution of Bayesian estimates is significantly similar to those of the non-Bayesian estimates. The optimal strategies in *"quiet"* environments with regard to different set of priorities are also slightly different. In a *"quiet"* environment, the optimal strategy for maximizing average performance display low degree of Bayesianess while the optimal strategy for capturing more big wins display low degree of non-Bayesianess.

	Low	Noisy /Maximize	Noisy/ Big wins
		• High degree of Bayesianess	• High degree of non-Bayesianess
Estimation accuracy		Quiet/Maximize	Quiet/ Big wins
		• Low degree of Bayesianess	• Low degree of non-Bayesianess
	High		

Maximize average performance   **Decision maker's preference**   Big wins

Figure 32: Conceptual framework for optimal decision-making strategy with regard to decision makers' preference and estimation accuracy

The optimal strategy for project portfolio selection decision therefore depends largely on the environment in which the decisions are made and the preferences of the DMs. *Figure 32* illustrates the framework for optimal decision-making strategy with regard to DMs' preference and estimate accuracy. Different settings of the environment (e.g. estimation accuracy, proportion of project proposals with truly positive performance, etc.) and different preferences (e.g. maximize average performance or capture big wins) will result in different optimal strategies. The DMs thus are recommended to understand the accuracy of their estimates in order to decide what degree of the *base rate* (i.e. prior mean) they should need to take into account. They are also recommended to identify their priorities prior to making decisions so that they can choose the appropriate approach (i.e. Bayesian or non-Bayesian).

7   Conclusion

This chapter serves as a synopsis of the entire thesis. Firstly, this chapter will consolidate the key findings by summarizing the answers for each of the research questions. Secondly, it will present the theoretical and managerial contribution of this research. And lastly, it will evaluate the research limitations and suggestions for future research.

7.1   Summary of key findings

This thesis seeks to study the differences between a Bayesian and a non-Bayesian DM in terms of investment strategies and the outcomes resulting from their strategies. The key results reveal that when fewer than 50% of project proposals have truly positive performance, a Bayesian DM invests in a lower proportion of proposals than a non-Bayesian DM does. Nevertheless, the average *ex post* performance resulting from the funded projects of a Bayesian DM is higher than that of a non-Bayesian DM. Thus, without psychological adjustments, a non-Bayesian DM will experience higher post-decision disappointment. Although the selected project portfolio resulting from non-Bayesian estimates yields a lower average *ex post* performance in this case, a non-Bayesian DM will have a higher proportion of funded projects that result in big wins. Furthermore, when more than 50% of project proposals have truly positive performance, all the results are reversed. However, the result for proportions of funded projects that results in high gains is only reversed when the estimate error is relatively high in relation to the distance between the threshold and the prior mean. Otherwise, the result in proportions of funded projects resulting in high gains are mostly equal between a Bayesian and non-Bayesian DM and are equal to the optimal one. However, in practice, the case where more than 50% of project proposals have truly positive performance is extremely unlikely due to the scarcity of resources, the harsh competition, and other constraints.

Furthermore, the differences between the investment outcomes of a Bayesian and a non-Bayesian DM decrease when estimation accuracy increases. The lower the estimation accuracy is, the higher the differences are.

The analytic results are testified empirically in the R&D portfolio selection decisions in the pharmaceutical industry. The characteristics of the decision-making environment of the pharmaceutical industry are analyzed with the theoretical lens of decision making under uncertainty. Accordingly, the decision-making environment of the pharmaceutical industry displays characteristics of a *noisy* environment with high estimate errors. The low estimation accuracy in this industry will amplify the differences in outcomes of a Bayesian DM's versus a non-Bayesian DM's investment decisions. As the DMs in this industry show quintessential characteristics of non-Bayesian DMs (i.e. ignore the *base rate*, make decisions by *gut feel*) and the observed empirical outcomes perfectly coincide with the theoretical outcomes of non-Bayesian investment decisions, our theoretical findings are well-reflected empirically.

In short, the answers to the research questions are as follows:

1. What is the difference between a Bayesian and a non-Bayesian DM in terms of:

Parameter	Summary of answer
a. The proportion of funded projects	When fewer than 50% of the project proposals have truly positive performance, a non-Bayesian DM will have a higher proportion of her alternatives funded than that of a Bayesian DM. The result is reversed when more than 50% of the projects proposals have truly positive performance.
b. The average performance of funded project	When fewer than 50% of the project proposals have truly positive performance, the funded portfolio of a Bayesian DM will yield higher average performance than that of a non-Bayesian DM. The result is reversed when more than 50% of the projects proposals have truly positive performance.

c.	*The proportion of funded project that result in loss*	When fewer than 50% of the project proposals have truly positive performance, the proportion of funded projects that results in loss of a Bayesian DM is lower than that of a non-Bayesian DM. The result is reversed when more than 50% of the projects proposals have truly positive performance.
d.	*The proportion of funded projects that result in very high gains*	When fewer than 50% of the project proposals have truly positive performance, the proportion of funded projects that results in very high gains of a Bayesian DM is lower than that of a non-Bayesian DM. When more than 50% of the project proposals have truly positive NPVs, the proportions of funded projects that results in very high gains of a Bayesian and a non-Bayesian DM have two possible results: (i) when the estimate error is relatively small in relation to the distance between the threshold and the prior mean, the proportions of funded projects that result in high gains are equal for a Bayesian and a non-Bayesian DM at near optimal level, (ii) when the estimate error is relatively large in relation to the distance between the threshold and the prior mean, the proportion of funded projects that results in high gains is higher for a Bayesian DM.

2. **How are these differences affected by:**

    a. <u>*The accuracy of the value estimates.*</u>

    The higher the accuracy of the value estimate, the less pronounced these differences are.

b. *The proportion of project proposals with truly positive performance*

Most of the answers, with two exceptions, are reversed when there are more versus when fewer than 50% of the proposals have truly positive performance. The first exception is the proportions of high gains which can be equal between a Bayesian and non-Bayesian DM when more than 50% of the project proposals have positive performance. The second exception is the expected portfolio values, as they are independent of the proportion of projects proposals with truly positive performance.

3. How well are these differences reflected in empirical data?

The differences are highly well-reflected in the pharmaceutical industry. As the pharmaceutical industry has an inherent high estimate errors due to various reasons (e.g. lack of statistical support, information asymmetry, etc.), the industry reflects well the differences between a Bayesian DM and a non-Bayesian DM. As DMs in this industry display distinctive characteristics of non-Bayesian DMs, theoretical findings suggest that the R&D investment decision outcomes in this industry would display the characteristics of non-Bayesian outcomes. As it is observed that the R&D investments in the pharmaceutical industry result in severe losses, low productivity, high attrition rates, and low success rates, our theoretical findings are validated. Earlier literature also recommends DMs in this industry to adopt a more Bayesian approach to improve the performance of R&D project portfolio selection.

7.2 Theoretical and managerial contributions

The theoretical contributions of this thesis include constructing a mathematical model that provides novel analytic results on the differences and the impact of various problem parameters in such differences. These analytic results are subsequently validated empirically. In terms of managerial contributions, this thesis provides insights into how estimation uncertainties affect investment decisions and their corresponding outcomes. Based on those insights, this thesis constructs a conceptual framework for practitioners to identify

appropriate optimal decision-making strategy in project portfolio selection. More specifically, due to a variety of different variables and the complex nature of decision making under uncertainty, the optimal strategy for project portfolio selection decisions differs depending on the context of the decision-making environment and the preferences of the DMs.

7.3   Limitations and suggestions for future research

In terms of research limitation, this thesis is based on the assumption that proposals will be funded as long as their estimates are higher than the threshold. Although this assumption allows the author to examine the differences between Bayesian and non-Bayesian DMs in a more general level without being limited to some certain constraints, it ignores the scarcity of resources and other constraints.

Secondly, due to the scope limit, part of the research questions needs to be answered with simulation. Furthermore, empirical testing proves to be extremely challenging due to the lack or even the absence of data availability. Companies do not usually publish records of failed projects or failed investments; thus, it is extremely difficult to examine the performance of their project portfolio selection process in a thorough and unbiased manner. It is also problematic to compare portfolio selection strategies across different industries due to the different parameters and the lack of standardized methods.

For future research, it would be interesting to study the impact of Bayesian model of estimate uncertainty in other industries. It would also be useful to develop a more thorough mathematical model that captures all aspects of the differences between a Bayesian and a non-Bayesian DM to increase the generalizability of the model in this thesis. Furthermore, it is also necessary to examine those differences under different constraints, such as financial constraint or time constraint. Additionally, with more time and financial resources to collect data, it would also be worthwhile to conduct a quantitative empirical study on the differences between a Bayesian and non-Bayesian DM in R&D project portfolio selection decisions.

Appendix A. All Proofs of the Propositions

Proof of proposition 4. *Normal distributions of true values, estimates and Bayesian-adjusted estimates.*

As we define in the model description, we have the following:

True values: $V = \mu + E, \qquad E \sim N(0, \sigma^2)$

Hence, $V \sim N(\mu, \sigma^2)$

Estimates: $(V^E | V = v) = v + \Delta, \qquad \Delta \sim N(0, \tau^2)$

for all v's $(V^E) = V + \Delta = \mu + E + \Delta$

Hence, $V^E \sim N(\mu, \tau^2 + \sigma^2)$

Bayesian estimates: $v^B = \mathbb{E}[V | V^E = v^E]$

for all v^E's $V^B = \mathbb{E}[V | V^E]$

In order to observe the distribution of Bayesian estimates, we have to calculate the covariance and correlation of V and V^E.

1. Calculate the covariance of V and V^E.

General formula for calculating the covariance of X_1 and X_2:

$$Cov(X_1, X_2) = \mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)]$$

In this case: $X_1 = V, \quad \sigma_1 = \sigma, \qquad \mu_1 = \mu$

$X_2 = V^E, \quad \sigma_2 = \sqrt{\sigma^2 + \tau^2}, \quad \mu_2 = \mu$

Therefore, we have the following:

$$Cov(V, V^E) = \mathbb{E}[(V - \mu)(V^E - \mu)]$$

Because $V^E = V + \Delta$ and $V = \mu + E$, the previous equation becomes:

$$Cov(V, V^E) = \mathbb{E}[(\mu + E - \mu)(\mu + E + \Delta - \mu)$$

$$Cov(V, V^E) = \mathbb{E}[E \cdot (E + \Delta)]$$

$$Cov(V, V^E) = \mathbb{E}[E \cdot E + E \cdot \Delta] \tag{i}$$

As E and Δ are independent (i.e., $cov(E, \Delta) = 0$), we have the following:

$$\mathbb{E}(E \cdot \Delta) = cov(E, \Delta) + \mu_E + \mu_\Delta$$

$$\mathbb{E}(E \cdot \Delta) = cov(E, \Delta) + 0 + 0$$

$$\mathbb{E}(E \cdot \Delta) = 0 \tag{ii}$$

Plug the result calculated from (ii) to (i), we can continue to calculate $Cov(V, V^E)$

$$Cov(V, V^E) = \mathbb{E}[E^2] + 0$$
$$Cov(V, V^E) = \mathbb{E}[(E - 0)^2]$$

$$Cov(V, V^E) = E[(E - \mu)^2]$$

By definition, square of the difference between observed values and mean is covariance. Therefore, the previous equation becomes:

$$Cov(V, V^E) = \sigma^2$$

2. **Calculate the correlation of V and V^E.**

General formula for calculating the correlation ρ of X_1 and X_2:

$$\rho = corr(X_1, X_2) = \frac{cov(X_1, X_2)}{std(X_1) * std(X_2)}$$

In this case: $X_1 = V, \ X_2 = V^E$

$$Cov(X_1, X_2) = Cov\ (V, V^E) = \sigma$$

$$std(X_1) = std(V) = \sigma$$

$$std(X_2) = std(V^E) = \sqrt{\sigma^2 + \tau^2}$$

Therefore, we have the following:

$$\rho^E = corr(V, V^E) = \frac{\sigma^2}{\sigma\sqrt{\sigma^2+\tau^2}} = \frac{\sigma}{\sqrt{\sigma^2+\tau^2}}$$

3. **Calculate $V^B = \mathbb{E}[V|V^E = v^E]$ using bivariate conditional expectation**

The general bivariate conditional expectation formula of X_1 given X_2 is as follows:

$$\mathbb{E}(X_1|X_2 = x_2) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)$$

In this case: $X_1 = V, \quad X_2 = V^E, \quad x_2 = v^E$

$$\mu_1 = \mu, \qquad \mu_2 = \mu$$

From the previous step: $\rho = \rho^E = \frac{\sigma}{\sqrt{\sigma^2+\tau^2}}$

Plugging the data into the general bivariate conditional expectation formula, we have the following:

$$v^B = \mathbb{E}[V|V^E = v^E]$$

$$v^B = \mu + \frac{\sigma}{\sqrt{\sigma^2+\tau^2}} \cdot \frac{\sigma}{\sqrt{\sigma^2+\tau^2}} \cdot (v^E - \mu)$$

$$v^B = \mu + \frac{\sigma^2}{\sigma^2+\tau^2}(v^E - \mu)$$

$$v^B = \left(1 - \frac{\sigma^2}{\sigma^2+\tau^2}\right)\mu + \frac{\sigma^2}{\sigma^2+\tau^2}v^E$$

$$v^B = \left(\frac{\tau^2}{\sigma^2+\tau^2}\right)\mu + \left(\frac{\sigma^2}{\sigma^2+\tau^2}\right)v^E \tag{3}$$

Let v^E be random V^E:

$$V^B = \left(\frac{\tau^2}{\sigma^2+\tau^2}\right)\mu + \frac{\sigma^2}{\sigma^2+\tau^2}(\mu + E + \Delta)$$

$$V^B = \left(\frac{\tau^2}{\sigma^2+\tau^2}\right)\mu + \left(\frac{\sigma^2}{\sigma^2+\tau^2}\right)\mu + \left(\frac{\sigma^2}{\sigma^2+\tau^2}\right)(E + \Delta)$$

$$V^B = \left(\frac{\tau^2+\sigma^2}{\sigma^2+\tau^2}\right)\mu + \left(\frac{\sigma^2}{\sigma^2+\tau^2}\right)(E + \Delta)$$

$$V^B = \mu + \frac{\sigma^2}{\sigma^2+\tau^2}(E + \Delta) \qquad (\mu: \text{fixed}, \frac{\sigma^2}{\sigma^2+\tau^2}(E + \Delta): \text{random})$$

If $Var(X) = \sigma^2$, $Var(aX)$, with a being a constant, can be calculated as follow:

$$Var(aX) = a^2\sigma^2$$

Therefore, with $\mathbb{E}[V^B] = \mu$, the variance of V^B can be calculated from the previous result:

$$Var(V^B) = Var\left(\frac{\sigma^2}{\sigma^2+\tau^2}(E + \Delta)\right)$$

$$Var(V^B) = \left(\frac{\sigma^2}{\sigma^2+\tau^2}\right)^2 Var(E + \Delta)$$

$$Var(V^B) = \left(\frac{\sigma^2}{\sigma^2+\tau^2}\right)^2 \cdot (Var(E) + Var(\Delta))$$

$$Var(V^B) = \frac{\sigma^4}{(\sigma^2+\tau^2)^2} * (\sigma^2 + \tau^2)$$

$$Var(V^B) = \frac{\sigma^4}{\sigma^2+\tau^2}$$

---

*3 This is the closed form expression of Bayesian estimates

Hence, we can prove that:

$$V^B \sim N\left(\mu, \frac{\sigma^4}{\sigma^2 + \tau^2}\right)$$

In short, the values of the three distributions are as follow:

$$V \sim N(\mu, \sigma^2)$$

$$V^E \sim N(\mu, \tau^2 + \sigma^2)$$

$$V^B \sim N(\mu, \frac{\sigma^4}{\sigma^2 + \tau^2})$$

Proof of Proposition 5. *Comparing $Pr(V^B > 0), Pr(V > 0),$ and $Pr(V^E > 0)$*

In order to compare $Pr(V^B > 0), Pr(V > 0),$ and $Pr(V^E > 0)$, we first have to standardize $V^B, V,$ and V^E into standard normal distribution.

1. **Standardize $V \sim N(\mu, \sigma^2)$ to $X \sim N(0, 1)$**

First, we have to convert V to X. In order to convert a value to standard score (i.e., "X-score"), we first subtract the mean and then divide by the standard deviation. Accordingly, we have the following:

$$X = \frac{V - \mu}{\sigma}$$

Subsequently, V can be illustrated in terms of X as follow:

$$V = X\sigma + \mu$$

The threshold $V > 0$ can then be standardized into:

$$V > 0 \iff X > \frac{-\mu}{\sigma}$$

Thus, $Pr(V > 0)$ can be rewritten and calculated as follows:

$$Pr(V > 0) = Pr^V \left(X > \frac{-\mu}{\sigma} \right)$$

$$Pr(V > 0) = 1 - \Phi \left(\frac{-\mu}{\sigma} \right)$$

$$Pr(V > 0) = \Phi \left(\frac{\mu}{\sigma} \right)$$

In order to simplify the expression, we set $= \frac{\mu}{\sigma}$. The previous expression hence becomes:

$$Pr(V > 0) = \Phi(t)$$

Standardize $V^E \sim N(\mu, \sigma^2 + \tau^2)$ to $X^E \sim N(0, 1)$

Similarly, we first subtract the mean and then divide by the standard deviation. Accordingly, we have the following:

$$X^E = \frac{V^E - \mu}{\sqrt{\sigma^2 + \tau^2}}$$

Subsequently, V^E can be illustrated in terms of X^E as follows:

$$V^E = X^E \sqrt{\sigma^2 + \tau^2} + \mu$$

The threshold $V^E > 0$ can then be standardized into:

$$V^E > 0 \iff X^E > \frac{-\mu}{\sqrt{\sigma^2 + \tau^2}}$$

Thus, $Pr(V^E > 0)$ can be rewritten and calculated as follows:

$$Pr(V^E > 0) = Pr^E \left(X^E > \frac{-\mu}{\sqrt{\sigma^2 + \tau^2}} \right)$$

$$Pr(V^E > 0) = 1 - \Phi \left(\frac{-\mu}{\sqrt{\sigma^2 + \tau^2}} \right)$$

$$Pr(V^E > 0) = \Phi \left(\frac{\mu}{\sqrt{\sigma^2 + \tau^2}} \right)$$

In order to simplify the expression, we set $t^E = \frac{\mu}{\sqrt{\sigma^2 + \tau^2}}$. The previous expression hence becomes:

$$Pr(V^E > 0) = \Phi(t^E)$$

2. **Standardize $V^B \sim N(\mu, \frac{\sigma^4}{\sigma^2 + \tau^2})$ to $X^B \sim N(0, 1)$**

We can find X^B with the same formula as when we standardize V and V^E:

$$X^B = \frac{(V^B - \mu)\sqrt{\sigma^2 + \tau^2}}{\sigma^2}$$

Subsequently, V^B can be illustrated in terms of X^B as follow:

$$V^B = \mu + \frac{X^B \sigma^2}{\sqrt{\sigma^2 + \tau^2}}$$

The threshold can also be standardized in the same way as in the previous two sections:

$$V^B > 0 \Leftrightarrow X^B > \frac{-\mu\sqrt{\sigma^2 + \tau^2}}{\sigma^2}$$

Thus, $Pr(V^B > 0)$ can be rewritten and calculated as follow:

$$Pr(V^B > 0) = Pr^B\left(X^B > \frac{-\mu\sqrt{\sigma^2 + \tau^2}}{\sigma^2}\right)$$

$$Pr(V^B > 0) = 1 - \Phi\left(\frac{-\mu\sqrt{\sigma^2 + \tau^2}}{\sigma^2}\right)$$

$$Pr(V^B > 0) = \Phi\left(\frac{\mu\sqrt{\sigma^2 + \tau^2}}{\sigma^2}\right)$$

In order to simplify the expression, we set $t^B = \frac{\mu\sqrt{\sigma^2 + \tau^2}}{\sigma^2}$. The previous expression hence becomes:

$$Pr(V^B > 0) = \Phi(t^B)$$

3. Comparing $Pr(V^B > 0), Pr(V > 0),$ and $Pr(V^E > 0)$

In summary, from the previous three steps, we find the following:

$$Pr(V > 0) = \Phi(t), \text{ with } t = \frac{\mu}{\sigma}$$

$$Pr(V^E > 0) = \Phi(t^E), \text{ with } t^E = \frac{\mu}{\sqrt{\sigma^2 + \tau^2}}$$

$$\Pr(V^B > 0) = \Phi(\text{t}^B), \text{ with } t^B = \frac{\mu \sqrt{\sigma^2 + \tau^2}}{\sigma^2}$$

As $\Phi(\text{t})$, $\Phi(\text{t}^E)$ and $\Phi(\text{t}^B)$ are strictly increasing in t, t^E and t^B respectively, we can compare $Pr(V^B > 0)$, $Pr(V > 0)$, and $Pr(V^E > 0)$ by comparing t^B, t, and t^E. In order to compare t^B, t, and t^E, we will first compare t^E and t, and then we will compare t^B and t.

a. <u>Comparing t^E and t</u>

$$t^E - t = \frac{\mu}{\sqrt{\sigma^2 + \tau^2}} - \frac{\mu}{\sigma} = \frac{\mu\left(\sigma - \sqrt{\sigma^2 + \tau^2}\right)}{\sigma\sqrt{\sigma^2 + \tau^2}}$$

As $\tau > 0$ and $\sigma > 0$, we have the following: $\sigma^2 < \sigma^2 + \tau^2$

Since both sides of the inequality are larger than zero, we can square root the two sides of the inequality without changing the direction of the inequality:

$$\sigma < \sqrt{\sigma^2 + \tau^2} \Leftrightarrow \sigma - \sqrt{\sigma^2 + \tau^2} < 0$$

- If $\mu < 0$, then $t^E - t > 0 \Leftrightarrow t^E > t$
- If $\mu > 0$, then $t^E - t < 0 \Leftrightarrow t^E < t$

b. <u>Comparing t^B and t</u>

We can compare t^B and t by observing the sign of the difference between t^B and t.

$$t^B - t = \frac{\mu\sqrt{\sigma^2 + \tau^2}}{\sigma^2} - \frac{\mu}{\sigma}$$

$$t^B - t = \frac{\mu\sqrt{\sigma^2 + \tau^2} - \mu\sigma}{\sigma^2}$$

$$t^B - t = \frac{\mu(\sqrt{\sigma^2 + \tau^2} - \sigma)}{\sigma^2}$$

In the previous steps, we already proved that $\sqrt{\sigma^2 + \tau^2} > \sigma$. Thus, we have the following:

$$\sqrt{\sigma^2 + \tau^2} > \sigma$$

$$\Leftrightarrow \sqrt{\sigma^2 + \tau^2} - \sigma > 0$$

- As and $\sigma^2 > 0$, if $\mu < 0$, we have the following:

$$\frac{\mu\left(\sqrt{\sigma^2 + \tau^2} - \sigma\right)}{\sigma^2} < 0$$

$$\Leftrightarrow t^B - t < 0$$

$$\Leftrightarrow t^B < t$$

- if $\mu > 0$, we have the following:

$$\frac{\mu\left(\sqrt{\sigma^2 + \tau^2} - \sigma\right)}{\sigma^2} < 0$$

$$\Leftrightarrow t^B - t > 0$$

$$\Leftrightarrow t^B > t$$

c. <u>Comparing t^B and t^E</u>

We can compare t^B and t^E by observing the sign of its difference:

$$t^B - t^E = \frac{\mu\sqrt{\sigma^2 + \tau^2}}{\sigma^2} - \frac{\mu}{\sqrt{\sigma^2 + \tau^2}}$$

$$t^B - t^E = \frac{\mu(\sigma^2 + \tau^2) - \mu\sigma^2}{\sigma^2\sqrt{\sigma^2 + \tau^2}}$$

$$t^B - t^E = \frac{\mu\tau^2}{\sigma^2\sqrt{\sigma^2 + \tau^2}}$$

- If $\mu < 0$, we have the following:

$$t^B - t^E < 0$$

$$\Leftrightarrow t^B < t^E.$$

- If $\mu > 0$, we have the following:

$$t^B - t^E > 0$$

$$\Leftrightarrow t^B > t^E.$$

      d. <u>Comparing t^B, t, and t^E</u>

From a, b, and c, we can conclude that:

- If $\mu < 0$: $t^B < \mathrm{t} < \mathrm{t}^{\mathrm{E}}$
- If $\mu > 0$: $t^B > \mathrm{t} > \mathrm{t}^{\mathrm{E}}$

      e. <u>Comparing $Pr(V^B > 0), Pr(V > 0)$, and $Pr(V^E > 0)$</u>

As $\Phi(\mathrm{t})$, $\Phi(\mathrm{t}^{\mathrm{E}})$ and $\Phi(\mathrm{t}^{\mathrm{B}})$ are strictly increasing in t, t^E and t^B respectively, we can conclude that:

- If $\mu < 0$, $\Pr(V^B > 0) < \Pr(V > 0) < \Pr(V^E > 0)$
- If $\mu > 0$, $\Pr(V^B > 0) > \Pr(V > 0) > \Pr(V^E > 0)$.

Proof of Proposition 6. *Observing the effect of signal-to-noise ratio $(\alpha = \frac{\sigma}{\tau})$ on the difference among $Pr(V^B > 0)$, $Pr(V^E > 0)$, and $Pr(V > 0)$*

Plugging $\alpha = \frac{\sigma}{\tau}$ into the difference between t^B and t^E, we have the following:

$$t^B - t^E = \frac{\mu\tau^2}{\sigma^2\sqrt{\sigma^2+\tau^2}}$$

$$t^B - t^E = \frac{\mu}{\frac{\sigma^2}{\tau^2}\sqrt{\frac{\sigma^2}{\tau^2}+1}}$$

$$t^B - t^E = \frac{\mu}{\alpha^2\sqrt{\alpha^2+1}}$$

As the signal-to-noise ratio increases (i.e. the "signal" gets stronger and/or the "noise" gets weaker), the denominator increases.

- **If $\mu < 0$**, then $(t^B - t^E) < 0$.
    - As the fraction is negative, increases in denominators will increase the fraction.
    - Thus, $(t^B - t^E) < 0$ and increases in α if $\mu < 0$.
    - Thus, $|t^B - t^E|$ decreases in α if $\mu < 0$.
- **If $\mu > 0$,** then $(t^B - t^E) > 0$.
    - As the fractions is negative, increases in denominators will decrease the fraction.
    - Thus, $(t^B - t^E) > 0$ and decreases in α if $\mu > 0$.
    - Thus, $|t^B - t^E|$ decreases in α if $\mu > 0$.
- As t is in the middle of t^B and t^E in both senarios, the smaller the difference between t^B and t^E means the smaller the differences between t and t^E and between t and t^B.

In both cases, if the signal-to-noise ratio increases, the differences among t, t^B, and t^E in terms of absolute values decrease in $\alpha = \frac{\sigma}{\tau}$.

We also have the following:

$$|\Pr(V > 0) - \Pr(V^B > 0)| = \Phi(t) - \Phi(t^B)$$

$$|\Pr(V > 0) - \Pr(V^E > 0)| = \Phi(t) - \Phi(t^E)$$

$$|\Pr(V^B > 0) - \Pr(V^E > 0)| = \Phi(t^B) - \Phi(t^E)$$

As $\Phi(t)$, $\Phi(t^B)$, and $\Phi(t^E)$ increase in t, t^B, and t^E respectively and the differences among t, t^B, and t^E decrease in $\alpha = \frac{\sigma}{\tau}$. Consequently, the differences among $\Pr(V > 0)$, $\Pr(V^B > 0)$, and $\Pr(V^E > 0)$ decrease in $\alpha = \frac{\sigma}{\tau}$.

Proof of Proposition 7. *Using Mill's inverse ratio to compare* $\mathbb{E}[V|V^B > 0]$ *and* $\mathbb{E}[V|V^E > 0]$

In this section, the author will use Mill's inverse ratio to calculate $\mathbb{E}[V|V^B > 0]$ and $\mathbb{E}[V|V^E > 0]$ in relative terms and subsequently compare them.

1. Calculate $\mathbb{E}[V|V^E > 0]$ using Mill's inverse ratio

From Proposition 4, we have the following conjugate normal distributions of values and estimates:

Values: $V \sim N(\mu, \sigma^2)$

Estimates: $V^E \sim N(\mu, \sigma^2 + \tau^2)$

       Similar to what we did in the Proof of Proposition 5, we have to standardize V and V^E. Detailed steps of the standardization can be found in Proof of Proposition 5.

<u>Values:</u> $z = \frac{V - \mu}{\sigma}$, and $V = z\sigma + \mu$. The threshold becomes: $V > 0 \Leftrightarrow z > \frac{-\mu}{\sigma}$

<u>Estimates:</u> $z^E = \frac{V^E - \mu}{\sqrt{\sigma^2 + \tau^2}}$, and $V^E = z^E\sqrt{\sigma^2 + \tau^2} + \mu$. The threshold becomes:

$$V_E > 0 \Leftrightarrow z_E > \frac{-\mu}{\sqrt{\sigma^2 + \tau^2}}$$

       We can also find the standardized real expected value of V as follow:

$$\mathbb{E}[V] = \mathbb{E}[z\sigma + \mu]$$

$$\mathbb{E}[V] = \sigma\mathbb{E}[z] + \mu$$

       Adding the conditional into the previous expression, we can calculate $\mathbb{E}[V|V^E > 0]$ using Mill's inverse ratio:

$$\mathbb{E}[V|V^E > 0] = \mathbb{E}[z|V^E > 0]\sigma + \mu$$

$$\mathbb{E}[V|V^E > 0] = \mathbb{E}\left[z\Big|z^E > \frac{-\mu}{\sqrt{\sigma^2+\tau^2}}\right]\sigma + \mu$$

$$\mathbb{E}[V|V^E > 0] = \left[\rho^E \frac{\varphi\left(\frac{-\mu}{\sqrt{\sigma^2+\tau^2}}\right)}{1-\Phi\left(\frac{-\mu}{\sqrt{\sigma^2+\tau^2}}\right)}\right]\sigma + \mu$$

$$\mathbb{E}[V|V^E > 0] = \left[\rho^E \frac{\varphi\left(\frac{\mu}{\sqrt{\sigma^2+\tau^2}}\right)}{\Phi\left(\frac{\mu}{\sqrt{\sigma^2+\tau^2}}\right)}\right]\sigma + \mu$$

2. Calculate $\mathbb{E}[V|V^B > 0]$ using Mill's inverse ratio

Also from Proposition 1, we have the following conjugate normal distributions of values and estimates:

Values: $V \sim N(\mu, \sigma^2)$

Bayesian estimates: $V^B \sim N(\mu, \frac{\sigma^4}{\sigma^2+\tau^2})$

The next step is to standardize V and V^B:

Values: $z = \frac{V-\mu}{\sigma}, V = z\sigma + \mu$, threshold: $V > 0 \Leftrightarrow z > \frac{-\mu}{\sigma}$

Bayesian estimates: $z^B = \frac{(V^B-\mu)\sqrt{\sigma^2+\tau^2}}{\sigma^2}, V^B = \mu + \frac{z^B\sigma^2}{\sqrt{\sigma^2+\tau^2}},$

threshold: $V^B > 0 \Leftrightarrow z^B > \frac{-\mu\sqrt{\sigma^2+\tau^2}}{\sigma^2}$

In a similar manner as in the previous section, we can calculate $\mathbb{E}[V|V^B > 0]$ using Mill's inverse ratio:

$$\mathbb{E}[V|V^B > 0] = \mathbb{E}[z|V^B > 0]\sigma + \mu$$

$$\mathbb{E}[V|V^B > 0] = \mathbb{E}\left[z\Big|z^B > \frac{-\mu\sqrt{\sigma^2+\tau^2}}{\sigma^2}\right]\sigma + \mu$$

$$\mathbb{E}[V|V^B > 0] = \left[\rho^B \frac{\varphi\left(\frac{-\mu\sqrt{\sigma^2+\tau^2}}{\sigma^2}\right)}{1-\Phi\left(\frac{-\mu\sqrt{\sigma^2+\tau^2}}{\sigma^2}\right)}\right]\sigma + \mu$$

$$\mathbb{E}[V|V^B > 0] = \left[\rho^B \frac{\varphi(\frac{\mu\sqrt{\sigma^2+\tau^2}}{\sigma^2})}{\Phi(\frac{\mu\sqrt{\sigma^2+\tau^2}}{\sigma^2})}\right]\sigma + \mu$$

$$\mathbb{E}[V|V^B > 0] = \left[\rho^B \frac{\varphi(\frac{\mu\sqrt{\sigma^2+\tau^2}}{\sigma^2})}{\Phi(\frac{\mu\sqrt{\sigma^2+\tau^2}}{\sigma^2})}\right]\sigma + \mu$$

3. Calculating ρ^B and ρ^E

In order to compare $\mathbb{E}[V|V^B > 0]$ and $\mathbb{E}[V|V^E > 0]$, we first need to calculate ρ^B and ρ^E. From Proof of Proposition 1, we have:

$$\rho^E = corr(V, V^E) = \frac{\sigma^2}{\sigma\sqrt{\sigma^2+\tau^2}} = \frac{\sigma}{\sqrt{\sigma^2+\tau^2}}$$

To calculate ρ^B, we will follow the exact steps we did to find ρ^E. Firstly, we will find the covariance of V and V^B. Then, we will find the correlation of V and V^B.

To begin with, we will calculate covariance of V and V^B.

$$Cov(V, V^B) = \mathbb{E}[(V - \mu)(V^B - \mu)]$$

$$Cov(V, V^B) = \mathbb{E}[(\mu + E - \mu)\left(\mu + \frac{\sigma^2(\Delta+E)}{\sigma^2+\tau^2} - \mu\right)]$$

$$Cov(V, V^B) = \mathbb{E}[E \cdot \frac{\sigma^2(E+\Delta)}{\sigma^2+\tau^2}]$$

$$Cov(V, V^B) = \left(\frac{\sigma^2}{\sigma^2+\tau^2}\right) \cdot \mathbb{E}[E \cdot (E + \Delta)]$$

From Proof of Proposition 4, we have $\mathbb{E}[E \cdot (E + \Delta)] = \sigma^2$. Hence, the previous expression becomes:

$$Cov(V, V^B) = \left(\frac{\sigma^2}{\sigma^2 + \tau^2}\right) \cdot \sigma^2$$

$$Cov(V, V^B) = \frac{\sigma^4}{\sigma^2 + \tau^2}$$

As we found $Cov(V, V^B)$, we can now proceed to finding the correlation between V and V^B.

$$\rho^B = corr(V, V^B)$$

$$\rho^B = \frac{cov(V, V^B)}{std(V) \cdot std(V^B)}$$

$$\rho^B = \frac{\sigma^4}{\sigma^2 + \tau^2} \cdot \frac{1}{\sigma} \cdot \frac{\sqrt{\sigma^2 + \tau^2}}{\sigma^2}$$

$$\rho^B = \frac{\sigma}{\sqrt{\sigma^2 + \tau^2}}$$

Therefore, we can conclude that: $\rho^B = \rho^E = \frac{\sigma}{\sqrt{\sigma^2 + \tau^2}} = \rho$

4. Comparing $\mathbb{E}[V|V^B > 0]$ *and* $\mathbb{E}[V|V^E > 0]$

As the two correlations are the same, the expected values can be compared by comparing the following two values:

For estimates: $\qquad \frac{\varphi\left(\frac{\mu}{\sqrt{\sigma^2 + \tau^2}}\right)}{\Phi\left(\frac{\mu}{\sqrt{\sigma^2 + \tau^2}}\right)}$ (iii)

For Bayesian estimates: $\frac{\varphi\left(\frac{\mu\sqrt{\sigma^2 + \tau^2}}{\sigma^2}\right)}{\Phi\left(\frac{\mu\sqrt{\sigma^2 + \tau^2}}{\sigma^2}\right)}$ (iv)

Recall from Proof for Proposition 5, we set t^E and t^B as follow:

$$t^E = \frac{\mu}{\sqrt{\sigma^2 + \tau^2}}$$

$$t^B = \frac{\mu\sqrt{\sigma^2 + \tau^2}}{\sigma^2}$$

(iii) and (iv) will then become:

For estimates: $\dfrac{\varphi(t^E)}{\Phi(t^E)}$

For Bayesian estimates: $\dfrac{\varphi(t^B)}{\Phi(t^B)}$

According to Sampford (1953, equation 4), $\dfrac{\varphi(t^E)}{\Phi(t^E)}$ and $\dfrac{\varphi(t^B)}{\Phi(t^B)}$ are decreasing in t^E and t^B respectively. Therefore, we can compare $\mathbb{E}[V|V^B > 0]$ and $\mathbb{E}[V|V^E > 0]$ by comparing t and t^E.

We have: $t^B - t^E = \dfrac{\mu\tau^2}{\sigma^2\sqrt{\sigma^2+\tau^2}}$

- If $\mu < 0$: $t^B - t^E < 0 \Leftrightarrow t^B < t^E$
- If $\mu > 0$: $t^B - t^E > 0 \Leftrightarrow t^B > t^E$

As $\dfrac{\varphi(t^E)}{\Phi(t^E)}$ and $\dfrac{\varphi(t^B)}{\Phi(t^B)}$ are decreasing in t^E and t^B respectively, we can conclude the following:

- If $\mu < 0$: $\mathbb{E}[V|V^B > 0] > \mathbb{E}[V|V^E > 0]$
- If $\mu > 0$: $\mathbb{E}[V|V^B > 0] < \mathbb{E}[V|V^E > 0]$

Proof of Proposition 8. *Observing the effect of signal-to-noise ratio $(\alpha = \frac{\sigma}{\tau})$ on the difference between $\mathbb{E}[V|V^B > 0]$ and $\mathbb{E}[V|V^E > 0]$*

$$|\mathbb{E}[V|V^B > 0] - \mathbb{E}[V|V^E > 0]| = \left|\left[\rho\frac{\varphi(t^B)}{\Phi(t^B)}\right]\sigma + \mu - \left[\rho\frac{\varphi(t^E)}{\Phi(t^E)}\right]\sigma - \mu\right|$$

$$\Leftrightarrow |\mathbb{E}[V|V^B > 0] - \mathbb{E}[V|V^E > 0]| = \left|\rho\sigma\left(\frac{\varphi(t^B)}{\Phi(t^B)} - \frac{\varphi(t^E)}{\Phi(t^E)}\right)\right|$$

Recall from Proof for Proposition 7, we have: $\frac{\varphi(t^B)}{\Phi(t^B)}$ decreases in t^B and $\frac{\varphi(t^E)}{\Phi(t^E)}$ decreases in t^E. Recall from Proof for Proposition 6: $|t^B - t^E| = \left|\frac{\mu}{\alpha^2\sqrt{\alpha^2+1}}\right|$ decreases in $\alpha = \frac{\sigma}{\tau}$.

Thus $|\mathbb{E}[V|V^B > 0] - \mathbb{E}[V|V^E > 0]|$ decreases in $\alpha = \frac{\sigma}{\tau}$.

References

Abbassi, M., Ashrafi, M., & Tashnizi, E. S. (2014). Selecting balanced portfolios of R&D projects with interdependencies: A Cross-Entropy based methodology. *Technovation*, *34*(1), 54-63.

Abernathy, W. J., & Clark, K. B. (1985). Innovation: Mapping the winds of creative destruction. *Research policy*, *14*(1), 3-22.

Aguilar, O., & West, M. (2000). Bayesian dynamic factor models and portfolio allocation. Journal of Business & Economic Statistics, 18(3), 338-357.

Antonijevic, Z. (2014). Optimization of Pharmaceutical R&D Programs and Portfolios.

Amaratunga, D., Baldry, D., Sarshar, M., & Newton, R. (2002). Quantitative and qualitative research in the built environment: application of "mixed" research approach. *Work study*, *51*(1), 17-31.

Archer, N. P., & Ghasemzadeh, F. (1999). An integrated framework for project portfolio selection. *International Journal of Project Management*, *17*(4), 207-216.

Ascher, W. (1978). Forecasting: An appraisal for policy-makers and planners. *Baltimore, MD Johns Hopkins University Press.*

Banbury, C. M., & Mitchell, W. (1995). The effect of introducing important incremental innovations on market share and business survival. *Strategic Management Journal*, *16*(S1), 161-182.

Bard, J. F., Balachandra, R., & Kaufmann, P. E. (1988). An interactive approach to R&D project selection and termination. *IEEE Transactions on Engineering Management*, *35*(3), 139-146.

Barry, C.B., & Winkler, R. L. (1976). Nonstationarity and portfolio choice. *The Journal of Financial and Quantitative Analysis,* 11, 217-235.

Bell, D. E. (1985). Disappointment in decision making under uncertainty. *Operations research*, *33*(1), 1-27.

Bok, S. (1979). *Lying: moral choice in public and private life* (New York: Vintage).

Brandt, M. W., Goyal. A., Santa-Clara, P., & Stroud, J.R. (2005). A simulation approach to dynamic portfolio choice with an application to learning about return predictability. *Review of Financial Studies,* 18, 831-873.

Briggs, A. H. (Ed.). (2003). *Statistical methods for cost-effectiveness research: a guide to current issues and future developments*. Office of Health Economics.

Brown, K. C. (1974). A note on the apparent bias of net revenue estimates for capital investment projects. *The Journal of Finance*, *29*(4), 1215-1227.

Brown, K. C. (1978). The rate of return of selected investment projects. *The Journal of Finance*, *33*(4), 1250-1253.

Brous, P., Hiedemann, B., & Schultz, T. (2009). Success Rates, Optimism Bias, and the Predictive Power of Alternative Performance Measures: A Post-Audit of One Company's Real Investment Decisions. *Journal of Financial Education*, 29-41.

Capen, E.C., Clapp, R.V., Campbell, W.M. (1971). Competitive bidding in high-risk situations. *Journal of Petroleum Technology*, 23, 6 (June 1971), 641-653.

Carlin, B. P., & Louis, T. A. (2000). Bayes and empirical Bayes methods for data analysis.

Carson, E., & Cobelli, C. (2013). *Modeling methodology for physiology and medicine*. Newnes.

Charnes, A., & Stedry, A. C. (1966). A chance-constrained model for real-time control in research and development management. *Management Science*, *12*(8), B-353.

Christensen, C. M., Grossman, J. H., & Hwang, J. (2009). The innovator's prescription. *A disruptive solution for health care. New York: McGraw-Hill*.

Cliffe, L., Ramsay, M., & Bartlett, D. (2000). *The politics of lying: Implications for democracy*. Springer.

Cool, K. O. (1985). Strategic Group Information and Strategic Group Shifts. A longitudinal analysis of the US pharmaceutical industry, 1963-1982 (Competition, Performance; United States).

Denrell, J., & Fang, C. (2010). Predicting the Next Big Thing: Success as a Signal of Poor Judgement. *Management Science* 56(10): 1653-1667.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The quarterly journal of economics*, 643-669.

Dym, C. (2004). *Principles of mathematical modeling*. Academic press.

Flyvbjerg, B. (2009). Survival of the unfittest: why the worst infrastructure gets built—and what we can do about it. *Oxford review of economic policy*, *25*(3), 344-367.

Flyvbjerg, B., Holm, M. S., & Buhl, S. (2002). Underestimating costs in public works projects: Error or lie?. *Journal of the American planning association*, *68*(3), 279-295.

Flyvbjerg, B., Skamris Holm, M. K., & Buhl, S. L. (2005). How (in) accurate are demand forecasts in public works projects?: The case of transportation. *Journal of the American Planning Association*, *71*(2), 131-146.

Freeman, C. (1982). The Economics of Industrial Innovation. *The MIT Press*, Cambridge, MA

Gershenfeld, N. A. (1999). *The nature of mathematical modeling*. Cambridge university press.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). Bayesian Data Analysis Chapman & Hall London Google Scholar.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., & Vehtari, A. (2004). Rubin DB. Bayesian data analysis.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Boca Raton, FL, USA: Chapman & Hall/CRC.

Grabowski, H. (1997). The effect of pharmacoeconomics on company research and development decisions. *Pharmacoeconomics*, *11*(5), 389-397.

Harrison, J. R., & March, J. G. (1984). Decision making and postdecision surprises. *Administrative Science Quarterly*, 26-42.

Harter, H. L. (1969). *Order statistics and their use in testing and estimation* (Vol. 1). Washington, DC: US Government Printing Office.

Hastings Jr, C., Mosteller, F., Tukey, J. W., & Winsor, C. P. (1947). Low moments for small samples: a comparative study of order statistics. *The Annals of Mathematical Statistics*, 413-426.

Held, S. (2009). *Impact of Big Pharma organizational structure on R&D productivity* (Vol. 17). BoD–Books on Demand.

Henderson, R. (1994). The evolution of integrative capability: innovation in cardiovascular drug discovery. *Industrial and corporate change*, *3*(3), 607-630.

Jørgensen, M. (2013). The influence of selection bias on effort overruns in software development projects. *Information and Software Technology*, *55*(9), 1640-1650.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 273-251.

Kahneman, D., & Tversky, A. (1979). Intuitive prediction: Biases and corrective procedures. In S. Makridakis & S.C. Wheelwright (Eds.). *Studies in the management sciences: Forecasting*. New York: Springer

Kavadias, S., & Loch, C. H. (2004). *Project selection under uncertainty: dynamically allocating resources to maximize value* (Vol. 69). Springer Science & Business Media.

Kola, I., & Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? *Nature reviews Drug discovery*, *3*(8), 711-716.

Li, Z. (2009). Stochastic optimization model and O (N 2) solution algorithm for highway investment decision making under budget uncertainty. *Journal of Transportation Engineering*, *135*(6), 371-379.

Li, Z., & Madanu, S. (2009). Highway project level life-cycle benefit/cost analysis under certainty, risk, and uncertainty: methodology with case study. *Journal of Transportation Engineering*, *135*(8), 516-526.

Li, Z., & Sinha, K. (2004). Methodology for multicriteria decision making in highway asset management. *Transportation Research Record: Journal of the Transportation Research Board*, (1885), 79-87.

Liamputtong, P., & Ezzy, D. (2005). Making sense of qualitative data: analysis process. *Qualitative research methods*, 257-285.

Liesiö, J., Mild, P., & Salo, A. (2007). Preference programming for robust portfolio modeling and project selection. *European Journal of Operational Research*, *181*(3), 1488-1505.

Liesiö, J., Mild, P., & Salo, A. (2008). Robust portfolio modeling with incomplete cost information and project interdependencies. *European Journal of Operational Research*, *190*(3), 679-695.

Liesiö, J., & Salo, A. (2012). Scenario-based portfolio selection of investment projects with incomplete probability and utility information. *European Journal of Operational Research*, *217*(1), 162-172.

Lindley, D. V. (1986). The reconciliation of decision analyses. *Operations Research*, *34*(2), 289-295.

Lindley, D. V., Tversky, A., & Brown, R. V. (1979). On the reconciliation of probability assessments. *Journal of the Royal Statistical Society. Series A (General)*, 142: 146-180.

Loch, C. H., & Kavadias, S. (2002). Dynamic portfolio selection of NPD programs using marginal returns. *Management Science*, *48*(10), 1227-1241.

Mahadevan, S. (1997). Monte Carlo simulation. Mechanical Engineering-New York and Basel-Marcel Dekker, 123-146.

Martino, J. P. (1995). *Research and development project selection*. Wiley.

Mavrotas, G., Diakoulaki, D., & Capros, P. (2003). Combined MCDA–IP approach for project selection in the electricity market. *Annals of Operations Research*, *120*(1-4), 159-170.

Medaglia, A. L., Graves, S. B., & Ringuest, J. L. (2007). A multiobjective evolutionary approach for linearly constrained project selection under uncertainty. *European journal of operational research*, *179*(3), 869-894.

Miller, E. M. (1978). Uncertainty Induced Bias in Capital Budgeting. *Financial Management.* Autumn 1978. pp. 12-18.

Miller, E. M. (1986). The Competitive Market Assumption and Capital Budgeting Criteria. *Presented to the Financial Management Association*, October 1986. New York.

Miller, P. (2005). Role of pharmacoeconomic analysis in R&D decision making. *Pharmacoeconomics*, *23*(1), 1-12.

Mooney, C. Z. (1997). Monte Carlo simulation (Vol. 116). Sage Publications.

Morris, P. A. (1974). Decision analysis expert use. *Management Science*, *20*(9), 1233-1241.

Nickerson, R. C., & Boyd, D. W. (1980). The use and value of models in decision analysis. *Operations Research*, *28*(1), 139-155.

Odeck, J. (2004). Cost overruns in road construction—what are their sizes and determinants?. *Transport policy*, *11*(1), 43-53.

O'Hagan, A., & Luce, B. R. (2003). A primer on Bayesian statistics in health economics and outcomes research. Medtap International. *Inc., Bethesda, Maryland.*

O'Hagan, A., & Stevens, J. W. (2001). Bayesian assessment of sample size for clinical trials of cost-effectiveness. Medical Decision Making, 21(3), 219-230.

O'Hagan, A., & Stevens, J. W. (2002). Bayesian methods for design and analysis of cost-effectiveness trials in the evaluation of health care technologies. *Statistical Methods in Medical Research*, *11*(6), 469-490.

O'Sullivan, O. P. (2015). Dialogues in Philosophy, Mental and Neuro Sciences.

Pearson, K. (1897). On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proc. Roy. Soc.* London Ser. A 60, 489-498.

Pharmaceutical Manufacturers Association (1989). *America's Pharmaceutical Research Companies: A Cost Effective Source of Important New Medicines*. November.

Polson, N.G., & Tew, B. V. (2000). Bayesian portfolio selection: An empirical analysis of the S&P 500 index 1970-1996. *Journal of Business & Economic Statistics,* 18. 164-173.

Pruitt, S. W., & Gitman, L. J. (1987). Capital budgeting forecast biases: evidence from the Fortune 500. *Financial Management*, 46-51.

Roussel, P. A., Saad, K. N., & Erickson, T. J. (1991). *Third generation R&D: managing the link to corporate strategy*. Harvard Business Press.

Salo, A., & Hämäläinen, R. P. (2010). Preference programming–multicriteria weighting models under incomplete information. In Handbook of Multicriteria Analysis (pp. 167-187). Springer Berlin Heidelberg.

Schlapp, J., Oraiopoulos, N., & Mak, V. (2015). Resource allocation decisions under imperfect evaluation and organizational dynamics. *Management Science*, *61*(9), 2139-2159.

Shakhsi-Niaei, M., Torabi, S. A., & Iranmanesh, S. H. (2011). A comprehensive framework for project selection problem under uncertainty and real-world constraints. *Computers & Industrial Engineering*, *61*(1), 226-237.

Shih, Y. C. T. (2003). Bayesian approach in pharmacoeconomics: relevance to decision-makers. *Expert review of pharmacoeconomics & outcomes research*, *3*(3), 237-250.

Smith, J. E., & Winkler, R. L. (2006). The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science*, *52*(3), 311-322.

Snedecor, George W. (1946). Statistical Methods, 4th ed. Ames, 10: *Iowa State College Press*.

Soyer, R., & Tanyeri, K. (2006). Bayesian portfolio selection with multi-variate random variance models. *European Journal of Operational Research,* 171, 977-990.

Thaler, R. H. (1992). *The Winner's Curse: Paradoxes and Anomalies of Economic Life*. Free Press, New York.

Thomas, C. (2011). Too many products: Decentralized decision making in multinational firms. *American Economic Journal: Microeconomics*, *3*(1), 280-306.

Toppila, A., Liesiö, J., & Salo, A. (2011). A Resource Allocation Model for R&D Investments: A Case Study in Telecommunication Standardization. In *Portfolio Decision Analysis* (pp. 241-258). Springer New York.

Vilkkumaa, E., Liesiö J., Salo, A. (2014). Optimal strategies for selecting project portfolios using uncertain value estimates. *European Journal of Operational Research*, 233, 772-783.

Vilkkumaa, E., Salo, A., Liesiö, J., & Siddiqui, A. (2015). Fostering breakthrough technologies—How do optimal funding decisions depend on evaluation accuracy?. *Technological Forecasting and Social Change*, *96*, 173-190.

Wachs, M. (1989). When planners lie with numbers. *Journal of the American Planning Association*, *55*(4), 476–479.

Wachs, M. (1990). Ethics and advocacy in forecasting for public policy. *Business and Professional Ethics Journal*, *9*(1–2), 141–157.

Wacker, J. G. (1998). A definition of theory: research guidelines for different theory-building research methods in operations management. *Journal of Operations Management*, 16(4), 361-385.

Wang, J., Xu, Y., & Li, Z. (2009). Research on project selection system of pre-evaluation of engineering design project bidding. *International Journal of Project Management*, *27*(6), 584-599.

Wey, W. M. (2008, October). A multiobjective optimization model for urban renewal projects selection with uncertainty considerations. In *Natural Computation, 2008. ICNC'08. Fourth International Conference on* (Vol. 6, pp. 423-427). IEEE.

Winkler, R. L. & Barry, C. B. (1975). A Bayesian model for portfolio selection and revision. *The Journal of Finance*, 30. 179-192