PUBLIC R&D PROJECT PORTFOLIO SELECTION PROBLEMS UNDER EXPENDITURE UNCERTAINTY AND SECTORAL BUDGET BALANCING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MUSA ÇAĞLAR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
INDUSTRIAL ENGINEERING

JULY 2016

Approval of the thesis:

PUBLIC R&D PROJECT PORTFOLIO SELECTION PROBLEMS UNDER EXPENDITURE UNCERTAINTY AND SECTORAL BUDGET BALANCING

submitted by **MUSA ÇAĞLAR** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Industrial Engineering Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences** ________________

Prof. Dr. Murat Köksalan
Head of Department, **Industrial Engineering** ________________

Assoc. Prof. Dr. Sinan Gürel
Supervisor, **Industrial Engineering Department, METU** ________________

Examining Committee Members:

Assoc. Prof. Dr. Canan Sepil
Industrial Engineering Department, METU ________________

Assoc. Prof. Dr. Sinan Gürel
Industrial Engineering Department, METU ________________

Assoc. Prof. Dr. İsmail Serdar Bakal
Industrial Engineering Department, METU ________________

Assist. Prof. Dr. Mustafa Alp Ertem
Industrial Engineering Department, Çankaya University ________________

Assist. Prof. Dr. Salih Tekin
Industrial Engineering Department, TOBB ETÜ ________________

Date: ________________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: MUSA ÇAĞLAR

Signature :

ABSTRACT

PUBLIC R&D PROJECT PORTFOLIO SELECTION PROBLEMS UNDER EXPENDITURE UNCERTAINTY AND SECTORAL BUDGET BALANCING

ÇAĞLAR, MUSA
Ph.D., Department of Industrial Engineering
Supervisor : Assoc. Prof. Dr. Sinan Gürel

July 2016, 144 pages

In this dissertation, we deal with the two issues that exist in the practice of public R&D funding program management. First one is *the underutilization of the funding budget* owing to several sources of expenditure uncertainty. Project cancellations and spending uncertainty of successfully completed projects cause to funding budget underutilization. In the first and second part of the dissertation, we propose new approaches to enhance the utilization of the total funding budget. Specifically, in the first part of the dissertation, we focus on incorporation of project cancellations into the decision making process. In the second part of the dissertation, we deal with stochastic modeling of the expenditures of both canceled and successfully completed projects. We showed that modeling those uncertainties can significantly improve the utilization of the funding budget. Increased public R&D budget utilization will help to support more R&D projects and hence achieve higher socio-economic impact. Second issue in the practice of R&D funding management is *balancing of the total funding budget among sectors*. In the third part of the dissertation, we deal with incorporation of sectoral impact assessment results into the decision making process to ensure sectoral budget balancing. We compare our proposed approach with some alternative policy options. We showed that proposed model enhances total impact of funding budget by keeping relative budget balancing among sectors.

Keywords: public R&D project portfolio selection, budget underutilization, project cancellations, sectoral balancing, sectoral impact assessment, mixed integer second-order cone programming, chance constrained stochastic programming, dynamic programming

ÖZ

HARCAMA BELİRSİZLİĞİ VE SEKTÖREL BÜTÇE DENGELEME ALTINDA KAMU AR-GE PROJE PORTFÖY SEÇİMİ PROBLEMLERİ

ÇAĞLAR, MUSA
Doktora, Endüstri Mühendisliği Bölümü
Tez Yöneticisi : Doç. Dr. Sinan Gürel

Temmuz 2016 , 144 sayfa

Bu tezde, kamu Ar-Ge fonlama programı yönetimi uygulamasında var olan iki sorun ile ilgileniyoruz. Birinci sorun, fon bütçesinin, Ar-Ge projelerinin bir takım harcama belirsizliği sebeplerinden dolayı yetersiz kullanımıdır. Proje iptalleri ve başarıyla tamamlanan projelerin harcama belirsizliği fon bütçesinin yetersiz kullanımına yol açmaktadır. Tezin birinci ve ikinci kısmında, toplam fon bütçesinin kullanımını artırmak için yeni yaklaşımlar öneriyoruz. Özel olarak, tezin birinci kısmında, proje iptallerinin karar verme sürecine dahil edilmesine odaklanıyoruz. Tezin ikinci kısmında, hem iptal olan hem de başarıyla tamamlanan projelerin harcamalarının rassal olarak modellenmesiyle ilgileniyoruz. Bu belirsizlikleri modellemenin fonlama bütçesinin kullanımını önemli ölçüde iyileştirebileceğini gösterdik. Artırılan kamu Ar-Ge bütçe kullanımı daha fazla Ar-Ge projesinin desteklenmesine yardımcı olacak ve böylece daha yüksek sosyo-ekonomik etki meydana getirecektir. Ar-Ge fonlama yönetimi uygulamasındaki ikinci sorun toplam fonlama bütçesinin sektörler arasında dengeli dağıtılmasıdır. Tezin üçüncü kısmında, sektörel etki değerlendirmesi sonuçlarının sektörel bütçe dengelemesini sağlamak için karar verme sürecine dahil edilmesiyle ilgileniyoruz. Önerdiğimiz yaklaşımı bazı alternatif politika opsiyonları ile karşılaştırıyoruz. Önerdiğimiz modelin görece sektörler arasında bütçe dengelemesini koruyarak fonlama bütçesinin toplam etkisini artırdığını gösterdik.

Anahtar Kelimeler: kamu Ar-Ge proje potföy seçimi, bütçenin yetersiz kullanımı, proje iptalleri, sektörel dengeleme, sektörel etki değerlendirmesi, karışık tam sayılı ikinci derece konik programlama, şans kısıtlı rassal programlama, dinamik programlama

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the advice, guidance and support of many nice people around me.

First and foremost, I would like to express my special gratitude and thanks to my PhD advisor Sinan Gürel for his invaluable support, encouragement, supervision and beneficial suggestions throughout this research. His moral support and continuous guidance enabled me to complete my work successfully. He is an excellent supervisor and a great mentor. He gives special importance to his Phd supervision. He is always reachable and arranged regular meetings for my research. In terms of academic research, I have learned many new perspectives from him. I am also grateful for his help and advices on my career whenever I need to make critical decisions.

I would like to thank my committee members, İsmail S. Bakal and M. Alp Ertem. Their invaluable advices played an important role in the quality of this research. I will never forget moral support of İsmail S. Bakal during my tough times in Phd program. I specifically thank to M. Alp Ertem for accepting the member of my Phd committee.

I would also like to thank the other two members of my defense committee, Canan Sepil and Salih Tekin for accepting the member of my defense jury and for their insightful comments.

I am grateful to Murat Köksalan for his advices during my Phd study. I would like to thank Özgür Özpeynirci for motivating me to become a researcher. I appreciate the kindness and support of my friend and colleague Ahmet Silav. It has always been a great pleasure working with you in METU IE department and TÜBİTAK.

I thank my mom, dad, and sister for your unconditional love. I could not have succeeded without feeling your support and encouragement. Last but not least, I would like to express my love and gratitude to my beloved wife, Özden, for her unfailing love and emotional support. I am grateful for her being a part of my life and sharing my happiness and sorrow. I should also mention the joy and happiness that our little kid, Burak, brought into our lives. This dissertation is dedicated to them.

TABLE OF CONTENTS

List of Tables

TABLES

List of Figures

FIGURES

LIST OF ABBREVIATIONS

AHP	Analytic Hierarchy Process
ANP	Analytical Network Process
Avg	Average
BEB	Bery Esseen Bound
BSC	Balance Scorecard
CCSP	Chance Constrained Stochastic Programming
CLT	Central Limit Theorem
CPU	Central Processing Unit
CS	Chemical Sciences
DEA	Data Envelopment Analysis
DM	Decision Maker
DP	Dynamic Programming
EMS	Engineering and Material Sciences
ES	Earth Sciences
GP	Goal Programming
IP	Integer Programming
IRR	Internal Rate of Return
IS	Information Sciences
LP	Linear Programmming
LS	Life Sciences
KM	Knapsack Model
KP	Knapsack Problem
MAUT	Multi Attribute Utility Theory
MILP	Mixed Integer Linear Program
MINLP	Mixed Integer Nonlinear Program
MISOCP	Mixed Integer Second-Order Cone Program
MPS	Mathematical and Physical Sciences
MS	Management Sciences

MSTI	Ministry of Science, Technology and Industry
MOEA	Multi-objective Evolutionary Algorithm
NIH	National Institutes of Health
NLP	Nonlinear Programmming
NPV	Net Present Value
NSF	National Science Foundation
PI	Principal Investigator
R&D	Research and Development
RDPPS	R&D Project Portfolio Selection
ROI	Return on Investment
SOCP	Second Order Cone Programming
TÜBİTAK	The Scientific and Technological Council of Turkey

CHAPTER 1

INTRODUCTION

In today's competitive and rapidly changing world, public research and development (R&D) activities play a crucial role for a nation's long term economic welfare and competitiveness. Government funding agencies provide considerable amount of public funds through funding programs to R&D projects conducted by researchers in universities, public R&D institutions, and private sector companies. Public R&D activities contribute to talent development, increase of knowledge base, development of new scientific instruments/methods, development of new products/processes, and generation of spin-off companies (Salter and Martin (2001)). Besides, Kroll and Stahlecker (2012) also emphasize the significance of public R&D funding as follows:

> ... "the motivation to publicly fund research and development is to yield added value to society and, in doing so, remain accountable to the taxpayer. While the link between funding allocated and actual added value felt by society will inevitably turn out to be time-lagged and indirect it should nonetheless remain the ultimate rationale for public intervention in the field of science and technology".

Therefore, governments, being aware of this fact, invest large amount of public funds to R&D activities. For example, funding budget of The Scientific and Technological Research Council of Turkey (TÜBİTAK) reached to 1 billion Turkish Lira in 2012. Besides, government funding agencies usually receive a large number of project proposals for national R&D programs to be supported out of the same pool of funds. For instance, national fundamental research funding program of Turkey, known as TÜBİTAK 1001 program, received 3218 projects for the two calls at different terms in 2012 (TÜBİTAK (2012), p. 47), which means that on the average 1609 proposals received per call. Similar national level funding programs in United States, Russia and Mexico also receive large sets of proposals as recently indicated in Litvinchev et al. (2010). This problem environment makes public sector R&D project selection decisions more complex than private sector setting, in which selection decisions are made out of much smaller sets of projects. Therefore, the large-scale portfolio optimization models with novel formulations can contribute to the decision making process and decision makers (DMs) can gain new managerial insights.

In this dissertation, we consider a one stage public R&D project portfolio selection (RDPPS) problem. In this problem, we have an R&D funding program. There is a maximum project duration in this program. Calls are periodically opened under this program. Once a call is announced, a large set of R&D project proposals apply to the call. After, submitted project proposals are officially checked for eligibility requirements of the program. Eligible project proposals are evaluated according to the evaluation criteria by peer reviewers in panel meetings. Announced call has a total available budget B. The budget b_i and the score s_i of each project i is determined in panel meetings. The budget b_i of project i covers all of the planned expenditure of project i during its duration. The project duration of a project i can be any value that is smaller than or equal to maximum project duration of the program. In this problem setting, the objective is to maximize the total score of the selected projects while satisfying the total available budget (B) constraint.

For example, in Figure 1.1, we present a time line of one stage public R&D Project portfolio selection problem setting for an R&D Program with project duration of maximum 3 years. Since maximum project duration of this example program is 3 years, the call budget B puts a bound on sum of the planned expenditures (budgets) of supported projects during 3 years. The funding decisions are made at the beginning of period $t = 0$ according to the scores and budgets of the projects, after the panel meetings are conducted. Consider two projects i and j, that are selected at the period $t = 0$. Assume that project i has a duration of 2 years and project j has a duration of 3 years. Therefore, funding budget of projects i and j cover all the planned expenditures during their project life. After projects are selected, contracts are signed between funding agency and the project principal investigator (PI)s. When the projects are started, only partial funding budgets (i.e. partial year or a full year

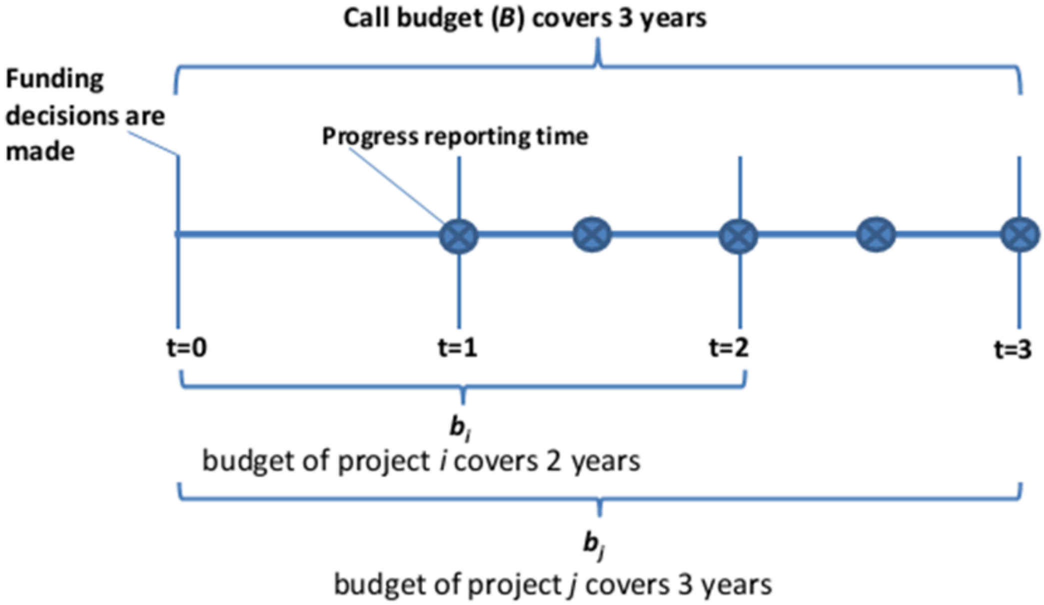

Figure 1.1: One Stage Public R&D Project Portfolio Selection Problem for an R&D Program with project duration of maximum 3 years

budgets) are transferred to the projects. There are reporting time points, in which projects are reviewed for compliance with the terms and conditions of the program. Decisions regarding to project cancellations and transferring partial funding budgets are made at those reporting time points by examining progress reports.

We are concerned with the two practical issues that exist in the public RDPPS problem. The first one is the *underutilization of available funding budget* due to several sources of expenditure uncertainty. The second one is the *sectoral balancing of available funding budget*. We first discuss practical motivations of those two issues in Section 1.1 and Section 1.2, respectively. Then, we give the outline of this dissertation in Section 1.3.

1.1 Underutilization of the funding budget

Public R&D organizations have limited funding budget appropriations, however, most of the time the funding budget cannot be spent fully and the budget underutilization occurs. When we look at the Turkey example, the leading R&D funding organization is TÜBİTAK. Between the years of 2010 and 2012, reported appropriations, expenditures and remainders of TÜBİTAK funding budget are presented in Figure 1.2. For instance, the funding budget remainder was 172 Million Turkish Lira in 2010 and it reached to 418 Million Turkish Lira in 2012. A recent study conducted by Ministry of Science, Technology and Industry (MSTI) also highlights that utilization of funding appropriation of TÜBİTAK is 58% in 2012 (Eser (2014)).

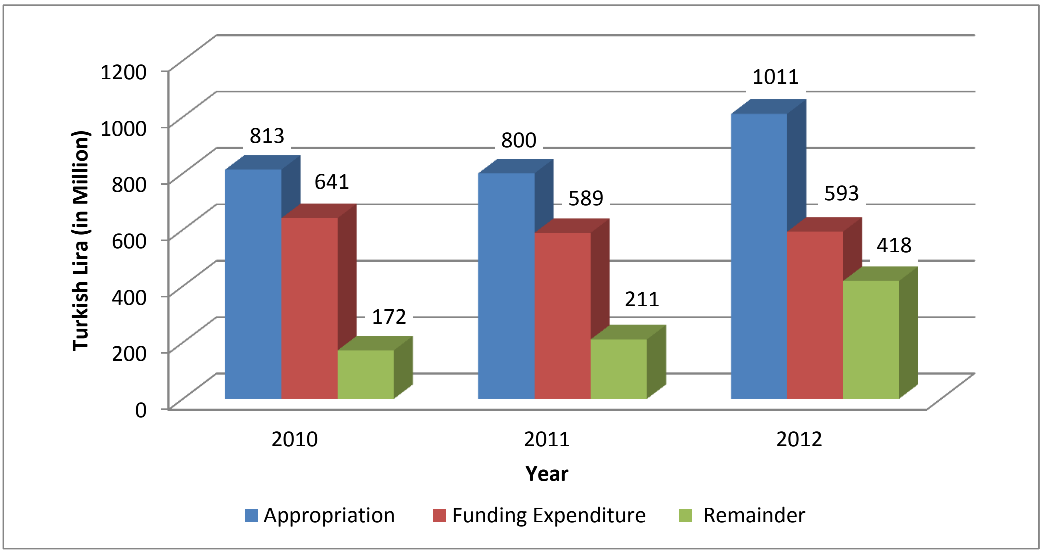

Figure 1.2: TÜBİTAK Funding Between 2010 and 2012
Source: Author's own compilation from activity reports of TÜBİTAK

DMs of public R&D funding agencies should carefully consider the factors that bring about underutilization of the funding budget. Practical life reveals that there are two main impediments. First one is the cancellation of a significant number of R&D projects before spending most of their budgets. R&D activities are inherently subject to uncertainties that can arise during the execution of an R&D project. Principal investigator (PI) and other project members may fail to comply with the approved project plan and there might also be some ethical reasons. Rules and conditions for cancellation of projects are usually arranged in the R&D program regulations (see NSF (2005); NIH (2013)). Canceled projects are regarded as unsuccessful and their expected scientific and technological benefits are not realized. A funding organization usually suspends a project that fails to comply with the terms and conditions of the R&D program. Unless a corrective action is implemented, the funding organization can cancel the project. As author of the dissertation is employed in an R&D funding organization, we can roughly estimate that 5-10% of the awarded projects may get into the cancellation process. The second factor is that a significant number of funded projects end up successfully while providing their expected benefits without spending their whole budget. Since planning of R&D activities is a complex process, estimating the budget of an R&D project is difficult. Although, detailed guidelines are provided for budgeting of R&D activities, researchers may overestimate the real budget of an R&D project.

The budget underutilization situation is apparent especially in emerging countries such as Turkey, which have been experiencing prominent public funding increases recently. Many new researchers have been applying to funding programs and yet have become familiar with formulating R&D projects. Budget releases due to the expenditure uncertainty could make a great opportunity for the relatively high scored projects, that were not awarded during the selection processes. Every R&D project successfully conducted and finalized brings its special scientific and socio-economic value to society. Therefore, DMs of funding organizations can benefit from considering the underutilization situations when they allocate funding budget to R&D projects.

In Chapter 3 and Chapter 4, we consider aforementioned sources of budget spending uncertainty in the public RDPPS problem to improve the utilization of the total funding budget. More specifically, in Chapter 3, we deal with incorporating project cancellations into decision making process. In Chapter 4, we deal with modeling of both cancellations and underspending of successfully completed projects. We show that considering those uncertainties can significantly improve the utilization of funding budget. Increased public R&D budget utilization will help to support more R&D projects and hence achieve higher socio-economic impact.

1.2 Sectoral balancing of the funding budget

In Chapter 5, we propose a sectoral budget balancing approach by employing the results of so-called "impact assessment" studies. In science, technology, and innovation policy context, impact assessment of an R&D funding program refers to measurement of scientific, economic, and social effects of an existing funding program in terms of predefined impact criteria and program objectives. Hence, impact assessment studies provide insights about added value of public R&D funding programs to society. Supporting R&D projects with public funds is regarded as a long term government investment to enhance competitiveness of a country. There is a significant interest to "impact assessment" studies in executive government departments to distribute limited funds effectively. Therefore, DMs of funding authorities encounter two major policy concerns in R&D funding management. Firstly, the impact assessment of an R&D funding program, as an evidence-based policy making tool, has recently received high level of attention from policy makers (Cervantes (2007)). Impact assessment studies assist to improve program contents (i.e. scope, funding mechanism, etc.) as well as to distribute public funds effectively.

Second prevalent policy concern is related to sectoral balancing of R&D funding allocations. Sector notion might express different meanings in various contexts. In our terminology, sector refers to funding program specific categorization of projects according to thematic areas, scientific disciplines, technological fields, industrial or economical classifications, etc. In practice, there are many classification systems for sectoral breakdowns. Funding authorities could adopt variety of them according to funding program scope and objectives. An R&D funding program could have different impacts on different sectors due to sectoral dynamics and different technological sophistication of sectors. Therefore, DMs of funding agencies want to effectively distribute limited program budget over different sectors according to sectoral impact assessment results. However, sectoral balancing does not mean to disproportionate allocations because science, technology, and innovation policies suggest that, especially in curiosity driven bottom up funding programs such as 1001 scientific and technological research projects funding program of TÜBİTAK, some level of expertise and critical mass should be maintained and developed in all sectors. In addition, this kind of funding programs receive significant number of project proposals from all sectors and disproportionate allocations could raise serious objections from research communities. For that reason, imbalanced project portfolios usually aren't approved by the DMs who assert that some sectors predominantly receive most of the funds, whereas other sectors receive very little (Karsu and Morton (2014)). On the other hand, some sectors that have high impact assessment values could get more allocations; whereas, sectors that have low impact assessment values could receive less allocations. Notion of more and less allocations could be nebulous. Therefore, in Chapter 5, we propose a flexible analytical approach to assist sectoral balancing decisions in the light of given sectoral impact assessment values.

As a result of those aforementioned policy shifts, government agencies, especially in emerging countries, have been strategically reorganized and new departments solely responsible for impact assessment studies have emerged. For example, MSTI of Turkey has recently founded the "Department of Impact Assessment". This new department will be responsible for impact assessment of all public R&D and innovation support programs, that exist in different funding organizations. It will share impact assessment results with policy makers. It will provide policy insights regarding revision and improvement of funding mechanisms. The department will also be responsible for developing analytical and accountable R&D resource allocation models in the light of impact assessment findings. Therefore, we believe that the proposed model in Chapter 5 will contribute to strategic allocation of R&D funding resources by government departments and provide policy implications about how impact assessment results could be incorporated into strategic fund allocation processes.

There is one issue that needs further elaboration. In panel review of R&D projects, there is usually a criterion called broader "impact". This criterion quantifies foreseen impact of a project to its sector by a panel peer review process. On the other side, sectoral impact assessment of a funding program refers to evaluation of realized outputs and results (by considering time lags of 5-10-15 years) of successfully completed projects in a funding program. For example, consider two projects from two different sectors such as earth sciences and information sciences. Assume that all attributes (all criterion scores, budget, number of researchers in the project etc.) of those two projects are the same. If we conduct an impact assessment of those two projects according to realizations of some output criteria after they finish, we might observe different results due to sectoral dynamics. Therefore, sectoral impact assessment studies deal with how impacts of different sectors could change according to their allocated resource allocations.

In Chapter 5, we incorporate sectoral impact assessment results into decision making process to ensure sectoral budget balancing in the public RDPPS problem. We compare our proposed approach with some alternative policy options. We show that proposed model enhances total impact of funding budget by keeping relative budget balancing among sectors.

1.3 Organization of the dissertation

The remainder of this dissertation is organized as follows.

The related literature review is presented in Chapter 2. Literature review consists of four sections. In the first section, we review studies on R&D project portfolio selection (RDPPS) problem. In the second section, we summarize the relevant research related to impact assessment based sectoral balancing in RDPPS problem. In the last two sections, we discuss chance constrained optimization and conic quadratic programming, respectively.

In Chapter 3, we address a public P-RDPPS problem with project cancellations. We consider two cases. In the first case, we assume that cancellation probability of a project cannot be assessed but the DM can estimate the number of projects that will be canceled. In the second case, we assume that for each project a cancellation probability can be assessed. For the first problem, we develop a mixed-integer linear programming (MILP) formulation and a dynamic programming (DP) algorithm. For the second problem, we develop a chance-constrained stochastic programming (CCSP) formulation that can be solved as a mixed-integer second-order cone program (MISOCP). Our computational results show that practical-size problems can be solved by the proposed solution approaches. In addition, we conduct additional analyses on proposed approaches to give managerial insights to the DM.

In Chapter 4, we formulate the expenditure of an R&D project with a mixture distribution by incorporating project cancellations and budget underspending of successfully completed projects. We develop a project portfolio selection model with a probabilistic budget constraint. We propose a solution method based on a normal approximation and a second-order cone programming. DMs of public funding agencies wonder quality of normal approximation to minimize risk of exceeding funding budget. Hence, we quantify convergence quality of normal approximation by applying Berry-Esseen theorem and propose ways to mitigate risk of budget constraint. Besides, we perform extensive analyses on the proposed model to give managerial implications to the DM. The proposed model can solve practical size instances in reasonable CPU times.

In Chapter 5, we consider impact assessment based sectoral budget balancing in the public RDPPS problem. We develop a two-stage model. In the first stage, the DM wants to maximize the total impact of the funding budget while assuring relative budget balancing among sectors. We propose a nonlinear objective function (i.e. parameterized social welfare function) in the first stage. We prove that nonlinearity in the objective function can be expressed by conic quadratic inequalities. In the second stage, the DM wants to maximize the total score of supported projects subject to alloted sectoral budgets. We propose a flexible decision making approach to show effect of sectoral allocation on various problem parameters. We solve an example problem with our proposed decision making approach. We demonstrate the value of the proposed approach by comparing it with some alternative policy options.

In Chapter 6, we present concluding remarks and future research directions.

CHAPTER 2

LITERATURE REVIEW

In this chapter, we review the studies related to our research. In particular, we review the literature on R&D project portfolio selection (RDPPS) problem in Section 2.1. Next, we review the research related to impact assessment based sectoral balancing in RDPPS problem in Section 2.2. We discuss chance constrained stochastic programming in Section 2.3 and second order cone programming in Section 2.4.

2.1 R&D Project Portfolio Selection Literature

In the RDPPS literature, one of the earliest extensive reviews is conducted by Baker (1974). After highlighting uncertain and unpredictable nature of R&D activities, he classifies the literature on RDPPS problem into three major categories:

(a) descriptive literature in which R&D project portfolio selection problem is considered as an organizational process flow model,

(b) estimation problems in which uncertainty, risk, cost, time to completion and interactions are discussed,

(c) normative models in which mathematical programming models and tools are described.

In addition, he underlines the importance of the normative models and defines the RDPPS problem as follows:

> “ Given set of alternatives (proposed projects) which require common scarce resources, determine that allocation of the resources to the alternatives which will maximize the benefit contribution (value) of the resulting program ”.

Second comprehensive overview is performed by Henriksen and Traynor (1999). They categorize the RDPPS methods as follows:

(a) Unstructured peer review,

(b) Scoring,

(c) Mathematical programming models (Linear Programming (LP), Nonlinear Programming (NLP), Integer Programming (IP), Goal Programming (GP), Dynamic Programming (DP)),

(d) Economic Models (Internal Rate of Return (IRR), Net Present Value (NPV), Return on Investment (ROI), Cost-Benefit Analysis, Option Pricing Theory),

(e) Decision analysis (Multi Attribute Utility Theory (MAUT), Decision Tree, Risk Analysis, Analytic Hierarchy Process (AHP),

(f) Interactive methods (Delphi, Q-sort, Behavioral Decision Aids, Decentralized Hierarchical Modeling),

(g) Artificial intelligence (Expert Systems, Fuzzy Sets),

(h) Portfolio optimization.

Besides, they emphasize that in the portfolio optimization method to R&D project selection, any combination of classified techniques indicated between (a)-(g) could be applied to find the best R&D portfolio. For instance, they state that integer programming models could be used to select the projects. Moreover, they also develop an improved scoring tool for the evaluation of R&D projects. Heidenberger and Stummer (1999) review quantitative approaches to RDPPS problem and classify methods as strategic management tools, benefits measurement methods and mathematical programming methods. Apparently, mathematical programming models have been specifically categorized by the review papers as a quantitative approach for the selection of R&D projects. We present the most relevant studies of RDPPS problem as follows.

Santhanam and Kyparisis (1995) address a general project selection model suitable for information technology and R&D domains in companies. Their model incorporates project failure risks, project benefits, project costs, several project resource requirements, and interaction among possible selected projects. They set three goals in terms of total failure risk, total benefit and total cost of selected projects. They define a risk score between 0 and 5 to quantify the likelihood of failure of a project and minimize the total risk of selected projects in their model. They develop a preemptive goal programming approach under various resource constraints and solve a real life example with 14 projects.

Coffin and Taylor (1996) develop an integrated RDPPS and scheduling problem. They state that private sector R&D managers may require an overall completion time threshold for the selected projects, and makespan of selected project portfolio might exceed that threshold. Therefore, they also incorporate scheduling decisions into their model. They also assume that probability of technical or commercial success of each project is known. They propose three objectives such as maximization of the expected return, maximization of probability of success of the portfolio, and minimization of the makespan. They develop a beam research heuristic for their proposed model. They solve an illustrative example with 20 projects under resource constraints. They set problem size of the illustrative example according to practice.

Beaujon et al. (2001) develop a mixed integer linear program (MILP) with a variety of resource and other constraints for the RDPPS problem. They assume that project values are uncertain. Therefore, they propose a method for quantifying the value of a project. They also discuss partial funding and implementation in their model. Moreover, they also conduct a sensitivity analysis of project values to make selection process more reliable. They solve an instance problem of a company with 390 projects by using the linear programming relaxation.

Meade and Presley (2002) underline the criticality of the RDPPS problem in companies and focus on project selection criteria. They state that R&D project evaluation model should be related to corporate strategy, qualitative aspects and risks, and needs of various stakeholders. Therefore, they introduce 15 evaluation criteria. Besides, they propose an analytical network process (ANP) for the RDPPS problem. Their proposed model admits the interactions among the selection criteria. They apply their proposed approach to select one of the two projects of a small technology company.

Ringuest et al. (2004) develop a two-parameter model for the selection of R&D projects in pharmaceutical companies. Their model depends on expected return and an intuitive measure of variation, called Gini coefficient. They assume zero or high economic return for the R&D projects. They assume that probability of high return is small (i.e. between 0.05 and 0.65). They generate all possible R&D project portfolio options with their expected return and Gini statistic. They apply their proposed method to a practical problem with 30 projects.

Eilat et al. (2006) decompose RDPPS problems into two main categories: static and dynamic problems. They state that in the dynamic setting, active projects that are already started are considered with proposed projects, and decision space includes both of them, namely continuity or cancellation of the active projects, and launching of the new projects. However, they state that in the static setting, only proposed projects are considered. They propose an extended data envelopment analysis (DEA) combined with a balance scorecard (BSC) approach for the static RDPPS problem with project interactions. Motivated by a governmental department responsible for selecting computer technology projects, they solve an example problem with two inputs and three outputs of 15 projects.

Medaglia et al. (2007) develop a multi-objective (i.e. minimize time-to-market and maximize economic return) mathematical model under resource constraints for the R&D and information technology project selection problems. They formulate cash flow and market share uncertainty of R&D projects with triangular, exponential, and Erlang random variables. They apply Monte Carlo simulation to deal with stochasticity. They also develop a multi-objective evolutionary algorithm (MOEA). They solve an example problem with 4 projects in a four-year planning horizon.

Koç et al. (2009) propose a multi-dimensional knapsack model for the RDPPS problem under budget and profit (i.e. net present value) uncertainty. They first propose two heuristics that depend on parametric changing of budget for forming priority list of projects. They also develop a stochastic programming approach with scenarios. They solve an example problem of a company with 16 projects in a five-year planning horizon.

Solak et al. (2010) discuss the need for more realistic RDPPS models for R&D intensive companies due to the full range of complexity. They define annual performance levels (i.e. financial returns) and required investment levels of R&D projects as discrete random variables and assume that associated probabilities are known. They develop a scenario based multi-stage mathematical model for the RDPPS problem. They simplify and solve the multi-stage model with a two-stage approach in every period (realizations with future decisions). They apply the Lagrangian relaxation method and a feasible dual conversion technique (to make the relaxed problem feasible). Their model considers optimal allocation of budgets to a small set of projects (i.e. 5 and 10 projects) in each time period by evaluating the realizations of annual financial returns. Their model also includes dynamic cancellation decisions of projects by periodically comparing their financial returns with budget expenditures.

Duzgun and Thiele (2010) develop a robust optimization model with multiple ranges to a RDPPS problem, in which cash flows are uncertain. They apply their proposed model to the example problem of Koç et al. (2009). Besides, they state that RDPPS problems possess high level of uncertainty. They also mention that probabilistic approaches are commonly adopted although it is too difficult to quantify distributions of unknown parameters and probabilities of project success. Hence, motivated by this observation, we develop a model with unknown cancellation probabilities in the first part of the Chapter 3.

Litvinchev et al. (2010) propose a fuzzy bi-objective optimization model for large-scale (i.e. more than a thousand projects) public RDPPS problem under resource constraints. They focus on the trade-off between objectives and generation of non-dominated solutions. In their model first objective is the maximization of total expected value of selected projects and second objective is the maximization of number of selected projects. However, cardinality type objective (i.e. maximization of number of selected projects) implies that poorly valued but low-budget projects may have a chance for getting selected and such a structure can be criticized by DMs. Therefore, in this dissertation, we do not consider incorporation of similar cardinality type objective into our proposed models.

Shakhsi-Niaei et al. (2011) propose a two-stage model for the RDPPS problem under uncertain project values. First stage provides the ranking of projects according to the uncertain project values, and second stage applies an integer programming model under variety of constraints to select the projects. They solve an illustrative example of a research center with 40 information technology projects.

Hassanzadeh et al. (2014) develop a robust optimization model to revise and reschedule a pharmaceutical R&D project portfolio. They obtain the net present value of each project by considering multiple phases of the drug development. They assume an interval uncertainty for the cost estimates of the projects. Their uncertainty modeling is similar to cash flow modeling of Duzgun and Thiele (2010). They also assume that duration, start, and possible tardiness of projects are known. They solve an example problem with 25 pharmaceutical R&D projects.

There are also a few isolated studies which propose models to maximize the success of R&D projects in the competitive industries. For example, Gerchak and Kilgour (1999) consider the failure probability of an R&D project and develop models for funding of parallel teams to increase the achievement probability of the single R&D project. Gerchak and Parlar (1999) study the strategic allocation of limited resources to the R&D activities when competitors' budget allocation to these R&D activities is unknown. They develop a model to find the optimal allocation between two activities conditional on the competitor's allocation to maximize the expected profit.

We notice that most of the studies in the literature focus on private sector RDPPS problems. Research on public (i.e. government sponsored) RDPPS problems is scant. RDPPS process in public funding organizations differs from RDPPS in private sector companies. We refer the interested reader to Bozeman and Rogers (2001) for a discussion on the differences between public and private sector domains.

A significant difference noted by Bozeman and Rogers (2001) is that RDPPS in private sector companies is a dynamic process, whereas RDPPS in public funding authorities has a static nature. They state that private sector companies can dynamically allocate/reallocate the resources between active projects (i.e. new and existing projects) and may accelerate or cancel some existing projects. They underline that public funding organizations support R&D projects via grants and contracts; hence, it is not possible to intervene them rapidly.

Another difference is related to the number of projects in an R&D project portfolio. Existing studies of RDPPS problem in private sector companies deal with small numbers of projects (i.e. between 2-40 projects), except Beaujon et al. (2001), whose example problem consists of 390 projects. On the other hand, the number of proposals applying to nationwide government R&D funding programs such as TÜBİTAK 1001 scientific and technological research projects funding program is in the scale of thousands. For example, a total of 3218 projects applied to two calls (i.e. the average 1609 projects per call) of this program in 2012 (TÜBİTAK (2012), p. 47). Litvinchev et al. (2010) state that large sets of project proposals exist in the national R&D funding programs of Mexico, Russia and United States. As a result of that, number of active funded (i.e. new funded and existing funded projects that are currently going on) projects in public funding organizations could reach to more than one thousand projects. For instance, the number of active funded projects in TÜBİTAK 1001 program was 1506 in 2012 (TÜBİTAK (2012), p. 45). Hence, dynamic allocation/reallocation and acceleration/cancellation decisions are impractical and static one-stage models are more convenient for public funding organizations. Therefore, in this dissertation, we develop mathematical programming models by considering static decision making environment of the public RDPPS problem.

To the best of our knowledge, in the public RDPPS literature, the project cancellations were not considered before. In Chapter 3, we deal with the uncertainty, that some awarded public R&D projects will be canceled during implementation. We incorporate project cancellations into the project selection process so that the limited amount of program budget can be effectively utilized. In Chapter 4, we develop a new approach to model underspending behavior of both canceled and successfully completed projects. We observed that proposed models in Chapter 3 and Chapter 4 significantly enhance utilization of the funding budget. To the best of our knowledge, we are the first to introduce the notion of underutilization of funding budget into the public RDPPS literature.

In the next section, we present the relevant literature related to the problem studied in Chapter 5.

2.2 Impact Assessment Based Sectoral Balancing Literature

Relevant work of the proposed model in Chapter 5 depends on three streams of research. The first one is the impact assessment literature, the second one is the sectoral balancing in RDPPS literature, and the last one is the literature on mathematical modeling with social welfare objective functions.

2.2.1 Impact Assessment

Many quantitative and qualitative impact assessment methods have been developed in the literature. Capron (1992) suggests mix of quantitative and qualitative methods for comprehensive impact assessment and reliability of results. Impact assessment studies also provide sectoral assessment details to assist policy makers. We refer the interested reader to Hughes and Martin (2012) for an extensive discussion on sectoral impact assessment findings of existing literature and their practical implications for government departments. There are diverse studies in the literature; hence, we cite relevant ones.

There are some intriguing studies which examine impact of general academic research in terms of defined output criteria on sectoral basis without controlling R&D funding programs. For instance, Jaffe (1989) finds a significant impact of academic research on patenting behavior of companies. He also investigates impact of five categorized technical areas (i.e. drugs and medical technology, chemical technology, electronics, optics, and nuclear technology, mechanical arts and other sectors) on patenting and show that patenting pattern of companies can change over different technical areas. Mansfield (1998) studies impact of academic research on drug and medical product, information processing, chemical, electrical, instruments, metals, and oil industries. He finds that percentage of new products and processes varies across different industrial sectors. Cohen et al. (2002) examine impact of public research activities over thirty four manufacturing sectors in terms of public research outputs such as research findings, prototypes, and new instruments/techniques. They find that impact of public research to industrial R&D varies across different sectors. Besides, they also investigate impact of ten scientific disciplines (i.e. biology, chemistry, physics, mathematics, computer science, material science, medical and health science, chemical engineering, electrical engineering, mechanical engineering) over thirty four manufacturing sectors. They indicate that there are clear differences among scientific disciplines in terms of their contribution to different manufacturing sectors. Moreover, they analyze how major industrial R&D projects could benefit from information sources of public research. They emphasize that publications, scientific reports, personal knowledge exchange, meetings/conferences, consulting, recently hired graduates, joint and cooperative ventures, patents, and licenses can contribute to industrial R&D activities. However, they find that contribution level of each information source is different.

There are studies which consider impact assessment of specific R&D funding programs to assist DMs on the effective allocation of limited public resources. For example, Lee et al. (2009) measure impact of six public R&D funding programs in South Korea by using following criteria; amount of funding, number of Ph.D. researchers, number of published papers, number of patents, and number of graduate students. They apply a data envelopment analysis method and obtain an aggregate measure between zero and one for each program to compare six R&D funding programs. They state that calculated aggregate scores can be used while allocating scarce resources to R&D funding programs. However, they do not provide any analytical method for this strategic allocation process. Besides, their model does not measure impact of R&D funding programs on sectoral level. Wang et al. (2013) have recently evaluated impacts of eighteen national funding programs of China by using expert judgments with paying special attention to sectoral (i.e. scientific disciplines) differences. They select 756 leading experts from seven scientific disciplines (i.e mathematical and physical sciences (MPS), chemical sciences (CS), life science (LS), earth sciences (ES), engineering and material sciences (EMS), information sciences (IS), and management sciences (MS)) and obtain valid responses from 460 leading experts. In their evaluation, each expert provides contribution on behalf of his/her scientific discipline.

They apply a vague set methodology and find that each program impact is different across seven scientific disciplines. They calculate an aggregate score between zero and one for each program by combining its sectoral impact scores. Moreover, they cluster eighteen programs according to their aggregate scores. Although they imply that obtained impact assessment values could be incorporated into budget distribution process, they do not give any technique for this significant budget allocation process.

A recent European Union FP7 project report point out that many impact assessment studies have been conducted so far; however, methodological incorporation of impact assessment results into policy formulation is still missing (EU- (2011), pp. 13-14). Therefore, in this study, we deal with how sectoral impact assessment values, calculated according to some qualitative or quantitative approach, could be analytically incorporated in to a public R&D project portfolio selection model.

2.2.2 Sectoral Balancing in RDPPS

In the R&D project portfolio selection literature, sectoral balancing concerns are incorporated into mathematical programming models in two ways. First one is defining sectoral balancing constraints so that resulting solutions provide desired balance of the DM. For example, Beaujon et al. (2001) define balance target constraints in a private sector RDPPS problem. They formulate those constraints by setting minimum and maximum fraction of funding to different project categories. Mavrotas et al. (2006) apply policy goal of balanced assignment concept by putting constraints to minimum and maximum number of selected projects in different sectors. Litvinchev et al. (2010) point out sectoral balancing of R&D projects in a public RDPPS problem. They assume that minimum and maximum amount of funding to different project areas are known and ensure sectoral balancing by introducing mathematical constraints.

Second one is defining an objective (i.e. minimize some deviation type measure from given balanced share of each sector) and presenting the non-dominated solutions to the DM to assess trade-off between total score objective and sectoral balancing objective. For instance, Stewart (1991) defines balance as desired proportion of resources allocated to different project categories in a private sector RDPPS problem. They assume that negative deviation from balance is more significant than positive deviation and formulate a nonlinear objective to minimize deviations accordingly. Karsu and Morton (2014) study balance concerns of R&D projects according to different project categories in a public RDPPS problem. They assume that balanced share of each project category in terms of budget is known. They introduce four imbalance indicators similar to Stewart (1991) and optionally minimize one of those balance indicators in the public RDPPS problem.

Existing approaches assume that balancing information (i.e. budget share of each sector, minimum and maximum amount of funding etc.) is perfectly known a priori by the DM. However, in our problem setting, the DM wants to decide on sectoral budget allocations according to sectoral impact assessment values. It is difficult to determine balanced budget share of each sector a priori without considering any measure about sectors. For instance, there has been a significant interest among government DMs about how to distribute funds to sectors in an R&D funding program. The author of the dissertation was asked to develop heuristic approaches for sectoral budget allocations in the absence of sectoral impact assessment results during his professional working life in TÜBİTAK. Besides, unlike existing bi-objective studies, sectoral budget allocation objective overrides total score objective because budget allocation decisions need to be made on a strategic level to enhance socio-economic impact of limited funding budget. Therefore, we model public RDPPS problem as a two-stage model. In the first stage, we deal with sectoral budget allocations. In the second stage, we consider maximization of total score of selected projects under given sectoral budgets. On the other side, sectoral budget balancing should be integrated with project selection decisions to give flexibility to the DM for evaluating how sectoral budget balancing can affect various sectoral decisions. Due to sectoral characteristics, sectors' reaction to various allocations might be different. For example, sectoral total score, average score and budget of selected projects in each sector, number of selected projects in each sector, and success rate in each sector and their relationships with each other could provide beneficial insights to the DMs.

2.2.3 Mathematical Modeling with Social Welfare Objectives

In the welfare economics literature, parameterized social welfare functions deal with fair distribution (i.e. balancing) of limited resources among the players. In this context, the player refers to problem specific entities such as sectors, regions, nations, clients, companies, commodities, etc. We refer the interested reader to Mas-Colell et al. (1995), (pp.825-829) for the notion of social welfare functions in the welfare economics.

Incorporation of social welfare objective functions into mathematical programming models has been recently studied by several researchers. For instance, Bertsimas et al. (2012) have used parameterized social welfare functions (i.e. α fairness scheme) in an air traffic flow management problem to effectively distribute total delay reduction among airline companies. They discuss that proposed mathematical programming models for the air traffic flow management problem have not been implemented in practice. They highlight that practical validity of such models is criticized by DMs, because traditional objective function of those models does not fairly distribute optimized total delay reduction among airline companies. Traditional objective function means maximizing sum of each airline's utility (i.e. delay reduction). They also underline that traditional objective function finds optimum solution at the expense of some airline companies. Consequently, some airline companies receive no utility (i.e. zero delay reduction) under the optimum solution. They have introduced the price of fairness (balancing) concept, which we will discuss later in Chapter 5.

Note that similar balancing concerns could arise in many problem domains. For instance, in our problem setting, the DM wants to maximize the total impact of the available funding budget. However, the optimum solution favors the sector with the highest impact, which may lead to a situation where no funding is available for some sectors or areas. That's why we adopt parametrized social welfare functions in our proposed model.

Under the parameterized social welfare function, referred as α fairness scheme, the DM wants to maximize following social welfare function F_α, parameterized by $\alpha \in [0, \infty)$.

$$F_\alpha(u) = \begin{cases} \frac{\sum_{j \in M} u_j^{1-\alpha}}{1-\alpha} & \text{for } \alpha \geq 0,\ \alpha \neq 1 \\ \sum_{j \in M} \log(u_j) & \text{for } \alpha = 1 \end{cases} \tag{2.1}$$

where u_j is the utility (i.e. the objective function) of player j, M is set of players, and α is an inequality aversion parameter. Traditional objective function corresponds to $\alpha = 0$, and it is referred as the utilitarian objective function. Besides, those functions are nonlinear for $\alpha > 0$. Bertsimas et al. (2012) focus on developing theoretical bound on what they called price of fairness and do not provide any tractable solution approach for the nonlinear model. By using the historical data, they record total delay reductions for $\alpha = 0, 0.5, 1, 2, \infty$, and discuss balancing patterns of different α values among players (i.e. airline companies). However, in our problem setting, we conduct preliminary experiments on different instances under various impact values and it turns out that $0 < \alpha < 1$ holds and increasing α only 0.1 amount makes a significant difference on sectoral budget allocations. Thus, we give a tractable solution method for any α value between 0 and 1, by employing the recent advances in second order cone programming.

In Chapter 5, we introduce the impact assessment based sectoral balancing in RDPPS problem. From a practical point of view, to the best of our knowledge, this is the first study (i) that incorporates sectoral impact assessment results into a public R&D project selection model, and (ii) that comprehensively addresses sectoral allocation and analysis in a public R&D project portfolio. Besides, from a methodological point of view, to the best of our knowledge, our contributions to literature include the following: (i) we consider sectoral budget balances as decision variables, (ii) we employ parameterized social welfare functions for relative fairness concerns (i.e. balancing of funding budget among sectors according to their impact values) in an RDPPS model, and (iii) we develop a tractable solution approach for our parameterized nonlinear social welfare objective. In the next section, we discuss chance constrained stochastic programming.

2.3 Chance Constrained Stochastic Programming

In Chapter 3 and Chapter 4, we propose chance constrained stochastic programming (CCSP) models for our public RDPPS problem. DMs want to mitigate risk of exceeding available funding budget even in the evident underutilization environment because total available funding budget depends on public resources. CCSP is a commonly used approach to deal with random variables in mathematical programming. It is introduced by Charnes and Cooper (1959). CCSP provides risk mitigation by quantifying a probability level for violation of the chance constraint. We refer to Luedtke et al. (2010) for a recent review on CCSP.

From a methodological standpoint, when a chance constraint includes sum of many non-identical random variables, it becomes usually computationally intractable. The derivation of the quantile function (i.e. inverse cumulative distribution function) of sum of many non-identical random variables is challenging for many distributions. Nemirovski and Shapiro (2006) state that most of the single linear chance constraints are computationally intractable since checking feasibility of a solution is NP-hard. Normal approximation is one of the tractable methods in the literature. Therefore, we apply normal approximation for our CCSP models. We refer to Shapiro et al. (2009), (pp. 141-144) for discussion of normal approximation of chance constraints.

Gurgur and Morley (2008) propose a CCSP for the yearly selection of infrastructure projects in a space company. They propose multi attribute utility theory for the valuation of projects. They assume expenditure of a project follows a normal distribution. They formulate chance constraints for the satisfaction of available resource limits. They solve a practical example with 267 projects by using commercial Excel Premium Solver. They do not employ second order cone programming approach while solving their model.

Our proposed CCSP models are in the form of static stochastic knapsack problem. Therefore, we cite related chance-constrained stochastic knapsack problem (CCSKP) literature. Several CCSKPs have been studied in the literature. Goel and Indyk (1999) develop a polynomial-time approximation schemes (PTAS) for the CCSKP where the item sizes are Poisson or exponentially distributed. They also propose a quasipolynomial-time approximation schemes (QPTAS) for the same problem where the item sizes are Bernoulli distributed. Kleinberg et al. (2000) study Bernoulli type item sizes (i.e. an item takes the value s with some common probability q and the value 0 with probability $(1 - q)$) and propose an approximation algorithm which is the function of some defined threshold probability ρ. Models studied by Goel and Indyk (1999) and Kleinberg et al. (2000) assume deterministic values in the objective function and propose approximation schemes to relax the chance constraint with some factor $1+\epsilon$. Goyal and Ravi (2010) provide a PTAS for the CCSKP with deterministic profits and normally distributed item sizes. They formulate a second order cone programming (SOCP) model with relaxed binary variables and transform it into a parametric Linear Program (LP) and apply their proposed PTAS to parametric LP.

In the next section, we discuss second order cone programming.

2.4 Second Order Cone Programming

In this dissertation, we benefit from the advances in second order cone programming (SOCP) in our proposed approaches. In particular, in Chapter 3 and Chapter 4, we transform the intractable chance constrained stochastic programming (CCSP) models to their equivalent tractable programs by employing the normal approximation and second order cone programming. In Chapter 5, we also prove that our proposed nonlinear social welfare function can be represented by second order conic inequalities. It was shown that continuous relaxations of the second order cone programming models are amenable to polynomial time interior-point algorithms (Nesterov and Nemirovski (1993)). Therefore, those models could be solved exactly by commercial solver packages such as CPLEX. As far as we know, SOCP methods have not been applied in RDPPS literature before. We refer the interested reader to Alizadeh and Goldfarb (2003) for comprehensive information on second order conic inequalities and the special functional forms that can be represented by conic quadratic inequalities.

In the next chapter, we study public RDPPS problem with project cancellations.

CHAPTER 3

PUBLIC RDPPS PROBLEM WITH PROJECT CANCELLATIONS

In this chapter, we consider a one-stage public RDPPS problem with project cancellations in two parts. In the first part, in Section 3.1, we assume that cancellation probabilities of projects cannot be assessed. In Section 3.1.1, we develop a mathematical model, which uses an estimated number of projects that will be canceled. The model maximizes the total score of selected projects within a given budget while taking into account that a number of them will be canceled leaving some residual budget. We transform our proposed model to a mixed-integer linear programming (MILP) model by using the duality theory and the McCormick's linearization approach. In addition, in Section 3.1.2, we also develop an efficient DP algorithm for our proposed model.

In the second part, in Section 3.2, we assume that cancellation probability of each project can be assessed. We formulate a mathematical programming model, which maximizes the total expected score of selected projects under a chance-constrained budget limit. We obtain a tractable nonlinear model in the form of mixed-integer second order cone programming (MISOCP). In Section 3.3, computational experiments on the proposed solution approaches for both problems are given. Firstly, in Section 3.3.1, separate computational analyses of the two proposed models are discussed. In particular, both the MILP and the DP approaches solve practical-size problems to optimum in short CPU times. Besides, the second proposed model can exactly solve 90% of the instances to optimality within a given time limit, and the average optimality gap for the instances that were only solved to feasibility is below 0.02%.

Secondly, in Section 3.3.2, we provide managerial insights to the DM on the proposed models. We make a connection between the two proposed models by using probability information of the second model. The DM wonders about risk of exceeding budget quantitatively even in the absence of cancellation probabilities. Hence, we obtain the budget risk of the first model by using the probability information of the second model for different factor settings. The risk of exceeding budget for the first model is very small (i.e. between 4.5% and 0.1%). In addition, we quantify the budget risk levels of different cancellation scenarios for the same instance by deviating number of cancellations from its expected value. In that case, if the number of cancellations is estimated 20% higher than its expected value, the generated solutions have still small budget risks. We also show that the first model gives a good approximation to the second model when we solve the second model with conservative budget risk of the first model. Furthermore, we quantify the value of second model for different factor settings and find that the second model generates better project portfolios than first model with anticipated risk levels. Moreover, we also show that the value of incorporating the cancellations into the decision making process by comparing it with the standard knapsack model setting. We show that the proposed models significantly improve both the budget utilization and total expected score of selected projects. The chapter is concluded with a summary in Section 3.4.

3.1 Model with Unknown Cancellation Probabilities

Every R&D funding program has a set of criteria that measure scientific merit and potential impact of project proposals. R&D funding organizations use a call-based system, in which once a call is announced, researchers apply to the public organization with their solicited R&D project proposals. After that, the research proposals are examined for eligibility, and the selected proposals are evaluated scientifically in research panels by peer reviewers. Besides, budgetary issues are also taken into account in order to make each project's budget plan more appropriate and realistic. Subsequently, research panels are conducted, where awarded projects are determined according to the panel scores and the total available budget. Awarded projects are announced and project funding contracts are signed between the public organization and project PIs. In signed contracts, the public organization commits that funding amounts will be given to projects according to the budget plans, as long as grantees comply with the terms and conditions of the R&D funding program. Hence, public organizations make irreversible funding decisions.

However, some of the awarded projects may be canceled during implementation. The progress and the expenditure rate of the projects that will be canceled usually slow down. Hence, if an R&D project is canceled, most of the budget of the project is unused and some expenditure until the cancellation decision can be paid back by the grantee. Under special circumstances, some of the expenditures can be exempted from repayment. Details of those situations are treated in the regulations of the R&D funding programs.

In this case, we assume that the DM has no information on the cancellation risk of individual projects. Based on past experience the DM can estimate a range for the number of projects that will be canceled in future. For possible scenarios on the number of projects to be canceled, we can formulate a problem that maximizes the total score of selected projects under a budget constraint, which assumes that canceled projects will return the remaining budget to the funding body. We make the following assumptions:

(1) The budget allocated to each project is determined by the funding authority just after the panel meetings.

(2) Each project's score is determined by the peer reviewers in panel meetings. The score aggregates all peer reviewers' individual scores regarding selection criteria. It does not include budgetary issues.

(3) Funding decisions are made at the end of each call. Unfunded projects in each call do not have any chance of funding from the same call budget. Hence, the single-stage problem environment is valid.

(4) Total available budget is limited and known during the funding period, and many projects compete for funding.

(5) A funded project has a work program and a multi-year financial schedule. Thereby, the planned annual expenditure of the project is known. Projects are

monitored by the progress reports. The reporting period is usually six months or one year. Compliance with the terms and conditions of the grant is verified at reporting time points. Cancellation decisions are made at those time points. If a project is canceled at a time point, its eligible payback spending up to this time point and its remaning budget after this time point are known. Therefore, by considering those time points and budget schedules, for a project i, we can estimate the fraction of spent budget (i.e. α_i) if it is canceled.

(6) Allocated whole budgets are not directly transferred to the projects in the beginning of the funding period. At a reporting time point, if the project satisfy the term and conditions, only partial grant (i.e. partial year budget or full year budget) is transferred to the project according to its financial schedule. Therefore, possibility of a cash shortage of funding budget due to incorporating cancellations is negligible.

In revenue and yield management literature, especially in the airline and lodging sectors, overbooking is a common strategy to deal with cancellations. In those sectors, the primary concern is increasing the revenue by finding the optimal overbooking limit. By trading off between cost of underutilized capacity (spoilage) and cost of exceeding capacity (offload), optimal overbooking limits are obtained. Although we do over-selection and this notion is similar to the overbooking concept, we cannot apply the methods of overbooking literature due to the non-monetary cost and benefits of public sector setting. For instance, score of the projects includes many non-monetary benefits which can be difficult to assess in monetary units. It is hard to measure the cost of the exceeding available budget or the cost of the underutilized budget in terms of benefits of the projects to society. Therefore, we limit the risk of exceeding budget in the budget constraint in a risk averse manner. More precisely, for a given number of cancellations, we assume that the worst case scenario will happen, in which the projects with the minimum residual budgets are canceled. In the next section, we develop an MILP model for the problem.

3.1.1 Mathematical Formulation

In the case of budget releases that stem from cancellations, we have the following nonlinear model with $\beta(x, \Gamma)$ denoting the cancellation function :

$$\text{(PC)} \ \max \quad \sum_{i \in N} s_i x_i \qquad \text{(Total Score)}$$

$$\text{s.t.} \ \sum_{i \in N} b_i x_i - \beta(x, \Gamma) \leq B \tag{3.1}$$

$$x_i \in \{0, 1\} \ \forall i \in N \tag{3.2}$$

where s_i and b_i represent the score and budget of project i, respectively. B is total available budget. N denotes set of all projects. x_i is 1, if project i is selected and 0, otherwise. Objective function is to maximize the total score of selected projects. Constraint (3.1) is the total budget limitation, where the cancellation function $\beta(x, \Gamma)$ represents the minimum residual budgets of canceled projects and has the following form:

$$\beta(x, \Gamma) = \min_{S \subseteq N, |S| = \Gamma} \sum_{i \in S} (1 - \alpha_i) b_i x_i \tag{3.3}$$

where x represents the $|N|$-tuple vector of x_i decision variables. S runs over all subsets of N with Γ elements, α_i represents the estimated fraction of spent budget for project i if it is canceled, and hence $\beta(x, \Gamma)$ equals the residual budget of canceled projects.

Proposition 3.1.1. *Model PC is $\mathcal{NP}$-hard.*

Proof. When $\Gamma = 0$, the problem is a knapsack problem. $\square$

Proposition 3.1.2. *Given a solution vector x^*, the cancellation function of the constraint* (3.1)*, equals the optimum objective function value of the following linear optimization problem:*

$$\textit{(MC)} \ \beta(x^*, \Gamma) = \min \sum_{i \in N} (1 - \alpha_i) b_i x_i^* y_i$$

$$\text{s.t.} \quad \sum_{i \in N} y_i = \Gamma \tag{3.4}$$

$$0 \leq y_i \leq x_i^* \qquad \forall i \in N \tag{3.5}$$

where y_i is the decision variable representing 1, if project i is canceled and 0, otherwise.

Proof. Among the selected projects we assume that Γ of them will be canceled. Objective $\beta(x^*, \Gamma)$ selects the ones that have minimum budgets. Constraint (3.4) ensures that Γ of them will be canceled. Constraint set (3.5) guarantees that only selected ones can be canceled. Note that coefficient matrix of constraint (3.4) is a unit row matrix and coefficient matrix of constraint set (3.5) is an identity matrix. This means that coefficient matrix of the model MC is totally unimodular because determinants of all square submatrices are either -1, 0, or 1. In addition, we have integral right-hand side values Γ and x_i^*s, so the model MC admits an integer-valued optimal solution y^*. Therefore, the objective value of the model MC equals cancellation function value in (3.3). □

As we later discuss in the second case, Γ can be modeled as a Poisson-Binomially distributed random variable and computing the cumulative distribution function of the Poisson binomial distribution is not possible without cancellation probabilities. Hence, modeling probabilistic Γ with a stochastic programming approach such as chance-constraint programming is challenging. Therefore, we do not consider probabilistic Γ value in the first case due to unknown cancellation probabilities. To give more insights on estimation, we present different gamma scenarios to the DM by considering the past data and also compare these scenarios with no cancellation situations.

Constraint set (3.5) has a significant implication in this formulation. Embedding inner optimization models into upper problems by using duality theory is used in robust optimization (Bertsimas and Sim (2004)). However, resulting models are always linear due to the different problem environment. However, in case of cancellations, we strictly enforce the cancellations from the selected ones due to the fact that only selected projects can be canceled. Since we have an inner minimization problem, we have to formulate constraint set (3.5) and this brings additional computational complexity when we take the dual. We will handle this issue in Proposition 3.1.3.

Proposition 3.1.3. *Model PC has an equivalent MILP as follows:*

$$
\begin{aligned}
\textit{(PCL)} \quad \max \ & \sum_{i \in N} s_i x_i && && \textit{(Total Score)} \\
\textit{s.t.} & \sum_{i \in N} b_i x_i - \Gamma z - \sum_{i \in N} w_i \leq B && && (3.6) \\
& z + p_i \leq (1 - \alpha_i) b_i x_i && \forall i \in N && (3.7) \\
& - w_i \leq M_i x_i && \forall i \in N && (3.8) \\
& - w_i \leq -p_i && \forall i \in N && (3.9) \\
& w_i \leq -M_i x_i + M_i + p_i && \forall i \in N && (3.10) \\
& w_i \leq 0 && \forall i \in N && (3.11) \\
& p_i \leq 0 && \forall i \in N && (3.12) \\
& x_i \in \{0, 1\} && \forall i \in N && (3.13) \\
& z \textit{ is urs} && && (3.14)
\end{aligned}
$$

Proof. If we take dual of the model MC, z and p_i are the dual variables of constraint (3.4) and constraint set (3.5), respectively and we obtain the following maximization problem:

$$
\begin{aligned}
\text{(D)} \ \max \ & \Gamma z + \sum_{i \in N} p_i x_i^* \\
\text{s.t.} \quad & z + p_i \leq (1 - \alpha_i) b_i x_i^* \ \ \forall i \in N && (3.15) \\
& p_i \leq 0 \ \ \forall i \in N && (3.16) \\
& z \text{ is urs} && (3.17)
\end{aligned}
$$

By strong duality, since we know that the model MC is feasible and bounded, the model D is also feasible and bounded and their objective values are the same at optimality. Hence, we can integrate the model D and PC, and we can obtain the following mixed integer nonlinear program (MINLP).

$$
\text{(PCNL)} \ \max \sum_{i \in N} s_i x_i \qquad \text{(Total Score)}
$$

$$
\begin{aligned}
\text{s.t.} \sum_{i \in N} b_i x_i - \Gamma z - \sum_{i \in N} p_i x_i \le B & & (3.18) \\
z + p_i \le (1-\alpha_i) b_i x_i & \quad \forall i \in N & (3.19) \\
p_i \le 0 & \quad \forall i \in N & (3.20) \\
z \text{ is urs} & & (3.21) \\
x_i \in \{0,1\} & \quad \forall i \in N & (3.22)
\end{aligned}
$$

Constraint set (3.18) has a nonlinear term $p_i x_i$, where p_i is a non-positive continuous, and x_i is a binary variable. In fact this nonlinearity has a special structure known as a bilinear variable in which there is a product of one continuous and one integer (binary) variable (Gupte et al. (2013)). Define a bilinear variable $w_i = p_i x_i$ and consider the mixed-integer bilinear set for a specific index i:

$$S^i = ((w_i, p_i, x_i) \in \mathbb{R}^- \times \mathbb{R}^- \times \mathbb{N}^0 : w_i = p_i x_i, p_i \ge -M_i, x_i \le 1) \tag{3.23}$$

where M_i is a sufficiently large number. We can linearize the bilinear variable w_i by replacing $p_i x_i$ by its convex and concave envelopes, which was developed by McCormick (1976) and it is also known as the McCormick envelopes method in the literature. Performing this operation on S^i gives us the following set

$$
\begin{aligned}
SL^i = ((w_i, p_i, x_i) \in \ & \mathbb{R}^- \times \mathbb{R}^- \times \mathbb{N}^0 : w_i \le 0, w_i \ge -M_i x_i, w_i \ge p_i, \\
& w_i \le -M_i x_i + M_i + p_i)
\end{aligned} \tag{3.24}
$$

it is straightforward to verify that $S^i = SL^i \ \forall i \in N$. For $x_i = 0$, w_i equals to 0, and for $x_i = 1$, w_i equals to p_i. Then, we obtain the model PCL, which completes the proof. □

M_i has a critical role in the model PCL. In general, arbitrarily large big-M values cause numerical challenges due to loose linear programming relaxations. We conduct extensive preliminary runs with arbitrarily large big-M values, and observe that solution time tends to increase extremely as Γ increases. A recent discussion on big-M notion and the importance of tighter formulations in MILPs can be found in Klotz and Newman (2013) for interested readers. Hence, if we cannot find a tight value for M_i, the model will be inefficient due to computational issues. Therefore, we obtain the smallest big M_i in Proposition 3.1.4. Note that M_i is identical $\forall i \in N$.

Proposition 3.1.4. *The smallest possible value for M_i is*

$$M_i = (1 - \alpha_{min})b_{max} \; \forall i \in N$$

where $\alpha_{min} = \min_{k\in N}\{\alpha_k\}$ and $b_{max} = \max_{k\in N}\{b_k\}$.

Proof. By using duality and complementary slackness conditions of the models MC and D,

$$(x_i^* - y_i^*)p_i^* = 0 \;\; \forall i \in N \tag{3.25}$$

$$((1-\alpha_i)b_i x_i^* - z^* - p_i^*)y_i^* = 0 \;\; \forall i \in N \tag{3.26}$$

are obtained. Let set $F_a^b = \{i \in N | x_i^* = a, y_i^* = b\} \;\; a, b \in \{0, 1\}$. $\forall i \in F_0^0$, if we plug $x_i^* = 0$ into constraints (3.7) through (3.12), we obtain $z^* \leq -p_i^*$ and $-p_i^* \leq M_i$. Hence $z^* \leq M_i \;\; \forall i \in F_0^0$. $\forall i \in F_1^0$, $p_i^* = 0$ from (3.25) and $z^* \leq (1-\alpha_i)b_i$ from (3.7). $\forall i \in F_1^1$, $-p_i^* = z^* - (1-\alpha_i)b_i$ from (3.26). We can obtain a tight upper bound on $-p_i^*$ as $-p_i^* = z^* - (1-\alpha_i)b_i \leq (1-\alpha_i)(b_{max} - b_i)$ since $z^* \leq (1-\alpha_i)b_i \;\; \forall i \in F_1^0$ and b_i can go to b_{max}. However, $M_i = (1-\alpha_i)(b_{max} - b_i)$ can destroy the feasibility of $z^* \leq M_i \;\; \forall i \in F_0^0$ in the two cases. In the first case, consider an index i_{max} such that there exists a $b_{max} = b_{i_{max}}$ and in this case $M_i = 0, \; z^* \leq 0$ and this leads to infeasible z^*. Therefore, we drop b_i and set $M_i = (1-\alpha_i)b_{max}$. However, consider another case, where all b_i equal to each other, namely $b_i = b_{max} \; \forall i \in N$. Then, there may exist an index $t \in F_0^0$ such that $z^* \leq (1-\alpha_t)b_t$ and $\alpha_{max} = \max_{k\in N} = \{\alpha_k\} = \alpha_t$ can destroy the feasibility of at least one of $z^* \leq (1-\alpha_i)b_i \;\; \in F_1^0$ where $\alpha_i < \alpha_t$. To remedy this, we set $M_i = (1-\alpha_{min})b_{max} \; \forall i \in N$. □

The MILP formulation allows us to add new constraints to the model easily. For instance, in the public RDPPS problem, there can be some sort of policy constraints (i.e. regional balance among projects etc.) for some funding programs, and they can be easily incorporated into the model. On the other side, we have a knapsack model oriented approach for modeling the cancellations, and we know that the standard knapsack optimization model is $\mathcal{NP}$-hard. However, there is a dynamic programming (DP) algorithm that solves the knapsack problem in pseudo-polynomial time. Motivated by this fact and from a methodological perspective, it can be interesting to extend DP to incorporate project cancellations. Therefore, as an alternative to the MILP formulation, in the next section we propose a DP algorithm.

3.1.2 DP Algorithm

In this section, we present an exact DP algorithm for the model PC. Our DP algorithm originates from the classical knapsack DP algorithm. We assume that the projects with minimum residual budget will be canceled. Therefore, projects are sorted in ascending order of their residual budgets (i.e.$(1-\alpha_i)b_i$) to make our DP work properly.

We define spent budget of project i if it is canceled as $b_i^r = \lfloor b_i\alpha_i \rfloor$ for the sake of computational tractability. For a given Γ parameter, let S be a subset of sorted projects such that $S \subseteq N$, $|S| \geq \Gamma$, and let i^Γ be the Γ^{th} project's index in S. Then, subset S is a feasible solution if the following condition is satisfied.

$$\sum_{i \in S | i \leq i^\Gamma} b_i^r + \sum_{i \in S | i > i^\Gamma} b_i \leq B \tag{3.27}$$

Due to the special structure in (3.27), our DP depends on two arrays. Utilization of multiple arrays in DPs has been recently proposed by Monaci et al. (2013) for robust optimization. Let $v_1[d][c][i]$ denote the highest score that is obtained with a feasible solution that has exactly budget d that is allocated to the projects selected from the set $\{1, 2, \ldots, i\} \subseteq N$ and exactly c of them are canceled. Let $v_2[d][i]$ denote the highest score that is obtained with a feasible solution that has exactly total budget d in which projects of $\{1, 2, \ldots, i\} \subseteq N$ are taken into account and exactly Γ of them are canceled. First array $v_1[d][c][i]$ considers the canceled projects with their used budgets. We define an upper bound B_Γ on the total budget of canceled projects in order to increase the efficiency of our DP. It can be formulated mathematically as follows:

$$B_\Gamma = \min \left[\max_{A \subseteq N | A| = \Gamma} \sum_{j \in A} b_j^r, B \right] \tag{3.28}$$

Note that for the first array $v_1[d][c][i]$, d=$0, 1, 2, \ldots, B_\Gamma$, c=$0, 1, 2, \ldots, \Gamma$, and i=0,1,2,...,|N|. For the second array $v_2[d][i]$, d=$0, 1, 2, \ldots, B$, and i=Γ, Γ+$1, \ldots, |N|$. After those notations, we can generate all elements of arrays $v_1[d][c][i]$ and $v_2[d][i]$ by the following DP recursions:

$$v_1[d][c][i] = max\{v_1[d][c][i-1], v_1[d-b_i^r][c-1][i-1] + s_i\} \tag{3.29}$$
$$for \;\; d = 0, 1, 2, \ldots, B_\Gamma, c = 1, 2, \ldots, \Gamma, i = 1, 2, \ldots, |N|.$$
$$v_2[d][i] = max\{v_2[d][i-1], v_2[d-b_i][i-1] + s_i\} \tag{3.30}$$
$$for \;\; d = 0, \ldots, B, i = \Gamma+1, \Gamma+2, \ldots, |N|.$$

The initializations are done as follows: $v_1[d][c][i]$=-∞ for $d = 0, 1, 2, \ldots, B_\Gamma$, $c = 0, 1, 2, \ldots, \Gamma$, i=0,1,2,...,|N|. $v_2[d][i] = -\infty$, $d = 0, 1, \ldots, B, i = 0, 1, 2, \ldots, |N|$. We set v_1[0][0][0]=0. The second array v_2 is initialized with the value of first array v_1 for $c = \Gamma$. We link the two arrays by setting $v_2[d][i]$=$v_1[d][\Gamma][i]$ for all d and $i \geq \Gamma$. The first array v_1 keeps track of selected but canceled projects, and the second array v_2 considers remaining non-canceled selected projects. The optimal solution of the problem is found by the maximum valued array as follows:

$$v_2^* = \max(v_2[d][|N|]|d = 1, \ldots, B)$$

with budget value $d^* \leq B$.

We give the steps of the DP algorithm in Algorithm 1.

First of all, projects are sorted in ascending order of their residual budgets. Steps 1 through 10 are the initialization steps. Steps 11 through 27 are the DP recursions. The DP algorithm begins with possible cancellations by considering the projects that have minimum residual budgets (steps 11 to 16). All possible cancellation combinations are determined by the recursion in step 14 by considering inclusion or non-inclusion of a specific project i. Total available budget is exhausted with canceled projects during step 14 and with selected and not canceled projects during step 23. The recursion in step 16 ensures that for a specific project i, for the solutions in which $d < b_i^r$ holds, project i is not included in those solutions. The recursion in step 19 links arrays v_2 and v_1 if array v_1 selects exactly Γ projects. The recursion in step 26 ensures that for

Algorithm 1 DP Algorithm

Require: Given set N, parameters B, Γ and $s_i, b_i, \alpha_i, b_i^r$, B_Γ for all $i \in N$.

Sort the projects in ascending order of their residual budgets

Initialization ;

for d=0 to B_Γ **do**
  for c=0 to Γ **do**
    for i=1 to $|N|$ **do**
      if $d = 0$ and $c = 0$ and $i = 0$; **then**
        $v_1[d][c][i]$=0;
      else
        $v_1[d][c][i]$=$-\infty$;
for d=0 to B **do**
  for i=0 to $|N|$ **do**
    $v_2[d][i]$=$-\infty$;

DP Recursions ;

Generating First Array ;

for i=1 to $|N|$ **do**
  for c=1 to Γ **do**
    for d=b_i^r to B_Γ **do**
      $v_1[d][c][i] = max\{v_1[d][c][i-1], v_1[d-b_i^r][c-1][i-1] + s_i\}$;
    for d=$[b_i^r - 1]^+$ to 0 **do**
      $v_1[d][c][i]$=$v_1[d][c][i-1]$;

Linking The Two Arrays ;

for i=Γ to $|N|$ **do**
  for d=0 to B_Γ **do**
    $v_2[d][i]$=$v_1[d][\Gamma][i]$;

Generating Second Array ;

for i=$\Gamma + 1$ to $|N|$ **do**
  for d=b_i to B **do**
    if $v_2[d][i] < max\{v_2[d][i-1], v_2[d-b_i][i-1] + s_i\}$; **then**
      $v_2[d][i] = max\{v_2[d][i-1], v_2[d-b_i][i-1] + s_i\}$;
  for d=b_i-1 to 0 **do**
    if $v_2[d][i]$ < $v_2[d][i-1]$; **then**
      $v_2[d][i]$=$v_2[d][i-1]$;
v_2^*=max ($v_2[d]|N|]$ | d=1,..,B)

a specific project i, for the solutions in which $d < b_i$ holds, project i is not included in those solutions. The recursion in step 23 provides all solution combinations by considering inclusion and non-inclusion of specific project i.

Computational complexity of the DP can be determined as follows: Steps 11 through 16 requires $|N|\Gamma(B_\Gamma + 1)$ operations, steps through 17 to 19 requires $(B_\Gamma + 1)(|N| - \Gamma + 1)$ operations, and steps 20 through 27 requires $(|N| - \Gamma)(B + 1)$ operations. Therefore, our DP solution time is the function of $|N|\Gamma(B_\Gamma + 1) + (B_\Gamma + 1)(|N| - \Gamma + 1) + (|N| - \Gamma)(B + 1)$ operations and upper bound on the solution time in the worst case has a computational complexity of $\mathcal{O}(|N|\Gamma B))$. In the proof of following proposition, we graphically describe our DP recursions.

We first introduce the concept of graph G together with an illustrative example before the providing the proposition on the optimality of the DP approach. We construct a directed acyclic graph $G = (V, A)$ with vertices V and arcs A. We present how a feasible solution is generated on G and show that the DP recursion finds the longest feasible path on G. Each vertex in the graph is indicated by the notation $[d][c][i]$ as introduced in the DP algorithm. Vertices can be partitioned into Γ subsets of vertices, referred to as blocks, by using $c = 0, 1, ...\Gamma$.

Arcs can be classified into three types. The first type, referred to as "zero-arc", connects vertex $[d][c][i-1]$ with vertex $[d][c][i]$ within the same block. Usage of a "zero-arc" means not selecting project i. The second type, referred to as "cancellation-arc", connects vertex $[d - b_i^r][c-1][i-1]$ in the block $(c-1)$ with vertex $[d][c][i]$ in the block c with score s_i if $b_i^r \leq d$. Usage of a "cancellation-arc" means selecting project i and assuming it will be canceled. Third type, referred to as "selective-arc", connects vertex $[d - b_i][\Gamma][i-1]$ with vertex $[d][\Gamma][i]$ within the block Γ with score s_i if $b_i \leq d$. Usage of a "selective-arc" means selecting project i and assuming it will not be canceled.

An illustrative example is given in Table 3.1 with $B = 13$ and $\Gamma = 2$. Generation of just three feasible solutions (i.e. feasible paths) of the example are presented in Figure 3.1 for the sake of demonstration. The source and dummy sink vertices are indicated by $[0][0][0]$ and z.

Table 3.1: Illustrative example for a graphical representation

Projects	**1**	**2**	**3**	**4**
s_i	25	24	23	22
b_i	5	6	8	10
α_i	1/5	1/3	3/8	2/5
$(1-\alpha_i)b_i$	4	4	5	6
$b_i^r = \lfloor b_i \alpha_i \rfloor$	1	2	3	4

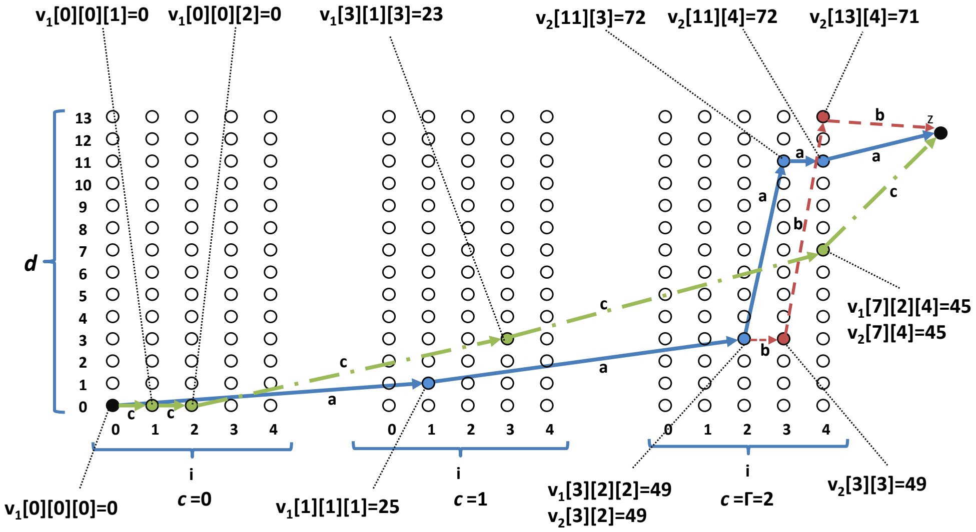

Figure 3.1: Graphical Representation of the DP Algorithm

Consider the path "a" (i.e. all solid arcs with label a) from the source vertex to the sink vertex. It reaches the sink vertex after using the two "cancellation-arcs" (due to $\Gamma = 2$), one "selective-arc", and one "zero-arc". First cancellation-arc connects vertices $[0][0][0]$ and $[1][1][1]$ with score 25. Second cancellation-arc connects vertices $[1][1][1]$ and $[3][2][2]$ with score 24 (i.e. total score becomes 49). Selective-arc connects vertices $[3][2][2]$ and $[11][2][3]$ with score 23 (i.e. total score becomes 72). Note that the DP-based array valuation notation is used in the graph for the algorithmic tractability. For instance, vertex $[3][2][2]$ is exploited by the DP arrays $v_1[3][2][2]$ and $v_2[3][2]$ because the two arrays are linked in vertex $[3][2][2]$ (i.e. $v_2[3][2] = v_1[3][2][2] = 49$).

Zero-arc connects vertices $[11][2][3]$ and $[11][2][4]$. Last arc is the dummy arc for the connection to sink vertex z. Hence, path "a" selects first three projects (i.e. projects 1, 2 and 3) and assumes the first two ones (i.e. projects 1 and 2) will be canceled. Note that for this path (i.e. selection of projects 1, 2 and 3), all other cancellation scenarios under $\Gamma = 2$ condition won't change feasibility of the solution due to the special structure given in (3.27). In fact this is valid for all feasible paths. Now, consider the "a-b" path that has the two arcs from path "a" and three arcs with label "b". The first two arcs with label "a" correspond to the same meaning as in the path "a". First arc with label "b" is a zero-arc connecting vertices $[3][2][2]$ and $[3][2][3]$ (i.e. $v_2[3][3] = v_2[3][2] = 49$). Second arc with label "b" is a selective-arc connecting vertices $[3][2][3]$ and $[13][2][4]$ (i.e. $v_2[13][4] = v_2[3][3] + 22 = 71$). Last arc with label "b" is the dummy arc for the connection to sink vertex z. Similarly, consider path "c" (i.e. all arcs with label "c") from source vertex to sink vertex. First two arcs are zero-arcs connecting vertices $[0][0][0]$, $[0][0][1]$, and $[0][0][2]$, respectively. Third and fourth arcs are cancellation-arcs connecting vertices $[0][0][2]$, $[3][1][3]$, and $[7][2][4]$. Last arc is the dummy one.

Proposition 3.1.5. *The DP recursions defined in* (3.29) *and* (3.30) *find an optimal solution for the model PC.*

Proof. The graphical illustration can be expressed mathematically as follows: for a given Γ, select any feasible subset $S \subseteq N$, $|S| \geq \Gamma$, and i^Γ is the Γ-th project's index in S. A corresponding path is generated on the graph as follows: for each index $i \leq i^\Gamma$, we use a zero-arc if $i \notin S$ and a cancellation-arc if $i \in S$, until reaching vertex $[\sum_{i \leq i^\Gamma} b_i^r][\Gamma][i^\Gamma]$ in the last block (i.e. block Γ). Then for each index $i > i^\Gamma$ if $|S| > \Gamma$, we use a zero-arc if $i \notin S$ and a selective-arc if $i \in S$, until reaching vertex $[\sum_{i \leq i^\Gamma} b_i^r + \sum_{i > i^\Gamma} b_i][\Gamma][n]$. Then, we link this vertex to sink vertex z.

Our DP recursion generates all the solutions in this manner and selects the one that has the longest path (highest score) in the graph. □

So far, we have considered the case that the DM has no information on cancellation probabilities of the projects but can make an estimate of number of projects that will be canceled. We have provided two solution approaches. The first one is the mathematical programming approach (The Model PCL) and the second one is the DP approach (Algorithm 1). In Section 3.3, we will present the computational performance of both solution approaches.

In the next section, we consider the case in which a cancellation probability could be assessed in advance for each project.

3.2 Model with Cancellation Probabilities

In this section, we develop a stochastic programming model by assuming that cancellation probability of each project (π_i) is known or can be assessed. In the peer review process of the projects, there is usually a criterion assessed, called the feasibility of the project. Each project is assigned a feasibility score. Therefore, there can be a close relationship between cancellation probability and feasibility, and this kind of information can be obtained by using the past data.

3.2.1 Stochastic Formulation

In this model, we want to maximize the expected total score of selected projects while total random budget spending is not exceeding total available budget for some given confidence level. Below, we give the mathematical formulation of the problem:

$$\text{(SP-P)} \max \sum_{i \in N} \mathbb{E}(\hat{s}_i) x_i \qquad \text{(Expected Total Score)}$$

$$\text{s.t. } P\left(\sum_{i \in N} \hat{b}_i x_i \leq B\right) \geq \theta \tag{3.31}$$

$$x_i \in \{0, 1\} \; \forall i \in N \tag{3.32}$$

where $\mathbb{E}(.)$ is the expectation operator, $\hat{s}_i$ and $\hat{b}_i$ indicate the random variables corresponding to score and budget of project i, respectively. x_i is 1, if project i is selected and 0, otherwise. N is the set of all projects. B denotes the total available budget. The objective function is to maximize the expected total score of selected projects.

In stochastic programming, constraint (3.31) is referred to as the chance-constraint, which states that the constraint is satisfied with a probability of at least θ.

The number of canceled projects can be modeled as a Poisson-Binomially distributed random variable (Hong (2013)). Each project is viewed as an independent Bernoulli trial with a certain cancellation probability. Let I_i be a Bernoulli random variable with cancellation probability π_i (i.e. $I_i = 1$, with probability π_i and $I_i = 0$, w.p. $(1 - \pi_i)$). We can obtain expected number of canceled projects and its variance as follows:

$$\Gamma = \sum_{i \in N} I_i \Rightarrow \ \mathbb{E}(\Gamma) = \sum_{i \in N} \mathbb{E}(I_i) = \sum_{i \in N} \pi_i \ , \ \text{Var}(\Gamma) = \sum_{i \in N} \text{Var}(I_i) = \sum_{i \in N} \pi_i(1 - \pi_i) \tag{3.33}$$

However, for the stochastic formulation, rather than the number of cancellations we need the distribution of budget ($\hat{b}_i$) and score ($\hat{s}_i$) of each project. We assume that if a project is canceled then its score becomes zero and its budget becomes $\alpha_i b_i$ where α_i denotes the estimated fraction of the budget spent and not paid back by the canceled project i. We can express the distribution of random variables ($\hat{b}_i$) and ($\hat{s}_i$) as follows:

$$\hat{b}_i = \begin{cases} b_i & \text{if } \ I_i = 0 \quad \text{w.p. } (1 - \pi_i) \\ \alpha_i b_i & \text{if } \ I_i = 1 \quad \text{w.p. } \pi_i \end{cases}$$

$$\hat{s}_i = \begin{cases} s_i & \text{if } \ I_i = 0 \quad \text{w.p. } (1 - \pi_i) \\ 0 & \text{if } \ I_i = 1 \quad \text{w.p. } \pi_i \end{cases}$$

Above random variables are called as Bernoulli type since they have two realizations and the realizations not necessarily have to be zero and one as in the case of classical Bernoulli random variable. Note that we need to derive quantile function (i.e. inverse cumulative distribution function) of sum of many non-identical $\hat{b}_i$ random variables to exactly solve the resulting CCSP. However, such a derivation is computationally intractable. Hence, we apply normal approximation with mean and variance of $\hat{b}_i$ random variables. For the expectation and variance of the above distributions, we apply some algebra:

$$
\begin{aligned}
\mathbb{E}(\hat{b}_i) &= b_i(1-\pi_i) + \alpha_i b_i \pi_i = b_i[1-\pi_i(1-\alpha_i)] && (3.34)\\
\text{Var}(\hat{b}_i) &= \mathbb{E}(\hat{b}_i^2) - [\mathbb{E}(\hat{b}_i)]^2 = b_i^2(1-\pi_i) + \alpha_i^2 b_i^2 \pi_i - (b_i[1-\pi_i(1-\alpha_i)])^2\\
&= b_i^2 - b_i^2\pi_i + \alpha_i^2 b_i^2 \pi_i - b_i^2(1-\pi_i)^2 - 2\alpha_i b_i^2 \pi_i (1-\pi_i) - \alpha_i^2 b_i^2 \pi_i^2\\
&= b_i^2[1-\pi_i-(1-\pi_i)^2)] + \alpha_i b_i^2 \pi_i[\alpha_i - 2(1-\pi_i) - \alpha_i\pi_i]\\
&= b_i^2[1-\pi_i-(1-2\pi_i+\pi_i^2)] + \alpha_i b_i^2 \pi_i[\alpha_i - 2(1-\pi_i) - \alpha_i\pi_i]\\
&= b_i^2\pi_i(1-\pi_i) + \alpha_i b_i^2 \pi_i[\alpha_i - 2(1-\pi_i) - \alpha_i\pi_i] && (3.35)\\
\mathbb{E}(\hat{s}_i) &= s_i(1-\pi_i) && (3.36)\\
\text{Var}(\hat{s}_i) &= \mathbb{E}(\hat{s}_i^2) - [\mathbb{E}(\hat{s}_i)]^2 = s_i^2(1-\pi_i) - s_i^2(1-\pi_i)^2\\
&= s_i^2[1-\pi_i-(1-2\pi_i+\pi_i^2)]\\
&= s_i^2\pi_i(1-\pi_i) && (3.37)
\end{aligned}
$$

Note that for $\pi_i = 0 \ \forall i \in N$, our problem reduces to a knapsack problem. This implies that the special case is NP-hard. Besides, Nemirovski and Shapiro (2006) state that most of the single linear chance constraints are computationally intractable for $\theta > 0.5$ since checking feasibility of a solution is NP-hard. They discuss that a single linear chance constraint is computationally tractable for some special distributions such as multivariate normal distribution. If we adopt normal approximation, and set $\theta = 0.5$, then we have a standard knapsack constraint with mean $\hat{b}_i$ values. If we assume $\theta > 0.5$, we have also additional quadratic terms due to variance of $\hat{b}_i$. Therefore, we can conclude that our problem is $\mathcal{NP}$-hard.

3.2.2 Deterministic Equivalent Formulation

In order to derive the deterministic equivalent formulation of the stochastic program we first prove a normal approximation of the total random budget spending in Proposition 3.2.1.

Proposition 3.2.1. *Let $\hat{b}_1, \hat{b}_2, \ldots \hat{b}_i$ with $\in K \subseteq N$ be independent non-identically distributed (i.ni.d.) Bernoulli type random variables with finite expectation $\mathbb{E}(\hat{b}_i)$ and* *positive variance Var*$(\hat{b}_i)$. *Then normalized summand*

$$S_{|K|} = \frac{\sum_{i\in K}\left[\hat{b}_i - \mathbb{E}(\hat{b}_i)\right]}{\sqrt{\sum_{i\in K} Var(\hat{b}_i)}} \tag{3.38}$$

converges to the standard normal distribution ($N(0,1)$) as $|K|$ goes to infinity.

Proof. A sequence of i.ni.d. random variables obeys the Central Limit Theorem (CLT) if the Lyapunov or Lindeberg condition is satisfied (Baurer (1996) and Shapiro et al. (2009)). The Lyapunov condition (which is stronger than the Lindeberg condition) can be stated as follows: If for some $\delta > 0$, the following condition

$$\lim_{|K|\to\infty} \frac{\sum_{i\in K}\mathbb{E}\left[\left|\hat{b}_i - \mathbb{E}(\hat{b}_i)\right|^{2+\delta}\right]}{\sum_{i\in K}\text{Var}(\hat{b}_i)\left[\sqrt{\sum_{i\in K}\text{Var}(\hat{b}_i)}\right]^{\delta}} = 0 \tag{3.39}$$

holds, then normalized summand $S_{|K|}$ in (3.38) converges to the standard normal distribution ($N(0,1)$). Firstly, we show that $(2+\delta)^{th}$ moment exists for $\delta > 0$. We write $\mathbb{E}(\hat{b}_i^{2+\delta}) = b_i^{2+\delta}(1-\pi_i) + \alpha_i^{2+\delta} b_i^{2+\delta}\pi_i = b_i^{2+\delta}[1-\pi_i(1-\alpha_i^{2+\delta})]$ for every $i \in K$ by using definition of moment. Then, $\sum_{i\in K}\mathbb{E}\left[\left|\hat{b}_i - \mathbb{E}(\hat{b}_i)\right|^{2+\delta}\right]$ exists. Let $b_{max} = \max_{i\in K}\{b_i\}$, where b_i is the realization of the random variable $\hat{b}_i$ if it is not canceled. Hence, $\left|\hat{b}_i - \mathbb{E}(\hat{b}_i)\right| \leq b_{max}$ is true for every $i \in K$, which shows that $\hat{b}_1, \hat{b}_2, \ldots, \hat{b}_{|K|}$ are uniformly bounded. Then for each $\delta > 0$, we can determine an upper bound (UB) for the term in (3.39) as follows:

$$\frac{\sum_{i\in K}\mathbb{E}\left[\left|\hat{b}_i - \mathbb{E}(\hat{b}_i)\right|^{2+\delta}\right]}{\sum_{i\in K}\text{Var}(\hat{b}_i)\left[\sqrt{\sum_{i\in K}\text{Var}(\hat{b}_i)}\right]^{\delta}} \leq \frac{b_{max}^{\delta}\sum_{i\in K}\mathbb{E}\left[\left|\hat{b}_i - \mathbb{E}(\hat{b}_i)\right|^{2}\right]}{\sum_{i\in K}\text{Var}(\hat{b}_i)\left[\sqrt{\sum_{i\in K}\text{Var}(\hat{b}_i)}\right]^{\delta}} \tag{3.40}$$

We also know that $\mathbb{E}\left[\left|\hat{b}_i - \mathbb{E}(\hat{b}_i)\right|^{2}\right] = \mathbb{E}\left[\hat{b}_i^2 - 2\hat{b}_i\mathbb{E}(b_i) + (\mathbb{E}(b_i))^2\right] = \mathbb{E}(\hat{b}_i^2) - [\mathbb{E}(\hat{b}_i)]^2$, which is the definition of $\text{Var}(\hat{b}_i)$ in fact. Then, we can simplify the UB in (3.40) as follows:

$$\frac{\sum_{i\in K}\mathbb{E}\left[\left|\hat{b}_i - \mathbb{E}(\hat{b}_i)\right|^{2+\delta}\right]}{\sum_{i\in K}\text{Var}(\hat{b}_i)\left[\sqrt{\sum_{i\in K}\text{Var}(\hat{b}_i)}\right]^{\delta}} \leq \left(\frac{b_{max}}{\sqrt{\sum_{i\in K}\text{Var}(\hat{b}_i)}}\right)^{\delta} \tag{3.41}$$

Then, for every $\delta > 0$ right-hand side of (3.41) converges to zero as $|K|$ goes to infinity. Since UB converges to zero, the term in (3.39) converges to zero. From CLT theory (Baurer (1996)), we know that if Lyapunov condition is satisfied, then Lindeberg condition is also satisfied. $\square$

Constraint set (3.31) can be expressed as follows:

$$P\left(\sum_{i\in N} \hat{b}_i x_i \leq B\right) \Rightarrow P\left(\frac{\sum_{i\in N}[\hat{b}_i - \mathbb{E}(\hat{b}_i)]x_i}{\sqrt{\sum_{i\in N} \text{Var}(\hat{b}_i)x_i^2}} \leq \frac{B - \sum_{i\in N}\mathbb{E}(\hat{b}_i)x_i}{\sqrt{\sum_{i\in N} \text{Var}(\hat{b}_i)x_i^2}}\right) \tag{3.42}$$

By Proposition 3.2.1, S_n converges to the standard normal distribution ($N(0,1)$). Then, the inequality in (3.42) by replacing left-hand side term with the standard normal random variable Z can be written as:

$$P\left(Z \leq \frac{B - \sum_{i\in N}\mathbb{E}(\hat{b}_i)x_i}{\sqrt{\sum_{i\in N} \text{Var}(\hat{b}_i)x_i^2}}\right) \geq \theta \Rightarrow \Phi\left(\frac{B - \sum_{i\in N}\mathbb{E}(\hat{b}_i)x_i}{\sqrt{\sum_{i\in N} \text{Var}(\hat{b}_i)x_i^2}}\right) \geq \theta$$
$$\Rightarrow \frac{B - \sum_{i\in N}\mathbb{E}(\hat{b}_i)x_i}{\sqrt{\sum_{i\in N} \text{Var}(\hat{b}_i)x_i^2}} \geq \Phi^{-1}(\theta) \tag{3.43}$$

where Φ is the cumulative distribution function (c.d.f) for the standard normal variable Z. We arrange (3.43) as:

$$\sum_{i\in N}\mathbb{E}(\hat{b}_i)x_i + \Phi^{-1}(\theta)\sqrt{\sum_{i\in N} \text{Var}(\hat{b}_i)x_i^2} \leq B \tag{3.44}$$

Constraint (3.44) is the deterministic equivalent form of (3.31). Note that we assume that $\theta \geq 0.5$, then $\Phi^{-1}(\theta) > 0$, which makes the constraint (3.44) a second-order conic inequality. Finally, we can formulate deterministic equivalent of the SP-P model as a MISOCP:

$$\max \sum_{i\in N} s_i(1-\pi_i) \qquad \text{(Expected Total Score)}$$
$$\text{s.t } y = \frac{B}{\Phi^{-1}(\theta)} - \frac{\sum_{i\in N} b_i[1-\pi_i(1-\alpha_i)]x_i}{\Phi^{-1}(\theta)} \tag{3.45}$$
$$\sum_{i\in N}\left[b_i^2\pi_i(1-\pi_i) + \alpha_i b_i^2 \pi_i[\alpha_i - 2(1-\pi_i) - \alpha_i\pi_i]\right] x_i^2 \leq y^2 \tag{3.46}$$

$$y \geq 0 \tag{3.47}$$

$$x_i \in \{0, 1\} \ \ \forall i \in N \tag{3.48}$$

The objective function is to maximize the expected total score of selected projects as obtained in equation (3.36). Constraint sets (3.45) and (3.46) are the conic reformulation of constraint (3.44) by using equations (3.34) and (3.35). y in equality (3.45) is an auxiliary decision variable corresponding to the linear part of the constraint (3.44). Constraint (3.46) is the cone generated according to the constraint (3.44). In the next section, we give our computational results.

3.3 Computational Results

In this section, we present the results of computational experiments for the proposed approaches. For the first case with no cancellation probabilities, we solve the model PCL using IBM ILOG CPLEX 12.6 via Concert Technology. We implemented the DP algorithm using C++ language. For the second case, we solve the MISOCP model using IBM ILOG CPLEX 12.6 via Concert Technology. We conduct all experiments on a computer with processor Intel Core i5 1.7 GHz, 8.00 GB memory (RAM), 64-bit operating system, and Windows 7 Professional. In CPLEX, we set the time limit to 10.800 CPU seconds.

3.3.1 Separate Analyses of the Two Models

In our computational experiments, we define a budget fraction (bf) parameter to determine the total available budget (B). bf indicates the ratio of total budget of all applying projects to the budget available for funding. We set B by using parameter bf as $B = (\sum_{i \in N} b_i) \times bf$. We also test our solution methods for different levels of number of applying projects, indicated by "size". Another experimental factor is α_i, indicating the ratio of the unused budget of canceled project i. We set $\alpha_i = \alpha, \ \forall \ i \in N$ for the sake of simplicity.

Table 3.2 summarizes computational results for the PCL model and the DP algorithm. For this case, Γ, the number of projects to cancel, is an experimental factor. For each experimental setting we solve 10 random instances. We choose the levels of size parameters in accordance with the number of funding applications to a typical program call by TÜBİTAK 1001 program. Score of each project is uniformly distributed in [10-25] as integers, and budget of each project is uniformly distributed in [5-30] (in 10,000 monetary units) as integers. Table 3.2 reports the average number of selected projects (ns), the average Γ/ns ratio, the average CPU times for the PCL and the DP.

From Table 3.2 we observe that as Γ increases the average CPU time spent by CPLEX increases significantly. α also affects CPU time of CPLEX. As α increases, CPU time of CPLEX decreases. Problem parameters do not have a significant effect on CPU time of the DP for $\alpha = 0$ as CPU times are too small (i.e. the average CPU time is 0.31 seconds). They have a significant effect on CPU time of the DP for $\alpha = 0.3$. As α increases, CPU time of the DP increases. This is due to the fact that if $\alpha > 0$, then rounded positive budgets are considered in the DP for the canceled projects and this increases the computations of recursions. In summary, as α increases, CPU time of CPLEX decreases, whereas CPU time of the DP increases. For instance, for problem size 3000, $bf = 20\%$ and Γ=40 and 70, CPLEX solves the instances faster than the DP. However, we observe that the average CPU time of the DP is 1.27 seconds and the average CPU time of CPLEX is 19.64 in overall.

These results indicate that practical-size instances can be solved by both the model PCL and the DP algorithm. Within reasonable CPU times these methods can solve several instances for different number of canceled projects levels.

Furthermore, we observe that increase in the average number of selected projects in the case of cancellation is far less than Γ. This may be due to the fact that small budget increases because of the estimated number of canceled projects makes an opportunity of selecting relatively higher scored and higher budget projects than that of situation with no cancellation. As we expected, when we increase Γ, ratio of Γ/ns increases because number of selected projects does not increase considerably. For instance, for $\alpha = 0$, problem size=1000, and $bf(\%) = 20$, when we set $\Gamma = 20$, the average $ns = 360.6$, $\Gamma/ns = 5.5$ and when we set $\Gamma = 40$, the average $ns = 367.7$, $\Gamma/ns = 10.9$

We next give the results for the second case, with cancellations probabilities available. Table 3.3 presents computational results for the MISOCP.

Table 3.2: Computational results for the PCL and the DP

α_i	size	bf (%)	Γ	ns	Γ/ns ratio (%)	cpu (sec.) PCL	cpu (sec.) DP
0	1000	10	0	216.8	0	0.11	0.04
			20	226.0	8.8	5.16	0.03
			40	235.5	17.0	86.79	0.03
		20	0	353.6	0	0.26	0.11
			20	360.6	5.5	3.41	0.08
			40	367.7	10.9	11.76	0.05
	2000	10	0	435.1	0	0.30	0.17
			20	444.4	4.5	2.73	0.15
			40	453.6	8.8	17.73	0.12
			70	467.5	15.0	303.44	0.12
		20	0	710.6	0	0.18	0.37
			20	717.0	2.8	0.91	0.37
			40	724.4	5.5	10.20	0.27
			70	735.0	9.5	29.56	0.24
	3000	10	0	651.5	0	0.29	0.32
			20	661.0	3.0	1.08	0.37
			40	670.2	6.0	22.40	0.30
			70	684.5	10.2	48.95	0.33
		20	0	1063.2	0	0.28	0.74
			20	1070.4	1.9	1.75	0.97
			40	1077.8	3.7	2.37	1.05
			70	1088.3	6.4	30.32	0.52
average						26.36	0.31
0.3	1000	10	20	224.3	8.9	2.62	0.14
			40	231.6	17.3	12.80	0.41
		20	20	358.9	5.6	1.43	0.25
			40	365.1	11.0	9.36	0.41
	2000	10	20	442.6	4.5	1.29	0.48
			40	450.2	8.9	14.06	1.00
			70	461.0	15.2	52.21	4.88
		20	20	716.0	2.8	1.44	0.77
			40	721.3	5.5	1.59	1.80
			70	729.7	9.6	19.76	5.14
	3000	10	20	659.3	3.0	2.23	0.99
			40	666.7	6.0	6.48	2.63
			70	678.2	10.3	30.59	8.47
		20	20	1068.9	1.9	2.15	1.60
			40	1075.2	3.7	2.30	3.43
			70	1082.7	6.5	6.05	9.26
average						10.40	2.60
total average						19.64	1.27

Table 3.3: Computational results for the MISOCP

Factors						CPU (sec.)				E(Γ)/ns
α_i	size	bf (%)	c-prob	opt	gap (%)	mean	std. dev.	ns	E(Γ)	ratio (%)
0	1000	10	U (0.01-0.1)	9	0.01	368.8	390.1	221.4	12.1	5.5
			U (0.01-0.2)	3	0.02	1743.6	660.3	228.5	23.7	10.4
		20	U (0.01-0.1)	9	0.01	835.3	2293.2	361.6	19.7	5.5
			U (0.01-0.2)	7	0.02	2154.1	2405.6	373.2	38.8	10.4
	2000	10	U (0.01-0.1)	7	0.01	349.3	493.6	446.6	24.3	5.5
			U (0.01-0.2)	6	0.02	1789.0	1543.1	462.5	48.2	10.4
		20	U (0.01-0.1)	10	-	8.7	12.6	729.7	40.1	5.5
			U (0.01-0.2)	10	-	78.4	168.8	754.6	79.0	10.5
	3000	10	U (0.01-0.1)	10	-	63.5	134.6	670.9	36.7	5.5
			U (0.01-0.2)	9	0.01	240.3	185.9	695.5	72.6	10.4
		20	U (0.01-0.1)	10	-	12.0	11.1	1095.0	59.9	5.5
			U (0.01-0.2)	10	-	16.0	6.3	1132.7	117.9	10.4
average				83%	0.02	638.2	692.1			
0.3	1000	10	U (0.01-0.1)	10	-	109.7	169.6	219.8	11.9	5.4
			U (0.01-0.2)	9	0.01	989.7	2534.8	224.4	22.9	10.2
		20	U (0.01-0.1)	10	-	4.4	4.3	358.7	19.4	5.4
			U (0.01-0.2)	10	-	41.8	75.2	366.3	37.5	10.2
	2000	10	U (0.01-0.1)	10	-	8.7	15.2	442.8	23.9	5.4
			U (0.01-0.2)	8	0.01	137.6	215.9	452.8	46.2	10.2
		20	U (0.01-0.1)	10	-	1.8	1.5	723.1	39.5	5.5
			U (0.01-0.2)	10	-	2.4	2.5	738.8	76.0	10.3
	3000	10	U (0.01-0.1)	10	-	3.9	5.2	663.9	36.1	5.4
			U (0.01-0.2)	10	-	12.2	13.8	679.6	69.5	10.2
		20	U (0.01-0.1)	10	-	2.4	3.5	1084.8	58.9	5.4
			U (0.01-0.2)	10	-	10.6	9.3	1109.1	113.6	10.2
average				98%	0.01	110.4	254.2			
total average				90%	0.01	374.3	473.2			

For each size and bf factors, we randomly generate and solve 10 instances for each experimental setting. In difference to the previous case, for each project we generate a cancellation probability (c-prob) value. In fact, we solve the same instances of the model PCL by adding probability information. We set $\alpha_i = \alpha, \ \forall i \in N$ and $\theta = 0.95$ for the sake of simplicity. The score and budget value of each project is generated using the same settings as in the first model. Table 3.3 reports the number of instances solved to optimum (opt), the average optimality gap (gap) of unsolved instances, the average (mean) and the standard deviation (std. dev.) of CPU time (CPU) for the instances solved to optimum, the average ns, the average expected number of cancellations (E (Γ)), and the average E(Γ)/ns ratio. We restrict our analysis to the set of instances solved within the imposed time limit. For that reason, we exclude all instances with a CPU time of 10.800 seconds. Note that the means and standard deviations of the CPU time given in Table 3.3 are conditional values calculated under the condition that the time limit is not attained.

We observe that as α increases, CPU time decreases considerably. For $\alpha = 0$, the average CPU time is 638.2 seconds with the standard deviation of 692.1, whereas it is 110.4 seconds with the standard deviation of 254.2 for $\alpha = 0.3$. In addition, for $\alpha = 0$, 83% of the instances can be solved to optimum and the average optimality gap of unsolved instances is 0.02%; for $\alpha = 0.3$, 98% of the instances can be solved to optimum and the average optimality gap of unsolved instances is 0.01%. This may be due to the fact that as α increases, less projects are selected due to the increase in the mean of budget random variable and the decrease in the variance of budget random variable. Besides, we assert that cancellation probability has a prominent effect on the hardness of the instances. Apparently, instances with cancellation probabilities uniformly distributed between 0.01-0.2 require much more CPU time for all instances. We also observe that some of the instances cannot be solved to optimum in 3 hours. As we may expect, expected number of cancellations increases as the cancellation probabilities increase. We observe that as bf increases, the number of selected projects increases linearly. Apparently, hardness of instances heavily depends upon problem data, hence CPU time from one replication to another can change dramatically. In most of the instances, the problem size has no effect at all. The average E(Γ)/ns ratio is between 5.4% and 10.4%, which is close to the real life cancellation rates. Finally, we conclude that 90% of the instances can be solved to optimum and the average optimality gap of remaining instances is 0.01%.

3.3.2 Managerial Insights

We conduct further analyses on the proposed approaches to provide practical implications to the DM. Firstly, the DM may want to find out the probability of not exceeding the budget (or the risk of exceeding budget) for the first model. Recall that in the first model, the number of cancellations is an input. For the problem size of 3000, we relate the two models by using the expected number of cancellations of the second model as an input to the first model as follows. We set Γ to $\lfloor E(\Gamma) \rfloor$ of the second model MISOCP. Then, we solve 10 instances of the model PCL by using E(Γ) values of second model for each factor setting. We give the results in Table 3.3. After that, we calculate the expected total score (E(score)) and probability level (θ_{PCL}) of each solution of the model PCL by using probability information of the second model MISOCP. For a given problem size of 3000, we have factors α, bf, and c-prob; thereby, we have 2^3x(10 instances)=80 calculations. We report the average values of E(score) and θ_{PCL} (in % units to show numerical precision) of solutions of the model PCL in Table 3.4 under the PCL section. In Table 3.4, each row gives the average of ten instances. When we examine the average probability level of the solutions of the model PCL for various factor combinations, probability level (θ_{PCL}) is between 95.5% and 99.9%. Note that we solve the model PCL without any probability information, but probability level of the model PCL is very high since we minimize the risk of exceeding the budget in the model PCL in a novel way.

Secondly, the DM may want to see the value of the second model over the first model (i.e. value of assessing cancellation probabilities). Therefore, we solve the MISOCP with the two different confidence levels; namely θ_{PCL} and $\theta = 0.95$ to compare the two models. We report the average expected total score (E(score)) and the expected number of cancellations (E(Γ)) of the model MISOCP in Table 3.4 under the MISOCP section for probability levels θ_{PCL} and $\theta = 0.95$. Then, we calculate the value of second model (VSM) in Table 3.4 under the section VSM by comparing the sections PCL and MISOCP with the following measures:

Table 3.4: Comparison of the PCL and the MISOCP

Parameters				PCL		MISCOP				VSM			
						θ_{PCL}		$\theta = 0.95$		θ_{PCL}		$\theta = 0.95$	
α_i	bf (%)	c-prob	Γ	E(score)	θ_{PCL}	E(score)	$E(\Gamma)$	E(score)	$E(\Gamma)$	Δ_0	$\%\Delta_0$	Δ_1	$\%\Delta_1$
0	10	U (0.01-0.1)	36	12003.2	98.8	12004.1	36.5	12054.8	36.7	0.8	0.01	51.6	0.43
		U (0.01-0.2)	72	11641.1	99.9	11644.4	70.9	11810.7	72.6	3.3	0.03	169.7	1.46
	20	U (0.01-0.1)	59	19109.5	99.9	19111.3	59.0	19301.7	59.9	1.8	0.01	192.2	1.01
		U (0.01-0.2)	117	18431.3	99.9	18435.4	113.8	18880.5	117.9	4.1	0.02	449.2	2.44
0.3	10	U (0.01-0.1)	36	11946.4	95.5	11948.8	36.1	11951.5	36.1	2.4	0.02	5.1	0.04
		U (0.01-0.2)	69	11513.1	99.5	11525.2	68.8	11590.8	69.5	12.1	0.11	77.8	0.68
	20	U (0.01-0.1)	58	19034.6	99.9	19040.9	58.5	19135.7	58.9	6.3	0.03	101.1	0.53
		U (0.01-0.2)	113	18279.7	99.9	18303.6	111.3	18537.0	113.6	23.9	0.13	257.2	1.41

$$\Delta_0 = (E(score) \text{ of the MISOCP with } \theta_{PCL}) - (E(score) \text{ of the PCL with } \theta_{PCL})$$

$$\%\Delta_0 = 100 \text{ x } \left(\frac{E(score) \text{ of the MISOCP with } \theta_{PCL} - E(score) \text{ of the PCL with } \theta_{PCL}}{E(score) \text{ of the PCL with } \theta_{PCL}} \right)$$

$$\Delta_1 = E(score) \text{ of the MISOCP with } \theta = 0.95 - E(score) \text{ of the PCL with } \theta_{PCL}$$

$$\%\Delta_1 = 100 \text{ x } \left(\frac{E(score) \text{ of the MISOCP with } \theta = 0.95 - E(score) \text{ of the PCL with } \theta_{PCL}}{E(score) \text{ of the PCL with } \theta_{PCL}} \right)$$

In terms of Δ_0 and $\%\Delta_0$ values, we observe that the model PCL is a good approximation of the model MISOCP when we use the probability level of the model PCL (θ_{PCL}) to solve MISOCP. We also observe that for the same probability level (i.e. θ_{PCL}), the MISOCP generates solutions with slightly better expected total scores. While the model PCL conservatively chooses to assume the least budget projects will be canceled, the MISOCP can generate solutions for anticipated risk levels.

Note that θ_{PCL} values are very close to 1 (except the average θ_{PCL} is 95.5% for one factor combination). However, the DM can find those probability levels too conservative and can prefer to use an alternative high level of probability level such as 0.95. For that alternative case, Δ_1 and $\%\Delta_1$ values increase prominently when compared to Δ_0 and $\%\Delta_0$ values. Besides, for the same level of α, the VSM increases as bf and c-prob increase. Therefore, using the model MISOCP when we can obtain probability information provides prominent benefits.

Thirdly, the DM may wonder about how the probability level of the first model PCL changes when we estimate Γ differently than its expected value. Therefore, we relate the two models by setting Γ=$\lfloor E(\Gamma) \rfloor$ and using the probability information of the second model MISOCP. However, we actually do not know the cancellation probabilities when we solve the model PCL. Therefore, we select the minimum three θ_{PCL} values in Table 3.4 under PCL section to understand how probability level changes especially for higher Γ values in the worst case. When we examine the average θ_{PCL} values for 8 factor combinations, the minimum three of them are 95.5%, 98.8%, 99.5%. Hence, we pick up one instance from those three factor combinations, solve the model PCL for integer Γ values in the range $[\lfloor E(\Gamma) \rfloor$x0.8, $\lfloor E(\Gamma) \rfloor$x1.2$]$ and then calculate corresponding probability levels. Figures 3.2, 3.3, and 3.4 show probability levels of corresponding Γ scenarios of an instance for those three factor settings. Note that Γ=$\lfloor E(\Gamma) \rfloor$ of each instance is indicated by red points in the figures.

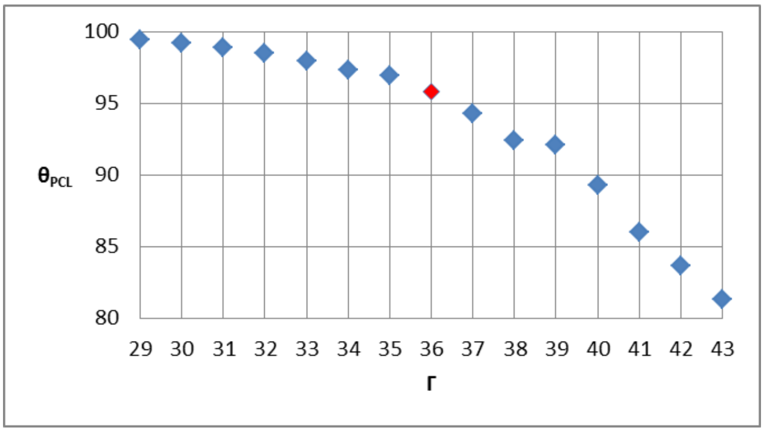

Figure 3.2: Probability level of an instance for setting $\alpha = 0.3$, $bf = 10\%$, c-prob $\in$ U(0.01-0.1)

The minimum probability levels are calculated for setting $\alpha = 0.3$, $bf = 10\%$, c-prob $\in$ U(0.01-0.1) in Table 3.4. Thus, we take an instance from this factorial setting, solve the model PCL by changing Γ values and show calculated probability levels in Figure 3.2. For this worst case factorial setting, if we estimate Γ parameter 20% higher than its expected value (i.e. estimate 43 instead of 36), then probability level drops from 95.5 % to around 82%. Apparently, estimating smaller Γ values than its expected value generates higher probability levels. For instance, if we estimate Γ parameter 20% smaller than its expected value (i.e. estimate 29 instead of 36), then we have probability level very close to 100%. The second minimum probability levels are generated for $\alpha = 0$, $bf = 10\%$, c-prob $\in$ U(0.01-0.1). Similarly, an instance is taken from this factorial setting and results are presented in Figure 3.3. For this instance, if we estimate Γ parameter 20% higher than its expected value, then we have probability level of 94%. The third minimum probability levels (albeit average probability level is 99.5%) are calculated for setting $\alpha = 0.3$, $bf = 10\%$, c-prob $\in$ U(0.01-0.2). An instance is also picked up from this setting, and probability levels for various Γ scenarios are given in Figure 3.4. For this case, Γ=$\lfloor E(\Gamma) \rfloor = 69$, so if we estimate Γ parameter as 82 (i.e. 20% higher than its expected value), then probability level is around 94.2%. Thus, even in the worst cases, we have still acceptable probability levels when we estimate Γ parameters higher than their expected values. Obviously, probability levels of remaining 5 factorial settings will be much better when we estimate higher Γ values.

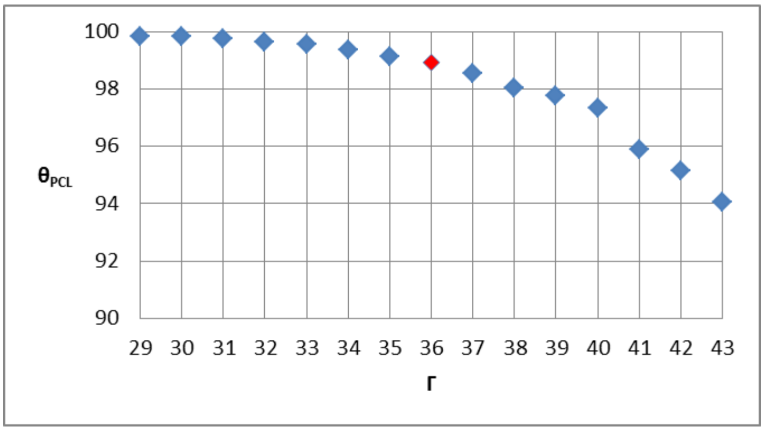

Figure 3.3: Probability level of an instance for setting $\alpha = 0$, $bf = 10\%$, c-prob $\in$ U(0.01-0.1)

Finally, the DM wonders about the value of incorporating the cancellations into the decision making process in terms of the budget utilization, the expected total score of selected projects, and the expected number of successfully completed projects. Note that the expected number of successfully completed projects ($E(nsc)$) can be obtained as follows:

$$nsc = ns - \Gamma \ \Rightarrow E(nsc) = ns - E(\Gamma) \tag{3.49}$$

If the cancellations aren't considered, a knapsack problem (KP) is solved. Therefore,

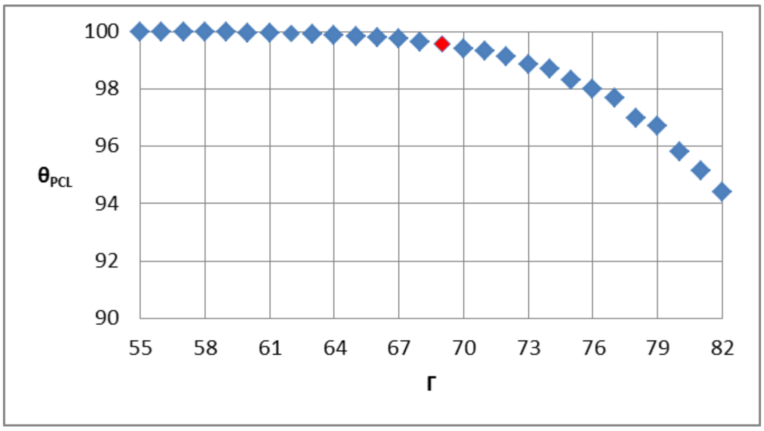

Figure 3.4: Probability level of an instance for setting $\alpha = 0.3$, $bf = 10\%$, c-prob $\in$ U(0.01-0.2)

for the problem size=3000 and for each factor combination, we solve 10 instances with the KP setting with s_i, b_i, and B of the instances. After that, with using probability information of the MISOCP, we obtain the expected total score, the expected budget expenditure (for probability level 95%) and the expected number of successfully completed projects of the solutions of the KP. Then, we calculate improvement values by using formula (3.50) and present improvements (%) in Table 3.5. For a given problem size=3000, we have 3 factors and 10 instances (i.e. 2^3 full factorial design with 10 replications). Hence, we obtain 2^3x10=80 improvement values for each measure. We have two levels for each factors, so each row in Table 3.5 indicates the minimum, the average and the maximum value of 40 improvement values for each measure.

$$\text{Improvement } (\%) = 100 \text{ x } \frac{\text{Expected Value of MISOCP-Expected Value of KP}}{\text{Expected Value of KP}} \tag{3.50}$$

$bf(\%)$ factor does not have a significant effect on improvement values, because we calculate improvement values in percent values. Thus, we notice that improvement values in terms of magnitude increase linearly with $bf(\%)$ factor. Factor c-prob has a significant effect on improvement values. As c-prob increases, improvement values increase more than two times. For instance, the average expected score improvement is 2.5% for c-prob $\in$ U(0.01-0.1), and it is 5.4% for c-prob $\in$ U(0.01-0.2). Similarly,

Table 3.5: Value of Modeling Cancellations

		Improvement (%)								
		E(score)			E(nsc)			Budget Utilization		
Factor	Levels	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
bf (%)	10	1.9	4.0	6.8	1.7	4.3	7.4	2.7	5.6	9.8
	20	1.9	3.9	6.4	1.9	4.2	7.1	3.0	5.9	9.9
c-prob	U(0.01-0.1)	1.9	2.5	3.1	1.7	2.7	3.6	2.7	3.6	4.6
	U(0.01-0.2)	4.1	5.4	6.8	4.4	5.8	7.4	6.0	7.9	9.9
α_i	0	2.7	4.6	6.8	2.7	5.0	7.4	3.9	6.9	9.9
	0.3	1.9	3.2	4.7	1.7	3.5	5.4	2.7	4.7	6.7

the average expected number of successfully completed projects increases from 2.7% to 5.8% and the average expected budget utilization increases from 3.6% to 7.9% as c-prob increases. Factor α_i has also a significant effect on improvement values. As it increases, improvement values decrease as we expected. This is due the fact that as α_i increases, the residual budget from the cancellations decreases. It is interesting to note that the expected budget utilization for probability level 95% is between 2.7% and 9.9%. Therefore, the budget utilization increases prominently with increase in the expected score and the expected number of successfully completed projects when we incorporate the cancellations into the decision making process.

Note that we assess the value of the modeling cancellations by comparing the knapsack model with our second proposed model MISOCP. However, the DM may raise concerns regarding the value of the first proposed model PCL. We can easily claim that the value of the model PCL will generate very close improvement values for the same probability level by examining the $\%\Delta_0$ results of Table 3.4 under the section VSM. To clarify, the second MISOCP model improves the expected score of first model for the same probability level between 0.01% and 0.13% (i.e. the model PCL is a tight approximation of the second model MISOCP for the same probability level as shown before). This implies that the improvement values of the model PCL will be very close to that of the MISOCP.

3.4 Conclusions

In this chapter, we have considered a one-stage P-RDPPS problem environment in the presence of cancellations. We considered the two cases. First, we assumed there exists no information regarding cancellation probabilities and proposed a model that handles cancellations. Our model maximizes the total score of the selected projects assuming that the unused budgets will occur due to the canceled projects. The model assumes that a given number of the least budget projects will be canceled, so that it mitigates the budget risk. We also developed an efficient DP algorithm to solve the problem. Both solution approaches could solve practical-size instances in short CPU times.

In the second case, we studied the problem with known cancellation probabilities. The problem is to maximize the expected total score of the selected projects while the risk of exceeding available budget is controlled via a chance-constraint on budget. In this case, we first showed that the number of cancellations follows Poisson binomial distribution. We next proved that the standardized sum of i.ni.d Bernoulli type budget random variables can be approximated by the standard normal distribution and formulate this problem as a CCSP. Then, we transformed it into its deterministic equivalent MISOCP and solved using IBM ILOG CPLEX. Our computational study showed that 90% of the instances could be solved to optimum in given time limit. The average gap for the remaining instances is below 0.02%.

We also made additional analyses to provide practical implications to the DM. For instance, we show that the budget risk of the first model is acceptable by using the probability information of the second model. Besides, we assess the budget risk levels of various cancellation situations for the same instance and observe that estimating higher number of cancellations at some degree still generates solutions with small budget risk. We also compare the two models and show that the first model is a tight approximation of the second model when we solve the second model with calculated budget risk of the first model. Moreover, we obtain the value of second model and observe that the second model generates better project portfolios with anticipated risk levels. Finally, we assess the value of incorporating the cancellations to the model by comparing it with the standard setting and show that the proposed approaches improve both the budget utilization and the total expected score of the selected projects.

An article written on the results of Chapter 3 is accepted with minor revisions for publication in OR Spectrum (Çağlar and Gürel (2016)).

In the next chapter, we propose a new model that incorporates not only cancellations but also budget expenditure pattern of successfully completed projects to improve funding budget utilization. In addition, we assume continuous expenditure distribution for cancellations and underspending of successfully completed projects. Moreover, we propose an alternative approach and compare it with our new model.

CHAPTER 4

PUBLIC RDPPS PROBLEM UNDER EXPENDITURE UNCERTAINTY

In this chapter, we consider a one-stage public RDPPS problem with several sources of expenditure uncertainty. As we present in the RDPPS literature, financial data (i.e. budget, cost, expenditure, profit, cash flow etc.) uncertainty is usually modeled by defining random variables and determining their associated probability distributions. For example, Medaglia et al. (2007) formulate cash flow and market share uncertainty of R&D projects with triangular, exponential, and Erlang random variables. Solak et al. (2010) model annual financial returns of R&D projects as discrete random variables and assume that associated probabilities are known. Note that both studies deal with very small number of projects (i.e. less than 10 projects) and formulate financial data of each project with a different continuous or discrete distribution.

In our public RDPPS problem, we consider thousands of project proposals with specific budgets. Since it is not practical and realistic to use a separate expenditure distribution for each project, for every project we have modeled the ratio of budget spent by a probability distribution. To be more clear, each project has an approved budget b_i, and we know that its expenditure $\hat{b}_i$ will be within the interval $(0, b_i]$. Thus, we define a ratio $R_i = \frac{\hat{b}_i}{b_i}$ to model budget spending behavior of a project i and focus on how R_i can be modeled in a tractable manner. In the literature, beta family of distributions is commonly used for ratio uncertainty modeling. They are also very flexible and may take many different forms according to the distribution parameters (Johnson et al. (1995), Chapter 25). Therefore, we utilize beta distributions in formulation of R_i.

In Section 4.1, we first state the problem and our assumptions. In Section 4.2, we formulate the underspending uncertainty of both the canceled projects and successfully completed projects. In Section 4.3, we propose a chance constrained stochastic programming (CCSP) model for our problem. From a methodological standpoint, chance constraint of budget is computationally intractable; hence, we apply a solution method based on a normal approximation and a second order cone programming. However, when the normal approximation is used, there can be some convergence error that can distort the desired risk level of the chance constraint of funding budget. Therefore, by using Berry-Esseen theorem, we quantify convergence quality of the normal approximation and propose ways to mitigate risk of probabilistic budget constraint. In Section 4.4, we present computational results. Proposed model could exactly solve 86% of the the instances to optimality within the given time limit, and the average optimality gap for the instances that were only solved to feasibility is below 0.01%. In Section 4.5, we also perform additional experiments to give practical insights to the DM. For instance, we investigate the case of unknown distribution of budget underutilization of canceled and successfully completed projects. For this case, we develop an alternative distribution modeling and compare it with the proposed approach on different problem settings. Finally, we compare the proposed approach with a standard setting and find that the proposed approach delivers an increase between 8.4% and 18.6% in utilization of the funding budget. Concluding remarks are presented in Section 4.6.

4.1 Problem Statement

We address a one-stage (i.e. funding decisions are made after all applying projects are evaluated and decisions are irreversible) public RDPPS problem. R&D funding agencies usually adopt a call-based system. A call is announced and researchers apply with their solicited research proposals for funding. Research proposals are analyzed for eligibility check and then eligible proposals are evaluated by peer reviewers in panel meetings. Scientific and technological benefits of projects are scored by peer reviewers according to the set of selection criteria. Budget of each project is not regarded as a selection criteria; however it is considered in research panels for a descent project budget plan. Funding decision of projects are made according to the panel scores, project budgets and total available funding budget. PIs of the funded projects sign a project funding contract with public R&D agency. Project funding contract commits that approved budget will be transferred to the project if project comply with the terms and conditions of grant policy statements. This is a typical process in a one-stage call-based programs such as 1001 program of TÜBİTAK.

After funding decisions, funded projects start to conduct their planned research agenda. However, during the research activities some of the R&D projects can be canceled by the R&D funding agency due to the violation of terms and conditions. If a cancellation occurs, most of the budget expenditure is usually paid back. In some cases, some expenditures can be excluded. Those details are provided in grant policy statements. Besides, expected scientific and technological benefit of canceled projects is not achieved. Therefore, cancellations not only affect expected benefits of project portfolio but also underutilization of funding budget. Moreover, some of the funded projects finish successfully but they do not expend their whole budget. This is an expected outcome since budgeting process is a complex issue and researchers focus on R&D content development rather than perfect budgeting of project activities. For that reason, expected benefits of those project come true with an underutilized budget.

In this section, we consider aforementioned cases and formulate expected budget expenditure and benefit (score) of projects with a probabilistic approach to improve the expected score of project portfolio as well as utilization of the funding budget. Before formulating the analytical developments, we herein state following assumptions for our problem:

(1) Budget of each project is clarified during research panels. Score of each project is rated by peer reviewers according to the set of selection criteria. Score includes scores of panel jury. Scoring does not evaluate budget of projects.

(2) The funding decisions are made at the end of each call. Therefore, one-stage problem environment is considered.

(3) Probability information of the aforementioned factors (i.e. canceled and successful projects with project expenditure uncertainty) can be assessed using

judgment or estimated using past data. For example, there is usually a criterion that evaluates project management, team and research possibilities. Therefore, there can be a close relationship between cancellation probability and that kind of criterion, and this information can be gathered from past canceled projects' data. Besides, proposals are gathered from different PIs and universities. There can also be a relation between cancellation risks and PIs, universities. Moreover, peer reviewers can comment on cancellation risk of each project and those comments can be integrated with past data to estimate cancellation probabilities. Probability information of successful projects with underutilized budget also can be obtained from past data of successfully finished projects.

(4) Large number of projects compete to be awarded for funding under limited available funding budget. Thus, project portfolio optimization should be used.

4.2 Proposed Model

In this section, we formulate aforementioned concepts (i.e. canceled and successful projects with underutilized budget case) by considering their effect on score and budget of projects. Let s_i denote the score of project i. We already state that expected scientific and technological benefits of the canceled projects is not fulfilled. Therefore, the score of project i becomes zero with cancellation probability p_i. Let the random variable $\hat{s}_i$ denote the real score of project i. Then, we can express $\hat{s}_i$ as follows:

$$\hat{s}_i = \begin{cases} s_i & \text{w.p. } (1 - p_i) \\ 0 & \text{w.p. } p_i \end{cases} \tag{4.1}$$

Expectation and variance of $\hat{s}_i$ are calculated as follows:

$$\mathbb{E}(\hat{s}_i) = s_i(1 - p_i) \tag{4.2}$$

$$\text{Var}(\hat{s}_i) = s_i^2 p_i(1 - p_i) \tag{4.3}$$

Let b_i denote the approved budget of project i. We examine budget expenditure uncertainty by using a confidential sample data of a call. According to our observation, there could be three different cases of budget spending behavior as follows:

(1) cancellation of the project i and underspending of its budget with probability p_i

(2) successful completion of the project i and underutilization of its budget with probability q_i

(3) successful completion of the project i and fully used its budget with probability $1 - p_i - q_i$

We also observe that budget expenditure ratio for the first case could be modeled by the beta distribution in the interval $(0, \tau_1)$ and similarly the budget expenditure ratio for the second case could be modeled by another beta distribution in interval $(\tau_2, 1)$. Thus, both canceled and successful projects with underutilized budget cause the budget expenditure uncertainty. Let the random variable $\hat{b}_i$ denote the real expenditure of project i. Then, we express $\hat{b}_i$ with a ratio random variable R_i defined on an interval (0,1] as follows:

$$\hat{b}_i = b_i R_i \tag{4.4}$$

Mixture distributions arise when statistical sample data can be categorized into separate subsamples. Therefore, R_i can be formulated as a mixture distribution according to aforementioned three cases of budget spending. We already state that budget expenditure for the first two cases can fit to truncated (bounded) beta distribution. Third case can be modeled by a degenerate distribution (i.e. R_i will be 1 with probability $1 - p_i - q_i$). In the literature, similar cases are modeled by a truncated inflated beta distribution. It is recently proposed by Pereira et al. (2012) to model unemployment insurance benefit ratio. Truncated inflated beta distribution is a mixture distribution of truncated standard beta distribution in some bounded open interval (a, b), and it is inflated in some points. Therefore, we formulate random variable R_i as a variant of a truncated inflated beta distribution. We can introduce the probability density function (p.d.f) of the random variable R_i by taking into account the three cases as follows:

$$f_{R_i}(r) = \begin{cases} p_i f(r; \alpha_1, \beta_1, 0, \tau_1) & \text{if } r \in (0, \tau_1) \\ q_i f(r; \alpha_2, \beta_2, \tau_2, 1) & \text{if } r \in (\tau_2, 1) \\ 1 - p_i - q_i & \text{if } r = 1 \end{cases} \tag{4.5}$$

where $f(r; \alpha_1, \beta_1, 0, \tau_1)$ is the p.d.f of a truncated beta distribution in interval $(0, \tau_1)$ with parameters α_1, β_1 and $f(r; \alpha_2, \beta_2, \tau_2, 1)$ is the p.d.f of a truncated beta distribution in interval $(\tau_2, 1)$ with parameters α_2, β_2. Note that we assume $\tau_1 \leq \tau_2 < 1$. Hence, R_i is the mixture distribution of two truncated beta random distributions and a degenerate distribution (inflated point at 1). In order to obtain cumulative distribution function (c.d.f) of the R_i, we need the definition of a mixture random variable and p.d.f of a truncated beta distribution. In the following definitions, we first give definition of a mixture random variable and then p.d.f of a truncated beta distribution for the sake of completeness.

Definition 4.2.1. *Given a finite set of (J) distribution functions, let $f_j(x)$ and $F_j(x)$ be the p.d.f. and c.d.f. of distribution j, respectively. Let w_j be the probability of selecting distribution $j \in J$ such that $w_j > 0$ and $\sum_j w_j = 1$. Then p.d.f and c.d.f of the mixture random variable X, $f(x)$ and $F(x)$ can be expressed as follows:*

$$f(x) = \sum_{j \in J} w_j f_j(x) \tag{4.6}$$

$$F(x) = \sum_{j \in J} w_j F_j(x) \tag{4.7}$$

Definition 4.2.2. *Let W be a truncated (bounded) beta distribution in open interval (a, b) with shape parameters α and β. Its p.d.f and c.d.f are given as follows (see Johnson et al. (1995), Chapter 25):*

$$f_W(w; a, b, \alpha, \beta) = \frac{\Gamma(\alpha+\beta)(w-a)^{\alpha-1}(b-w)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)(b-a)^{\alpha+\beta-1}} \quad \textit{for } a < W < b \tag{4.8}$$

$$P(W \leq w) = F_W(w) = \int_a^w \frac{\Gamma(\alpha+\beta)(w-a)^{\alpha-1}(b-w)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)(b-a)^{\alpha+\beta-1}} dw \tag{4.9}$$

where the Gamma function is defined as: $\Gamma(u) = \int\limits_0^\infty s^{u-1}e^{-s}ds.$

By using Definitions 4.2.1 and 4.2.2, we can obtain the cumulative distribution function (c.d.f.) of R_i explicitly as follows:

$$P(R_i \le k) = F_{R_i}(k) = p_i \int_0^{min(\tau_1,k)} \frac{\Gamma(\alpha_1+\beta_1)r^{\alpha_1-1}(\tau_1-r)^{\beta_1-1}}{\Gamma(\alpha_1)\Gamma(\beta_1)\tau_1^{\alpha_1+\beta_1-1}}dr$$
$$+ q_i \int_{\tau_2}^{min(1,k)} \frac{\Gamma(\alpha_2+\beta_2)(r-\tau_2)^{\alpha_2-1}(1-r)^{\beta_2-1}}{\Gamma(\alpha_2)\Gamma(\beta_2)(1-\tau_2)^{\alpha_2+\beta_2-1}}dr + (1-p_i-q_i)\mathbb{I}_1(k) \quad (4.10)$$

where $k \in (0,1]$ and $\mathbb{I}_1(k)$ is the indicator function and takes the value of 1 if $k = 1$ and 0 otherwise.

So far, we obtain the expected score ($\mathbb{E}(\hat{s}_i)$) and p.d.f and c.d.f of real budget expenditure ($\hat{b}_i$) of each project. We are now ready to formulate the CCSP model of the public RDPPS problem as follows:

$$\text{(SP)} \ \max \ \sum_{i\in N} s_i(1-p_i)x_i \qquad \text{(Expected Total Score)}$$
$$\text{s.t.} \ P\left(\sum_{i\in N} b_i R_i x_i \le B\right) \ge \theta \qquad (4.11)$$
$$x_i \in \{0,1\} \ \forall i \in N \qquad (4.12)$$

where $P(\cdot)$ denotes the probability measure. x_i is a decision variable representing 1, if project i is selected and 0, otherwise. N indicates the set of all projects. B is the total available budget. Objective function is to maximize the expected total score of selected projects. Chance constraint in (4.11) provides that sum of random budget expenditure of selected projects does not exceed the total available budget with a probability level of θ. Some policy, geographical, and sectoral constraints can be also incorporated into our modeling framework; however, in this study, our main emphasis is modeling of uncertain project expenditures of canceled and successfully completed projects.

When $p_i = 0$ and $q_i = 0 \ \forall \ i$, the problem is a knapsack problem. So, this special case is NP-hard. In addition, Nemirovski and Shapiro (2006) discuss that most of the individual linear chance constraints are computationally intractable because checking feasibility of a solution is NP-hard for $\theta > 0.5$. They state that a linear chance constraint is computationally tractable for some special distributions such as multivariate normal distribution. If we adopt normal approximation, and set $\theta = 0.5$, then we have a knapsack capacity type constraint with mean $\hat{b}_i$ values. If we assume $\theta > 0.5$, we have additional quadratic terms for the variance of $\hat{b}_i$. Therefore, we could state that our problem is $\mathcal{NP}$-hard.

In order to solve model SP with an exact approach, we need to transform the chance constraint in (4.11) to its deterministic counterpart. Thus, we have to derive quantile function (i.e. inverse cumulative distribution function) of sum of random budget expenditure in the probabilistic constraint (4.11). We derive the c.d.f of a single random variable R_i in (4.10). Note that its c.d.f has no closed-form expression. Therefore, derivation of its quantile function is a challenging issue. Moreover, to obtain c.d.f. and quantile function of term $\sum_{i \in N} b_i R_i$ in (4.11), convolution of many non-identical distributions should be derived. Therefore, obtaining true c.d.f. of summation of many random terms in probabilistic constraints is usually computationally intractable. That is why convex approximations such as normal approximation are usually employed. In order to apply normal approximation to sum of many independent non-identically distributed (i.ni.d.) random variables, their standardized sum should obey the central limit theorem. There are two version of CLT theorem adopted for normal approximation, namely Lyapunov and Lindeberg CLTs. Lyapunov CLT is stronger than Lindeberg. We refer to Shapiro et al. (2009), pp. 141-144 for recent discussion of different versions of CLT for i.ni.d. random variables. In the following Theorem, we first give the Lyapunov CLT. After then in Proposition 4.2.1, we show that general family of truncated distributions obeys Lyapunov CLT.

Theorem 4.2.1. *Lyapunov's Central Limit Theorem (CLT) for i.ni.d. random variables(Baurer (1996) and Shapiro et al. (2009)): Let $X_1, X_2, ..., X_n$ be i.ni.d. random variables with finite expectation $\mathbb{E}(X_i)$, positive variance Var(X_i) and finite moments $\mathbb{E}(X_i^{2+\delta})$ for δ > 0. Then, if for some δ > 0, the following condition*

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{i=n} \mathbb{E}\left[|X_i - \mathbb{E}(X_i)|^{2+\delta}\right]}{\left[\sqrt{\sum_{i=1}^{i=n} Var(X_i)}\right]^{2+\delta}} = 0 \tag{4.13}$$

is satisfied, then normalized summand $S_n = \frac{\sum_{i=1}^{i=n}[X_i - \mathbb{E}(X_i)]}{\sqrt{\sum_{i=1}^{i=n} Var(X_i)}}$ *converges to standard normal distribution (*$N(0,1)$*) as* n *goes to infinity.*

Proposition 4.2.1. *Let* $\{T_i\}_{i=1}^{i=n}$ *be an i.ni.d. truncated random variables defined on bounded interval* $[l,u]$ *such that* $0 \leq l < u$*. Then, Lyapunov's CLT theorem applies to sequence* $\{T_i\}_{i=1}^{i=n}$ *as* n *goes to infinity.*

Proof. See Appendix A.1. □

In the following corollary, we show that standardized version of term $\sum_{i\in N} b_i R_i x_i$ in constraint (4.11) converge to standard normal distribution.

Corollary 4.2.1.1. *Let* $b_1R_1, b_2R_2, ..., b_iR_i \in N$ *be i.ni.d. random variables as defined in equations* (4.4) *and* (4.11)*. Then* $H_n^x = \frac{\sum_{i\in N}[b_iR_i - \mathbb{E}(b_iR_i)]x_i}{\sqrt{\sum_{i\in N} Var(b_iR_i)x_i^2}}$ *converges to* $N(0,1)$ *if the solution of model SP includes statistically significant number of supported projects (i.e.* $x_i = 1$*).*

Proof. See Appendix A.2. □

Convergence to standard normal distribution is provided in Corollary 4.2.1.1. Therefore, we can transform CCSP model of public RDPPS problem to its deterministic equivalent formulation by using normal approximation. However, as shown in Corollary 4.2.1.1, we need the mean and variance of random variable $\hat{b}_i = b_i R_i$ to approximate true distribution of H_n^x . By using the properties of expectation and variance operator, we obtain:

$$\mathbb{E}(\hat{b}_i) = \mathbb{E}(R_i b_i) = b_i \mathbb{E}(R_i) \tag{4.14}$$

$$\text{Var}(\hat{b}_i) = \text{Var}(R_i b_i) = b_i^2 \text{Var}(R_i) \tag{4.15}$$

Since approved budget of each project b_i is known, we need to have $\mathbb{E}(R_i)$ and Var(R_i) to apply normal approximation. Since R_i is a mixture distribution including truncated beta distribution, we first define the derivation of moment of mixture distributions and then we give the moment information of truncated beta distribution in order to obtain mean, variance and n^{th} moment of random variable R_i.

Definition 4.2.3. *Let X be a mixture random variable as defined in Definition 4.2.1. The n^{th} moment of X is formulated as follows:*

$$\begin{aligned}\mathbb{E}[X^n] &= \int_{-\infty}^{+\infty} x^n f(x)dx = \int_{-\infty}^{+\infty} x^n \sum_{j\in J} w_j f_j(x)dx \\ &= \sum_{j\in J} w_j \int_{-\infty}^{+\infty} x^n f_j(x)dx = \sum_{j\in J} w_j m_j^n \end{aligned} \tag{4.16}$$

where m_j^n is the n^{th} moment of distribution j.

In the following proposition, we derive expectation, variance and n^{th} moment of a truncated beta distributed random variable.

Proposition 4.2.2. *The mean, variance and n^{th} moment of the truncated beta random variable W defined in open interval (a, b) are given as follows:*

$$\mathbb{E}(W) = \frac{\alpha b + \beta a}{\alpha + \beta} \tag{4.17}$$

$$Var(W) = \frac{(b-a)^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{4.18}$$

$$\mathbb{E}(W^n) = \begin{cases} \sum_{k=0}^{k=n} \frac{n!}{k!(n-k)!} a^k (b-a)^{n-k} \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+n-k)}{\Gamma(\alpha)\Gamma(\alpha+\beta+n-k)} & \textit{if } 0 < a < b \\ b^n \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+n)}{\Gamma(\alpha)\Gamma(\alpha+\beta+n)} & \textit{if } a = 0 \textit{ and } b > 0 \end{cases} \tag{4.19}$$

Proof. See Appendix A.3. □

We give the necessary definitions and derivations to obtain mean, variance and n^{th} moment of random variable R_i. We are now ready to derive its moment information in the following proposition.

Proposition 4.2.3. *The mean, variance and n^{th} moment of the truncated inflated beta distributed random variable R_i are given as follows:*

$$
\begin{aligned}
\mathbb{E}(R_i) &= p_i \frac{\alpha_1 \tau_1}{\alpha_1 + \beta_1} + q_i \frac{\alpha_2 + \beta_2 \tau_2}{\alpha_2 + \beta_2} + (1 - p_i - q_i) && (4.20)\\
Var(R_i) &= p_i \left[\left(\frac{\alpha_1 \tau_1}{\alpha_1 + \beta_1} \right)^2 + \frac{\tau_1^2 \alpha_1 \beta_1}{(\alpha_1 + \beta_1)^2 (\alpha_1 + \beta_1 + 1)} \right] \\
&\quad + q_i \left[\left(\frac{\alpha_2 + \beta_2 \tau_2}{\alpha_2 + \beta_2} \right)^2 + \frac{(1 - \tau_2)^2 \alpha_2 \beta_2}{(\alpha_2 + \beta_2)^2 (\alpha_2 + \beta_2 + 1)} \right] + (1 - p_i - q_i) \\
&\quad - \left[p_i \frac{\alpha_1 \tau_1}{\alpha_1 + \beta_1} + q_i \frac{\alpha_2 + \beta_2 \tau_2}{\alpha_2 + \beta_2} + (1 - p_i - q_i) \right]^2 && (4.21)\\
\mathbb{E}(R_i^n) &= p_i \left(\tau_1^n \frac{\Gamma(\alpha_1 + \beta_1) \Gamma(\alpha_1 + n)}{\Gamma(\alpha_1) \Gamma(\alpha_1 + \beta_1 + n)} \right) + (1 - p_i - q_i) \\
&\quad + q_i \left(\sum_{k=0}^{k=n} \frac{n!}{k!(n-k)!} \tau_2^k (1 - \tau_2)^{n-k} \frac{\Gamma(\alpha_2 + \beta_2) \Gamma(\alpha_2 + n - k)}{\Gamma(\alpha_2) \Gamma(\alpha_2 + \beta_2 + n - k)} \right) && (4.22)
\end{aligned}
$$

Proof. See Appendix A.4. □

4.3 Deterministic Equivalent Formulation

We derive mean and variance of random variable R_i in Proposition 4.2.3. In Corollary 4.2.1.1, we show that total random budget spending obeys the central limit theorem. Thus, in this section, we formulate the deterministic equivalent programming of model SP by using normal approximation and second order conic inequalities.

Constraint set (4.11) can be expressed as follows:

$$
P\left(\sum_{i \in N} b_i R_i x_i \leq B \right) \Rightarrow P\left(\frac{\sum_{i \in N} [b_i R_i - b_i \mathbb{E}(R_i)] x_i}{\sqrt{\sum_{i \in N} b_i^2 \text{Var}(R_i) x_i^2}} \leq \frac{B - \sum_{i \in N} b_i \mathbb{E}(R_i) x_i}{\sqrt{\sum_{i \in N} b_i^2 \text{Var}(R_i) x_i^2}} \right) \tag{4.23}
$$

In Corollary 4.2.1.1, we show that H_n^x converges to standard normal distribution. Therefore, we can obtain the following probabilistic constraint:

$$
P\left(Z \leq \frac{B - \sum_{i \in N} b_i \mathbb{E}(R_i) x_i}{\sqrt{\sum_{i \in N} b_i^2 \text{Var}(R_i) x_i^2}} \right) \geq \theta \Rightarrow \Phi \left(\frac{B - \sum_{i \in N} b_i \mathbb{E}(R_i) x_i}{\sqrt{\sum_{i \in N} b_i^2 \text{Var}(R_i) x_i^2}} \right) \geq \theta
$$

$$\Rightarrow \frac{B - \sum_{i\in N} b_i \mathbb{E}(R_i) x_i}{\sqrt{\sum_{i\in N} b_i^2 \text{Var}(R_i) x_i^2}} \geq \Phi^{-1}(\theta) \quad (4.24)$$

where Z is the standard normal random variable and $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ are its cumulative distribution function (c.d.f) and quantile function, respectively. We reorganize (4.24) as follows:

$$\sum_{i\in N} b_i \mathbb{E}(R_i) x_i + \Phi^{-1}(\theta) \sqrt{\sum_{i\in N} b_i^2 \text{Var}(R_i) x_i^2} \leq B \quad (4.25)$$

Deterministic equivalent formulation of the constraint (4.11) is obtained in Constraint (4.25). We assume that $\theta \geq 0.5$, then $\Phi^{-1}(\theta) > 0$, which makes the constraint set (4.25) convex and it can be reformulated by second-order conic inequalities. Resulting deterministic equivalent reformulation of the SP model is a mixed integer second-order cone program (MISOCP):

$$\text{(SP-1)} \ \max \sum_{i\in N} s_i (1 - p_i) x_i \qquad \text{(Expected Total Score)}$$

$$\text{s.t } \eta = \frac{B}{\Phi^{-1}(\theta)} - \frac{\sum_{i\in N} b_i \mathbb{E}(R_i) x_i}{\Phi^{-1}(\theta)} \quad (4.26)$$

$$\sum_{i\in N} b_i^2 \text{Var}(R_i) x_i^2 \leq \eta^2 \quad (4.27)$$

$$\eta \geq 0 \quad (4.28)$$

$$x_i \in \{0, 1\} \ \ \forall i \in N \quad (4.29)$$

The objective function is to maximize the expected total portfolio score. Conic reformulation of constraint (4.25) is derived in constraint sets (4.26) and (4.27). η in equation (4.26) is an auxiliary variable for linear portion of the constraint (4.25). Constraint (4.27) is a second-order cone generated by constraint (4.25). $\mathbb{E}(R_i)$ is derived in equation (4.20), and Var(R_i) is derived in equation (4.21). An efficient frontier of model SP-1 can be obtained by solving it for different values of θ.

Regarding solutions of the proposed model, DMs may wonder two kinds of information. The first one is the expected number of canceled projects since canceled projects are regarded as unsuccessful. Therefore, DMs need to know the expected number of cancellations. By using cancellation probability of each project (p_i), we can obtain expected total number of cancellations. Let I_i be a Bernoulli random variable with success parameter p_i and define $\Gamma = \sum_{i \in N} I_i x_i$ that denotes number of cancellations. Γ follows Poisson-Binomial distribution (Hong (2013)). We derive expected number of canceled projects and its variance for a solution vector x as follows:

$$\mathbb{E}(\Gamma) = \sum_{i \in N} \mathbb{E}(I_i) x_i = \sum_{i \in N} p_i x_i \tag{4.30}$$

$$\text{Var}(\Gamma) = \sum_{i \in N} \text{Var}(I_i) x_i = \sum_{i \in N} p_i (1 - p_i) x_i \tag{4.31}$$

Secondly, convergence quality of true unknown distribution to normal approximation is a significant concern for DMs. They want to gain managerial insight into the quality of approximation since project portfolio decisions rely on public financial resources. In the CCSP model for the public RDPPS problem, θ provides the confidence level and conversely $1 - \theta$ specify the risk level. DMs prefer high level of confidence level (i.e. very low risk level) in the CCSP models. However, this confidence level is not a precise value for true unknown distribution and in fact it represents exact confidence level of standard normal distribution. For that reason, if there is some convergence error between true distribution and standard normal distribution and that error can distort risk level of $1 - \theta$. For example, Hong (2013) demonstrate that convergence rate of normal approximation can be slower especially at the tails of distribution for the Poisson-Binomial distribution case. Since high level of confidence refers to the right tail of distributions (i.e. right boundary points), convergence error concerns can be critically significant. Therefore, derivation of convergence error with probability metrics can assist DMs to assess real risk level of probabilistic constraints when approximation methods are used. In the following, Berry–Esseen theorem, maximum error of normal approximation is given by Esseen (1956) in terms of probability levels. Berry–Esseen theorem states that for any realization on the probability space, maximum difference between the true unknown distribution and the standard normal distribution in terms of probability levels (i.e. confidence levels in CCSP context) has a bound. Hence, this theorem assist to quantify the real confidence level error when normal approximation is employed in chance constrained programming.

Theorem 4.3.1. *Berry–Esseen theorem for the quality of normal approximation: Let $Y_1, Y_2, ...Y_n$ be i.ni.d. random variables with $\mathbb{E}(Y_i) = 0$, positive second moment $\mathbb{E}(Y_i^2)$ and a finite third moment $\mathbb{E}(Y_i^3) < \infty$. Let $Q_n = \frac{\sum_{i=1}^{i=n} Y_i}{\sqrt{\sum_{i=1}^{i=n} \mathbb{E}(Y_i^2)}}$, G_n is the c.d.f of Q_n, Φ is the c.d.f of the standard normal distribution. Then, for the Kolmogorov distance is defined by*

$$D_{_{Kol}} = \sup_{z \in \mathbb{R}} |G_n(z) - \Phi(z)| . \tag{4.32}$$

there exists a constant C, such that $D_{_{Kol}} \leq C\psi$ where

$$\psi = \left(\sum_{i=1}^{i=n} \mathbb{E}(|Y_i^3|) \right) \left(\sum_{i=1}^{i=n} \mathbb{E}(Y_i^2) \right)^{-3/2} . \tag{4.33}$$

Remark 4.3.1. *Esseen (1956) theoretically showed that the constant C satisfies*

$$7.59 \geq C \geq \frac{\sqrt{10}+3}{6\sqrt{2\pi}} \approx 0.4097$$

However, the best estimate on the upper bound has substantially been improved by researchers over past decades. It is 0.56 *obtained recently by Shevtsova (2010).*

Remark 4.3.2. *Berry-Esseen's theorem depends on only the first three moments to give the upper bound on the maximum error of normal approximation.*

In Theorem 4.3.1, we give the convergence error for a general family of random variables. For our case, we need to derive some moment information of random variable R_i to apply Berry–Esseen theorem. In Corollary 4.3.1.1, we derive Berry-Esseen bound for any solution obtained by model SP-1.

Corollary 4.3.1.1. *Let $b_1R_1, b_2R_2, ..., b_iR_i \in N$ be i.ni.d. random variables as defined in equations 4.4 and 4.11. Let G_n^x be the true c.d.f of $H_n^x = \frac{\sum_{i \in N}[b_iR_i - b_i\mathbb{E}(R_i)]x_i}{\sqrt{\sum_{i \in N} b_i^2 Var(R_i)x_i^2}}$ for a specific solution vector x. Then Kolmogorov distance between G_n^x and Φ^x (Normal approximation for the solution vector x) satisfies:*

$$D_{_{Kol}}^x \leq C \left(\sum_{i=1}^{i=n} \left| b_i^3\mathbb{E}(R_i^3) - 3b_i^3\mathbb{E}(R_i^2)\mathbb{E}(R_i) + 2b_i^3[\mathbb{E}(R_i)]^3 \right| x_i \right) \left(\sum_{i=1}^{i=n} b_i^2 Var(R_i)x_i \right)^{-3/2} \tag{4.34}$$

where $\mathbb{E}(R_i)$, and $\mathbb{E}(R_i^2)$ are determined in equations (4.20) *and* (4.21). *$\mathbb{E}(R_i^3)$ is derived in proof by using equation in* (4.22).

Proof. See Appendix A.5. □

We can calculate an upper bound for the maximum error of normal approximation by using recent best estimate of C.

In the next section, we present computational study.

4.4 Computational Results

In this section, we report the results of computational experiments for the proposed model. We solve model SP-1 by using IBM ILOG CPLEX 12.6 via Concert Technology and C++ coding language. All experiments are conducted on a computer with processor Intel Core i5 1.7 GHz, 8.00 GB memory (RAM), 64-bit operating system, and Windows 7 Professional. We set the time limit to 3 hours (i.e. 10800 CPU seconds). We conduct a 2^k full factorial design to assess impacts of different problem parameters. Since there are many parameters in our problem, we only consider most important factors in our full factorial design. The six important factors and their levels are presented in Table 4.1.

Table 4.1: Factor Values

		Levels	
Factor	Description	Low	High
Problem size	Number of project applications	1000	2000
p_i	Cancellation probability	U(0.01-0.1)	U(0.01-0.2)
$q_i = q \; \forall i \in N$	Probability of underutilized budget for successfully completed projects	0.4	0.5
$\mathbb{E}(W_1)$; $\mathbb{E}(W_2)$	Mean of random variables W_1 and W_2	0.3 ; 0.8	0.35 ; 0.85
Var(W_1)	Variance of random variable W_1	0.01	0.05
Var(W_2)	Variance of random variable W_2	0.01	0.05

Problem size represents number of projects applying for funding. We use two problem size (i.e. 1000 and 2000) factor levels that are set according to the real life examples. Recall that random variable R_i depends on many parameters. We next give factorial design of parameters of R_i. Cancellation probability of project i (p_i) is generated according to uniform distribution (U) between 0.01-0.1 (at low level) and

0.01-0.2 (at high level) to assess probability level affect. We first set $q_i = q \; \forall i \in N$ because it is difficult to assign different underspending probabilities to successfully completed projects. Common probability (q) of successful completion of projects with underspending budget is set to 0.4 and 0.5, respectively. Note that W_1 is random budget expenditure ratio of canceled projects and W_2 is the random budget expenditure ratio of successfully completed projects. Recall that we model them as truncated beta distributions. We set mean of truncated beta distributions as pairs such that $\mathbb{E}(W_1) = 0.3$, $\mathbb{E}(W_2) = 0.8$ and $\mathbb{E}(W_1) = 0.35$, $\mathbb{E}(W_2) = 0.85$. We know means have similar effects so that we take them as pairs and thereby we obtain more tractable factorial design. We use a minimum and a maximum value for variance of truncated distributions. We set both $Var(W_1)$, and $Var(W_2)$ to 0.01 and 0.05.

We solve 5 random instances for each factor combination. Project instances, namely replications, are generated according to Table 4.2.

Table 4.2: Other Problem Parameters

Parameter	Description	Value
s_i	Score of project i	U(10-25)
b_i	Budget of project i	U(5-30)
bf (%)	Ratio for available budget over sum of all project budgets	10
B	Available funding budget	$(\sum_{i \in N} b_i) \times bf$
τ_1	Range parameter of random variable W_1	0.5
τ_2	Range parameter of random variable W_2	0.5
θ	Probability level of chance constraint	0.95

As we mentioned before, an aggregate score approach (i.e. sum of scores of evaluation criteria) is considered since it is currently adopted in many funding bodies such as TÜBİTAK. Project score (s_i) is, as compatible with real life instances, uniformly distributed in interval [10-25] as integers, and project budget (b_i) is, as compatible with real life instances, uniformly distributed in interval [5-30] (in 10,000 monetary units) as integers. We define a $bf(\%)$ parameter representing ratio for available budget over sum of all projects budgets. Hence, total available budget funding (B) is determined as $B = (\sum_{i \in N} b_i) \times bf$. Range parameters τ_1 and τ_2 of random variables W_1 and W_2 are set to 0.5. We set probability level of chance constraint (θ) to 0.95.

By using mean and variance of truncated distributions, shape parameters (i.e. α_1, β_1 of W_1, and α_2, β_2 of W_2) of them are calculated using equations (4.35) through (4.38). We refer to Johnson et al. (1995), Chapter 25, Section 25.2 for parameter setting in beta family of distributions. For given values of $\mathbb{E}(W_1)$, $Var(W_1)$, and τ_1, α_1 parameter of random variable W_1 is calculated according to (4.35) and then α_1 is plugged in equation (4.36) to find β_1. Similar calculations are performed by using parameters of random variable W_2 in equations (4.37) and (4.38) to obtain α_2 and β_2.

$$\alpha_1 = \left(\frac{\mathbb{E}(W_1)}{\tau_1}\right)^2 \left(1 - \frac{\mathbb{E}(W_1)}{\tau_1}\right) \left(\frac{Var(W_1)}{\tau_1^2}\right)^{-1} - \left(\frac{\mathbb{E}(W_1)}{\tau_1}\right) \tag{4.35}$$

$$\beta_1 = \frac{\left(\frac{\mathbb{E}(W_1)}{\tau_1}\right)\left(1 - \frac{\mathbb{E}(W_1)}{\tau_1}\right)}{\left(\frac{Var(W_1)}{\tau_1^2}\right)} - 1 - \alpha_1 \tag{4.36}$$

$$\alpha_2 = \left(\frac{\mathbb{E}(W_2) - \tau_2}{1 - \tau_2}\right)^2 \left(1 - \frac{\mathbb{E}(W_2) - \tau_2}{1 - \tau_2}\right) \left(\frac{Var(W_2)}{(1 - \tau_2)^2}\right)^{-1} - \left(\frac{\mathbb{E}(W_2) - \tau_2}{1 - \tau_2}\right) \tag{4.37}$$

$$\beta_2 = \frac{\left(\frac{\mathbb{E}(W_2)-\tau_2}{1-\tau_2}\right)\left(1 - \frac{\mathbb{E}(W_2)-\tau_2}{1-\tau_2}\right)}{\left(\frac{Var(W_2)}{(1-\tau_2)^2}\right)} - 1 - \alpha_2 \tag{4.38}$$

Since problem size linearly affects expected total scores, number of selected projects and total budget variances, we present results for each problem size separately, in Tables 4.3 and 4.4. Before analyzing the results, we first explain formats (i.e. column fields) of both tables: $\mathbb{E}(score)$ represent expected total score of selected projects in the portfolio, opt (%) is percentage of instances solved to optimum, gap (%) measures the optimality gap of not optimally solved instances in a given time frame (if any), cpu (sec.) measures mean and standard deviation of CPU time for optimally solved instances in a given time limit. We exclude CPU time of not optimally solved instances in CPU time calculations to prevent dominance effect of 10800 seconds on CPU time of exactly solved instances. Number of selected projects is represented by ns, E(Γ) is expected number of canceled projects, E(Γ)/ns indicates fraction of expected number of cancellations in the portfolio. In fact, we can estimate expected number of successfully completed projects ($E(nscp)$) by using ns and E(Γ) as follows:

$$E(nscp) = ns - E(\Gamma) \tag{4.39}$$

Table 4.3: Comparison of factor effects for problem size of 1000

Factor	Levels	E(score)	opt (%)	gap (%)	cpu (sec.) mean	cpu (sec.) std. dev.	ns	E(Γ)	E(Γ)/ns (%)	E($nscp$)	TBV	BEB
p_i	U(0.01-0.1)	4141.7	93	0.01	663.1	1465.5	231.2	12.6	5.4	218.6	683.4	0.08
	U(0.01-0.2)	4017.6	76	0.02	1778.3	2348.4	236.5	24.1	10.2	212.3	942.1	0.07
q	0.4	4052.0	88	0.02	1121.2	1706.1	232.2	18.2	7.8	214.0	789.3	0.08
	0.5	4107.4	81	0.01	1320.3	2313.8	235.5	18.5	7.8	217.0	836.2	0.07
$\mathbb{E}(W_1)$;$\mathbb{E}(W_2)$	0.3 ; 0.8	4119.6	80	0.02	1295.7	2069.3	236.3	18.6	7.9	217.7	881.4	0.07
	0.35 ; 0.85	4039.7	89	0.01	1145.7	1997.8	231.4	18.1	7.8	213.3	744.1	0.08
Var(W_1)	0.01	4081.0	85	0.01	1153.6	1984.6	233.9	18.4	7.8	215.6	786.4	0.07
	0.05	4078.4	84	0.02	1287.8	2082.5	233.7	18.3	7.8	215.4	839.1	0.08
Var(W_2)	0.01	4087.3	91	0.01	817.4	1480.8	234.2	18.4	7.8	215.9	657.9	0.09
	0.05	4072.0	78	0.02	1624.1	2401.3	233.4	18.3	7.9	215.1	967.6	0.06
average		4079.7	84	0.01	1220.7	2035.2	233.8	18.4	7.8	215.5	812.8	0.07

Table 4.4: Comparison of factor effects for problem size of 2000

Factor	Levels	E(score)	opt (%)	gap (%)	cpu (sec.) mean	cpu (sec.) std. dev.	ns	E(Γ)	E(Γ)/ns (%)	E($nscp$)	TBV	BEB
p_i	U(0.01-0.1)	8408.6	91	0.01	414.5	1318.0	468.5	25.6	5.5	442.9	1352.2	0.06
	U(0.01-0.2)	8165.5	84	0.01	589.2	1722.7	479.9	49.3	10.3	430.6	1884.1	0.05
q	0.4	8230.5	88	0.01	735.1	1993.7	470.9	37.2	7.9	433.7	1570.8	0.06
	0.5	8343.7	88	0.01	268.5	797.7	477.6	37.7	7.9	439.8	1665.5	0.05
$\mathbb{E}(W_1)$;$\mathbb{E}(W_2)$	0.3 ; 0.8	8369.9	86	0.01	532.1	1480.5	479.1	37.9	7.9	441.2	1756.5	0.05
	0.35 ; 0.85	8204.2	89	0.01	471.5	1589.4	469.3	37.0	7.9	432.3	1479.7	0.06
Var(W_1)	0.01	8288.9	90	0.01	549.7	1641.2	474.4	37.5	7.9	436.9	1565.0	0.05
	0.05	8285.2	85	0.01	453.9	1421.9	474.0	37.4	7.9	436.6	1671.2	0.06
Var(W_2)	0.01	8298.0	95	0.01	523.6	1646.8	474.7	37.5	7.9	437.2	1311.4	0.06
	0.05	8276.2	80	0.01	480.1	1416.8	473.7	37.4	7.9	436.3	1924.8	0.05
average		8287.1	88	0.01	501.8	1536.2	474.2	37.5	7.90	436.7	1618.1	0.05

We also report $E(nscp)$ values in the tables. TBV calculates total budget variance of selected projects in the portfolio, BEB measures Berry-Esseen bound for quality of normal approximation. For each problem size factor, we have 2^5 factor combinations and for each factor combination we conduct 5 replications. Thus, each row of tables reports average of 2^5 x 5 = 160 replications in terms of measures of column fields.

Problem size factor conspicuously affects expected total score of projects and expected number of successfully completed projects as expected. Those values increase approximately 2 times as problem size is increased from 1000 to 2000. Problem size factor does not have a significant effect on number of instances solved to optimum; $opt(\%)$ value is 84% for problem size 1000 and it is 88% for problem size 2000. gap (%) values for not optimally solved instances are 0.01% for both problem size settings. However, mean and standard deviation of CPU time of optimally solved instances surprisingly decreases as problem size increases.

Cancellation probability (p_i) affects expected total score of projects in the portfolio as expected. For each problem size, as p_i increases $E(score)$ value decreases. However, effect of p_i on $E(score)$ becomes more prominent for problem size 2000 due to nearly two times increase in $E(nscp)$ values. Cancellation probability (p_i) has a significant effect on hardness of instances. Clearly, as p_i increases, $opt(\%)$ value decreases from 93% to 76% for problem size 1000 and it decreases from 91% to 84% for problem size 2000. However, $gap(\%)$ values of not optimally solved instances are too small (i.e. maximum 0.02%). As p_i increases, mean and standard deviation of CPU time of optimally solved instances increase more than two times for problem size 1000 and they increase moderately for problem size 2000. Note that standard deviation of CPU time is very high and this shows that CPU time from one instance to another can dramatically change. Although ns value increases as p_i increases, $E(nscp)$ value decreases. This is due to fact that $E(\Gamma)$ and $E(\Gamma)/ns$ values increase compatibly as p_i increases. Finally, as p_i increases TBV value increases as expected and BEB value decreases 0.01 point. Hence, as p_i increases quality of normal approximation increases by a probability level of 0.01 for both problem sizes.

As q (i.e. probability of successful completion of projects with underutilized budget) increases from 0.4 to 0.5, expected total score of projects in the portfolio increases as expected; because in that case ns and $E(nscp)$ values increase due to chance of more available room in budget constraint. As q increases, $opt(\%)$ value decreases for problem size 1000; whereas it remains the same for problem size 2000. $gap(\%)$ values of not optimally solved instances are very small with a maximum level of 0.02 percent as in p_i factor case. As q increases, average and standard deviation of CPU time of optimally solved instances increase for problem size 1000 and conversely they decrease for problem size 2000. As q increases TBV value increases and BEB value decreases 0.01 point. Thus, convergence to normal approximation improves with a probability level of 0.01 in both problem size factor settings.

Effect of $\mathbb{E}(W_1)$ and $\mathbb{E}(W_2)$ shows significance of modeling budget underutilization. Recall that $\mathbb{E}(W_1)$ is mean budget spending ratio of a canceled project and $\mathbb{E}(W_2)$ is mean budget spending ratio of a successful project that has an underutilized budget. When $\mathbb{E}(W_1)$ and $\mathbb{E}(W_2)$ are decreased only 0.05 point, $E(score)$ and $E(nscp)$ increase clearly. For instance, $E(score)$ increases from 4039.7 to 4119.6 for problem size 1000 and it increases from 8204.2 to 8369.9 for problem size 2000. Similarly, $E(nscp)$ increases from 213.3 to 217.7 for problem size 1000 and it increases from 432.3 to 441.2 for problem size 2000. This implication shows the importance of modeling of real past data accurately for DMs dealing with large scale project portfolios.

Effect of $Var(W_1)$ is limited on $E(score)$ and $E(nscp)$ values. Recall the derivation of $Var(R_i)$ in Proposition 4.2.3. Underutilized budgets of canceled projects are characterized by parameters p_i, $E(W_1)$ and $Var(W_1)$. Since cancellation probability (p_i) of each project is generated according to uniform distribution in some range, variability in p_i values can smooth out effect of $Var(W_1)$. In the next subsection, we conduct an analysis to see if there exists a smooth out effect of p_i. Number of instances solved to optimum is slightly affected by $Var(W_1)$. As $Var(W_1)$ increases, $opt(\%)$ value decreases from 85% to 84% for problem size 1000, and it decreases from 90% to 85% for problem size 2000. As $Var(W_1)$ increases, mean and standard deviation of CPU time of optimally solved instances increase for problem size 1000 and in contrast they decrease for problem size 2000. When $Var(W_1)$ is increased from 0.01 to 0.05, TBV value increases and surprisingly BEB value increases 0.01 point for both problem sizes. Hence quality of normal approximation deteriorates with a probability level of 0.01 in both problem sizes.

Variance of W_2 has a mild effect on $E(score)$. When $Var(W_2)$ is increased from 0.01 to 0.05, expected score slightly decreases. It also has a prominent effect on number of instances solved to optimum. As $Var(W_2)$ increases, $opt(\%)$ value decreases from 91% to 78% for problem size 1000, and it decreases from 95% to 80% for problem size 2000. As $Var(W_2)$ increases, average and standard deviation of CPU time of optimally solved instances considerably increase for problem size 1000 and oppositely, they decrease for problem size 2000. As $Var(W_2)$ increases, TBV value increases and BEB value decreases as expected. BEB value decreases 0.03 points for problem size 1000 and it decreases 0.01 point for problem size 2000. Therefore, convergence to normal approximation improves more prominently for problem size 1000.

When we evaluate effects of factors, we observe effects of p_i, q, $\mathbb{E}(W_1)$ and $\mathbb{E}(W_2)$ are significant for $E(score)$. Therefore, those factors must be carefully determined. For instance, as we stated before, peer reviewers can comment on p_i value of each project and these comments can be integrated with past data to estimate p_i values. Variance of W_1, and W_2 have limited effect on $E(score)$. Besides, we observe that the proposed model solves practical size instances in reasonable amount of CPU time.

We observe that quality of normal approximation can change according to factor levels. In Figure 4.1, mean BEB value of 5 replications for each factor combination is presented. First 32 factor combinations correspond to instances with problem size 1000, remaining factor combinations from 33 to 64 correspond to instances with problem size 2000. We can clearly observe that factor levels have a significant effect on quality of normal approximation. BEB value is between 0.12 and 0.04 for problem size 1000 and it is between 0.08 and 0.03 for problem size 2000. Therefore, parameters of problem on hand have an impact on quality of convergence, that is probably wondered by DMs.

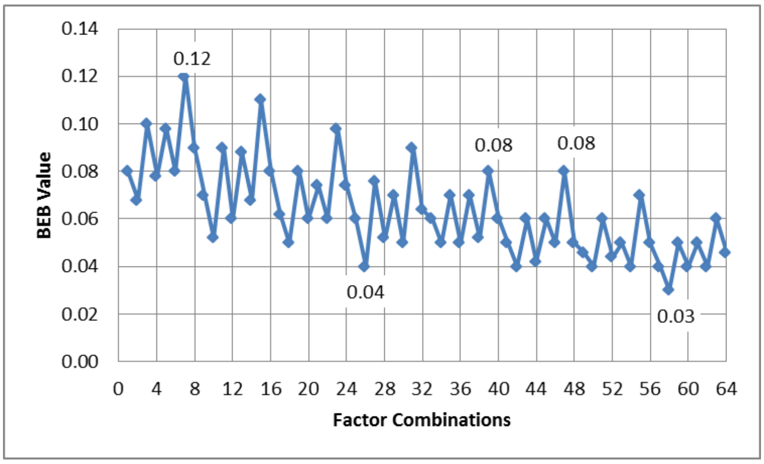

Figure 4.1: BEB Values of Factor Combinations

In our problem, we are motivated by TÜBİTAK 1001 program environment. Hence, we set problem size and bf values accordingly. However, some funding programs might not receive that many proposals. Therefore, we also analyze the Berry-Esseen bound of much smaller problem size setting. In practice, when the problem size gets smaller, bf values usually increase. For example, Karsu and Morton (2014) study R&D project selection problem with 150 project proposals and set bf value to $\%50$. Hence, we conduct a full factorial design with problem size=150 and $bf = \%50$ to examine how BEB values behave for smaller sets. Average BEB values of each factor are presented in Table 4.5. Average BEB values lies between 0.11 and 0.14. So, it turns out that BEB values of much smaller sets can be still acceptable for DMs.

Table 4.5: Berry-Esseen bound for problem size=150

Factor	Levels	BEB
p_i	U(0.01-0.1)	0.14
	U(0.01-0.2)	0.11
q	0.4	0.13
	0.5	0.12
$\mathbb{E}(W_1)$;$\mathbb{E}(W_2)$	0.3 ; 0.8	0.11
	0.35 ; 0.85	0.13
Var(W_1)	0.01	0.11
	0.05	0.14
Var(W_2)	0.01	0.14
	0.05	0.11
	average	0.12

4.5 Managerial Implications

In this section, we provide additional analyses to give practical implications to the DM. Firstly, in the next subsection, what Berry-Esseen bound offers to DMs are discussed.

4.5.1 How a Berry-Esseen bound can assist to decision making?

Berry-Esseen bound can assist DMs in two ways as presented in Table 4.6. Firstly, for a given θ value and a solution, DM is informed about the BEB value. Then, DM subtract BEB value from given θ value and obtain a new adjusted value, call it $\theta^I = \theta-$BEB. If this θ^I satisfy him/her, s/he admits this solution with its anticipated worst case risk level of 1-θ^I. Otherwise, DM prefer a new solution with a higher level of θ such that calculated θ^I satisfy him/her. For example, assume that a problem instance is solved with $\theta = 0.95$ and resulting calculated BEB value is 0.09, then $\theta^I = 0.86$. Thus, probability of not exceeding of funding budget is 0.86 in the worst case and assume that the DM is not satisfied with this probability level and requires a probability level of 0.9. Hence, the DM wants to see a new solution of the instance that will be solved with the probability level of 0.99.

Secondly, DM is first informed about properties of the problem at hand. Then according to average BEB value for given parameter settings in Tables 4.3 and 4.4, DM is informed for average BEB values of past solved instances. DM specify its desired θ value first and then average BEB value of past instances with same parameter settings added to θ as follows; $\theta^{II} = \theta$+BEB. Then problem is solved with this θ^{II} value and solutions are presented to DM. Finally, DM is satisfied this solution since his/her preferred confidence level is updated according to maximum convergence error. For instance, assume that we inform the DM about properties of the instance that will be solved and average BEB value of similar instances. Assume that average BEB value=0.08 and DM determines desired probability level as 0.9 in the worst case, then $\theta^{II} = 0.98$. Thus, the instance will be solved with probability level of 0.98. Thereby, convergence error of normal approximation is mitigated for the DM.

Table 4.6: How Berry-Esseen bound can assist to DMs

	Ways	
	I	II
Given confidence level	θ	θ
BEB value	Solved model	Similar past instances
Adjusted confidence level	$\theta^I = \theta-$BEB	$\theta^{II} = \theta+$BEB
Action	Present θ^I to DM	Solve model with θ^{II}
Outcomes	DM satisfies or increase θ until DM is satisfied with θ^I and solve model with new θ^I	DM satisfies

In the next subsection, we investigate if there is any smooth out effect of p_i values.

4.5.2 Do cancellation probabilities smooth out of effect of Var(W_1)?

Instead of generating cancellation probabilities from a range, we set p_i values to 0.1, 0.2, and 0.3 to see whether they smooth out effect of Var(W_1). In this case, for each p_i and Var(W_1) value, we conduct a 2^2 factorial design with factors q and Var(W_2). We take problem size=2000, $\mathbb{E}(W_1) = 0.3$, $\mathbb{E}(W_2) = 0.8$. For each factor combination, we conduct 5 replications. In Table 4.7, each row (for each p_i and Var(W_1) value) represents mean of 20 replications (2^2 x 5 replications). We observe as constant p_i value increases, effect of Var(W_1) on expected score tends to slightly increase. We can conclude that some effect of Var(W_1) on expected score is smoothed out by different cancellation probabilities. On the other hand, when we scrutinize effects of Var(W_1), and Var(W_2) on expected score, we can observe that their effects are smaller than that of p_i, q, $\mathbb{E}(W_1)$ and $\mathbb{E}(W_2)$. This result can be explained by effect of truncation. Since our random variables are truncated, maximum value that a variance can take strictly decreases. For instance, in our problem, variance can be maximum 0.05. Hence, accurate truncation with a good mean estimate mitigates risk of imperfect information of variances. It is also interesting to note that when p_i values are the same, we are dealing with independent identically distributed (i.i.d) random variables and BEB values sharply decreases. This is an expected result of CLT since convergence to normal distribution improves with i.i.d random variables. Note that for the i.i.d case, the best estimate of C in Theorem 4.3.1 drops to 0.4748 due to Shevtsova (2011) and BEB values in Table 4.7 are calculated accordingly.

Table 4.7: Effect of constant p_i values

p_i	Var(W_1)	E(score)	ns	E(Γ)	E(Γ)/ns (%)	E($nscp$)	TBV	BEB
0.1	0.01	8271.9	484.8	48.5	10	436.3	2000.5	0.03
	0.05	8267.1	484.8	48.5	10	436.3	2140.1	0.04
0.2	0.01	7778.0	514.2	102.8	20	411.3	3134.3	0.02
	0.05	7770.8	513.6	102.7	20	410.8	3448.4	0.03
0.3	0.01	7244.9	548.5	164.6	30	384.0	4122.9	0.01
	0.05	7234.8	547.7	164.3	30	383.4	4652.4	0.02

In the next subsection, effect of confidence level (θ) is discussed.

4.5.3 What if we use different θ values?

DMs may wonder how various probability levels (θ) could affect obtained solutions. Therefore, an efficient frontier of any measure of interest can be presented to DM by varying confidence levels numerically. For example, effect of θ on expected score of one problem instance is presented in Figure 4.2. We can observe as θ increases expected score decreases as expected. Increasing θ from 0.90 to 0.99 results in considerably mitigating budget risk thereby constructing more conservative portfolios.

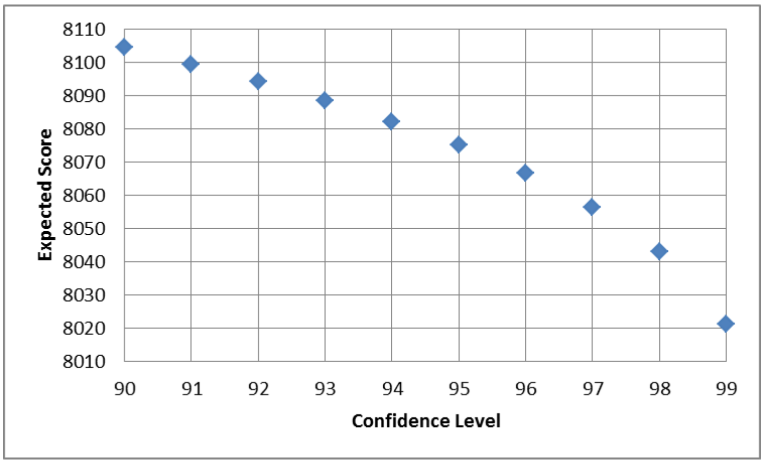

Figure 4.2: Effect of θ on expected score

In the next subsection, effect of distribution modeling is discussed.

4.5.4 What if we don't know distribution of W_1, and W_2?

In stochastic programming, identifying uncertain parameters accurately is a costly task. For our problem, even identification of p_i and q values are demanding. Moreover, we determine distribution of budget spending for different underutilization cases. In this subsection, we assume distribution of W_1, W_2 and their parameters are unknown and only mean values are known. Expected values of W_1, and W_2 could be obtained by analyzing average spending of past canceled and successfully completed projects. Budget spending can still be modeled by using mean values at expense of some error in true objective value of model SP-1. In this case, random variable R_i

can be modified to $\bar{R}_i$ as follows;

$$\bar{R}_i = \begin{cases} \mathbb{E}(W_1) & \text{w.p. } p_i \\ \mathbb{E}(W_2) & \text{w.p. } q \\ 1 & \text{w.p. } 1 - p_i - q \end{cases} \tag{4.40}$$

Expectation, variance and third moment of $\bar{R}_i$ are calculated as follows:

$$\mathbb{E}(\bar{R}_i) = p_i\mathbb{E}(W_1) + q\mathbb{E}(W_2) + (1 - p_i - q) \tag{4.41}$$

$$Var(\bar{R}_i) = \mathbb{E}(\bar{R}_i^{\,2}) - [\mathbb{E}(\bar{R}_i)]^2 = p_i(\mathbb{E}(W_1))^2 + q(\mathbb{E}(W_2))^2 + (1 - p_i - q) - [\mathbb{E}(\bar{R}_i)]^2 \tag{4.42}$$

$$\mathbb{E}(\bar{R}_i^{\,3}) = p_i(\mathbb{E}(W_1))^3 + q(\mathbb{E}(W_2))^3 + (1 - p_i - q) \tag{4.43}$$

In Proposition 4.5.1, we show that modeling only with mean values gives an upper bound on optimal objective value of model SP-1.

Proposition 4.5.1. *Let z_1^* be an optimal objective value of model SP-1, and z_2^* be an optimal objective value of a modified model when mean and variance of $\bar{R}_i$ are used in SP-1. Then $z_2^* \geq z_1^*$ holds.*

Proof. See Appendix A.6. □

Let SP-2 be a modified model in which mean and variance of $\bar{R}_i$ are used in model SP-1. We solve model SP-2 and compare it with model SP-1 to assess effect of beta distribution modeling. We use problem size=2000, $\mathbb{E}(W_1) = 0.3$ and $\mathbb{E}(W_2) = 0.8$ in below experiments. We calculate following difference measures for comparison:

$$\Delta E(score) = E(score) \text{ of the model SP-1} - E(score) \text{ of the model SP-2} \tag{4.44}$$

$$\Delta E(nscp) = \Delta E(nscp) \text{ of the model SP-1} - \Delta E(nscp) \text{ of the model SP-2} \tag{4.45}$$

$$\Delta TBV = TBV \text{ of the model SP-1} - TBV \text{ of the model SP-2} \tag{4.46}$$

$$\Delta BEB = BEB \text{ of the model SP-1} - BEB \text{ of the model SP-2} \tag{4.47}$$

Note that we use moments of $\bar{R}_i$ in inequality (4.34) for BEB value calculation of model SP-2. We first analyze effect of formulating only cancellations with beta distribution in Table 4.8 (e.g. $p_i > 0\ \forall i$ and q=0). Each row represent average difference of 5 instances. As we show in Proposition 4.5.1, expected score of model SP-2 is greater than that of model SP-1. Besides, ΔE(score) values are small. Hence, objective value of model SP-2 gives a good upper bound on objective value of SP-1. ΔE(nscp) values are close to zero. TBV value of model SP-2 is smaller than that of model SP-1 due to using $\bar{R}_i$ in model SP-2. BEB value of SP-2 is smaller than BEB value of model SP-1, meaning that when $\bar{R}_i$ is used for modeling, BEB value of true model SP-1 is calculated erroneously with a probability level of at maximum 0.02.

Table 4.8: Effect of beta distribution for q=0

p_i	Var(W_1)	ΔE(score)	ΔE(nscp)	ΔTBV	ΔBEB
U(0.01-0.1)	0.01	-0.5	0.2	17.5	0.01
	0.05	-3.9	-0.2	82.5	0.02
U(0.01-0.2)	0.01	-1.1	-0.3	27.9	0.01
	0.05	-5.5	-0.1	150.7	0.02

In Table 4.9, we examine effect of beta distribution modeling by assuming there is no cancellation in model (e.g. p_i=0 $\forall\ i$ and $q > 0$). Similarly, expected score of model SP-2 is greater than that of model SP-1. Besides, difference between objective values begins to increase. Magnitude of ΔTBV, and ΔBEB values are much higher than that of Table 4.8. Thus, effect of beta distribution modeling becomes more prominent when there is no cancellation.

Table 4.9: Effect of beta distribution for p_i=0 $\forall i$

q	Var(W_2)	ΔE(score)	Δns	ΔTBV	ΔBEB
0.4	0.01	-9.2	-0.8	128.6	0.03
	0.05	-38.4	-2.0	637.4	0.04
0.5	0.01	-12.4	-0.4	164.7	0.03
	0.05	-47.0	-2.4	819.5	0.03

We next examine mixed effect of beta distributions in the full model (e.g. $p_i > 0\ \forall i$ and q>0). We conduct a 2^4 factorial design with factors p_i, q, $Var(W_1)$ and $Var(W_2)$. We use 5 replications (i.e.2^4x$5 = 80$ runs) and calculate differences between model SP-1 and SP-2. Effects of parameters p_i and $Var(W_1)$ on Δ values are presented in

Table 4.10. Each row represents average of 20 replications. As $Var(W_1)$ increases, magnitude of ΔE(score) and ΔE(nscp) values slightly increase and ΔTBV value increases. Effects of parameters q and $Var(W_2)$ are shown in Table 4.11. As both q and $Var(W_2)$ increase, magnitude of ΔE(score), ΔE(nscp) and ΔTBV values increase. It is interesting to observe BEB values of both models SP-1 and SP-2 are very close to each other in Tables 4.10 and 4.11. When we assess effects of all four factors, effects of q and $Var(W_2)$ are more prominent. If we scrutinize all results in Tables 4.8, 4.9, 4.10 and 4.11, we make following observation:

Observation 4.5.1. *If budget underutilization occurs due to only cancellations, model SP-2 gives a good approximation to model SP-1. However, if some considerable fraction of projects successfully complete with underutilized budget, quality of approximation of model SP-2 decreases.*

Table 4.10: Mixed effect of beta distributions for p_i and $Var(W_1)$

p_i	Var(W_1)	ΔE(score)	ΔE(nscp)	ΔTBV	ΔBEB
U(0.01-0.1)	0.01	-18.5	-0.8	477.5	-0.005
	0.05	-21.4	-0.9	550.7	0.001
U(0.01-0.2)	0.01	-16.3	-1.1	516.0	-0.004
	0.05	-20.8	-1.5	659.2	0.006

Table 4.11: Mixed effect of beta distributions for q and $Var(W_2)$

q	Var(W_2)	ΔE(score)	ΔE(nscp)	ΔTBV	ΔBEB
0.4	0.01	-7.6	-0.7	215.8	0.006
	0.05	-26.5	-1.4	762.2	-0.007
0.5	0.01	-9.4	-0.6	260.4	0.005
	0.05	-33.5	-1.7	965.0	-0.007

4.5.5 What does modeling of uncertainty offer to DMs?

In this section, we quantify value of proposed model (PM) by assuming it captures uncertainty behavior of practical life. If aforementioned PM is not applied, a well known knapsack model (KM) will be solved. Hence, for problem size 2000, we solve our 5 instances as a KM with s_i and b_i parameters. Then, we calculate expected score, expected spent budget (for θ=0.95) and expected number of successfully completed projects of solutions of KM by using probability information of PM. After, we compare these values with those of PM for different stochastic parameters by using following formula and report improvements (%) in Table 4.12. Since we have 5 factors and 5 instances, we have calculate 2^5x5=160 improvement values. Each row in Table 4.12 reports information of 80 comparisons.

$$\text{Improvement } (\%) = 100 \text{ x } \frac{\text{Expected Value of PM-Expected Value of KM}}{\text{Expected Value of KM}} \tag{4.48}$$

Table 4.12: Value of Modeling Uncertainty

		Improvement Values (%)								
		E(score)			Budget Utilization			E($nscp$)		
Factor	Levels	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
p_i	U(0.01-0.1)	5.8	7.7	10.1	8.4	11.1	14.2	6.5	8.8	11.6
	U(0.01-0.2)	8.2	10.4	13.2	11.9	15.0	18.6	9.4	12.2	15.6
q	0.4	5.8	8.3	11.3	8.4	12.0	15.8	6.5	9.7	13.8
	0.5	6.9	9.8	13.2	10.1	14.2	18.6	7.7	11.3	15.6
$\mathbb{E}(W_1)$;$\mathbb{E}(W_2)$	0.3 ; 0.8	7.5	10.1	13.2	11.0	14.7	18.6	8.4	11.7	15.6
	0.35 ; 0.85	5.8	7.9	10.5	8.4	11.4	14.7	6.5	9.3	12.8
Var(W_1)	0.01	5.8	9.0	13.2	8.4	13.1	18.6	6.5	10.5	15.6
	0.05	5.8	9.0	13.1	8.4	13.0	18.5	6.5	10.5	15.6
Var(W_2)	0.01	6.0	9.2	13.2	8.8	13.3	18.6	6.7	10.6	15.6
	0.05	5.8	8.9	12.9	8.4	12.8	18.2	6.5	10.4	15.6

Factors p_i and q have significant effect on improvement values. Therefore, value of PM increases as those probability parameters increase. Factors $E(W_1)$, and $E(W_2)$ have also important effect on improvement values. It is interesting to observe that very small changes in means can result in notable impacts on improvement values. Small returning budgets of many projects can constitute large amounts that make additional room for budget constraint. As expected, $Var(W_1)$ has no significant effect on improvement values since p_i values smooth out its effect. Factor $Var(W_2)$ has small effects on improvement values. As it increases all improvement measures slightly decreases (except max value of $E(nscp)$ which remains the same). In overall, adopting the PM yields notable increases on objective value, budget utilization and expected number of successfully completed projects. For instance, it provides minimum 5.8%, average 9%, and maximum 13.2% improvement in expected score.

Besides, budget utilization improvement is minimum 8.4%, average 13%, and maximum 18.6%. Moreover, improvement of expected number of successfully completed projects is minimum 6.5%, average 10.5%, and maximum 15.6%.

4.6 Conclusions

In this chapter, we consider a one-stage public RDPPS problem with project expenditure uncertainty. Analysis of empirical data leads to modeling of budget expenditure uncertainty with a mixture distribution (i.e. truncated inflated beta distribution). We develop a CCSP model for public RDPPS. However, it is computationally intractable since derivation of quantile function for total random budget expenditure is challenging. Therefore, after showing sum of general truncated random variables converges to normal distribution (i.e. CLT applies to total random budget spending of our problem), we use Gaussian (normal) approximation scheme and apply second order cone programming reformulation to exactly solve approximated model. DMs wonder convergence quality of normal approximation for financial decision making problems. Therefore, we derive maximum probability error of normal approximation for any solution by applying Berry-Esseen theorem to our problem. We also propose alternative ways to incorporate Berry-Esseen bound to decision making process. Note that Berry-Esseen bound can be calculated for any distribution which has a finite mean, variance and third moment. Therefore, our proposed approach can be effectively adopted in CCSP formulation of any problem for which quantile function of a similar type chance constraint cannot be derived.

We also conduct further analyzes for our problem to give managerial insights to the DM. For instance, we analyze effect of probability level of CCSP on expected score and identify that increasing probability level leads to more conservative portfolios. We also analyze the case of unknown distribution of budget underspending of canceled and successfully completed projects. For this case, we give an alternative distribution modeling and compare it with proposed distribution modeling on different problem settings. We observe that alternative modeling gives a good approximation to proposed modeling if there is only cancellations in the problem. However, approximation quality diminishes in case of some considerable fraction of successfully completed projects underspend their budgets. Finally, we quantify the value of proposed model by comparing it with the standard project selection model. Our proposed model delivers an increase between 8.4% and 18.6% in utilization of funding budget and an increase between 5.8% and 13.2% in total expected score of selected projects in the portfolio.

In the next chapter, we study sectoral balancing problem in a public RDPPS model by incorporating impact assessment results.

CHAPTER 5

IMPACT ASSESSMENT BASED SECTORAL BALANCING IN PUBLIC RDPPS PROBLEM

In this chapter, we study the sectoral budget balancing problem in the public RDPPS model. We consider an R&D funding program to which many project proposals from various sectors apply. We assume that sectoral impacts of the funding program are known. The DM wants to distribute limited funding budget among sectors in the light of sectoral impacts.

In Section 5.1, we state the problem setting. In Section 5.1.1, we develop a two-stage public RDPPS model. In the first stage, the DM deals with sectoral budget decisions to maximize the impact of the funding budget while ensuring relative budget balancing among the sectors. In the second stage, the DM wants to maximize the total score of supported projects under allocated sectoral budgets. We propose a nonlinear social welfare objective function in the first stage. In Section 5.1.2, we prove that nonlinearity in the objective function can be expressed by second-order conic inequalities. In Section 5.1.3, we also develop an informed decision making approach to show the effects of sectoral allocation on various indicators such as the total impact of funding budget, total score of selected projects, sectoral scores and budgets, sectoral number of supported projects, and sectoral success rates. To the best of our knowledge, incorporation of sectoral impacts into a public project selection model and in-depth sectoral analysis of projects have not been studied before. In Section 5.2, we apply the proposed approach on an example problem. We generate the example problem by using sectoral public data of TÜBİTAK 1001 program. For the example problem, we also obtain proxy sectoral impact assessment values from the literature. In

Section 5.2.5, we show the value of the proposed approach by comparing it with alternative policy options to give insights to the DM. In Section 5.3, we present the concluding remarks on the proposed approach.

5.1 Proposed Model and Solution Approach

In this problem setting, we have an R&D funding program. Many projects from various sectors apply to the funding program. Budget (b_{ij}) and score (s_{ij}) of project i in sector j are clarified in panel review meetings. We assume that the impact of the R&D funding program over different sectors are quantified according to past completed projects by some quantitative and/or qualitative method. Let e_j and B_j be the impact assessment value and allocated budget (decision variable) of sector j, respectively. Selection of project i in sector j is indicated by a binary decision variable x_{ij}. We develop a two stage model. In the first stage, the DM wants to maximize the impact of the funding budget according to the impact assessment values of sectors by considering sectoral balancing decisions (i.e. deciding on B_j values). In the second stage, the DM wants to maximize the total score of selected projects (i.e. deciding on x_{ij} values) in the portfolio under allocated sectoral budgets. Next, we discuss our two stage model.

5.1.1 Two Stage Solution Procedure

In the first stage, formulation of the objective function is critical. The DM wants to maximize the impact of the funding budget, whereas s/he also wants to ensure sectoral balancing of the funding budget in a fair way. The DM wants to relatively balance sectoral budget allocations according to the impact assessment values. In previous studies, parameterized social welfare functions are applied for general balancing concerns, and problem specific players (i.e. sectors, etc.) have not been compared according to some measure. In our problem, we assume that we have the impact assessment values and the DM wants to see a sectoral distribution compatible with those impact assessment values. In addition, the DMs usually perceive sectoral impact assessment findings as a result of allocated one unit budget to sectors. Therefore, we formulate the objective function of the first stage model as maximization of "$\sum_j e_j B_j$". This objective function maximizes the total impact of the funding budget by considering sectoral impact values (e_j) and sectoral allocated budgets (B_j). However, this objective function is in the form of utilitarian objective function, which we discuss in Section 2.2. Therefore, this objective function may fail to properly balance the funding budget among sectors because the utilitarian objective function may dominantly favor the best sector at the expense of other sectors. Resulting sectoral budget allocations may not be accepted by the DM due to relative balancing (i.e. impact value based budget allocations to sectors in a fair way) concerns. Therefore, we adopt the parameterized social welfare function in the objective function and develop the following model:

$$\text{Stage I. } (O_1)\ \max\ \sum_j \frac{(e_j B_j)^{1-\alpha}}{(1-\alpha)} \qquad \text{(Total } \alpha\text{-balanced Impact of the Budget)} \tag{5.1}$$

$$\text{s.t.}\ \sum_j B_j \leq B \qquad \text{(Sectoral Budget Allocation)} \tag{5.2}$$

$$\sum_i b_{ij} \geq\ B_j \geq 0\ \ \forall\, j \tag{5.3}$$

The objective function O_1 in (5.1) is formulated according to the parameterized social welfare function in (2.1). Note that we denote the u_j as utility of sector j, and $u_j = e_j B_j$. The objective function maximizes the total impact of funding budget while ensuring balancing of sectoral budgets with the parameter α. Constraint (5.2) ensures that total allocated budget to sectors is limited by the funding budget B. Constraint set (5.3) ensures that sectoral budgets cannot be negative and the budget of a sector cannot be greater than sum of the budgets of the applying projects in that sector. We don't put any range restriction on the impact assessment values (i.e. e_j values). For example, as reviewed in Section 2.2, Lee et al. (2009), and Wang et al. (2013) obtain impact assessment values between 0 and 1. Besides, Jaffe (1989) obtains propensity to patent (i.e. patent elasticity), Mansfield (1998) measures percentage of new products and processes, and Cohen et al. (2002) consider categorical data indicating sectoral importance and transform it into percentage units. Therefore, all of those measures could be used as e_j values in the first stage model.

After deciding on B_j values in the first stage, we incorporate them into the second stage model as sectoral budget limits. The second stage model is formulated as follows:

$$\text{Stage II. } (O_2) \ \max \sum_i \sum_j s_{ij} x_{ij} \qquad \text{(Total Score of R\&D Project Portfolio)} \tag{5.4}$$

$$\text{s.t.} \sum_i b_{ij} x_{ij} \leq B_j \ \ \forall\, j \qquad \text{(Sectoral Budget Limit)} \tag{5.5}$$

$$x_{ij} \in \{0,1\} \ \forall\, i,j \tag{5.6}$$

The objective function O_2 in (5.4) maximizes total score of selected projects in all sectors. Constraint set (5.5) ensures that total budget of selected projects in sector j cannot exceed allocated sectoral budget B_j. Constraint set (5.6) indicates the selection decision of project i in sector j. In our two stage model, deciding on objective of the first stage model and its associated sectoral allocations is equivalent to selection of single α parameter. In the next section, we first discuss how α parameter of our problem could be set. After that, we solve an example problem and give insights to the DM about implications of different values of α.

Note that the objective function (O_1) of the first stage model is nonlinear for $\alpha > 0$. We obtain the tractable reformulation of the first stage model in Section 5.1.2. In particular, we prove that this nonlinearity can be expressed by conic quadratic inequalities for $0 < \alpha < 1$, hence the first stage model is easily solvable by commercial solvers such as CPLEX. We also provide a reformulation example for $\alpha = 0.4$ in Example 5.1.1. Therefore, we can obtain a tractable reformulation for any α value between 0 and 1.

5.1.2 Second Order Cone Programming (SOCP) Reformulation of the Nonlinear Objective

In the objective function of the first stage model given in (5.1), for a positive α, we have the following nonlinear term for each sector j :

$$(e_j B_j)^{1-\alpha} \tag{5.7}$$

We first introduce an auxiliary variable t_j and revise the objective function in (5.1) as follows:

$$\max \ \frac{1}{1-\alpha} \sum_j t_j \tag{5.8}$$

and then we add the following constraint to the model.

$$t_j \leq (e_j B_j)^{1-\alpha} \ \ \forall j \tag{5.9}$$

We have a maximization objective; hence, the resulting reformulation with linear objectives and non-linear constraint set is equivalent to original model.

Next, in the following Proposition 5.1.1, we show that resulting reformulation can be expressed by conic quadratic inequalities. We drop index j of t_j, e_j, and B_j for the sake of simplicity.

Proposition 5.1.1. *Inequality* (5.9)

$$t \leq (eB)^{1-\alpha}$$

with $e \geq 0, B \geq 0, t \geq 0$ *is SOCP representable for* $\alpha = \frac{k_1}{k_2}$ *where* k_1 *and* k_2 *are integers such that* $0 < k_1 < k_2$.

Proof. Inequality (5.9) can be written as $t \leq (eB)^{\frac{k_2-k_1}{k_2}} \Rightarrow t^{k_2} \leq (eB)^{k_2-k_1}$. Then, there exists a natural number $k_3 \geq 0$ such that

$t^{2^k} \leq (eB)^{k_2-k_1} \times t^{k_3}$ where $k = log_2(k_2 + k_3)$, which can be further written as

$t^{2^k} \leq (eB)^{k_2-k_1} \times t^{k_3} \times 1^{k_1}$.

As discussed in Alizadeh and Goldfarb (2003), an inequality of the following form

$$t^{2^k} \leq s_1 s_2 .....s_{2^k} \tag{5.10}$$

for $t \geq 0$, $s_1 \geq 0$, ...,$s_{2^k} \geq 0$ can be expressed by at most 2^{k-1} inequalities of the form

$$w_m^2 \leq u_m v_m, \;\; w_m, u_m, v_m \geq 0, \; 1 \leq m \leq 2^{k-1} \tag{5.11}$$

Moreover, each inequality $w_m^2 \leq u_m v_m$ can be expressed by following conic quadratic inequality:

$$\|2w_m, u_m - v_m\| \leq u_m + v_m \tag{5.12}$$

□

Example 5.1.1. *Consider a problem setting with $\alpha = 0.4 = \frac{2}{5}$, so $k_1 = 2$, and $k_2 = 5$.*

Then we have $t \leq (eB)^{\frac{3}{5}} \Rightarrow t^5 \leq (eB)^3$, which is equivalent to:

$t^8 \leq (eB)^3 \times t^3$, which can be further expressed as:

$t^{2^3} \leq (eB)^3 \times t^3 \times 1^2$. Then we can reformulate this inequality with the form in (5.11) *with following inequalities:*

$$w_1^2 \leq eB \times t \tag{5.13}$$

$$w_2^2 \leq w_1 \times 1 \tag{5.14}$$

$$t^2 \leq w_1 \times w_2 \tag{5.15}$$

where $w_1, w_2 \geq 0$*.*

Moreover, these inequalities can be expressed by following conic quadratic inequalities:

$$\|2w_1, eB - t\| \leq eB + t \tag{5.16}$$

$$\|2w_2, w_1 - 1\| \leq w_1 + 1 \tag{5.17}$$

$$\|2t, w_1 - w_2\| \leq w_1 + w_2 \tag{5.18}$$

We reformulate all remaining α levels used in the example problem accordingly.

5.1.3 Informed Decision Making Approach for the Selection of α

Determining the α value depends on problem environment and concerns of the DM. Therefore, we conduct preliminary experiments with the first stage model. Our preliminary study suggests that allocated sectoral budgets become almost equal for $\alpha \geq 0.9$. Clearly, this is not desirable for the DM because s/he wants to relatively balance funding budget according to the sectoral impact values. Therefore appropriate operational α value for our problem setting is at most 0.9. But, how can the DM decide on a suitable α value between 0 and 0.9? In real-world resource allocation problems, the DMs are usually informed with related problem properties to gain managerial implications and select the desired portfolio. To adopt an appropriate α level for our problem setting, the DM is interested in how following problem properties behave for various α values:

(1) Degradation of total impact of funding budget (stage I) objective (i.e. price of sectoral balancing) and associated sectoral budget distribution: Bertsimas et al. (2011) have recently introduced the price of fairness (balancing) concept, which measures relative loss in the linear objective (i.e. utilitarian objective)

value under positive α value, compared to $\alpha = 0$. For our problem setting, price of balancing for any α (i.e. POB(α)) can be expressed as;

$$\text{POB}(\alpha) = \frac{O_1(\alpha_0) - O_1(\alpha)}{O_1(\alpha_0)} \tag{5.19}$$

where $O_1(\alpha_0)$ is the optimum objective value of O_1 when we solve the model for $\alpha = 0$ and $O_1(\alpha)$ is the corresponding linear objective value (i.e. $\sum_j e_j B_j$) when we solve the model for positive α. As we increase α, corresponding linear objective value decreases because of sectoral balancing. However, distribution of total impact objective (O_1) and funding budget (B) among sectors changes and relative balancing is achieved. Hence, as we increase α, the DM will evaluate POB(α) value and resulting distribution of O_1 and B. Thereby, the DM will have an idea about how α affects sectoral budget allocations and total impact of funding budget.

Empirical derivation of POB(α) is given in (5.19). On the other side, Bertsimas et al. (2011) provide theoretical bound on POB(α) for the case $\alpha = 1$ and $\alpha \to \infty$. Moreover, Bertsimas et al. (2012) underline that practical setting of α can be demanding for any value of $\alpha > 0$. Hence, to facilitate selection of problem specific α, they generalize the characterization of theoretical bound on POB(α) for any $\alpha > 0$ as a function of number of players and ratio of maximum utility over minimum utility. Derivation of a bound on POB(α) needs additional technical effort which is presented in Appendix B.1. These bounds reflect strategic long term insights to DM for the selection of α parameter. For instance, if the DM is not willing to exceed the price of balancing more than 45% in any problem instance in the worst case, DM can select corresponding α value accordingly. To decide on the α value, Bertsimas et al. (2011) benefit from POB(α) and compare it with equal balancing case (i.e. $\alpha \to \infty$) in a different problem setting. However, in our problem setting, we know $\alpha \leq 0.9$ and we have additional problem properties such as second stage objective function distribution among sectors, average score and budget of the supported projects in each sector, the number of supported projects and success rate in each sector. The DM may arise different concerns and preferences about those problem properties. Thus, s/he can benefit from analysis of those properties to decide on the α value. Next, we discuss how to examine them.

(2) Effect of sectoral balancing on total score (stage II) objective: The DM wants to see how total score objective could change, as α increases. In addition, the DM wants to know how total score distribution among sectors could change, as we increase α value. There can be many interesting patterns for different α values. For example, the DM may wonder whether there is any sector that has low impact value but dominates total sectoral score objective, as we increase α value.

(3) The average score and the budget of selected projects in each sector: The DM may wonder how the average score and the budget of supported projects in each sector may behave, as we increase the α value. There might be some undesirable cases under different α values. For instance, the DM may want to learn whether there is any low impact valued sector whose average score decreases as well as whose average budget increases notably, as we increase the α value. Because, this kind of pattern implies that probability of selecting low scored projects could increase in a low impact valued sector as we increase the α value, which may not be desirable for the DM.

(4) The success rate and the number of selected projects in each sector: In public R&D funding organizations, one of the most used indicators for assessing sectoral competition among proposals is the success rate of each sector. It is the ratio of the number of selected (i.e. supported) projects over the number of submitted proposals in each sector. For example, the DM may want to require an R&D project portfolio in which low impact valued sectors must have lower success rates; whereas, high impact valued sectors must have higher success rates. Thus, the DM wants to learn the success rate of each sector under various α setting. Besides, the number of submitted project proposals could notably change in each sector and the DM may also wonder about the number of supported projects in each sector. Therefore, evaluating success rate with the number of supported projects is more beneficial. For instance, the DM may want to learn whether there is any low impact valued sector whose success rate and number of projects increase considerably, as we increase α. Because, the

DM may not want to support too many projects in a low impact valued sector.

We addressed problem properties (i.e. sectoral indicators) that can assist to the DM while selecting appropriate α value. In the next section, we conduct in-depth analysis of those steps by solving a practical-size example problem and give managerial insights for deciding sectoral allocations (i.e. selection of α).

5.2 An Example Problem

We are motivated by TÜBİTAK 1001 scientific and technological research projects funding program. However, impact assessment of this funding program has not yet been conducted by the Department of Impact Assessment of MSTI. In their web site, there isn't any published report about assessment findings of funding programs. Therefore, we adopt proxy impact assessment values from the literature. In Table 5.1, we obtain the average impact of seven scientific disciplines of Wang et al. (2013) by processing their data across eighteen funding programs and applying their vague set methodology. We use their preferred aggregation parameter $\lambda = 0.5$ for their vague set approach. However, scientific disciplines of TÜBİTAK 1001 program are slightly different. Thus, we also try to match them as in Table 5.1.

Table 5.1: Proxy impact assessment values for the example problem

TÜBİTAK 1001 program Scientific disciplines (j)	Wang et al. (2013) Scientific Disciplines	impact assessment values (e_j)
Environment, Atmosphere, Earth and Marine Sciences	Earth sciences (ES)	0.395
Electrical, Electronics and Informatics	Information sciences (IS)	0.354
Engineering	Engineering and material sciences (EMS)	0.391
Health Sciences	Life science (LS)	0.333
Social Sciences and Humanities	Management sciences (MS)	0.232
Basic Sciences	Mathematical and physical sciences (MPS)	0.331
Agriculture, Forestry and Veterinary	Chemical sciences (CS)	0.417

For each sector, we calculate the number of project proposals and the budget range of project proposals in Table 5.2. We use real descriptive statistics of TÜBİTAK 1001 program given in Appendix B.2. As shown in Table B.1 of Appendix B.2, a total of 3218 project proposals apply to two calls of TÜBİTAK 1001 program in 2012 and the average number of project proposals per call is greater than 1600 proposals. We also know that the number of project proposals increases every year. Therefore, we set the number of project proposals to 2000. We determine the number of project proposals in each sector by using sectoral percentage distribution given in Table B.1 of Appendix B.2. For instance, the sectoral percentage of environment, atmosphere, earth and marine sciences is 7.46 %, then we obtain 149 proposals in this sector in Table 5.2 (2000x0.0746 and rounding it to nearest integer). We randomly generate the budget (b_{ij}) of project i in sector j according to a uniform distribution between $[U_{min}, U_{max}]$. U_{max} cannot be grater than 36 (in 10,000 monetary units) because this is the current maximum project budget upper bound in TÜBİTAK 1001 program. We assume that U_{min} cannot be lower than 7 (in 10,000 monetary units). Note that sectoral U_{min} and U_{max} values are set according to average budget column of Table B.1 in Appendix B.2. Apparently, the average budget of each sector is different. We randomly generate the score (s_{ij}) of project i in sector j according to uniform distribution between [10,25]. We determine the total available funding budget (B) by setting $B = (\sum_{ij} b_{ij})$ x bf, where the budget fraction (bf) value is set to 0.17 by using the success rate given in Appendix B.2 as a proxy. We also assign a code to each sector in Table 5.2 for the sake of brevity.

Table 5.2: The number of proposals and the budget range in each sector

Sector code	TÜBİTAK 1001 program Scientific disciplines (j)	Number of Proposals	Project Budget U_{min}	U_{max}
s1	Environment, Atmosphere, Earth and Marine Sciences	149	21	36
s2	Electrical, Electronics and Informatics	143	9	36
s3	Engineering	349	7	36
s4	Health Sciences	257	23	36
s5	Social Sciences and Humanities	305	7	22
s6	Basic Sciences	470	11	36
s7	Agriculture, Forestry and Veterinary	327	14	36
	Total	2000		

We generate our problem instance and solve it for $\alpha = 0, 0.1, ..., 0.9$. We use IBM ILOG CPLEX 12.6.2 with Concert Technology and C++ programming language for computational experiments. Then, we apply steps of our informed decision making approach given in Section 5.1.3 to assist to the DM for selection of α.

5.2.1 Degradation of total impact of funding budget and associated sectoral budget distribution

The distribution of the total impact of the funding budget (O_1) and the distribution of the funding budget (B) among sectors is given in Table 5.3. POB (α) values are also given at the bottom of table. Note that sector s7 receives the whole funding budget

Table 5.3: Budget impact, sectoral budget allocations, and POB (α)

		Budget Impact (O_1)									
	e_j	α=0	α=0.1	α=0.2	α=0.3	α=0.4	α=0.5	α=0.6	α=0.7	α=0.8	α=0.9
s1	0.395	0.0	711.0	621.7	560.1	522.6	500.1	479.1	465.7	455.8	447.5
s2	0.354	0.0	237.5	359.3	389.1	397.2	399.3	399.3	398.3	397.2	396.1
s3	0.391	0.0	642.4	591.2	541.9	509.1	489.1	471.2	459.0	450.0	442.6
s4	0.333	0.0	128.9	264.7	317.4	341.0	352.0	360.6	365.0	368.0	370.3
s5	0.232	0.0	3.5	43.4	95.1	138.0	172.1	197.4	217.9	234.3	247.8
s6	0.331	0.0	121.5	256.9	310.8	335.6	346.9	356.8	361.8	365.4	367.7
s7	0.417	3258.9	1222.6	815.7	671.4	598.4	551.7	524.2	503.3	487.5	475.4
	Total	**3258.9**	**3067.4**	**2952.9**	**2885.7**	**2841.9**	**2811.2**	**2788.7**	**2770.9**	**2758.2**	**2747.5**

		Budget Allocations (B_j)									
	e_j	α=0	α=0.1	α=0.2	α=0.3	α=0.4	α=0.5	α=0.6	α=0.7	α=0.8	α=0.9
s1	0.395	0	1800	1574	1418	1323	1266	1213	1179	1154	1133
s2	0.354	0	670	1015	1099	1122	1128	1128	1125	1122	1119
s3	0.391	0	1643	1512	1386	1302	1251	1205	1174	1151	1132
s4	0.333	0	387	795	953	1024	1057	1083	1096	1105	1112
s5	0.232	0	15	187	410	595	742	851	939	1010	1068
s6	0.331	0	367	776	939	1014	1048	1078	1093	1104	1111
s7	0.417	7814	2932	1956	1610	1435	1323	1257	1207	1169	1140
	Total	**7814.0**	**7814.0**	**7815.0**	**7815.0**	**7815.0**	**7815.0**	**7815.0**	**7813.0**	**7815.0**	**7815.0**
	POB(α)	**0.000**	**0.059**	**0.094**	**0.114**	**0.128**	**0.137**	**0.144**	**0.150**	**0.154**	**0.157**

for $\alpha = 0$, because it has the highest impact value. Thus, $\alpha = 0$ value is impractical. As we increase α, the distributions of O_1 and B quickly change and relative balancing is achieved. Note that the price of balancing increases as we increase α. For instance, the total impact of the funding budget drops from 3258.9 to 2841.9 for 0.4, and associated the price of balancing value becomes 0.128. However, the distribution of the funding budget becomes more balanced. We also plot the distributions of O_1 and B in Figure 5.1 and 5.2, respectively to give a concrete representation about how sectoral balancing works.

In Figure 5.1, as we increase α, the contributions of sectors s7, s1, and s3 to objective O_1 decrease, whereas those of sectors s2, s4, s6, and s5 increase. Note that the impact

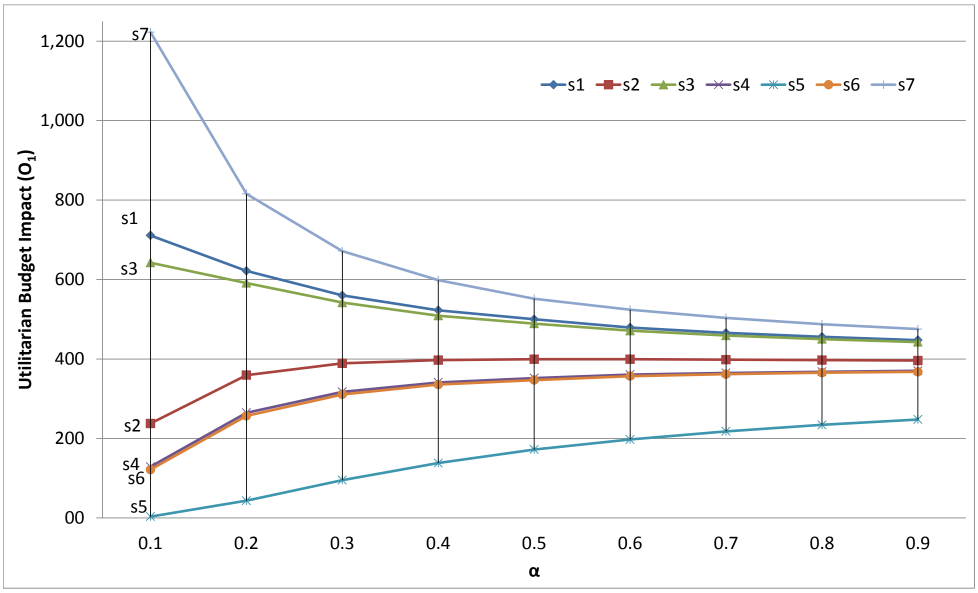

Figure 5.1: Sectoral budget impact distribution under different α levels

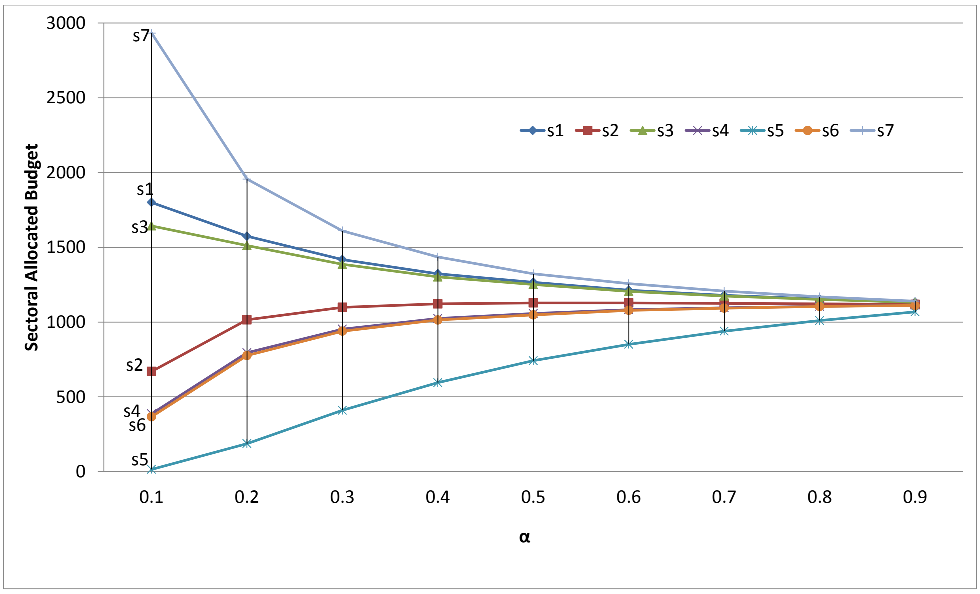

Figure 5.2: Sectoral budget distributions under different α levels

values of sectors s7, s1, and s3 are greater than the median impact value, sector s2 has median impact value, and the impact values of sectors s4, s6, and s5 are lower than the median impact value. As we increase α, the distribution of objective O_1 among sectors is adjusted according to sectoral impact values. Hence, distribution under various α levels constitutes a funnel shape. Note that sectors s4 and s6 follow the same increasing pattern because their impact values are almost the same. Similarly, sectors s1 and s3 follow very similar decreasing pattern since their impact values are very close. Decrease in sector s7 is sharper than any other sector since it has the highest impact value. Note that shares of sectors in objective O_1 are strictly compatible with sectoral impact values. In other words, share of the highest impact valued sector is the highest, and share of the lowest impact valued sector is the lowest. This structure shows how social welfare objective splits total impact of funding budget among sectors, as we increase α level.

Associated allocated sectoral budgets (i.e. B_j values) are presented in Figure 5.2. Similar adjustment pattern and funnel shape are also observed in sectoral allocated budgets. However, adjustment of funding budget is not that strict as in O_1. For instance, sectoral allocated budgets converge to almost the same funding level for $\alpha = 0.8, 0.9$. Note that sector s5 gets funding amount very close to zero for $\alpha = 0.1$, which could be undesirable for the DM. Therefore, for our example problem, those α values (i.e. 0.1, 0.8, and 0.9) turn out be impractical for the DM. Note that sectors s4 and s6 receive almost the same amount of funding due to impact values. Likewise, sectors s1 and s3 get nearly the same amount of funding. As we increase α, distribution of B is rapidly adjusted according to impact values. However, after some α value, allocated amounts to sectors turn out be almost the same, which contradicts with the idea of relative balancing according to impact values. That's why determining α value in social welfare objectives depends on problem environment and concerns of the DM.

We also present price of balancing and its theoretical bound in Figure 5.3. As we expected, price of balancing (POB(α)) increases as we slightly increase α. However, sectoral balancing prominently improves. For instance, if the DM selects $\alpha = 0.3$, price of balancing will be around 0.11. This means that total impact of funding budget (O_1) will be 11% lower than its optimal value (i.e. maximum value of O_1 for $\alpha = 0$); however, resulting sectoral allocated budgets will be practical for the DM.

We illustrate our approach on the example problem. Price of balancing is not greater than 15% for $\alpha = 0.7$. (Recall that $\alpha \geq 0.8$ is impractical for this example problem). In other words, the loss in the optimum budget impact objective due to sectoral balancing is smaller than 15%. Note that calculated theoretical bound is around 45% for $\alpha = 0.7$. This theoretical bound implies that price of balancing of any problem instance will not be worse than 45% as long as we keep number of sectors and ratio of maximum impact value over minimum impact value (see R value in Appendix B.1) as constant. As discussed in Bertsimas et al. (2012), worst case bound provides long term implications to the DM for the selection of α. For our problem setting, the DM could benefit from theoretical bound as follows. Let us assume that impact assessment of an R&D funding program is conducted in every five years. Hence, sectoral impact assessment values will be valid during five years. Let us also assume that sectoral classification will not be changed during five years (i.e. number of sectors is constant). During those five years, total number of project proposals, its distribution among sectors, budget and score scheme of projects as well as available funding budget could change in different calls of the program. Therefore, this bound indicates the price of balancing of any call instance in the worst case. For example, if the DM wants to ensure that price of balancing will be smaller than 45% in any call instance in the future, s/he perceives from Figure 5.3 that α could be at most 0.7.

We also examine price of balancing of the example problem under alternative e_j values. Bertsimas et al. (2012) discuss that the bound value may not be pathologic and price of balancing can be close to the bound value. Therefore, we change impact values of sectors used in the example problem, because, for our problem setting, only impact values affect price of balancing under the same number of sectors (see Appendix B.1 for details). We generate two alternative cases by keeping the same level theoretical bound (i.e. R parameter, see Appendix B.1 for details) and relative ranking of impact values. We also define relative total distance (RTD)= $\sum_j (\max_{j \in M} e_j - e_j)$ where M is set of sectors. Alternative impact values are given in Table 5.4. Then, we solve our problem instance with those impact values and also plot associated price of balancing values in Figure 5.3 as alternative cases (i.e. ac1 and ac2). We observe that price of balancing for alternative cases is below than 30% (i.e. it increases from 15% to 27% for $\alpha = 0.7$). We also observe that as RTD increases, price of balancing

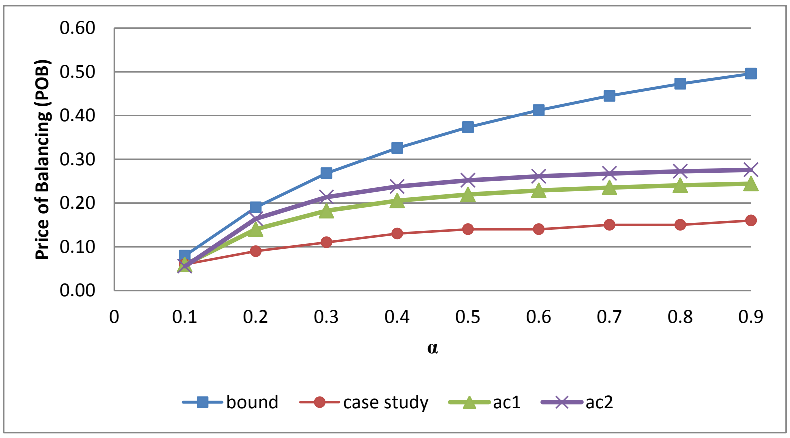

Figure 5.3: Price of balancing

increases. Because RTD measures total distance from the highest impact value. Note that RTD value of the example problem is 0.47 (i.e. impact values data of Wang et al. (2013)).

Table 5.4: Two alternative cases of e_j levels

sectors	ac1	ac2
s1	0.9	0.8
s2	0.7	0.7
s3	0.8	0.75
s4	0.67	0.65
s5	0.55	0.55
s6	0.65	0.6
s7	1	1
RTD	1.73	1.95

We present distribution of O_1 and B to the DM with price of balancing information. It seems that the DM could prefer α values between 0.2 and 0.7 by evaluating given information so far. However, in our problem setting, sectoral balancing could affect many problem properties, which are wondered by the DM. Hence, the DM could refine possible α values by considering those problem properties. Next, we illustrate and discuss them respectively.

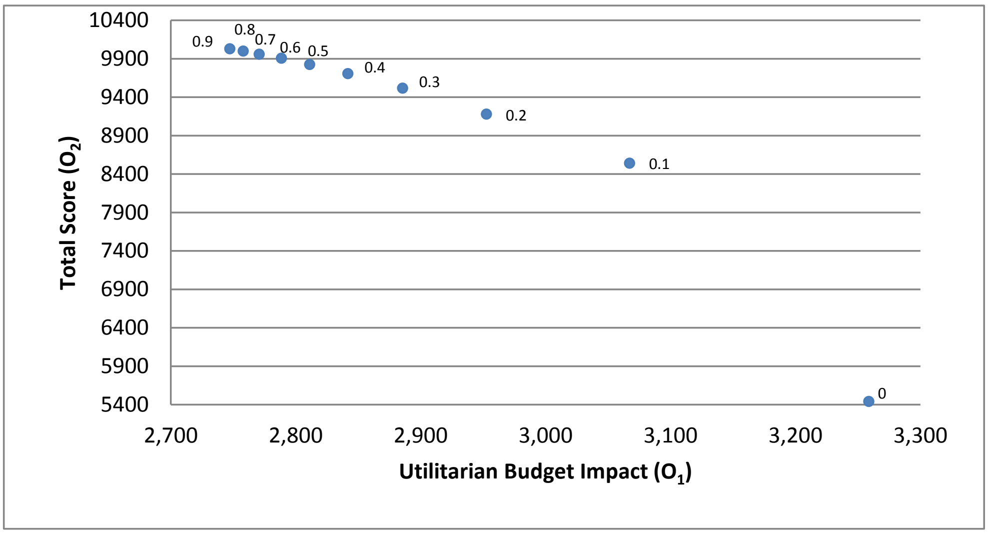

Figure 5.4: Relationship between objective O_1 and O_2

5.2.2 Effect of sectoral balancing on the total score

Relationship between objective O_1 and O_2 is given in Figure 5.4. As we increase α, objective O_1 decreases due to price of balancing, whereas objective O_2 improves. Because, as sectoral balancing improves, all sectors get more funding and more high scored projects of each sector are selected in the portfolio. We also measure the total score increase (i.e. TSI(α)) by using the following formula:

$$\text{TSI}(\alpha) = \frac{\text{TS}(\alpha) - \text{TS}(\alpha_0)}{\text{TS}(\alpha_0)} \tag{5.20}$$

where $\text{TS}(\alpha_0)$ is the total score of selected projects when we solve our two stage model for $\alpha = 0$ and $\text{TS}(\alpha)$ is the total score of selected projects when we solve the model for positive α. In Figure 5.5, we present the total score increase and associated price of balancing to give more insights to the DM. For instance, if the DM selects $\alpha = 0.2$, price of balancing will be around 9.5% (i.e. O_1 decreases from 3258.9 to 2952.9), but total score increase will be around 68% (i.e. O_2 increases from 5438

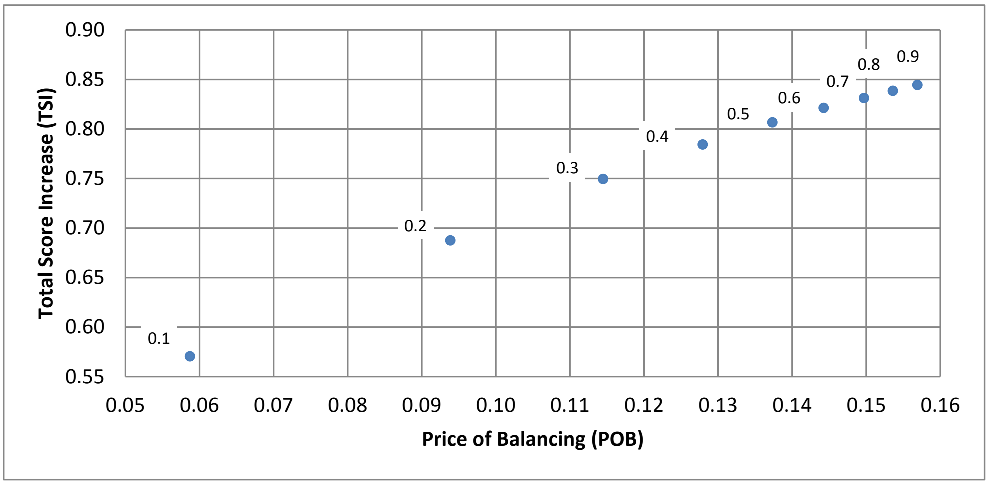

Figure 5.5: Price of balancing and total score increase

to 9176). Therefore, total score improves notably, as we slightly increase α level. This observation shows significance of sectoral balancing and supports the idea that imbalanced portfolios are not practical for the DMs.

Distribution of total score (O_2) among sectors is presented in Table 5.5. Recall that sector s7 gets total funding budget for $\alpha = 0$; hence, total score is 5438, which equals to sectoral score of s7. As we increase α, distribution of O_2 notably changes and sectoral balancing occurs. We also plot distribution of O_2 under various α values in Figure 5.6 to give implications about how sectoral budget balancing affects sectoral scores. As we increase α, sectoral score of s7, s1, and s3 decreases, whereas that of sectors s2, s4, s6, and s5 increases (Recall that impact value of sectors s7, s1, and s3 is greater than that of s2, s4, s6, and s5). This is the similar pattern observed both in distribution of stage I objective O_1 and total funding budget (B). On the other side, due to different budget range of each sector (recall that sectoral budgets are different, see Table 5.2 for sectoral budget details), reaction of sectors to different allocated budgets could notably change in terms of sectoral scores. Distribution of O_1 and (B) is always compatible with sectoral impact values. Namely, lower impact valued sec-

Table 5.5: Distribution of total score objective (O_2) among sectors

	e_j	Total Score (O2)									
		α=0	α=0.1	α=0.2	α=0.3	α=0.4	α=0.5	α=0.6	α=0.7	α=0.8	α=0.9
s1	0.395	0	1469	1317	1209	1142	1101	1063	1037	1018	1004
s2	0.354	0	885	1180	1244	1261	1266	1266	1264	1261	1259
s3	0.391	0	2382	2251	2120	2031	1975	1924	1889	1863	1841
s4	0.333	0	344	680	804	860	885	905	916	922	927
s5	0.232	0	48	507	976	1314	1562	1737	1873	1977	2061
s6	0.331	0	631	1200	1406	1495	1536	1571	1588	1601	1609
s7	0.417	5438	2780	2041	1754	1599	1498	1437	1390	1354	1327
	Total	**5438**	**8539**	**9176**	**9513**	**9702**	**9823**	**9903**	**9957**	**9996**	**10028**
	POB(α)	**0.000**	**0.059**	**0.094**	**0.114**	**0.128**	**0.137**	**0.144**	**0.150**	**0.154**	**0.157**

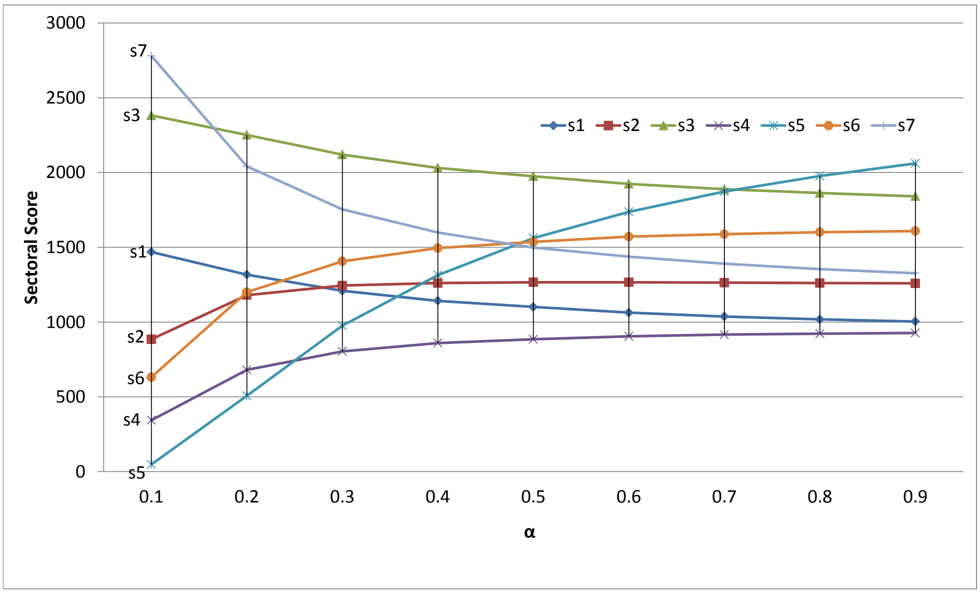

Figure 5.6: Distribution of total score under various α levels

tor allocation in distribution of O_1 and (B) cannot be greater than allocation of higher impact valued sector. However, when it comes to distribution of sectoral scores, allocated sectoral scores in each sector are not necessarily compatible with impact values because of different sectoral budgets (Later, we see that similar behavior will also be observed for other remaining problem parameters such as average sectoral score, sectoral success rate and number of supported projects in each sector). This situation can be clearly observed in Figure 5.6 since there are many intersections of sectoral lines under different values of α. In Figure 5.6, as we increase α, we observe that change in sectoral score of sector s5 is the most prominent. For instance, its sectoral score is close to zero for $\alpha = 0.1$, whereas, its sectoral score gets maximum share in O_2 for $\alpha = 0.9$. This kind of behavior implies that as we increase α, the lowest impact valued sector might dominate other sectors in terms of sectoral score, which may not be desirable for the DM. Later, we also discuss that there are different implications of this behavior of sector s5 in terms of other sectoral indicators.

The DM may arise different concerns in terms of sectoral scores. Let us assume that the DM wants to keep share of the lowest impact valued sector in total score small. For instance, the DM may require that sectoral score of the lowest impact valued sector (i.e. sector s5) should be smaller than that of median impact valued sector. This would suggest α value of at most 0.3 (Recall that sector s2 has median impact value). Note that price of balancing is 11.4% and total score increase is 75% for $\alpha = 0.3$ (see Figure 5.5). Similarly, the DM may prefer that sectoral score of the lowest impact valued sector must be smaller than that of highest impact valued sector (i.e. sector s7). This preference would suggest α value of at most 0.4. Note that price of balancing is 12.8% and total score increase is around 78% for $\alpha = 0.4$. Therefore, the DM could give different concerns according to distribution of total sectoral score and could select corresponding α value. We illustrate and discuss how sectoral balancing could affect total score objective and its distribution among sectors. We also give illustrative concerns of the DM to assist for selecting α.

Next, we discuss average score and budget of supported projects in each sector under various α values.

5.2.3 Average score and budget of supported projects in each sector

Average score and budget of supported projects in each sector are given in Table 5.6. As we increase α, overall average score increases. For instance, overall average score is 17.4 for $\alpha = 0$, it becomes 20.1 for $\alpha = 0.2$ and remains almost the same for other α values. In addition, overall average budget decreases, as we increase α. For example, overall average budget is 25 for $\alpha = 0$, it becomes 18.2 for $\alpha = 0.1$, and it goes on decreasing for other α values. Therefore, overall average score increases and overall average budget decreases, as we increase α. However, note that sectoral average score and sectoral average budget behave differently. In particular, the average budget and score of some sectors decrease, whereas the average budget and score of some sectors increase. For that reason, we also plot average score and budget of supported projects in each sector under various α values in Figures 5.7 and 5.8 respectively to give more concrete insights about how sectoral balancing affects those indicators.

Table 5.6: Average score and budget of supported projects under various α levels

		Average Score									
	e_j	α=0	α=0.1	α=0.2	α=0.3	α=0.4	α=0.5	α=0.6	α=0.7	α=0.8	α=0.9
s1	0.395	0	21	21.6	21.6	21.5	21.6	21.7	22.1	22.1	21.8
s2	0.354	0	18.1	18.4	18	18.3	18.1	18.1	18.1	18.3	18.2
s3	0.391	0	19.2	19.1	19.1	19.2	19.4	19.4	19.5	19.4	19.4
s4	0.333	0	22.9	21.9	21.7	21.5	21.6	21.5	21.3	21.4	21.1
s5	0.232	0	24	22	21.2	20.5	20.3	20	19.9	20	19.8
s6	0.331	0	22.5	21.1	20.7	20.5	20.5	20.4	20.4	20.3	20.4
s7	0.417	17.4	19.7	20	20.4	20.5	20.5	20.5	20.7	20.8	20.7
	avg	**17.4**	**19.9**	**20.1**	**20.1**	**20.1**	**20.1**	**20.0**	**20.1**	**20.1**	**20.0**

		Average Budget									
	e_j	α=0	α=0.1	α=0.2	α=0.3	α=0.4	α=0.5	α=0.6	α=0.7	α=0.8	α=0.9
s1	0.395		25.7	25.8	25.3	25	24.8	24.8	25.1	25.1	24.6
s2	0.354		13.7	15.9	15.9	16.3	16.1	16.1	16.1	16.3	16.2
s3	0.391		13.3	12.8	12.5	12.3	12.3	12.2	12.1	12	11.9
s4	0.333		25.8	25.6	25.8	25.6	25.8	25.8	25.5	25.7	25.3
s5	0.232		7.5	8.1	8.9	9.3	9.6	9.8	10	10.2	10.3
s6	0.331		13.1	13.6	13.8	13.9	14	14	14	14	14.1
s7	0.417	25	20.8	19.2	18.7	18.4	18.1	18	18	18	17.8
	avg	**25.0**	**18.2**	**17.1**	**16.5**	**16.2**	**16.0**	**15.8**	**15.8**	**15.7**	**15.6**
	POB(α)	**0.000**	**0.059**	**0.094**	**0.114**	**0.128**	**0.137**	**0.144**	**0.150**	**0.154**	**0.157**

In Figure 5.7, as we increase α, we observe that average sectoral score of s5, s4, and s6 notably decreases until $\alpha = 0.4$. Recall that impact value of those sectors is strictly

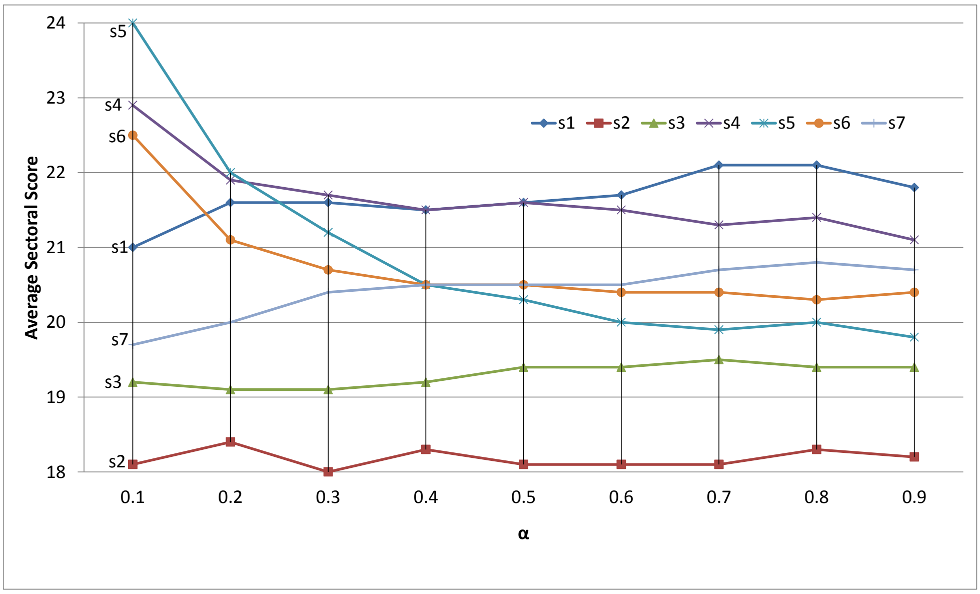

Figure 5.7: Sectoral average score under various α values

smaller than median impact value. The DM may arise some specific concerns regarding average sectoral scores. Let us assume that the DM is inclined to increase project competition in low impact valued sectors and s/he wants to keep average project score in low impact valued sectors relatively higher. For instance, the DM may want that average score in the lowest impact valued sector (i.e. sectors s5) should not be smaller than that of the highest impact valued sector (i.e. sector s7). This concern would suggest α value of at most 0.4. In Figure 5.8, as we increase α, average sectoral budget values do not notably change except sector s7 (i.e. the highest impact valued sector) and s5 (the lowest impact valued sector). As we increase α, average budget in sector 7 decreases. Note that in Figure 5.7, average score of sector s7 increases, as we increase α. Hence, higher scored projects are selected with lower budgets in sector s7, which could be desirable for the DM. On the other side, as we increase α, average budget in sector 5 increases. Recall that in Figure 5.7, average score of sector s5 notably decreases, as we increase α. Hence, more low scored projects are selected with higher budgets in sector s5 (i.e. the lowest impact valued sector), which may not be desirable for the DM. Therefore, if the DM evaluates Figure 5.7 and 5.8, s/he could

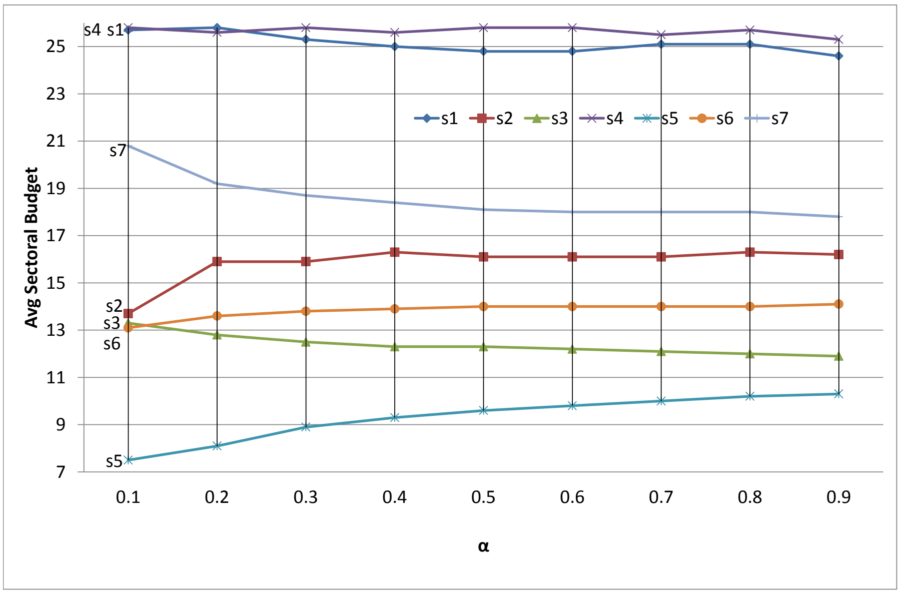

Figure 5.8: Sectoral average budget under various α values

select α value of at most 0.4.

Next, we discuss success rate and number of selected projects in each sector under various α values.

5.2.4 Success rate and number of selected projects in each sector

Success rate (i.e. ratio of number of supported projects over number of submitted project proposals) and number of supported projects in each sector are presented in Table 5.7. As we slightly increase α, overall success rate and total number of selected projects notably increase due to sectoral balancing. For instance, overall success rate is 0.16 and total number of selected projects is 313 for $\alpha = 0$ (Recall that only sector s7 is supported for $\alpha = 0$, which is not practical), overall success rate rises to 0.23 and total number of selected projects reaches to 456 for $\alpha = 0.2$. This pattern also demonstrates importance of sectoral balancing and supports the practical validity of sectoral balancing to some extent. However, after some α value, sectoral balancing might negatively affect the R&D project portfolio in terms of success rate and number of selected projects. For that reason, we also plot sectoral success rates and sectoral numbers of supported projects in Figure 5.9 and 5.10, respectively to capture possible negative effects of sectoral balancing on the R&D project portfolio.

Table 5.7: Success rate and number of selected projects under various α levels

		Success Rate									
	e_j	α=0	α=0.1	α=0.2	α=0.3	α=0.4	α=0.5	α=0.6	α=0.7	α=0.8	α=0.9
s1	0.395	0	0.47	0.41	0.38	0.36	0.34	0.33	0.32	0.31	0.31
s2	0.354	0	0.34	0.45	0.48	0.48	0.49	0.49	0.49	0.48	0.48
s3	0.391	0	0.36	0.34	0.32	0.30	0.29	0.28	0.28	0.28	0.27
s4	0.333	0	0.06	0.12	0.14	0.16	0.16	0.16	0.17	0.17	0.17
s5	0.232	0	0.01	0.08	0.15	0.21	0.25	0.29	0.31	0.32	0.34
s6	0.331	0	0.06	0.12	0.14	0.16	0.16	0.16	0.17	0.17	0.17
s7	0.417	0.96	0.43	0.31	0.26	0.24	0.22	0.21	0.20	0.20	0.20
	avg	**0.16**	**0.21**	**0.23**	**0.24**	**0.24**	**0.24**	**0.25**	**0.25**	**0.25**	**0.25**

		Number of Selected Projects									
	e_j	α=0	α=0.1	α=0.2	α=0.3	α=0.4	α=0.5	α=0.6	α=0.7	α=0.8	α=0.9
s1	0.395	0	70	61	56	53	51	49	47	46	46
s2	0.354	0	49	64	69	69	70	70	70	69	69
s3	0.391	0	124	118	111	106	102	99	97	96	95
s4	0.333	0	15	31	37	40	41	42	43	43	44
s5	0.232	0	2	23	46	64	77	87	94	99	104
s6	0.331	0	28	57	68	73	75	77	78	79	79
s7	0.417	313	141	102	86	78	73	70	67	65	64
	Total	**313**	**429**	**456**	**473**	**483**	**489**	**494**	**496**	**497**	**501**
	POB(α)	**0.000**	**0.059**	**0.094**	**0.114**	**0.128**	**0.137**	**0.144**	**0.150**	**0.154**	**0.157**

In Figure 5.9, as we increase α, success rate of sector s1, s7, and s3 (i.e. high impact valued sectors) decreases; whereas success rate of sector s2, s4, s6, and s5 increases. In addition, decreasing pattern is notable in the highest impact valued sector s7. Increasing pattern of the lowest impact valued sector s5 is the most prominent. Success rate of sector s1 and s2 is higher than other sectors because number of project proposals in those sectors are low (see Table 5.2) when compared to other sectors (except that success rate of sector s5 exceeds that of sector s1 for $\alpha > 0.7$). Let us assume that the DM wants to see lower success rate in low impact valued sectors (i.e. sectors s4, s6, and s5) and higher success rate in high impact valued sectors (i.e. sectors s1, s7, and s3). This preference would suggest α value of at most 0.4. If the DM selects $\alpha = 0.4$, success rate of sectors s4, s6, and s5 will be greater than 0.16, and success rate of sectors s1, s7, and s3 will be greater than 0.24. Therefore, for $\alpha = 0.4$, even

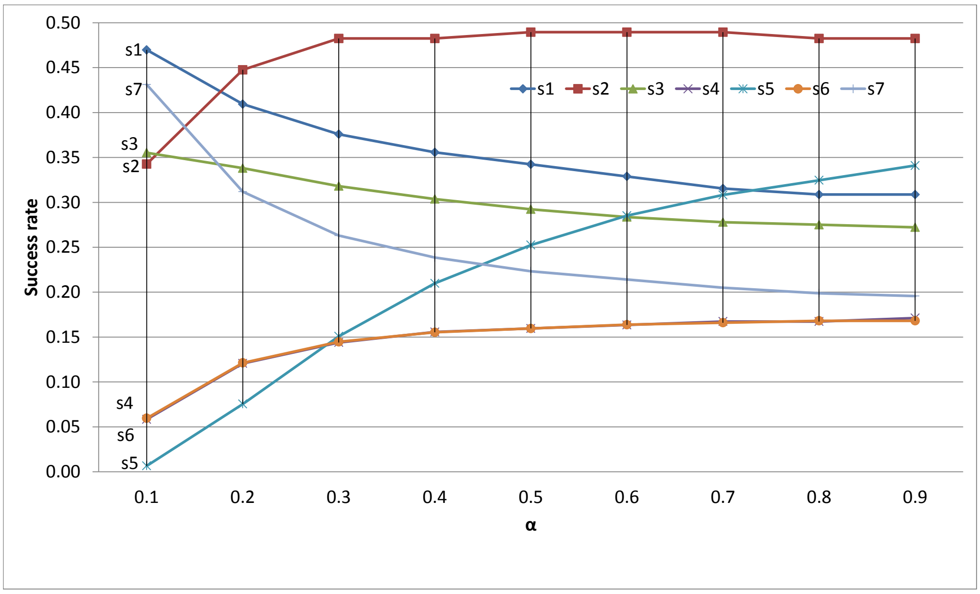

Figure 5.9: Sectoral success rate under various α values

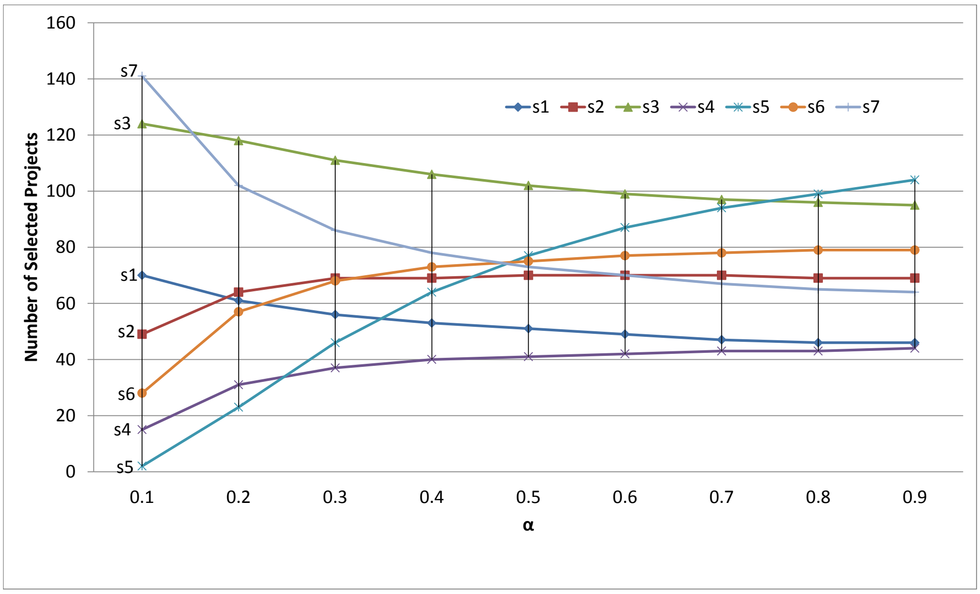

Figure 5.10: Number of selected projects under various α values

success rates of low impact valued sectors also exceed some specific threshold value such as 0.15, which could be desirable for the DM due to sectoral balancing concerns.

Analyzing success rate with number of supported projects could give a more meaningful picture since number of project proposals can vary among sectors. For example, sectoral success rates of sectors s4 and s6 are the same in Figure 5.9; however, number of supported projects in those sectors are not equal due to different number of project proposals (see Table 5.2). In Figure 5.10, as we increase α, number of supported projects of sector s7, s3, and s1 (i.e. high impact valued sectors) decreases, but number of supported projects of sector s2, s6, s4, and s5 increases. Like in success rate, change in number of supported projects is more notable in sector s7 (i.e. the highest impact valued sector) and s5 (i.e. the lowest impact valued sector). We also observe that number of supported projects of sector s5 becomes greater than that of sector s7 for $\alpha > 0.4$, which could be undesirable for the DM. Let us assume that the DM wants to keep success rate and number of supported projects of the lowest impact valued sector relatively small. Therefore, when we examine sectoral success rates and sectoral number of supported projects, it turns out that practical α value could be at most $\alpha = 0.4$. Because, for $\alpha \geq 0.5$, too many projects are supported in the lowest impact valued sector s5 as well as its success rate exceeds that of the highest impact valued sector, which could be refused by the DM.

Next, we present value of our proposed approach.

5.2.5 Value of the Proposed Sectoral Allocation Model

We compare our proposed sectoral allocation model with some policy options that can be adopted for sectoral allocation if impact assessment values are not available. We define three policy options that could be practically applied as heuristics for this purpose. First one is maximization of only total score objective with no sectoral constraints (Max-Score). Second option is dividing funding budget among sectors equally (Eq-Sec-Alloc). Third option is dividing funding budgets among sectors according to direct proportion of normalized number of project proposal $\times$ average budget in each sector (Weighted-Demand). We compare our proposed approach under $\alpha = 0.4$ with those policy options. We select $\alpha = 0.4$ for comparison (Sec-Balancing)

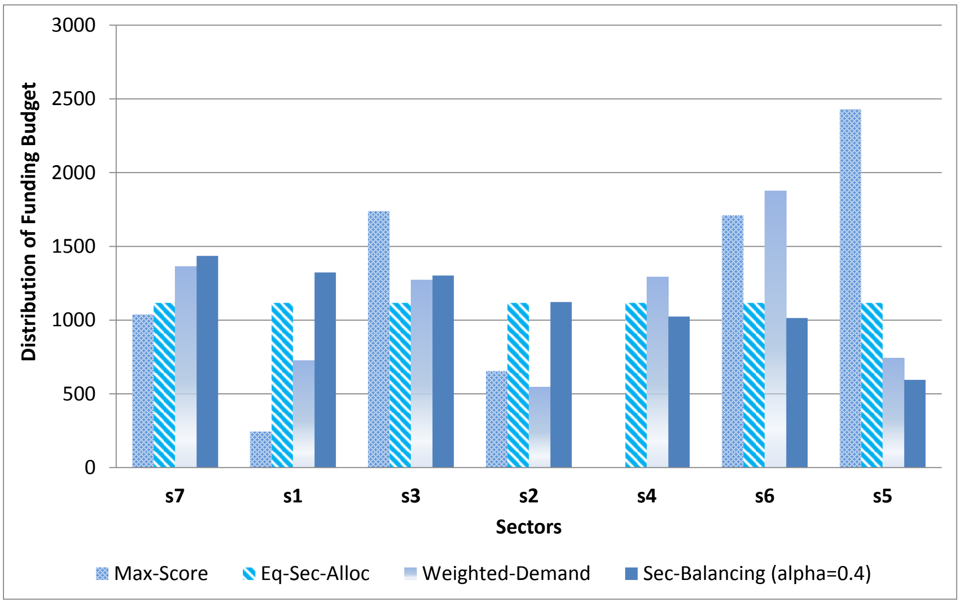

Figure 5.11: Sectoral allocations under different policy options

since our informed decision making approach suggests this α value several times according to illustrative preferences of the DM.

In Figure 5.11, we present allocated sectoral budgets (in 10,000 monetary units) under each policy option and sectoral balancing with $\alpha = 0.4$. Note that we sort the sectors according to their impact values to give a more concrete representation. We observe that allocated sectoral budgets might notably change according to adopted policy option. For example, if the DM selects only maximization of score objective without any sectoral restrictions (i.e. Max-Score), sector s5 will receive serious amount of funding (i.e. nearly one over third of funding budget). However, note that it will get the minimum amount of funding if the DM applies impact assessment based sectoral balancing approach (i.e. Sec-Balancing). We also observe that sector s4 receives no funding under Max-Score policy option since its sectoral budget range is the maximum (see Table 5.2). We also observe that sector s1, one of high impact valued sectors, receives very little under Max-Score policy option. Sectoral budget allocation under Max-Score policy option demonstrates that sectoral budget decisions should be incorporated into decision making process.

Under Eq-Sec-Alloc policy option each sector receives the same amount. Note that allocated budget of sector s5 nearly doubles under Eq-Sec-Alloc policy option when we compare it with its allocation under Sec-Balancing policy. Recall that number of projects of sector s5 increases considerably if it exceeds allocated budget under Sec-Balancing policy with $\alpha = 0.4$. Therefore, negative portfolio effect of sector s5 still remains under Eq-Sec-Alloc policy option. This observation also shows significance of proposed informed decision making approach. Because, deciding on sectoral budgets with several portfolio indicators in an interactive manner unveils many interesting patterns of sectors and gives more insights to the DM.

Under Weighted-Demand policy option each sector receives funding proportion calculated by normalizing number of project proposal $\times$ average budget. We observe that sector s1 and s5 get similar amount of funding under Weighted-Demand policy option. However, they receive very different amounts under Sec-Balancing option due to their impact values. In addition, we observe that allocated budgets of sector s1 and s2 under Weighted-Demand policy option are notably smaller than their allocated budgets under Sec-Balancing option. We also observe that allocated budget of sector s6 under Weighted-Demand policy option considerably exceeds its allocated budget under Sec-Balancing option. Clearly, alternative policy options in the absence of impact assessment values could result in different allocations when we compare them with proposed Sec-Balancing approach.

In addition, in Table 5.8, we calculate total impact of funding budget (O_1) and associated price of balancing for each policy option. For instance, total impact of funding budget (O_1) under Max-Score policy option is 2570.4 with price of balancing $\%21.1$. Total impact of funding budget (O_1) are 2737.5 (with price of balancing $\%16.0$) and 2768.3 (with price of balancing $\%15.1$) under Eq-Sec-Alloc and Weighted-Demand policy options, respectively. However, total impact of funding budget (O_1) under proposed approach ($\alpha = 0.4$) is 2841.9 with price of balancing $\%12.8$. Overall, our proposed approach enhances the total impact of funding budget by appropriately balancing sectoral budgets.

Table 5.8: Total impact under various policy options

	Max-Score	Eq-Sec-Alloc	Weighted-Demand	Sec-Balancing ($\alpha = 0.4$)
Total Impact (O_1)	2570.4	2737.5	2768.3	2841.9
Price of balancing	21.1%	16.0%	15.1%	12.8%

5.3 Conclusions

In this chapter, we consider a public RDPPS problem in which the DM wants to allocate sectoral budgets according to sectoral impact assessment results. We develop a two stage model. In the first stage, the DM deals with sectoral budget decisions in the light of sectoral impact assessment values to maximize impact of available funding budget while ensuring sectoral budget balancing concerns. We propose a social welfare objective parameterized by single parameter to incorporate sectoral impact assessment values into decision making process. In the second stage, the DM wants to maximize total score of selected projects under allocated sectoral budgets. Besides, the DM concerns with effects of allocated sectoral budgets on several sectoral indicators such as sectoral total score, mean score and budget of supported projects in each sector, success rate and number of supported projects in each sector. Therefore, we develop a flexible informed decision making approach to assist the DM on sectoral budget decisions. We solve the example problem, illustrate some preferences of the DM, present associated sectoral budgets and their effects on aforementioned sectoral indicators. To the best of our knowledge, incorporation of sectoral impact assessment results into a public RDPPS model and comprehensive sectoral analysis of projects have not been considered in previous studies. Finally, we compare proposed approach with some policy options which could be applied in the absence of impact assessment values. We show that sectoral allocations under alternative policy options could be notably different than sectoral allocation of proposed approach. Consequently, proposed approach enhances the total impact of funding budget by considering sectoral budget allocations. R&D funding agencies can apply our proposed approach easily while making sectoral budget allocation decisions.

CHAPTER 6

CONCLUSION

We summarize the conducted research in this dissertation in Section 6.1 and give potential future research directions in Section 6.2.

6.1 Concluding Remarks

In this dissertation, we are motivated by the two practical issues that exist in public R&D funding management. Our first motivation is the underutilization of the funding budget. We find that project cancellations and underspending of successfully finished projects could cause to the budget underutilization.

We first consider cancellation situations in Chapter 3 in two cases. In the first case, we assume that cancellation probabilities of projects are not known. A significant contribution of this chapter is the development of models with unknown cancellation probabilities. We develop a mathematical programming model, which uses an estimated number of canceled projects as an input parameter. Objective to maximize is the total score of selected projects under given funding budget while incorporating that a number of them will be canceled leaving some remaining budget. By considering the worst case scenario, we formulate a nonlinear cancellation function and integrate it to the budget constraint by using the duality theory and the McCormick's linearization technique. However, resulting linearized model has a big M value (i.e. large enough positive value to accurately solve the model). We observe arbitrarily large big M values can make the model computationally challenging. Therefore, we also obtain the smallest big M value by using complementary slackness and feasibility conditions. Resulting model has a mixed-integer linear programming (MILP) form, which could be solved easily by commercial solvers such as CPLEX.

On the other side, in Chapter 3, we have a knapsack problem oriented model for modeling the cancellations, and we know that there is a DP algorithm that solves the knapsack problem in pseudo-polynomial time. From a methodological standpoint, it is interesting to develop a DP algorithm that considers estimated number of canceled projects as an input parameter. Therefore, we also develop an efficient DP algorithm that solves the proposed model in seconds for practical size problem instances.

In the second case of Chapter 3, we assume that cancellation probabilities are known. We propose a mathematical programming model, which maximizes the total expected score of selected projects under probabilistic budget constraint. The second proposed model could exactly solve 90% of the instances to optimality within given time limit, but the average optimality gap for the instances that were only solved to feasibility is below 0.02%. Moreover, we provide practical insights to the DM on the proposed models. For instance, we obtain the budget risk of the first model by using the probability information of the second model. The risk of exceeding budget for the first model is between 4.5% and 0.1%. Besides, we obtain the budget risk levels of various cancellation scenarios for the same instance by changing the number of cancellations by $\pm$%20. We also show that the first model gives a tight approximation to the second model when we solve the second model with conservative budget risk of the first model. Moreover, we find the value of second model under different factorial settings and observe that the second model generates better project portfolios than first model with anticipated risk levels. We also show that the proposed models significantly enhances the budget utilization. Results in Chapter 3 of this dissertation is accepted for publication in OR Spectrum (Çağlar and Gürel (2016)).

In Chapter 4, we consider the budget underutilization situation with several sources of expenditure uncertainty. More specifically, we focus on stochastic modeling of the expenditure of project cancellations and successfully finished projects. Since we consider thousands of project proposals with different budgets, it is impractical to use a different expenditure distribution for each project. For that reason, we have modeled the ratio of budget spent by a probability distribution. In particular, each project has an approved budget b_i, and its stochastic expenditure $\hat{b}_i$ will be within the interval $(0, b_i]$. Therefore, we introduce a ratio $R_i = \frac{\hat{b}_i}{b_i}$ for the budget spending of a project i and focus on tractable modeling of R_i. We first consider cancellation situation. By using a confidential call data set from TÜBİTAK, we observe that budget spending ratio of canceled projects can be modeled by the same beta distribution. Besides, we assume cancellation probability of each project is different since projects from different PIs and research institutions may have different cancellation risks. Secondly, we consider budget spending behavior of successfully completed projects that spend less budget than planned. We also observe that budget spending ratio of those projects could be formulated with the same beta distribution. In addition, we assume that probability of successful completion with an underutilized budget is the same for all projects. Lastly, we observe that some successfully finished projects spend their whole budgets. Hence, spending ratio of 100% should be included while modeling R_i. Motivated by this environment, we formulate budget spending R_i with a mixture distribution. Therefore, an important contribution of this chapter is tractable modeling of several practical underutilization cases with a probabilistic approach under large set of projects.

In Chapter 4, after formulation of R_i, we develop a chance-constrained stochastic programming (CCSP) model. From a methodological perspective, chance constraint of budget is computationally intractable; thus, we propose a solution method based on a normal approximation and a conic quadratic programming. On the other hand, when the normal approximation is employed, there could be some convergence error that can distort desired risk level of chance constraint of funding budget. For that reason, by applying Berry-Esseen theorem, we obtain the convergence quality of the normal approximation and propose ways to mitigate risk level of probabilistic budget constraint. We think that proposed ways could be effectively applied in CCSP formulation of any problem in which probabilistic total resource capacity type constraint is computationally intractable. Proposed model could exactly solve 86% of the the instances to optimality within given time limit, but the average optimality gap for the instances that were only solved to feasibility is below 0.01%.

In Chapter 4, we also conduct additional analyses to give practical implications to the DM. Firstly, we observe that there is some smooth out effect of cancellation probabilities on the variance of budget spending of canceled projects. Secondly, we observe that increasing probability level of budget constraint leads to decreasing expected total score of selected projects. Thirdly, we examine the case of unknown distribution of budget underspending of canceled and successfully completed projects. For this case, we propose an alternative distribution modeling and compare it with proposed approach on different problem settings. We observe that alternative modeling provides a good approximation to proposed approach if there is only cancellations in the problem. On the other hand, approximation quality deteriorates if some considerable fraction of successfully completed projects spend less budget than planned. Finally, we compare proposed approach with standard setting and find that proposed approach delivers a notable increase in utilization of funding budget. Chapter 4 of this dissertation is currently under review.

Our second practical motivation is sectoral balancing of funding budget according to sectoral impact findings. In Chapter 5, we introduce sectoral budget balancing problem in the public RDPPS model. We consider an R&D funding program that receives many project proposals from various sectors. We assume that sectoral impact assessment values of the funding program are available. The DM wants to balance available funding budget among sectors according to the sectoral impact assessment values. We propose a two-stage model. In the first stage, the DM is concerned with sectoral budget allocations to maximize the impact of funding budget while keeping the relative budget balancing among sectors. In the second stage, the DM deals with the maximization of the total score of selected projects under determined sectoral budgets. We propose a social welfare objective function in the first stage. This social welfare function is governed by a single parameter α and it is nonlinear for $\alpha > 0$. We prove that nonlinearity in the objective function can be handled by conic quadratic inequalities. We also propose an informed decision making approach to show effect of sectoral budget decisions on various indicators such as total impact of funding budget, total score of selected projects, sectoral scores and budgets, sectoral number of supported projects, and sectoral success rates. From a practical standpoint, a significant contribution of this chapter is incorporation of sectoral impact assessment values into a public project selection model and in-depth sectoral analysis of projects. We illustrate our proposed approach on an example problem. We generate our example problem by using sectoral public data of TÜBİTAK 1001 program. For our example problem, we also derive proxy sectoral impact assessment values from the literature. Finally, we show the value of the proposed approach by comparing it with some alternative policy options to give managerial insights to the DM. Chapter 5 of this dissertation is currently under review.

6.2 Future Research Directions

We assume a deterministic call budget B in our proposed models throughout the dissertation. As a future research, extension to uncertain call budget case could be studied. We know call budget cover several fiscal years. However, in Turkish budgeting system, only current year's budget is known and there are upper bounds on budget of upcoming years. Hence, some portion of call budget is subject to uncertainty. Uncertain call budget extension will make all proposed models more challenging. In addition, in all our proposed models, we assume an aggregated score measure (i.e. sum of scores of evaluation criteria) for the panel evaluation of projects, which is currently applied in funding organizations like TÜBİTAK. It can be interesting to study multiple criteria case explicitly as a future research to give alternative insights to DMs.

Regarding the proposed models in Chapter 3 and Chapter 4, possible bi-objective models of public RDPPS problem can be investigated. For instance, there can be a trade-off between expected number of cancellations and expected scores. Therefore, a bi-objective model that maximize total expected scores and minimize expected number of cancellations while satisfying probabilistic budget constraint can be formulated. Moreover, applicability of refined normal approximation in CCSP could be examined to minimize Berry-Esseen bound.

In Chapter 5, we consider sectoral budget balancing within a funding program. As a future research, proposed first stage model could be applied to distribute total budget of a funding agency among different R&D funding programs. However, scale (i.e. upper budget limit of a project) and the objective of each R&D funding program can be totally different. For example, in TÜBİTAK, one funding program supports small scale projects up to 30.000 Turkish Lira, another funding program supports big scale projects up to 2.5 million Turkish Liras. Apparently, scope and objectives of the programs are different. If impact assessment results of those programs close to each other, then social welfare objective would allocate close amounts. However, this would result in supporting several projects in the big scale program and supporting too many projects in the small scale program. Obviously, this kind of allocation is not practical. Therefore, incorporation of scale and scope concerns into our first stage model with novel analytical approaches could be studied as a future work.

The proposed two stage model in Chapter 5 has a limitation. For instance, in stage II, allocated budget to any sector may not be fully utilized. It may be better to calculate the total impact of funding budget by using sum of budgets of selected projects in each sector. Therefore, integration of stage I and stage II model could be studied as a future research.

REFERENCES

Optimizing the research and innovation policy mix: The practice and challenges of impact assessment in Europe. EU FP7 OMC-net project report 234501, 2011.

F. Alizadeh and D. Goldfarb. Second-order cone programming. *Mathematical Programming*, 95(1):3–51, 2003.

N. R. Baker. R and D project selection models - assessment. *IEEE Transactions on Engineering Management*, EM21(4):165–171, 1974.

H. Baurer. *Probability Theory*. De Gruyter studies in mathematics 23. Walter de Gruyter, Transl. by Burckel, Robert B., Berlin, New York, 1996.

G. J. Beaujon, S. P. Marin, and G. C. McDonald. Balancing and optimizing a portfolio of R&D projects. *Naval Research Logistics*, 48(1):18–40, 2001.

D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, 52(1): 35–53, 2004.

D. Bertsimas, VF. Farias, and N. Trichakis. The price of fairness. *Operations Research*, 59(1):17–31, 2011.

D. Bertsimas, VF. Farias, and N. Trichakis. On the efficiency-fairness tradeoff. *Management Science*, 58(12):2234–2250, 2012.

B. Bozeman and J. Rogers. Strategic management of government-sponsored R&D portfolios. *Environment and Planning C: Government and Policy*, 19:413–442, 2001.

H. Capron. *Economic Quantitative Methods for the Evaluation of the Impact of R&D Programmes: A State-of-the-art*. Monitor-Spear Series, European Community Commission, Brussels, 1992.

M. Çağlar and S. Gürel. Public R&D project portfolio selection problem with cancellations. *Accepted with minor revision in OR Spectrum*, 2016.

M. Cervantes. Background note for panel 3: Implications for TIP and NESTI. Presented in the Joint Workshop on Innovation Indicators for Policy Making and Impact Assessment, OECD, Paris, 2007.

A. Charnes and W. W. Cooper. Chance constrained programming. *Management Science*, 6(1):73–79, 1959.

M.A. Coffin and B.W. III Taylor. Multiple criteria R&D project selection and scheduling using fuzzy logic. *Computers & Operations Research*, 23(3):207 – 220, 1996.

W. M. Cohen, R. R. Nelson, and J. P. Walsh. Links and impacts: The influence of public research on industrial R&D. *Management Science*, 48(1):1–23, 2002.

R. Duzgun and A. Thiele. Robust optimization with multiple ranges: theory and application to R&D project selection. *Technical Report, Lehigh University, Bethlehem, PA, USA*, 2010.

H. Eilat, B. Golany, and A. Shtub. Constructing and evaluating balanced portfolios of R&D projects with interactions: a DEA based methodology. *European Journal of Operational Research*, 172:1018–1039, 2006.

A. Eser. Ar-ge ve yenilik ekosistemimiz ve mevcut kamu Ar-Ge destekleri. *Kalkınmada Anahtar Verimlilik*, 301:15–17, 2014.

C. G. Esseen. A moment inequality with an application to the central limit theorem. *Skand. Aktuarietidskr.*, 39:160–170, 1956.

Y. Gerchak and D. M. Kilgour. Optimal parallel funding of research and development projects. *IIE Transactions*, 31(2):145–152, 1999.

Y. Gerchak and M. Parlar. Allocating resources to research and development projects in a competitive environment. *IIE Transactions*, 31(9):827–834, 1999.

A. Goel and P. Indyk. Stochastic load balancing and related problems. *Proceedings of the 40thAnnual Symposium on Foundations of Computer Science*, page 579–586, 1999.

V. Goyal and R. Ravi. A PTAS for the chance-constrained knapsack problem with random item sizes. *Operations Research Letters*, 38:161–164, 2010.

A. Gupte, S. Ahmed, M. S. Cheon, and S. Dey. Solving mixed integer bilinear problems using MILP formulations. *SIAM Journal on Optimization*, 23(2):721–744, 2013.

C.Z. Gurgur and T.C. Morley. Lockheed Martin space sytems company optimizes infrastructure project portfolio selection. *Interfaces*, 38(4):251–262, 2008.

F. Hassanzadeh, M. Modarres, H. R. Nemati, and K. Amoako-Gyampah. A robust R&D project portfolio optimization model for pharmaceutical contract research organizations. *International Journal of Production Economics*, 158:18–27, 2014.

K. Heidenberger and C. Stummer. Research and development project selection and resource allocation: a review of quantitative modelling approaches. *International Journal of Management Reviews*, 1(2):197–224, 1999.

A. D. Henriksen and A. J. Traynor. A practical R&D project-selection scoring tool. *IEEE Transactions on Engineering Management*, 46(2):158–170, 1999.

Y. Hong. On computing the distribution function for the Poisson binomial distribution. *Computational Statistics and Data Analysis*, 59:41–51, 2013.

A. Hughes and B. Martin. *Enhancing Impact, The Value of Public Sector R&D*. Leadership for Business and Higher Education and UK-Innovation Research Center, 2012.

A. B. Jaffe. Real effects of academic research. *American Economic Review*, 79(5): 957–970, 1989.

N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions*. New York: Wiley, 2 edition, 1995.

O. Karsu and A. Morton. Incorporating balance concerns in resource allocation decisions: A bi-criteria modelling approach. *Omega*, 44:70–82, 2014.

J. Kleinberg, Y. Rabani, and E. Tardos. Allocating bandwidth for bursty connections. *SIAM J. Comput.*, 30(1):191–217, 2000.

E. Klotz and A. M. Newman. Practical guidelines for solving difficult mixed integer linear programs. *Surveys in Operations Research and Management Science*, 18 (1-2):18–32, 2013.

A. Koç, D. P. Morton, E. Popova, S. M. Hess, E. Kee, and D. Richards. Prioritizing project selection. *Engineering Economist*, 54(4):267–297, 2009.

H. Kroll and T. Stahlecker. Global review of competitive R&D funding- a project commissioned by the World Bank. *Synthesis Report*, 2012.

H. Lee, Y. Park, and H. Choi. Comparative evaluation of performance of national R&Dprograms with heterogeneous objectives: A DEA approach. *European Journal of Operational Research*, 196:847–855, 2009.

I. S. Litvinchev, F. Lopez, A. Alvarez, and E. Fernandez. Large-scale public R&D portfolio selection by maximizing a biobjective impact measure. *IEEE Transactions on Systems Man and Cybernetics Part a-Systems and Humans*, 40(3):572–582, 2010.

J. Luedtke, S. Ahmed, and G. L. Nemhauser. An integer programming approach for linear programs with probabilistic constraints. *Mathematical Programming*, 122 (2):247–272, 2010.

E. Mansfield. Academic research and industrial innovation: An update of empirical findings. *Research Policy*, 26:773–776, 1998.

A Mas-Colell, M. D. Whinston, and J. R. Green. *Microeconomic Theory*. Oxford University Press, New York, 1995.

G. Mavrotas, D. Diakoulaki, and Y. Caloghirou. Project prioritization under policy restrictions. a combination of mcda with 0–1 programming. *European Journal of Operational Research*, 171:296–308, 2006.

G. P. McCormick. Computability of global solutions to factorable nonconvex programs .1. convex underestimating problems. *Mathematical Programming*, 10(2): 147–175, 1976.

L. A. Meade and A. Presley. R&D project selection using the analytic network process. *IEEE Transactions on Engineering Management*, 49(1):59–66, 2002.

A. L. Medaglia, S. B. Graves, and J. L. Ringuest. A multiobjective evolutionary approach for linearly constrained project selection under uncertainty. *European Journal of Operational Research*, 179(3):869–894, 2007.

M. Monaci, U. Pferschy, and P. Serafini. Exact solution of the robust knapsack problem. *Computers & Operations Research*, 40(11):2625–2631, 2013.

A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2006.

Y. Nesterov and A. Nemirovski. Interior-point polynomial algorithms for convex programming. *SIAM, Philadelphia*, 1993.

NIH. National Institutes of Health, Grant Policy Statements. 2013.

NSF. National Science Foundation, Grant Policy Manual. 2005.

G. H. A. Pereira, A. B. Denise, and Mônica C. S. The truncated inflated beta distribution. *Communications in Statistics-Theory and Methods*, 41(5):907–919, 2012.

J. L. Ringuest, S. B. Graves, and R. H. Case. Mean–gini analysis in R&D portfolio selection. *European Journal of Operational Research*, 154(1):157–169, 2004.

A. J. Salter and B. R. Martin. The economic benefits of publicly funded basic research: a critical review. *Research Policy*, 30(3):509–532, 2001.

R. Santhanam and J. Kyparisis. A multiple criteria decision model for information system project selection. *Computers & Operations Research*, 22(8):807 – 818, 1995.

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

A. Shapiro, D. Dentcheva, and A. Ruszczynski. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia, 1 edition, 2009.

I. Shevtsova. On the absolute constants in the berry esseen type inequalities for identically distributed summands. *arXiv:1111.6554*, 2011.

I. G. Shevtsova. An improvement of convergence rate estimates in the lyapunov theorem. *Doklady Mathematics*, 82:862–864, 2010.

S. Solak, J.P. B. Clarke, E. L. Johnson, and E. R. Barnes. Optimization of R&D project portfolios under endogenous uncertainty. *European Journal of Operational Research*, 207(1):420–433, 2010.

T. J. Stewart. A multi-criteria decision support system for R&D project selection. *The Journal of the Operational Research Society*, 42(1):17–26, 1991.

TÜBİTAK. Activity Report. 2012.

J. Wang, W. Xu, J. Ma, and S. Wang. A vague set based decision support approach for evaluating research funding programs. *European Journal of Operational Research*, 230:656–665, 2013.

APPENDIX A

SUPPLEMENTARY MATERIAL FOR CHAPTER 4

A.1 Proof of Proposition 4.2.1

Proof. Since the sequence is truncated, we can write $\mathbb{E}(T_i^k) < u^k < \infty$ for $i = 1, ..n$. $k = 1, 2, 3....$ We can also write, $|T_i - \mathbb{E}(T_i)| \leq u - l$ is true for every $i = 1, ..n.$. Then for each $\delta > 0$, we can determine an upper bound for the term in (4.13) as follows:

$$\frac{\sum_{i=1}^{i=n} \mathbb{E}\left[|T_i - \mathbb{E}(T_i)|^{2+\delta}\right]}{\left[\sqrt{\sum_{i=1}^{i=n} \mathrm{Var}(T_i)}\right]^{2+\delta}} \leq \frac{(u-l)^\delta \sum_{i=1}^{i=n} \mathbb{E}\left[|T_i - \mathbb{E}(T_i)|^2\right]}{\sum_{i=1}^{i=n} \mathrm{Var}(T_i) \left[\sqrt{\sum_{i=1}^{i=n} \mathrm{Var}(T_i)}\right]^\delta} \tag{A.1}$$

We also know that $\mathbb{E}\left[|T_i - \mathbb{E}(T_i)|^2\right] = \mathbb{E}\left[T_i^2 - 2T_i\mathbb{E}(T_i) + (\mathbb{E}(T_i))^2\right] = \mathbb{E}(T_i^2) - [\mathbb{E}(T_i)]^2$, which is the definition of $\mathrm{Var}(T_i)$. Then we can simplify the upper bound in (A.1) as follows:

$$\frac{\sum_{i=1}^{i=n} \mathbb{E}\left[|T_i - \mathbb{E}(T_i)|^{2+\delta}\right]}{\left[\sqrt{\sum_{i=1}^{i=n} \mathrm{Var}(T_i)}\right]^{2+\delta}} \leq \left(\frac{u-l}{\sqrt{\sum_{i=1}^{i=n} \mathrm{Var}(T_i)}}\right)^\delta \tag{A.2}$$

Then, for every $\delta > 0$ right hand side of (A.2) converges to zero as n goes to infinity. Since the upper bound converges to zero, then term in the left hand side of (A.2) converges to zero. Therefore, Lyapunov CLT theorem is satisfied and $\frac{\sum_{i=1}^{i=n}[T_i - \mathbb{E}(T_i)]}{\sqrt{\sum_{i=1}^{i=n} \mathrm{Var}(T_i)}}$ converges to $N(0, 1)$ as n goes to infinity, which completes proof. $\square$

A.2 Proof of Corollary 4.2.1.1

Proof. As defined in equation (4.4), $\hat{b}_i = b_i R_i$. We know that R_i is bounded in the interval (0,1] and we also know that usually R&D programs have a specified budget upper bound (let call it b_{max}) for the applying projects. We can conclude that $\hat{b}_i$ is a truncated random variable in the interval $(0, b_{max}]$ for every i. Hence, any subset of sequence $b_1R_1, b_2R_2, ..., b_iR_i \in N$ that has substantial number of elements satisfies Lyapunov CLT theorem. $\square$

A.3 Proof of Proposition 4.2.2

Before beginning the proof, we first give the following property to facilitate derivations.

Property A.3.1. *Let W be a truncated beta distribution as defined in Definition 4.2.2. Then define a transformed random variable T such that*

$$T = \frac{W - a}{b - a} \tag{A.3}$$

where T is the standard beta distribution in open interval $(0, 1)$. This property facilitate the derivation of the mean, variance and n^{th} moment of the random variable W, since the moments of standard beta distribution are readily available (see Johnson et al. (1995),Chapter 25). Note that n^{th} moment is derived for using in the Berry-Esseen theorem.

Proof. By using Property A.3.1, we can write that $W = a + (b - a)T$. We know the mean, variance and n^{th} moment of standard random variable T (see Johnson et al. (1995),Chapter 25). Therefore, we can obtain the mean, variance and n^{th} moment of truncated beta random variable W by using expectation or variance operator as follows:

$$\mathbb{E}(W) = \mathbb{E}(a + (b - a)T) = a + (b - a)\mathbb{E}(T) = a + \frac{(b - a)\alpha}{\alpha + \beta} = \frac{a\alpha + a\beta + b\alpha - a\alpha}{\alpha + \beta}$$
$$= \frac{\alpha b + \beta a}{\alpha + \beta}$$

$$
\begin{aligned}
Var(W) &= Var(a+(b-a)T) = Var((b-a)T) = (b-a)^2 Var(T) \\
&= \frac{(b-a)^2\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}
\end{aligned}
$$

for $0 < a < b$

$$
\begin{aligned}
\mathbb{E}(W^n) &= \mathbb{E}((a+(b-a)T)^n) = \mathbb{E}\left(\sum_{k=0}^{k=n}\binom{n}{k}a^k(b-a)^{n-k}T^{n-k}\right) \\
&= \sum_{k=0}^{k=n}\frac{n!}{k!(n-k)!}a^k(b-a)^{n-k}\mathbb{E}(T^{n-k}) \\
&= \sum_{k=0}^{k=n}\frac{n!}{k!(n-k)!}a^k(b-a)^{n-k}\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+n-k)}{\Gamma(\alpha)\Gamma(\alpha+\beta+n-k)}
\end{aligned}
$$

for $a = 0$ and $b > 0$

$$
\mathbb{E}(W^n) = \mathbb{E}((a+(b-a)T)^n) = \mathbb{E}((bT)^n) = b^n\mathbb{E}(T^n) = b^n\frac{\Gamma(\alpha+\beta)\Gamma(\alpha+n)}{\Gamma(\alpha)\Gamma(\alpha+\beta+n)}
$$

□

A.4 Proof of Proposition 4.2.3

Proof. By using equation (4.16), we can write $\mathbb{E}(R_i) = \sum_{j=1}^{j=3} w_j m_j^1$ where m_1^1 is the mean (first moment) of the truncated distribution in open interval $(0, \tau_1)$, m_2^1 is the mean of the truncated distribution in open interval $(\tau_2, 1)$, m_3^1 is the mean of the degenerate distribution at point 1, by using equation in (4.17), we can obtain first moments (means) so that $m_1^1 = \frac{\alpha_1\tau_1}{\alpha_1+\beta_1}$ and $m_2^1 = \frac{\alpha_2+\beta_2\tau_2}{\alpha_2+\beta_2}$, since mean of the constant value is itself, then $m_3^1 = 1$. Apparently, $w_1 = p_i$, $w_2 = q_i$, and $w_3 = (1 - p_i - q_i)$. Therefore, we obtain: $\mathbb{E}(R_i) = p_i\frac{\alpha_1\tau_1}{\alpha_1+\beta_1} + q_i\frac{\alpha_2+\beta_2\tau_2}{\alpha_2+\beta_2} + (1 - p_i - q_i)$. We know $Var(R_i) = \mathbb{E}(R_i^2) - [\mathbb{E}(R_i)]^2$. By using equation (4.16), we can write $\mathbb{E}(R_i^2) = \sum_{j=1}^{j=3} w_j m_j^2$ where m_1^2 is the second moment of the truncated distribution in open interval $(0, \tau_1)$, m_2^2 is the second moment of the truncated distribution in open interval $(\tau_2, 1)$, m_3^2 is the second moment of the degenerate distribution at point 1. Let W_1 is the truncated beta random variable in interval $(0, \tau_1)$ and W_2 is the truncated beta random variable in interval $(\tau_2, 1)$. We know that $m_k^2 = \mathbb{E}(W_k^2) =$

$[\mathbb{E}(W_k)]^2 + Var(W_k)$ for $k = 1, 2$ and $m_3^2 = 1$. Then we can write:

$$\begin{aligned} Var(R_i) &= \mathbb{E}(R_i^2) - [\mathbb{E}(R_i)]^2 = p_i \mathbb{E}(W_1^2) + q_i \mathbb{E}(W_2^2) + (1 - p_i - q_i) - [\mathbb{E}(R_i)]^2 \\ &= p_i \left([\mathbb{E}(W_1)]^2 + Var(W_1) \right) + q_i \left([\mathbb{E}(W_2)]^2 + Var(W_2) \right) + (1 - p_i - q_i) - [\mathbb{E}(R_i)]^2 \end{aligned} \tag{A.4}$$

Hence, by using equations in (4.17), (4.18) and (4.20), we derive:

$$\begin{aligned} Var(R_i) = p_i &\left[\left(\frac{\alpha_1 \tau_1}{\alpha_1 + \beta_1} \right)^2 + \frac{\tau_1^2 \alpha_1 \beta_1}{(\alpha_1 + \beta_1)^2 (\alpha_1 + \beta_1 + 1)} \right] \\ &+ q_i \left[\left(\frac{\alpha_2 + \beta_2 \tau_2}{\alpha_2 + \beta_2} \right)^2 + \frac{(1 - \tau_2)^2 \alpha_2 \beta_2}{(\alpha_2 + \beta_2)^2 (\alpha_2 + \beta_2 + 1)} \right] + (1 - p_i - q_i) \\ &- \left[p_i \frac{\alpha_1 \tau_1}{\alpha_1 + \beta_1} + q_i \frac{\alpha_2 + \beta_2 \tau_2}{\alpha_2 + \beta_2} + (1 - p_i - q_i) \right]^2 \end{aligned}$$

By directly applying equations in (4.16) and (4.19) together, we derive:

$$\begin{aligned} \mathbb{E}(R_i^n) = p_i &\left(\tau_1^n \frac{\Gamma(\alpha_1 + \beta_1)\Gamma(\alpha_1 + n)}{\Gamma(\alpha_1)\Gamma(\alpha_1 + \beta_1 + n)} \right) + (1 - p_i - q_i) \\ &+ q_i \left(\sum_{k=0}^{k=n} \frac{n!}{k!(n-k)!} \tau_2^k (1 - \tau_2)^{n-k} \frac{\Gamma(\alpha_2 + \beta_2)\Gamma(\alpha_2 + n - k)}{\Gamma(\alpha_2)\Gamma(\alpha_2 + \beta_2 + n - k)} \right) \end{aligned}$$

□

A.5 Proof of Corollary 4.3.1.1

Proof. Define $Y_i = \hat{b}_i - \mathbb{E}(\hat{b}_i)$. Then,

$$\mathbb{E}(Y_i) = 0 \tag{A.5}$$

$$\begin{aligned} \mathbb{E}(Y_i^2) &= \mathbb{E}\left([\hat{b}_i - \mathbb{E}(\hat{b}_i)]^2 \right) = \mathbb{E}\left[\hat{b}_i^2 - 2\hat{b}_i \mathbb{E}(\hat{b}_i) + (\mathbb{E}(\hat{b}_i))^2 \right] = \mathbb{E}(\hat{b}_i^2) - [\mathbb{E}(\hat{b}_i)]^2 \\ &= \text{Var}(\hat{b}_i) = b_i^2 \text{Var}(R_i) \end{aligned} \tag{A.6}$$

$$\begin{aligned} \mathbb{E}(|Y_i^3|) &= \mathbb{E}\left(\left| [\hat{b}_i - \mathbb{E}(\hat{b}_i)]^3 \right| \right) = \mathbb{E}\left[\left| \hat{b}_i^3 - 3\hat{b}_i^2 \mathbb{E}(\hat{b}_i) + 3\hat{b}_i [\mathbb{E}(\hat{b}_i)]^2 - [\mathbb{E}(\hat{b}_i)]^3 \right| \right] \\ &= \left| \mathbb{E}(\hat{b}_i^3) - 3\mathbb{E}(\hat{b}_i^2)\mathbb{E}(\hat{b}_i) + 2[\mathbb{E}(\hat{b}_i)]^3 \right| = \left| b_i^3 \mathbb{E}(R_i^3) - 3b_i^3 \mathbb{E}(R_i^2)\mathbb{E}(R_i) + 2b_i^3 [\mathbb{E}(R_i)]^3 \right| \end{aligned} \tag{A.7}$$

By using equation (4.22) and property of gamma function such that $\Gamma(t+1) = t\Gamma(t)$, we derive:

$$
\begin{aligned}
\mathbb{E}(R_i^3) &= p_i \left(\tau_1^3 \frac{\Gamma(\alpha_1+\beta_1)\Gamma(\alpha_1+3)}{\Gamma(\alpha_1)\Gamma(\alpha_1+\beta_1+3)} \right) + (1-p_i-q_i) \\
&+ q_i \left(\sum_{k=0}^{k=3} \frac{3!}{k!(3-k)!} \tau_2^k (1-\tau_2)^{3-k} \frac{\Gamma(\alpha_2+\beta_2)\Gamma(\alpha_2+3-k)}{\Gamma(\alpha_2)\Gamma(\alpha_2+\beta_2+3-k)} \right) \qquad \text{(A.8)} \\
&= p_i \left(\tau_1^3 \frac{(\alpha_1+2)(\alpha_1+1)\alpha_1}{(\alpha_1+\beta_1+2)(\alpha_1+\beta_1+1)(\alpha_1+\beta_1)} \right) + (1-p_i-q_i) \\
&+ q_i(1-\tau_2)^3 \frac{(\alpha_2+2)(\alpha_2+1)\alpha_2}{(\alpha_2+\beta_2+2)(\alpha_2+\beta_2+1)(\alpha_2+\beta_2)} \\
&+ 3q_i\tau_2(1-\tau_2)^2 \frac{(\alpha_2+1)\alpha_2}{(\alpha_2+\beta_2+1)(\alpha_2+\beta_2)} \\
&+ 3q_i\tau_2^2(1-\tau_2) \frac{\alpha_2}{(\alpha_2+\beta_2)} + q_i\tau_2^3 \qquad \text{(A.9)}
\end{aligned}
$$

By using equations (4.32), (4.33) and Remark 4.3.1, we can obtain inequality (4.34). This completes the proof. □

A.6 Proof of Proposition 4.5.1

Proof. Let P_1 be the feasible polyhedron of true model SP-1, and P_2 be the feasible polyhedron of modified model in which mean and variance of $\bar{R}_i$ are used. To show $z_2^* \geq z_1^*$ always holds, we have to prove $P_1 \subseteq P_2$. For a sufficiently small ϵ_1 and a given θ value, pick any feasible solution vector x in P_1 such that x satisfies following equation:

$$\sum_{i \in N} b_i \mathbb{E}(R_i) x_i + \Phi^{-1}(\theta) \sqrt{\sum_{i \in N} b_i^2 \mathrm{Var}(R_i) x_i^2} = B - \epsilon_1 \qquad \text{(A.10)}$$

which means that picked solution vector x is on the edge of polyhedron P_1. This solution vector x always satisfies following inequality

$$\sum_{i \in N} b_i \mathbb{E}(\bar{R}_i) x_i + \Phi^{-1}(\theta) \sqrt{\sum_{i \in N} b_i^2 \mathrm{Var}(\bar{R}_i) x_i^2} \leq B \qquad \text{(A.11)}$$

Since $\mathbb{E}(R_i) = \mathbb{E}(\bar{R}_i)\ \forall\ i$ due to equations (4.20) and (4.41)

and $Var(R_i) - Var(\bar{R}_i) = p_i Var(W_1) + qVar(W_2)\ \forall\ i$ due to equations (A.4) and (4.42). Hence, solution vector x always will be in P_2. Similarly, for a sufficiently small ϵ_2 and a given θ value pick any solution vector y in P_2 such that y satisfies following equation:

$$\sum_{i \in N} b_i \mathbb{E}(\bar{R}_i) y_i + \Phi^{-1}(\theta) \sqrt{\sum_{i \in N} b_i^2 \text{Var}(\bar{R}_i) y_i^2} = B - \epsilon_2 \tag{A.12}$$

which means that picked solution vector y is on the edge of polyhedron P_2. This solution vector y always will not be in P_1 because

$$\sum_{i \in N} b_i \mathbb{E}(R_i) y_i + \Phi^{-1}(\theta) \sqrt{\sum_{i \in N} b_i^2 \text{Var}(R_i) y_i^2} > B \tag{A.13}$$

holds due to $Var(R_i) > Var(\bar{R}_i)$ for $p_i > 0$ and $q > 0$. Therefore, $P_1 \subseteq P_2$ is valid and $z_2^* \geq z_1^*$ holds. $\square$

APPENDIX B

SUPPLEMENTARY MATERIAL FOR CHAPTER 5

B.1 Derivation of POB(α) bound

We need to introduce some notation for applying bound formula of Bertsimas et al. (2012) to our problem setting. Let M is set of sectors and let $X \subset R^{|M|}$ is a set of all feasible allocations of funding budget to $|M|$ sectors. Let $x \in X$ represents any feasible allocation. Specific contribution of sector j to objective O_1 can be associated with a utility function $f_j : X \to R_+$. Then utility set U can be expressed as:

$U = \{u \in R_+^{|M|} \mid \exists x \in X : f_j(x) = u_j, \forall j \in M\}$ and let $u_j^* = sup\{u_j \mid u \in U\}$

For a compact and convex set U, Bertsimas et al. (2012) show:

$$\text{POB}(\alpha) \leq 1 - \min_{y \in [1,|M|]} \frac{R^{1/\alpha} y^{(\alpha+1/\alpha)} + |M| - y}{R^{1/\alpha} y^{(\alpha+1/\alpha)}(|M| - y) R y} \tag{B.1}$$

where $R = \dfrac{\max_{j \in M} u_j^*}{\min_{j \in M} u_j^*}$ and $\max_{j \in M} u_j^* \geq \min_{j \in M} u_j^* > 0$.

They also show that bounds are strong and near-tight as $R \to 1$. POB(α) increases as R and |M| increase. Thus, we can obtain bounds on price of balancing of any problem instance, by using impact assessment values and number of sectors.

In the following corollary, we show that R value depends only impact assessment parameters of given problem instance.

Corollary B.1.0.2. *For our problem setting,* $R = \dfrac{\max_{j \in M} e_j}{\min_{j \in M} e_j}$.

Proof. Specific contribution of sector j to objective O_1 can be written as $f_j(x) = u_j = e_j B_j$. Consider a feasible allocation vector x^j where sector j receives B amount of budget, and all other remaining sectors receives zero. That implies $u_j^* = sup\{e_j B_j \mid u \in U\} = e_j B$. Then,

$$R = \frac{\max\limits_{j \in M} u_j^*}{\min\limits_{j \in M} u_j^*} = \frac{B \max\limits_{j \in M} e_j}{B \min\limits_{j \in M} e_j} = \frac{\max\limits_{j \in M} e_j}{\min\limits_{j \in M} e_j}. \qquad \square$$

B.2 Data from 2012 Activity Report of TÜBİTAK

In Table B.1, number of proposal represents number of project applications that apply to two calls of TÜBİTAK 1001 program during 2012. Number of supported represents total number of supported projects in those two calls. Amount of funding indicates total approved committed funding budget to supported projects in 1,000 Turkish Liras. By using those figures, in Table B.1, we calculate approximated average budget (in 10,000 Turkish Liras) by dividing amount of funding to number of supported projects and use this as a proxy for generation of sectoral budgets. In Table B.1, we also calculate number of proposals in percent units in each sector, because we use it as a proxy for generation of number of project proposals in each sector for a given total number of applying projects in our example problem. We also find success rate as 0.17 by dividing total number of supported to total number of proposals and use this as a proxy for budget fraction in our example problem.

Table B.1: Data derived from TÜBİTAK (2012)

TÜBİTAK 1001 program Scientific disciplines	Number of Proposals	Number of Supported	Amount of Funding	Approximated Average Budget	Number of Proposals (%)
Environment, Atmosphere, Earth and Marine Sciences	240	61	17.486	28.5	7.46%
Electrical, Electronics and Informatics	230	54	12.197	22.5	7.15%
Engineering	562	86	18.795	21.5	17.46%
Health Sciences	413	55	16.315	29.5	12.83%
Social Sciences and Humanities	491	68	9.686	14.5	15.26%
Basic Sciences	756	133	31.137	23.5	23.49%
Agriculture, Forestry and Veterinary	526	78	19.304	25	16.35%
Total	3218	535	124.92		100%

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Çağlar, Musa
Nationality: Turkish (TC)
Date and Place of Birth: 06.09.1982, Denizli
Marital Status: Married
Phone: +1 (518) 596 48 80
Email: musacaglar@gmail.com

EDUCATION

Degree	Institution	Year
M.S.	METU / Graduate School of Natural and Applied Sciences / Industrial Engineering	2009
B.S.	ITU / Faculty of Management / Industrial Engineering	2005
High School	Denizli Türk Eğitim Vakfı (TEV) Anatolian High School	2001

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
Aug, 2015 - Aug, 2016	Dept. of STI Policy / TÜBİTAK	Unpaid Vacation
Feb, 2011 - Aug, 2015	Dept. of STI Policy / TÜBİTAK	Expert
Apr 2009 - Feb, 2011	Dept. of STI Policy / TÜBİTAK	Assist. Expert
Mar, 2006 - Apr, 2009	Dept. of Strategic Management / TÜBİTAK	Assist. Expert
Jun, 2004 - July, 2004	Headquarters /Arcelik	Intern
Jul, 2003 - Aug, 2003	Dept. of Manufacturing / Assan Aluminyum	Intern