A CLASS OF TWO-STAGE STOCHASTIC QUADRATIC PROGRAMS WITH APPLICATIONS IN OIL MARKET

YUN SHI

PhD

The Hong Kong Polytechnic University

2019

The Hong Kong Polytechnic University

Department of Applied Mathematics

A Class of Two-stage Stochastic Quadratic Programs with Applications in Oil Market

Yun Shi

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

Mar 2019

Certificate of Originality

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____(Signature)

Yun Shi (Name of student)

Dedicated to my family.

Abstract

In this thesis, two applications in mathematical finance and economics are modelled by a class of two-stage stochastic programs. The focuses are on the modeling of the practical problems as well as their real life implementations with numerical implementations. The first application on portfolio selection is approached by the construction of a relaxed second-order stochastic dominance (SSD) constrained stochastic optimization problem. The second model concerns the equilibrium in the world oil market share, which is treated as a solution of a two-stage stochastic Nash equilibrium problem. General framework of a class of two-stage convex stochastic optimization problems is analyzed in details, with special attentions paid to the constructions of the Lagrangian together with its saddle-point characterization of optimality.

The problem of portfolio selection aims to construct optimal assets allocation strategies subject to constraints on risk management. More specifically, a portfolio optimization model with relaxed SSD constraints is proposed and solved, and its solution is the portfolio of choice. The proposed model uses Conditional Value at Risk (CVaR) constraints at probability level $\beta \in (0, 1)$ to relax SSD constraints. The relaxation is justified by theoretical convergence results based on sample average approximation (SAA) method when sample size $N \to \infty$ and CVaR probability

level β tends to 1. SAA method is used to reduce infinite number of inequalities of SSD constraints to finite ones and also to calculate the expectation value. The proposed relaxation on the SSD constraints in portfolio optimization problem is achieved when the probability level β of CVaR takes value less than but close to 1, and the model can be readily solved by cutting plane method. The performance and characteristics of the constructed portfolios are tested empirically on three sets of real market data, and the results obtained from the numerical experiments are analyzed and discussed in details. It is shown that with appropriate choices of CVaR probability level β, the constructed CVaR-SSD portfolios are sparse and outperform both the benchmark portfolios and the portfolios constructed by solving the portfolio optimization problems with SSD constraints.

The second application is an attempt to provide explanation and mechanism about the stable patterns observed in market share of world's crude oil trading over the last several decades via a two-stage stochastic model of Nash equilibrium for Cournot competition. To summarize, a convex two-stage non-cooperative multi-agent equilibrium problem under uncertainty is formulated as a two-stage stochastic variational inequality (SVI). Under standard assumptions, sufficient conditions for the existence of solutions of the two-stage SVI are provided. A regularized SAA method is proposed to solve it. The convergence of the method is proved as the regularization parameter tends to zero and the sample size tends to infinity. In order to explain the oil market share observation, a two-stage stochastic production and supply planning problem with homogeneous commodity in an oligopolistic market is constructed under the framework of two-stage SVI. Numerical experiments are performed based on historical data. The data are used in-sample to aid the selection

of parameters as well as the modification of the game model structure. Out-of-sample tests are presented to demonstrate the effectiveness of the proposed model in its ability for describing the market share of oil producing agents.

Publications Arising from the Thesis

- M. Xue, Y. Shi and H. Sun. Portfolio Optimization with Relaxation of Stochastic Second Order Dominance Constraints via Conditional Value at Risk. *Journal of Industrial and Management Optimization*, accepted, 2019.

- J. Jiang, Y. Shi, X.Wang and X. Chen. Regularized Two-stage Stochastic Variational Inequalities for Cournot-Nash Equilibrium under Uncertainty. Invited submission to *Journal of Computational Mathematics*.

Acknowledgements

First and foremost, I would like to express my thanks to my chief supervisor Professor Xiaojun Chen. Originated from the discipline of theoretical physics, it is not until later stage of my PhD study that I became aware of the risk she took for accepting me to pursue the degree on applied mathematics under her supervision. I also have to admit that I was not completely prepared, at the early stage of my study, to confront the challenges brought up by the switch of academic areas nor the depth of thinking required for this journey of knowledge pursuit. It was the patience and the ever lasting guidance of Professor Xiaojun Chen, which helped me along this precious experience of life. Her allowance for trying out different ideas and encouragement to think much more deeply into the problems have shaped my views in ways otherwise would never have occurred. Both in academic and personal lives, it is my greatest honour to be able to call myself a graduate student of her.

I am greatly grateful to join the research group of Professor Xiaojun Chen and my thanks go to all my academic brothers and sisters: Professor Congpei An, Professor Wei Bian, Professor Xin Liu, Professor Yafeng Liu, Professor Hailin Sun, Professor Shulin Wu, Professor Zaikun Zhang, Professor Xile Zhao, Dr. Yanfang Zhang, Dr. Yang Zhou, Dr. Hong Wang, Dr. Tianxiang Liu, Dr. Qiyu Wang, Dr. Bo Wen, Dr. Dingtao Peng, Dr. Lin Chen, Dr. Liaoyuan Zeng, Dr. Ning Zhang, Ms. Lili Pan,

Mr. Andy Yat Ming Cheung, Mr. Guidong Liu, Mr. Jie Jiang, Ms. Fang He, Miss. Chao Li, Mr. Zhenhua Peng, Mr. Xiaozhou Wang, Mr. Jianfeng Luo and Mr. Rui Zhang for their inspiring discussions and substantial supports. I also consider myself very lucky to be able to jointly study with Professor Defeng Sun along with members of his research group, Dr. Liang Chen, Dr. Xiaoliang Song and Miss Ying Sun. His safari on vast range of problems with distinguished angles always surprise me and there are some many things one can learn from a deep thinker like him. I truly hope this list of academic family tree of mine can grow even larger with many more fruitful achievements from it.

My special thanks also go to my uncle, Professor Jianming Shi. It was him who introduced to me the research group of Professor Xiaojun Chen's and without his encouragement my study may never start in the very beginning.

I am also grateful to Professor Alexander Shapiro. I was lucky enough to be his companion for attending several summer schools and workshops. It was from the conversation and discussion with him that the field of stochastic optimization becomes my interests of research. His sharing of personal experience and knowledge also helped me in the makings of many life-changing decisions. Specially, it was him who encouraged me to focus on problems that were not only academically attractive but also practically applicable. My gratitude has to go to Professor Hailin Sun, who helped me through out my course of study from so many aspects. I would also like to thank Dr. Guofeng Zhang, who was to be my co-supervisor but did not happen due to my change in research field. I would also like to thank Dr. Zhonghua Qiao, who lectured me not only on academic courses but also on philosophies of life. I would also like to express my gratitude to Professor Xingqiu Zhao for acting as the chair

of my oral defense committee and thank Professor Li-Zhi Liao and Professor Akiko Tekeda for their kind acceptance for joining the board of my external examiners.

I would also like to thank all the supporting staffs and technicians in the department of applied mathematics for their supports through out my study. I am also deeply honoured for the trust of all members of the SIAM student chapter at PolyU who elected me to serve as the president of this wonderful student organization from 2017-2019.

My deepest gratitude and love goes to my parents, Jin Ji and Weiming Shi. Their understanding and support always provide me with endless courage and confidence. I have to apology for the loneliness they have to endure caused my absence during my pursuing of dreams over the years. I will ever devote myself to live up to their expectations and much beyond.

Last but not least, I would like to thank my wife Fei Fang for her company, understanding and support throughout my PhD study. It really is a once-a-lifetime experience to have her, also pursuing her PhD degree, along side (literally) with me in this wonderful journey of exploration.

Contents

List of Figures

List of Tables

List of Notations

$\mathbb{R}$	The set of real numbers
$\mathbb{R}^m$ ($\mathbb{R}^m_+$)	The set of m-dimensional real vectors (with nonnegative entries)
$\mathbb{R}^{m\times n}$ ($\mathbb{R}^{m\times n}_+$)	The set of $m \times n$ real matrices (with nonnegative entries)
$\|\boldsymbol{x}\|$	The Euclidean norm of the vector $\boldsymbol{x}$
$(x_+)_i = \max(0, x_i)$	The plus function for $x \in \mathbb{R}^n$ with $i = 1, ..., n$.
0	Zero matrix, zero vector of corresponding dimensions.
$\boldsymbol{e}$	Vector of 1.
$\boldsymbol{I}_X$	Indicator function of set X.
$\mathfrak{L}_n$	Lebesgue space of $\mathbb{R}^n$-valued functions.
$\boldsymbol{\xi}/\boldsymbol{\xi}(\omega)$	Random vector $\boldsymbol{\xi} : \Omega \to \Xi \subseteq \mathbb{R}^d$. To ease of notation, $\boldsymbol{\xi}$ is used to denote the random vector $\boldsymbol{\xi}(\omega)$.
ξ	Realization of random vector $\boldsymbol{\xi}$.
Ξ	Support set of random vector.
$\mathcal{Y}$	space of measurable function defined on Ξ.

$\mathbb{E}$ — The expectation operator $\mathbb{E}[\boldsymbol{\xi}]$, the Lebesgue integral of the $\mathcal{A}$-measurable function $\boldsymbol{\xi}$ with respect to the measure P, denoted as

$$\int_{\Omega} \xi(\omega) P(d\omega).$$

$\langle \cdot, \cdot \rangle$ — Inner products defined in the Hilbert space of $\mathbb{R}^n \times \mathcal{Y}$ with L_2-norm, i.e.,

$$\langle (x, y), (u, v) \rangle := x^T u + \int_{\Xi} y(\xi)^T v(\xi) P(d\xi),$$

where P denotes the probability distribution of $\boldsymbol{\xi}$.

a.e., a.s. — Abbreviations for almost every and almost surely respectively.

x, u, μ — Tender or offer from first-stage to second-stage.

$y(\boldsymbol{\xi}), v(\boldsymbol{\xi}), \lambda(\boldsymbol{\xi})$ — Second-stage decision.

ψ — Second-stage value function defined on tenders, feasible second-stage decisions with random argument, $\psi(x, \boldsymbol{\xi}, y(\boldsymbol{\xi}))$.

Ψ — Optimal second-stage value function defined on tenders and with random argument,

$$\Psi(x, \boldsymbol{\xi}(\omega)) = \sup(\inf)_{y(\boldsymbol{\xi})} \psi(x, \boldsymbol{\xi}, y(\boldsymbol{\xi})).$$

ϵ — Positive regularization parameter.

$\mathcal{T}_X(x)$ — Bouligand tangent cone to a set at a point $x \in X \subset \mathbb{R}^n$, i.e.,

$$\mathcal{T}_X(x) := \{u \in \mathbb{R}^n : d(x + tu, X) = o(t), t \geq 0\}.$$

$\mathcal{N}_X(x)$	Normal cone defined as the polar of the tangent cone, i.e., $$\mathcal{N}_X(x) := \{\zeta \in \mathbb{R}^n : \zeta^T u \leq 0, \forall u \in \mathcal{T}_X(x)\}.$$

Chapter 1

Introduction

Embracing the challenges brought by the era of Big Data comes the great demand in processing stochastic informatics. Techniques in the field of stochastic optimization are of great demand if one intends to confront real life applications modelled by stochastic models. Before discussing the solution process of the stochastic problems, the construction of the mathematical models may be of even greater challenges. Indeed, researches have seen many practical applications formulated under uncertainty. For example, mathematical models have been widely used to describe observations in finance and economics. Optimal solutions to these models are often characterized by the corresponding multi-stage stochastic variational inequalities (SVIs) and stochastic complementarity problems (SCPs). SVIs and SCPs were studied extensively over the last two decades [15, 17, 40, 70, 71, 103]. In this thesis, two applications are formulated and studied under the framework of a class of two-stage stochastic optimization problems.

The first application under consideration is a popular topic in mathematical finance, namely the *portfolio selection* problem. The adopted approach involves the

construction of the model whose solution concepts are of special properties required or restricted by the practitioners. The goal of portfolio selection intends to provide better assets allocation choices for investors.

The second application explores the mechanism behind the observation on the stability global oil market share via a two-stage stochastic Nash equilibrium problem (SNEP). Needless to say, deterministic non-cooperative game has a long history of being an effective model to describe market behaviour. In order to explore characteristics of real markets, the presence of uncertainty is considered by the model in this thesis. Note that the stochasticity plays vital roles not only in problem modelling but also in proceedings of historical data.

1.1 Expected return maximization with relaxed second-order stochastic dominance constraints

As per described, the problem of portfolio selection, one of the most popular and widely discussed research topics in mathematical finance, is studied first. By means of modelling and solving of a portfolio selection problem, investors seek to construct one appropriate bundle of assets, the so-called *portfolio*, often based on solutions of the proposed problem. Although the requirements and/or restrictions are usually user dependent, there exist many general rules that guide the construction of the optimization model. For example, investors often evaluate the performance of a portfolio by its anticipated profitability subject to the satisfaction of costumed constraints. On the other hand, to specific investors, e.g., insurance companies, pension funds, the minimization of risk or anticipated loss is of much greater importance than pursuing high returns. A rule of thumb is that, for any general risk-averse investors,

the appropriately chosen problem constructions should be those with their formulation set to balance the expected return and the risk simultaneously. In addition, the modelling of such problems shall reflect distinctive characteristics of the targeted market of assets, financial selection rules or objectives specified by the investors, etc.

It is no surprise then that the so-called mean-risk model is one of the most widely adopted types of portfolio selection problems. Rooting from the mean-variance (MV) rule proposed by Markowitz [58], its modifications are among the mostly recognized classes of mean-risk models in financial mathematics research community. To list a few, there are mean semi-variance model [59], mean-absolute deviation model [55], mean-absolute deviation-skewness model [56]. While the expected returns over a given length of investment horizon is widely adopted as the gauge for profitability, numerous risk measures besides variance have been proposed to evaluate and quantify the risks or uncertainties other than variances. For example, Value-at-Risk (VaR), which indicates the maximum possible loss when certain proportion of the left tail loss distribution is ignored, is a conventional adaptation among banks and investment firms alike [48]. The Conditional-Value-at-Risk (CVaR) [1, 78] is also widely used, which is more sensitive to the shape of the tail regarding the loss distribution.

For the first application in this thesis, a stochastic optimization problem with a relaxed second-order stochastic dominance (SSD) constraint is proposed and shown to be an effective way for portfolio selection. SSD is a fundamental concept in decision theory and economics, which has seen many uses in portfolio selection problems as the choice of constraint controlling the risks, see Mosler and Scarsini [61], Whitmore and Findlay [99]. The relation between mean-semi-deviation model and SSD was analyzed by Ogryczak and Ruszczyński [65] for the first time. The consistency

of SSD with several other risk measures was demonstrated further in [66] under the context of the mean-risk model. Stochastic programming problem with stochastic dominance constraints was proposed by Dentcheva and Ruszczyński and they studied optimality and duality along with other aspects of both the problem and its solution structure, see [22–24]. As will be shown in later parts in the thesis, the computation of expectation is among several major difficulties when dealing with stochastic programming problems. To this end, the convergence analysis of stochastic optimization problem with SSD constraints under the sample average approximation (SAA) has been studied in several papers [43, 57, 102], along with many efficient methods proposed for solving it. For example, a smoothing penalized SAA method was proposed by Sun et. al. [100] to solve this problem. Homem-de-Mello and Mehrota [41] proposed a sample average cutting surface algorithm for optimization problems with multidimensional polyhedral linear SSD constraints. Rudolf and Ruszczyński [89], Fábián et. al. [28] and Sun et. al. [101] proposed cutting plane methods for solving stochastic programs with SSD constraints.

In the regime of portfolio selection, adoption of stochastic dominance constrained optimization was firstly introduced by Dentcheva and Ruszczyński [25, 27]. Roman et. al. [86] studied multi-objective portfolio selection models with SSD constraints and target return distribution. The notion of stochastic dominance can be intuitively understood as a restriction on the feasible set of portfolios "better" than the underling or pre-specified *benchmark*. It follows that the reason behind the usage of stochastic dominance constraints is to restrict a feasible set of solutions, stochastically dominate that of the benchmark. Under the concept of portfolio selection, any portfolio in the feasible set dominates the benchmark portfolio. It is worth emphasising that

the dominance is viewed only from the perspective of risk control. In a mean-risk model, the solution process is to find the portfolio strategy whose distribution of the random returns dominates that of a benchmark portfolio strategy. For example, when the benchmark portfolio is a financial index, the corresponding portfolio selection problem becomes an enhanced indexation model. An enhanced indexation portfolio manages to track and outperform a stock market index with limited number of stocks, which is desired from a portfolio management prospective [104]. An enhanced indexation model based on SSD was proposed by Roman et. al. [87] and produced consistently good returns, sparsity and with minimal rebalancing required in the constructed portfolios.

Viewed as one motivation of the constraint relaxation in the proposed model, recall that the SSD constraints are *equivalent* to a continuum of CVaR constraints for all probability level $\beta \in (0, 1]$, see [25]. The equivalent result was used by Dentcheva and Ruszczyński [26], who proposed an equivalent stochastic optimization problem with inverse dominance constraint and designed an inverse cutting plane methods for solving it. The proposed portfolio selection problem in this thesis cooperates the SAA method and CVaR approach. More specifically, the SAA method is used to overcome the complicated calculations of expected value as well as to handle the infinite number of inequalities resulted from SSD constraints, followed by the CVaR approximation which later leads to possible relaxation. Recall the definition on equivalence result between SSD and CVaR, the relaxation model is constructed when the CVaR reformulation of SSD is used with *some* probability level $\beta < 1$. In later part of this thesis, the proposed model is referred as *SAA-based CVaR-SSD relaxation model*. Also note that SAA-based CVaR-SSD relaxation model is

related to aforementioned stochastic optimization problem with inverse dominance constraint studied in [26]. The key differences are that

1. the stochastic optimization problem with inverse dominance constrain reformulation is based on the equivalence relation between SSD constraints and a continuum of CVaR constraints;

2. the proposed SAA-based CVaR-SSD relaxation model is a relaxation model based on SAA scheme and CVaR relaxation of a max function proposed in [2] with specific choices of probability level $\beta < 1$.

For a particular choice of probability level, the equivalence result between SSD and CVaR no longer holds and the dominance constraints are relaxed. The idea behind the modelling is that by solving the relaxed model, a less conservative portfolio shall be obtained from an enlarged feasible set. For ease of explanation and illustration, Example 4.1 will be shown in Chapter 4. The performance of the portfolio constructed by solving the relaxed model was put to test under extensive empirical studies.

1.2 Stable global oil market share via a two-stage stochastic variational inequality

It is long known that many economical phenomena may be casted as a solution of stochastic optimization problem. Researchers have studied various real markets through SNEPs in recent years. Jofré, Rockafellar and Wets [47] investigated various economic equilibria using stochastic variational inequalities. A scenario-based multistage oligopolistic market equilibrium problem under uncertainty was discussed by

the authors of [33]. A two-settlement oligopolistic equilibrium with uncertainty in the future market was presented in [106]. For practical applications, electricity markets with hydro-electric distribution have been studied by Philpott, Ferris and Wets [72] in which the levels of water reserves were modelled under uncertainty. The contemporary treatment of classical equilibrium problems are investigated through finite-dimensional variational inequalities (VIs) and complementarity problems (CPs) with a wide range of applications under the assumption of deterministic and single-stage decision, see [29, 39] and references therein. Chen and Fukushima in [11] considered the stochastic linear CP by expected residual minimization (ERM) procedure. The quasi-Monte Carlo method was adopted to generate scenarios of observation and thus to obtain the discrete approximation problem. Chen, Wets and Zhang [17] investigated SVI problems by the ERM procedure, and the SAA method was employed to approximate the expected smoothing residual function. More recently, an extension from single-stage to multi-stage decision processes was discussed by Rockafellar and Wets [85]. In [12], Chen, Pang and Wets introduced the ERM procedure for two-stage SVI problems and the Douglas-Rachford splitting method was used to demonstrate numerical results. Chen, Sun and Xu [15] considered a two-stage stochastic linear CP. Structural properties of the problem were studied under the assumption of strong monotonicity, and a discrete scheme was conducted by partition of the support set and the corresponding convergence assertion was established. More generally, Chen, Shapiro and Sun [13] investigated the SAA of two-stage stochastic nonlinear generalized equations, which included two-stage SVI problems as special cases. Exponential rate of convergence was derived by using the technique of perturbed partial linearization.

From the perspective of the computational implementation, the equilibria can usually be rewritten as solutions of the corresponding minimization problems, mostly nonsmooth. For this class of problems, the smoothing techniques can be employed so that differentiable methods, e.g., Newton's method, become applicable in solving the smoothing problem, see for instance [9, 10, 16, 18, 19]. Moreover, Rockafellar and Sun extended the well-known progressive hedging method (PHM) for multi-stage stochastic programming problems to multi-stage SVIs in [77].

The second application of interests concerns about one observation of world oil market. More specifically, the stability of world oil market share is studied and explained as a result of market equilibrium of a two-stage non-cooperative multi-agent game under uncertainty. The game is used to describe a multi-agent homogenous commodity production and supply planning problem in an oligopolistic market. The focus is of a J-player SNEP of Cournot competition, whose solution concept is characterized by stochastic Cournot-Nash (C-N) equilibria. The decision vectors of production and supply plan problems, so-called *strategies*, are distinguished into two categories,

1. those of *"here-and-now"* type, which do not depend on outcomes of random events in the future, and

2. those treated as responses, occur at a *"wait-and-see"* stage.

The goal of the second application of the thesis is to establish a model which describes the market mechanism up to its ability of reproducing stable oil market share. It will be shown that the real observation on market share can be reproduced when solution concepts of the model assemble the strategies adopted by participants of the market

to achieve C-N equilibrium.

1.3 Overview of the thesis

The main objective of this thesis is to propose and solve two models arise in mathematical finance, portfolio selection problem, and economics, oil market share description, respectively. Both problems can be viewed as from a class of two-stage stochastic programs. In order to study the proposed models, the techniques in analyzing and solving the two-stage stochastic optimization problem are firstly studied and developed to the specifics of proposed models. Starting with detailed descriptions of the constructions of individual problems, along with appropriate assumptions, the models aim to reflect distinguished features of practicality. Theoretical properties of the models are studied as well as the properties of the corresponding solutions. The solution concepts along with their structures are captured, in part describing the insights about the corresponding problems. The theoretical results later lead to developments of effective solution methodology of the stochastic models. Extensive numerical experiments are performed on both applications. In the case of portfolio selection, the portfolios obtained by solving the proposed model lead to effective and better assets allocation. For the application in crude oil market, the stable market share can be reproduced by obtaining the equilibrium of the proposed game. The further research directions along with potential collaborations across other disciplines are discussed in final part of the thesis.

1.4 Contribution of the thesis

The properties of a class of two-stage stochastic optimization problems are studied. In particular, different expressions of this class of two-stage stochastic optimization problems are compared from the aspects of corresponding necessary and/or sufficient optimality conditions. The focuses are then moved to conduct explorations on practical applications.

In the first application of portfolio selection problem,

- A relaxation model of portfolio optimization problem with SSD constraints is proposed based on the SAA method and CVaR approximation, where the relaxation is achieved with a specific choice of probability level $\beta < 1$. This is motivated by the observation on the conservative performance of portfolios constructed by solving the SSD constrained portfolio optimization problems, and solving the relaxation model may lead to portfolios with better returns under adequate risk control. This is called the SAA-based CVaR-SSD relaxation model, whose solution is called the *CVaR-SSD portfolio*.

- The convergence analysis of the SAA-based CVaR-SSD relaxation model to the original SSD constrained problem is presented based on theoretical results of stochastic optimization problem with SSD constraints [102] with the re-proposed conditions in convex cases and CVaR approximation [2]. The convergence is also demonstrated empirically when the CVaR probability level β approaches 1 and the SAA sample size goes to infinity.

- The advantages of CVaR-SSD portfolios are demonstrated empirically based

on three sets of historical data. The performance of the constructed portfolios with probability level $\beta < 1$ is comparable to that of the SSD portfolio. It is observed in the numerical experiments that, with appropriate choices of β, the performance of CVaR-SSD portfolio is always better than the SSD portfolio using the same benchmark portfolio. More specifically, empirical studies are carried out with the benchmark portfolios being both the index portfolio and the MV portfolio for three sets of historical market data separately, while all CVaR-SSD portfolios and SSD portfolio outperform the corresponding benchmarks in all the data sets. In addition, the resulting CVaR-SSD portfolios are sparse which is advantageous from a portfolio management perspective since a smaller sized trading basket generates less transaction costs.

For the second application on world oil market share,

- A two-stage stochastic Cournot-Nash equilibrium problem is proposed to simulate production and supply competition of a homogenous product under uncertainty in an oligopolistic market. The model is recasted as two-stage SVIs whose solutions characterize a C-N equilibrium. The practicality of global oil market has the distinctive feature that the practitioners are not very sensitive to changes in spotting price. Therefore, the oil suppliers do not compete with lower selling prices but rather for greater market share.

- The equilibrium is described by a two-stage SVI and a regularized SAA method is proposed to solve the two-stage SVIs with convergence properties under mild assumptions. The inconvenience of non-unique second-stage solutions is dealt with a partial regularization approach.

- The model is tested numerically for its effectiveness. Moreover, it is used to reproduce the market share observation in the global market of crude oil. It is observed numerically that the model is not only able to reproduce historical in-sample market share but also capable of making out-of-sample predictions based on real data sets.

- The construction of the model involves the estimation of parameters, both deterministic and stochastic. The deterministic parameters are learned using the training data set, while the stochastic parameters are constructed empirically in the sense of preserving distributions of observables.

1.5 Organization of the thesis

This thesis consists of the following parts.

- In Chapter 1, introductions and backgrounds of two applications are briefly described. Literatures in relevant areas of researches are listed. The main contributions and findings in both problems are summarized with the outline of the contents presented.

- In Chapter 2, basic notations used in the thesis are presented and some of the well-known theoretical results are presented.

- In Chapter 3, a general framework of a class of two-stage stochastic optimization problem is demonstrated. Two different but related expressions of the problem are analyzed and explained together with their Lagrangian formulation and saddle-point characterization of optimality.

- In Chapter 4, the problem of portfolio selection is developed. Theoretical results of the problem structure are analyzed and the convergence analysis is illustrated for demonstrating the approximation and relaxation results of the proposed model.

- In Chapter 5, a two-stage SVI model is developed to characterize the SNEP for the concerned application in oil market. A regularized SAA approach is proposed to solve the SVI. Solution structure of the second-stage problem is presented for analysis and the connection between the first-stage and the second-stage problem is developed via the solution function of linear complementarity problem (LCP). The convergence analysis of the regularization approach is presented.

- In Chapter 6, the empirical studies of both problems are shown collectively, including the algorithms used for numerical implementations and the experiments results.

- In Chapter 7, the main results in this thesis are summerized and potential future research directions are given.

Chapter 2

Notation and Preliminaries

In this chapter, some basic notations used throughout this thesis are presented, and some preliminary theoretical results in the literature are recalled.

2.1 Notation

Through out the thesis, unless otherwise stated, an optimal value or solution is indicated by * while other given (possibly) non-optimal values or solutions are marked by $^0, \hat{}, \bar{}$. 0 may denote zero matrix, zero vector of corresponding dimensions referring to relevant contexts. $\boldsymbol{e}$ is used to present vector of 1. $\boldsymbol{I}_X$ denotes the indicator function of set $X \subset \mathbb{R}^n$. Let Ω be the set of all random events with ω represents random event i.e., $\omega \in \Omega$. $\mathcal{A}, \mathcal{F}$ denote collections of subsets, while an event or a subset is denoted by A or F, e.g., $A \in \mathcal{A} \subset \Omega$. Probability distribution of random event is denoted by P. A random vector is denoted by $\boldsymbol{\xi} : \Omega \to \Xi \subseteq \mathbb{R}^d$. To ease of notation, $\boldsymbol{\xi}$ is used to denote the random vector $\boldsymbol{\xi}(\omega)$ and realization of a random vector $\boldsymbol{\xi}$ is denoted by ξ with its support set presented by Ξ. $\mathcal{Y}$ denotes the space of measurable function defined on Ξ. The expectation operator $\mathbb{E}\big[\boldsymbol{\xi}\big]$ is used through

out this thesis, i.e., the Lebesgue integral of the $\mathcal{A}$-measurable function $\boldsymbol{\xi} := \boldsymbol{\xi}(\omega)$ with respect to (w.r.t.) the measure P, is denoted by

$$\int_{\Omega} \xi(\omega) P(d\omega).$$

Special symbols are in used for clearer demonstration of stochastic processes, specially in dealing with two-stage problems. Tenders from first-stage to second-stage are denoted by x, u, μ, and the second-stage stochastic decisions are presented by $y(\boldsymbol{\xi}), v(\boldsymbol{\xi}), \lambda(\boldsymbol{\xi})$. ψ denotes the second-stage value function defined on tenders, feasible second-stage decisions with random argument, e.g., $\psi(x, \boldsymbol{\xi}, y(\boldsymbol{\xi}))$. Optimal second-stage objective function defined on tenders and with random argument is given by $\Psi(x, \boldsymbol{\xi}) = \sup(\inf)_{y(\boldsymbol{\xi})} \psi(x, \boldsymbol{\xi}, y(\boldsymbol{\xi}))$. Inner products in the Hilbert space of $\mathbb{R}^n \times \mathcal{Y}$ with L_2-norm is denoted as follows.

$$\langle (x, y), (u, v) \rangle := x^T u + \int_{\Xi} y(\xi)^T v(\xi) P(d\xi).$$

For numerical experiments and analysis of the SAA method, N is used to present sample size. $\mathscr{C}\big([a, b]\big)$ denotes the space of continuous functions defined on $[a, b]$ with maximum norm with $\mathscr{C}^*([a, b])$ the dual space of $\mathscr{C}\big([a, b]\big)$, space of regular countably additive measures on $[a, b]$ having finite variation. Let $\mathscr{C}\big([a, b]\big)$ denote the space of continuous functions defined on $[a, b]$ with maximum norm. By the Riesz' representation theorem, the space dual to $\mathscr{C}\big([a, b]\big)$, denoted by $\mathscr{C}^*([a, b])$, is the space of regular countably additive measures on $[a, b]$ having finite variation.Let $\mathscr{C}^*_+([a, b])$ denote the subset of $\mathscr{C}^*([a, b])$ of positive measures and $\|\mu\|$ the induced norm of the mapping $\int_a^b \cdot \mu(d\eta) : \mathscr{C}([a, b]) \to \mathbb{R}$ with $\mu \in \mathscr{C}^*([a, b])$.

2.2 Probability space

Let Ω be a nonempty set of points ω. $\mathcal{A}$ denotes a σ-field of subset of Ω, or $\mathcal{A}$-*measurable* subset. Denote a probability measure $P : \mathcal{A} \to [0, 1]$ be σ-finite Borel regular, complete, non-atomic on $\mathcal{A}$. $\boldsymbol{\xi} : \Omega \to \mathbb{R}^d$ denotes a random vector defined on $(\Omega, \mathcal{A}, P)$ with support set $\Xi \subset \mathbb{R}^d$. $\mathcal{Y}$ is the space of measurable functions on Ξ. Let $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be the real line with the system $\mathcal{B}(\mathbb{R})$ of Borel sets with

$$P_{\boldsymbol{\xi}}(B) = P\{\omega : \boldsymbol{\xi}(\omega) \in B\}, B \in \mathcal{B}(\mathbb{R}),$$

and denote the probability distribution of $\boldsymbol{\xi}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. The distribution function of $\boldsymbol{\xi}$ is denoted by the function

$$F_{\boldsymbol{\xi}}(x) = P\{\omega : \boldsymbol{\xi}(\omega) \leq x\}, x \in \mathbb{R}.$$

It follows that the support of distribution function of $\boldsymbol{\xi}$, Ξ is closed. In later part in this thesis, the subscript $\boldsymbol{\xi}$ is sometimes omitted without causing confusion. Let $T \subset \mathbb{R}$ be a subset of the real line. A set of random variable $\boldsymbol{X} = (\boldsymbol{\xi}_t)_{t \in T}$ is called a *random process* with time domain T. For each given $\omega \in \Omega$ the function $(\xi_t(\omega))_{t \in T}$ is said to be a *realization* of the process corresponding to the outcome ω. The expectation $\mathbb{E}[\boldsymbol{\xi}]$, the Lebesgue integral of the $\mathcal{A}$-measurable function $\boldsymbol{\xi} = \boldsymbol{\xi}(\omega)$ w.r.t. the measure P, is denoted as

$$\int_{\Omega} \xi(\omega) P(d\omega).$$

A property holds "P-almost surely" (a.s.) if there is a set $N \in \mathcal{A}$ with $P(N) = 0$ such that the property holds for every point ω of $\Omega \setminus N$.

Theorem 2.1 (Lebesgue's Theorem on Dominated Convergence)**.** *Let* $\boldsymbol{\eta}, \boldsymbol{\xi}, \boldsymbol{\xi_1}, \boldsymbol{\xi_2}, \ldots$ *be random variables such that* $|\boldsymbol{\xi}_n| \leq \boldsymbol{\eta}, \mathbb{E}\big[\boldsymbol{\eta}\big] < \infty$ *and* $\boldsymbol{\xi}_n \to \boldsymbol{\xi}$ *a.s.. Then,*

$$\mathbb{E}\big[|\boldsymbol{\xi}|\big] < \infty,$$

$$\mathbb{E}\big[\boldsymbol{\xi}_n\big] \to \mathbb{E}\big[\boldsymbol{\xi}\big],$$

and

$$\mathbb{E}\big[\boldsymbol{\xi}_n - \boldsymbol{\xi}\big] \to 0$$

as $n \to \infty$*.*

Recall that the cumulative distribution function of a random vector $\boldsymbol{\xi}$ evaluated at η is denoted as $F_1(\boldsymbol{\xi}; \eta)$, i.e.,

$$F_1(\boldsymbol{\xi}; \eta) := P(\boldsymbol{\xi} \leq \eta),$$

and for any $\eta \in \mathbb{R}$ define

$$F_2(\boldsymbol{\xi}; \eta) := \int_{-\infty}^{\eta} F_1(\boldsymbol{\xi}; t) dt.$$

Then, for two random variables $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2 : \Omega \to \mathbb{R}^d$, the stochastic dominance is described as follows.

1. $\boldsymbol{\xi}_1$ is said to dominate $\boldsymbol{\xi}_2$ in first order, denoted by $\boldsymbol{\xi}_1 \succeq_1 \boldsymbol{\xi}_2$, if

$$F_1(\boldsymbol{\xi}_1; \eta) \leq F_1(\boldsymbol{\xi}_2; \eta), \forall \eta \in \mathbb{R}.$$

2. $\boldsymbol{\xi}_1$ is said to dominate $\boldsymbol{\xi}_2$ in second order, denoted by $\boldsymbol{\xi}_1 \succeq_2 \boldsymbol{\xi}_2$, if

$$F_2(\boldsymbol{\xi}_1; \eta) \leq F_2(\boldsymbol{\xi}_2; \eta), \forall \eta \in \mathbb{R}. \tag{2.1}$$

It is not hard to observe that first order stochastic dominance implies second order stochastic dominance.

SSD (2.1) can be reformulated, see [24], as

$$\mathbb{E}[(\eta - \boldsymbol{\xi}_1)_+] \leq \mathbb{E}[(\eta - \boldsymbol{\xi}_2)_+], \ \forall \eta \in \mathbb{R}, \tag{2.2}$$

where $(x)_+ := \max(0, x)$.

Chapter 3

A class of two-stage stochastic optimizations

In this chapter, the general frameworks of a class of two-stage stochastic optimization problems are studied. More specifically, the order of decision execution is of special interests when dealing with two-stage stochastic programming problems. Two inter-related expression of two-stage stochastic optimization problems are presented and analyzed. The framework of this class of two-stage stochastic optimization problems will be specified in later chapters to construct applicable models that describe real applications in finance and economics.

Duality theory for stochastic optimization problems with convex objective and convex constraints was developed in a series of research works published in 1970s by Rockafellar and Wets [79], [80] , [81], [82]. Recently, subsequent researches have been focusing on the studies of more general systems of SVIs ([11], [17], [12], [85], etc.) and *stochastic generalized equations* (SGEs) [13]. More specifically, since deterministic VIs have been intensively studied and used in describing optimality conditions/ equilibria of deterministic problems, its stochastic analogy ought to be functioning

for the corresponding stochastic optimization problems. Indeed, as will be presented in later chapters, a specifically structured SVI is used to represent an economical equilibrium and to explain the mechanism behind the stable oil market share observation. The processes that bridge the optimization problems and the SVIs, namely the optimality conditions will also be sketched in this chapter.

3.1 Elementary stochastic models

Starting with an *elementary optimization* problem that aims to minimize a continuously differentiable function $f(x)$ over feasible set D, and consider its first-order necessary condition of a stationary point x, i.e.,

$$-\nabla f(x) \in \mathcal{N}_D(x). \tag{3.1}$$

To explore more into the structure of the problem, let D be further specified by a constraints system,

$$D = \{x \in X \mid G(x) \in K\} \text{ with } G(x) = \big(g_1(x), \ldots, g_m(x)\big) \tag{3.2}$$

for a closed convex set $X \subseteq \mathbb{R}^{n_1}$ and a closed convex cone $K \subseteq \mathbb{R}^m$, with g_i being continuously differentiable for all $i = 1, ..., m$. It is standard to express the elementary problem by its *Lagrangian* formulation,

$$L(x, \mu) = f(x) + \sum_{i=1}^{m} \mu_i g_i(x),$$

and the feasibility condition (3.2) takes the form,

$$-\nabla_x L(x, \mu) \in \mathcal{N}_X(x), \qquad \nabla_\mu L(x, \mu) \in \mathcal{N}_{K^*}(\mu).$$

where K^* denotes the polar of K. Under appropriate constraint qualification (CQ), the above condition on Lagrangian can be interpreted as to have some *multiplier* $\mu = (\mu_1, ..., \mu_m) \in K^*$ with $G(x) \in \mathcal{N}_{K^*}(\mu)$, the following holds.

$$v \in \mathcal{N}_D(x),$$

where $v = \sum_{i=1}^{m} \mu_i \nabla g_i(x) + u$ for some $u \in \mathcal{N}_X(x)$. Therefore, in deterministic setting, the Lagrangian characterization of optimization problem can lead to optimality conditions in the form of variational inequality.

Remark 3.1. *The solution to a VI problem may not only involve that of the decision variables in the corresponding optimization problems but also contains the "multipliers" related to constraints.*

A solution x^* to problem (3.1) must exists when the feasible set D is bounded, and thus the set of all solutions is closed. For an intersection of closed convex sets D_1 and D_2 the following relation holds

$$\mathcal{N}_{D_1 \cap D_2}(x) \supseteq \mathcal{N}_{D_1}(x) + \mathcal{N}_{D_2}(x) = \{v_1 + v_2 \mid v_1 \in \mathcal{N}_{D_1}(x), v_2 \in \mathcal{N}_{D_2}(x)\}.$$

In addition, equality holds if both D_1 and D_2 are polyhedral, or the intersection of relative interiors of D_1 and D_2 is nonempty, $ri\ D_1 \cap ri\ D_2 \neq \emptyset$, see [80].

A natural expression of stochastic counterpart of the elementary optimization, with slight abuse of notation, for almost every (a.e.) $\xi \in \Xi$, considers the following problem.

$$\begin{aligned} \max_{y(\xi)} \quad & \psi\big(\xi, y(\xi)\big) \\ \text{s.t.} \quad & y(\xi) \in D(\xi), \end{aligned} \tag{3.3}$$

where for any given $\xi \in \Xi$, $\psi(\xi, \cdot)$ is continuously differentiable and $D(\xi)$ is closed and convex. More rigorously, a general decision to stochastic problem i.e., a *response*, has to be measurable and defined as

$$y(\cdot) : \xi \mapsto y(\xi) \text{ and for a.e. } \xi \in \Xi, y(\xi) \in \mathbb{R}^{n_2},$$

which responses to known information on realization of the random variable $\boldsymbol{\xi}$. For the general description of two-stage stochastic optimization problems, one needs to consider the mixed type of decision variables, especially when faced with practical applications.

Remark 3.2. *The term "two-stage" does not necessarily correspond to the orders of decision executions in time. Instead, it emphasizes the fact that the decision vectors are divided into two catagories. For example, if the decision vector $y \in \mathcal{Y}$ needs to be made after observing the realization of $\boldsymbol{\xi}$, the constraints restriction in (3.3) needs to hold a.e. $\xi \in \Xi$. On the other hand, if the decision vector is finalized before an observation of $\boldsymbol{\xi}$ is made, the response y cannot differ with different realizations ξ. More specifically, in the later case, additional constraints are needed to ensure that $y(\xi) = y$ for a.e. $\xi \in \Xi$. This is the well-known requirement of* nonanticipativity*, which plays an important role in describing the dynamics of the multi-stage stochastic optimization problem.*

Note that the extension to multi-stage problems gives rise to additional restriction on orders of information income described by a dynamic process $(\boldsymbol{\xi})_t$ which will not be studied in this thesis. Another modification of the stochastic problem is when multi-agents are involved in the structures of stochastic programming, in which case the

analytical aspects require more detailed analysis. In the remaining of this thesis, the first-stage decision vector is denoted as x while the second-stage decision "vector", which responses to realizations of random vector $\boldsymbol{\xi}$, is denoted as y.

3.2 Two-stage stochastic optimization problems

In this section, two natural expressions of a class of two-stage stochastic optimization models will be presented with their relations explained in details.

3.2.1 Intrinsic first-stage model

Consider the following "two-stage" stochastic optimization problem.

$$\begin{aligned} \min_{x\in\mathbb{R}^{n_1}} \quad & f(x) + \mathbb{E}\big[\Psi(x,\boldsymbol{\xi})\big] \\ \text{s.t.} \quad & x \in D_1, \end{aligned} \tag{3.4}$$

where $\Psi(x,\boldsymbol{\xi})$ is the stochastic optimal value of the second-stage problem for a given realization $\xi \in \Xi$, i.e.,

$$\begin{aligned} \Psi(x,\xi) := \inf_{y(\xi)\in\mathbb{R}^{n_2}} \quad & \psi\big(x,\xi,y(\xi)\big) \\ \text{s.t.} \quad & y(\xi) \in D_2(\xi). \end{aligned} \tag{3.5}$$

A natural concern about the feasibility of the problem arises if it is seen from the perspective of the entire problem.

- If there exist a first-stage decision $\bar{x}$ with realized observation $\bar{\xi}$, so that the second-stage problem (3.5) becomes infeasible, then $\psi\big(\bar{x},\bar{\xi},y(\bar{\xi})\big) = +\infty$ by definition.

- It may also be that the second-stage problem becomes unbounded from below and the second-stage optimal value can be improved infinitely. It follows that $\psi\big(\bar{x}, \bar{\xi}, y(\bar{\xi})\big) = -\infty$.

The pathological properties, e.g., the above cases, will lead to ill-defined problem and should be avoided in the modelling process. However, there may exist problem with features that cannot be altered w.r.t. real applications, e.g., second-stage problem (3.5) has multiple solutions. In these cases, specific assumptions need to be imposed as well as the adoption of, per say, regularization on the model.

3.2.2 Two-stage stochastic optimization problem with recourse

Consider the following description of a two-stage stochastic optimization problem with mixed decision strategies. At the first-stage, $x \in \mathbb{R}^{n1}$ is chosen to minimize a cost represented by the expression $f(x)$ subject to the first-stage constraints

$$x \in D_1 \text{ and } g_{1i}(x) \leq 0, i = 1, ..., m_1. \tag{3.6}$$

Later on in the event horizon, one realization of uncertain event $\xi \in \Xi$ is observed. In response to this observation, which is called the second-stage, a vector $y(\xi) \in \mathbb{R}^{n2}$ is chosen subject to second-stage constraints.

$$y(\xi) \in D_2 \text{ and } g_{2j}\big(x, \xi, y(\xi)\big) \leq 0, j = 1, ..., m_2, \text{ for a.e. } \xi \in \Xi, \tag{3.7}$$

at a cost $\psi\big(x, \xi, y(\xi)\big)$. Conventionally, one would expect that the choice of response $y(\xi)$ being subject to optimality rule, i.e., choose $y^*(\xi)$ such that the corresponding cost $\psi\big(x, \xi, y^*(\xi)\big)$ is minimized.

Remark 3.3. *Note that by imposing the choice $y(\xi)$ to be made* after *the observation $\xi \in \Xi$, one needs to further characterize the "dynamics" in deriving optimality conditions. Moreover, the expression of the solution structures should also take into account the effects of facts that the decisions are made at different stages along the event horizon.*

Under the same setups, the two-stage stochastic optimization problem can be viewed from a perspective other than the intrinsic first-stage approach. Let the strategies of solving the two-stage stochastic optimization problem be viewed collectively as to choose a decision pair

$$\big(x \in \mathbb{R}^{n1}, y : \Xi \to \mathbb{R}^{n2}\big),$$

which satisfies the constraints (3.6) and (3.7), and minimizes the total cost view *at* the first-stage, i.e.,

$$f(x) + \mathbb{E}\big[\psi\big(x, \boldsymbol{\xi}, y(\boldsymbol{\xi})\big)\big], \tag{3.8}$$

where y denotes a measurable response function, and the expectation $\mathbb{E}\big[\,\cdot\,\big]$ is well-defined, e.g., if $y \in \mathfrak{L}^{\infty}_{n2} := \mathfrak{L}^{\infty}(\Xi, \mathcal{F}, P; \mathbb{R}^{n2})$,

$$\mathbb{E}\big[\psi\big(x, \boldsymbol{\xi}, y(\boldsymbol{\xi})\big)\big] = \int_{\Xi} \psi\big(x, \xi, y(\xi)\big) P(d\xi).$$

This is a *stochastic programming problem with recourse*[1] and the response function y specifies the *recourse decision*.

Remark 3.4. *The most noticeable difference between the two different models is the objective functions (3.4) and (3.8). In (3.4), the expectation is acting upon the*

[1] This notation follows from the works of Rockafellar and Wets, while other scholars may refer to different problem types.

stochastic optimal value functions defined by the corresponding optimal second-stage decision as for the objective function (3.8) the decisions are made simultaneously for both stages.

In order to ensure the well-defineness of the expectation calculation in (3.8), y is required to be *essentially bounded, measurable* and the second-stage constraint (3.7) is required to hold a.s.. For ease of expression, in the remaining of this thesis, $y(\xi) \in \mathbb{R}^{n2}$ is reserved to specify the recourse function value where $y \in \mathfrak{L}_{n2}^{\infty}$ represents the essentially bounded response function. Furthermore, the focus of this thesis is mainly on stochastic convex programming where the assumptions on convexity apply to both x and y, and Lagrangian function is introduced in terms of multipliers μ and λ for the first-stage and second-stage constraints, respectively. The analysis is then based on the Lagrangian formulation which leads to a saddle-point condition for optimality.

Remark 3.5. *The optimal multipliers are solutions to a certain dual problem and can be interpreted as "equilibrium prices" relative to perturbations in constraints.*

3.2.3 Problem setting for two-stage stochastic optimization problem

In this subsection, the structure of the two-stage stochastic optimization problem is specified.

- D_1 and D_2 are convex, closed and nonempty subsets of $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$, respectively.

- g_{1i} on $\mathbb{R}^{n1}$ and $g_{2j}(\cdot, \xi, \cdot)$ on $\mathbb{R}^{n1} \times \mathbb{R}^{n2}$ are convex, everywhere-defined and finite for a.e. $\xi \in \Xi$.

- For each decision pair $\big(x, y(\xi)\big)$, the functions $g_{2j}\big(x, \cdot, y(\xi)\big)$ are measurable on Ξ, and bounded for $j = 1, ..., m_2$.

The above specifications imply that, with slight abuse of notation, if a strategy of two-stage response functions

$$x : \Xi \to \mathbb{R}^{n1} \text{ and } y : \Xi \to \mathbb{R}^{n2}$$

are arbitrary bounded, measurable functions, then

$$\xi \mapsto g_{2j}\big(x(\xi), \xi, y(\xi)\big), j = 1, ..., m_2,$$

are measurable, where $x(\xi)$ are defined similarly to the second-stage response function values $y(\xi)$ with additional constraint $x(\xi) = x$ for a.e. $\xi \in \Xi$. Similar line of argument can be seen in, for example [85], [93]. Moreover, the objective functions f and ψ are summable and the constraint functions are essentially bounded. In particular, the joint objective value in (3.8) is well-defined if $y(\xi)$ is a bounded, measurable function of a.e. $\xi \in \Xi$.

3.2.4 Lagrangian formulation of two-stage stochastic optimization problem

Let $Z = \mathbb{R}^{n1} \times \mathfrak{L}^{\infty}_{n2}$ denote the set of strategies and let the set of perturbations be denoted by $U = \mathbb{R}^{m1} \times \mathfrak{L}^{\infty}_{m2}$. Then, the perturbative functional $F : Z \times U \to \mathbb{R} \cup \{+\infty\}$

can be defined in the following. If $z = (x, y) \in Z$ and $u = (u_x, u_y) \in U$ satisfy

$$\begin{aligned} & x \in D_1 && \text{and } g_{1i}(x) \leq u_x, && i = 1, ..., m_1, \\ & y(\xi) \in_{a.s.} D_2 && \text{and } g_{2j}\big(x, \xi, y(\xi)\big) \leq_{a.s.} u_y(\xi), && j = 1, ..., m_2, \end{aligned}$$

then $F(z, u)$ is the original objective function (3.8); otherwise, $F(z, u) = +\infty$. Specify a space $\Lambda = \mathbb{R}^{m_1} \times \mathfrak{L}^1_{m_2}$ paired with U, and define the respective inner product[2] as in [85], [15]

$$\langle u, \lambda \rangle := u_x^T \lambda_x + \int_\Xi \big(u_y(\xi) \cdot \lambda_y(\xi)\big) P(d\xi),$$

with

$$\lambda = (\lambda_x, \lambda_y) \in \Lambda.$$

Recall that the two-stage stochastic programming problem with recourse under consideration is to find a strategy $z = (x, y) \in Z$ which minimizes the cost function (3.8), i.e.,

$$\begin{aligned} \min \quad & f(x) + \mathbb{E}\big[\psi\big(x, \boldsymbol{\xi}, y(\boldsymbol{\xi})\big)\big], \\ \text{s.t.} \quad & x \in D_1 && \text{and } g_{1i}(x) \leq 0, && i = 1, ..., m_1, \\ & y(\xi) \in_{a.s.} D_2 && \text{and } g_{2j}\big(x, \xi, y(\xi)\big) \leq_{a.s.} 0, && j = 1, ..., m_2. \end{aligned} \tag{3.9}$$

Remark 3.6. *The fact that "constraint on dynamics" is not explicitly expressed in problem (3.9) leads to a "disconnection" between optimality conditions based on Lagrangian approach and that of the (basic) Karush–Kuhn–Tucker (KKT) condition [80]. Nevertheless, the Lagrangian based saddle-point condition serves as a valid condition for analysis later in this thesis.*

[2] complementary results follows if $\mathfrak{L}^1_{m_2}$ is replaced by the dual Banach space $(\mathfrak{L}^\infty_{m_2})^*$

The Lagrangian associated with the system (3.9) is the function:

$$L : Z \times \Lambda \to \mathbb{R} \cup \{\pm\infty\}$$

which can be expressed as

$$L(z,\lambda) = \begin{cases} L_1(x,\lambda_x) + \int_\Xi L_2\big(x,\xi,y(\xi),\lambda_y(\xi)\big)P(d\xi) & \text{if } z \in Z_0 \text{ and } \lambda \in \Lambda_0, \\ -\infty & \text{if } z \in Z_0 \text{ and } \lambda \notin \Lambda_0, \\ +\infty & \text{if } z \notin Z_0, \end{cases}$$

where

$$Z_0 = \{z = (x,y) \in Z \mid x \in D_1 \text{ and } y(\xi) \in_{a.s.} D_2\},$$

$$\Lambda_0 = \{\lambda = (\lambda_x,\lambda_y) \in \Lambda \mid \lambda_x \geq 0 \text{ and } \lambda_y(\xi) \geq_{a.s.} 0\},$$

$$L_1(x,\lambda_x) = f(x) + \sum_{i=1}^{m_1} (\lambda_x)_i g_{1i}(x),$$

$$L_2(x,\xi,y(\xi),\lambda_y(\xi)) = \psi\big(x,\xi,y(\xi)\big) + \sum_{j=1}^{m_2} (\lambda_y)_j(\xi) g_{2j}(x,\xi,y(\xi)).$$

It follows that the primal problem reads

$$\min_{z\in Z_0} \sup_{\lambda\in\Lambda_0} L(z,\lambda), \tag{$\boldsymbol{P}$}$$

while the dual problem is set to be

$$\max_{\lambda\in\Lambda_0} \inf_{z\in Z_0} L(z,\lambda). \tag{$\boldsymbol{D}$}$$

It is easily seen that the properties relating the optimal values and optimal solutions of primal problem (***P***) and dual problem (***D***) correspond to minimax properties of the Lagrangian $L(z,\lambda)$. From the standard definitions of primal and dual problems, the following results on saddle-point condition can be derived.

Definition 3.1 (Saddle-points of the Lagrangian formulation)**.** *The pair* $(z^*, \lambda^*) \in Z \times \Lambda$ *is a saddle-point of the Lagrangian* $L(z, \lambda)$ *if and only if* z^* *gives the minimum in (**P**),* λ^* *gives the maximum in (**D**), and there holds* $\min \boldsymbol{P} = \max \boldsymbol{D}$.

Remark 3.7. *The* KKT *condition of the solutions of the primal problem (**P**) can be derived from a duality theorem if the primal-dual relation of the type* $\inf \boldsymbol{P} = \sup \boldsymbol{D}$ *can be established.* $z^* \in Z$ *gives the minimum in (**P**) if and only if there exists a* $\lambda^* \in \Lambda$ *such that* (z^*, λ^*) *is a saddle-point of* $L(z, \lambda)$*. One special case is that if* D_1 *and* D_2 *are bounded, a solution* λ *of the dual problem* $\boldsymbol{D}$ *may be attained, so that* $\min \boldsymbol{P} = \max \boldsymbol{D}$ *holds and the KKT characterization of* z *to* $\boldsymbol{P}$ *is valid. In general, the derivation of optimality based on Lagrangian system does not lead to the KKT conditions but rather a saddle-point characterization of the solutions of the Lagrangian formulation.*

The detailed derivation of duality follows from standard "perturbation analysis" (under certain restriction) that embeds the problem in a class of perturbed problems, and the interested readers are referred to [80] and references therein.

3.3 Different perspectives of two-stage stochastic optimization problem

Note that the intrinsic first-stage problem described in Subsection 3.2.1 is not suitable for deriving the Lagrangian since the two components in the strategy are not simultaneously decided. Moreover, the space in which the decisions needed to be chosen from differs with different perspectives of the modelling approaches (3.4)-(3.5) and (3.9). More specifically, when the observation of random vector is not known at

the stage of strategy decision, as in problem (3.9), the second-stage decision vector is chosen from the space of measurable functions $y \in \mathcal{Y}$. On the other hand, if the two-stage model has explicit order in terms of the realization of random vector, as in problems (3.4)-(3.5), the second-stage decision vector is chosen such that $y(\xi) \in \mathbb{R}^{n2}$. Thus, before proceeding, the interchangeability of sup/inf and expectation operators needs to be addressed.

A function h on $\Xi \times \mathbb{R}^n$ is called a *normal convex integrand* if, for each $\xi \in \Xi$, $h(\xi, \cdot)$ is a lower semicontinuous convex function on $\mathbb{R}^n$ with values in $(-\infty, \infty]$, not identically $+\infty$, and if furthermore there exists a sequence of measurable functions

$$y^k : \Xi \to \mathbb{R}^n, k = 1, 2, ...,$$

such that $h\big(\xi, y^k(\xi)\big)$ is measurable in ξ for each k, while for each fixed ξ the set of points of the form $y^k(\xi)$ lying in $dom\ h(\xi, \cdot) = \{y \in \mathbb{R}^n \mid h(\xi, z) < +\infty\}$ are dense in the latter set. The above definition is given in [80] along with criteria for ensuring the normality. One consequence of the normality is that $h\big(\xi, y(\xi)\big)$ is measurable in ξ when $y(\xi)$ is measurable, and the integral functional

$$I_h = \int_{\Xi} h\big(\xi, y(\xi)\big) P(d\xi)$$

has a well-defined value for every measurable function y such that $h\big(\xi, y(\xi)\big)$ is majorized by a summable function of ξ. The following proposition is crucial in relating problems (3.4)-(3.5) and (3.9).

Proposition 3.1. *[76, Theorem 4] Let h be a normal convex integrand on $\Xi \times \mathbb{R}^n$.*

Then the infimum of $h(\xi,\cdot)$ over $\mathbb{R}^n$ is measurable as a function of ξ and one has

$$\inf_{y\in\mathfrak{L}_n^p}\mathbb{E}\big[h\big(\boldsymbol{\xi},y(\boldsymbol{\xi})\big)\big]=\int_{\Xi}\Big[\inf_{y(\xi)\in\mathbb{R}^n}h(\xi,y(\xi))\Big]P(d\xi),$$

for any $p\in[1,+\infty]$ such that the expectation is well-defined.

Note that in cases where the inf / sup can be obtained, i.e., be replaced by min / max, the optimality conditions leads to the concepts of subdifferentiability. The issue of subdifferentiability in stochastic optimization is related to various definitions on optimality conditions. Recently, the equivalence between alternative order of approaches was investigated in depth by Burke et. al. [7]. Recall that the essential objective function in (3.9) is expressed as

$$F(x,y):=f(x)+\mathbb{E}\big[\psi\big(x,\boldsymbol{\xi},y(\boldsymbol{\xi})\big)\big]\ \text{for}\ (x,y)\in Z=\mathbb{R}^{n_1}\times\mathfrak{L}_{n_2}^{\infty}. \tag{3.10}$$

As seen from the first-stage, the *induced first-stage problem* by (3.9) is to minimize the function

$$J(x):=\inf_{y\in\mathfrak{L}_{n_2}^{\infty}}F(x,y) \tag{3.11}$$

over $\mathbb{R}^{n_1}$ subject to constraints (3.6) and (3.7).

Take a step back and consider the problem as seen from second-stage, at when for any first-stage decision $x\in\mathbb{R}^{n_1}$ and realization of random vector $\xi\in\Xi$, let

$$\begin{aligned}\Psi(x,\xi):=&\inf_{y(\xi)\in\mathbb{R}^{n_2}}\psi\big(x,\xi,y(\xi)\big),\\ &\text{s.t. second-stage constraints (3.7) hold,}\end{aligned} \tag{3.12}$$

i.e., given any pair (x,ξ), $\Psi(x,\xi)$ is the infimum of the second-stage cost over all recourse y satisfying the second-stage constraints. Thus, the intrinsic first-stage

problem is recovered, with the second-stage objective specifically expressed in terms of (x, ξ), as to solve

$$\begin{aligned} &\min j(x) := f(x) + \mathbb{E}\big[\Psi(x, \boldsymbol{\xi})\big], \\ &\text{s.t. first-stage constraints (3.6) hold.} \end{aligned} \tag{$\boldsymbol{Q}$}$$

Remark 3.8. *From the above process, the intrinsic first-stage problem requires that the second-stage decision needs to be made after x has been decided, hence includes the aforementioned constraints on decision dynamics, commonly known as the requirement of nonanticipativity.*

Remark 3.9. *Note that from the expression (3.12), the optimal second-stage has its formulation with given (x, ξ). However, in problem ($\boldsymbol{Q}$), the optimal first-stage decision variable x needs to be obtained by solving the problem. This "cyclic expression" may lead to not-well-defined problem setup if its structure is not arranged carefully.*

The following results are of basic needs in dealing with different approaches of solving these types of two-stage stochastic problems, especially for the exploration of relation between induced and intrinsic first-stage problems.

Proposition 3.2. *[80, Proposition 4] For each $x \in \mathbb{R}^{n_1}$, $\Psi(x, \xi)$ is measurable as a function of $\xi \in \Xi$.*

Theorem 3.1. *[80, Theorem 1] Suppose that for each intrinsically feasible first-stage decision x,*

$$\rho(x, \xi) := \inf\{|y| \mid \psi\big(x, \xi, y(\xi)\big) < \infty, \text{s.t. second-stage constraints (3.7) hold}\}$$

is essentially bounded in ξ. *Then,*

$$j(x) = \inf_{y \in \mathfrak{L}_{n_2}^{\infty}} F(x, y) \text{ for all } x \in \mathbb{R}^{n_1}. \tag{3.13}$$

In particular, x *gives the minimum in the intrinsic first-stage problem (**Q**) if and only if it gives the solution of the induced problem of minimizing (3.11).*

More specifically, the following results hold true under mild conditions.

- Solving the primal problem (***P***) may yield the same optimal value as to solve the intrinsic problem (***Q***).
- The optimal solution of the intrinsic first-stage problem agrees with that of the induced first-stage problem.

It then follows that the solution of the intrinsic first-stage problem (***Q***) can be obtained by analyzing the Lagrangian formulation of the two-stage stochastic optimization problem with recourse (***P***), provided that the solutions can be obtained.

The rationale behind the formulation of intrinsic first-stage relies on the concept of *complete recourse*, see [93].

Definition 3.2.

- Simple recourse*: the feasible set of second-stage decisions is always non-empty.*
- Complete recourse*: for any* $x \in \mathbb{R}^{n_1}$*, the feasible set of second-stage decisions is non-empty.*
- Relative complete recourse*: for any feasible first-stage decision* $x \in X_0$*, the feasible set of second-stage decisions is non-empty.*

The assumptions on relative complete recourse are usually made for ease of analysis as well as in the modelling stage of the problem, especially in building up the inter-relation between first-stage and second-stage decisions. For the purpose of modelling two-stage stochastic optimization problems ($\boldsymbol{Q}$) and ($\boldsymbol{P}$), the following results are given by Rockafellar and Wets.

Theorem 3.2. *[80, Theorem 2] Let D_2 be bounded, then the infimum in (3.13) is attained for each $x \in \mathbb{R}^{n_1}$ by at least one $y \in \mathfrak{L}^{\infty}_{n_2}$, i.e.,*

$$j(x) = \min_{y \in \mathfrak{L}^{\infty}_{n_2}} F(x, y) \text{ for all } x \in \mathbb{R}^{n_1}.$$

Corollary 3.1. *[80, Corollary] Let D_2 be bounded, then $x \in \mathbb{R}^{n_1}$ gives the minimum in problem ($\boldsymbol{Q}$) if and only if there exists $y \in \mathfrak{L}^{\infty}_{n_2}$ such that $z = (x, y)$ gives the minimum in problem ($\boldsymbol{P}$).*

Chapter 4

Portfolio selection problem with relaxed SSD constraints

In this chapter, the application of portfolio selection problem with relaxed SSD constraints is considered under the framework of the class of two-stage convex programming developed in the previous chapter. For portfolio selection, the returns of the assets are not known at the stage of portfolio selection so that the investment decision is seen as a first-stage variable, while the random returns are realized at the second-stage. Note that there are no second-stage decision variables in this special case. For any given time in a random process of portfolio selection, investor solves a portfolio selection problem whose solution will be his/her trading basket until a realization of random return in future is revealed. The process of portfolio selection can be repeated towards a finite investment horizon. Since the transaction costs are not considered in this model, a side benefit would be that if the size of the basket is small. Namely, a sparse portfolio is always favourable.

4.1 CVaR Approximation of Portfolio Optimization with SSD constraints

Based on the classic SSD constrained portfolio optimization problem, the proposed *SAA-based CVaR-SSD relaxation model* is constructed in this section. In short, the CVaR constraint under SAA is treated as an approximation of the SSD constraints when CVaR probability level β tends to 1 under. It then follows that the relaxation of CVaR constraint can be realized by lowering the CVaR probability level β to a pre-specified value smaller than 1.

Remark 4.1. *It is clear from the above expression, one interpretation of SSD is the partial order of distribution functions. Then, one can expect the advantage of SSD, see for example [25], over other comparing criteria from the perspective that it can fully use the information of random variables, thus "reasonably" choose more realistically than, e.g., MV model.*

The portfolio optimization problem with SSD constraints is formulated as follows.

$$\begin{array}{ll} \max\limits_{x} & \mathbb{E}[x^T\boldsymbol{\xi}] \\ \text{s.t.} & x \in X_0, \\ & x^T\boldsymbol{\xi} \succeq_2 Y(\boldsymbol{\xi}), \end{array} \tag{4.1}$$

where $x \in \mathbb{R}^n$ presents the fractions of capital invested in n assets, i.e., a *portfolio*. In this setting, portfolio can only take *long* position, and then set of feasible portfolios is denoted as $X_0 = \{x \in \mathbb{R}^n_+, x_1 + \ldots + x_n = 1\}$. The constraint in (4.1) is the second order stochastic dominance of random return $x^T\boldsymbol{\xi}$ over a given random return $Y(\boldsymbol{\xi})$, refereed to as a *benchmark*. Intuitively, the benchmark $Y(\boldsymbol{\xi})$ corresponds to

a pre-specified *benchmark portfolio* $\bar{x} \in \mathbb{R}^n$, i.e., $Y(\boldsymbol{\xi}) := \bar{x}^T\boldsymbol{\xi}$. Note that there may exist multiple benchmark portfolios for a given benchmark. The portfolio selection problem with SSD constraints (4.1) seeks a portfolio x^* in the sense of maximizing the expected return while the SSD constraints, $(x^*)^T\boldsymbol{\xi} \succeq_2 \bar{x}^T\boldsymbol{\xi}$, is satisfied.

For purposes of approximation, SSD constrained portfolio optimization problem (4.1) is subject to reformulation.

Let $H(\boldsymbol{\xi}, x, \eta) := (\eta - x^T\boldsymbol{\xi})_+ - (\eta - Y(\boldsymbol{\xi}))_+$ and $h(x, \eta) = \mathbb{E}[H(x, \eta, \boldsymbol{\xi})]$. Recall the equivalence relation (2.2) and the definition of SSD, problem (4.1) can be rewritten as

$$\begin{array}{ll} \min\limits_{x} & \mathbb{E}[-x^T\boldsymbol{\xi}] \\ \text{s.t.} & x \in X_0, \\ & h(x, \eta) \leq 0, \ \forall \eta \in \mathbb{R}. \end{array} \tag{4.2}$$

Problem (4.2) is a stochastic semi-infinite problem as well as problem (4.1). From a numerical optimization perspective, the form of the constraints in problem (4.2) does not provide much convenience since it does not satisfy the checkable Slater-type constraint qualification (CQ). Consequently, one often considers a relaxed form of problem (4.2) [22]:

$$\begin{array}{ll} \min\limits_{x} & -\mathbb{E}[x^T\boldsymbol{\xi}] \\ \text{s.t.} & x \in X_0, \\ & h(x, \eta) \leq 0, \ \forall \eta \in [a, b], \end{array} \tag{4.3}$$

with $a \leq b \in \mathbb{R}$. In this chapter, the above problem is denoted as the *SSD problem*.

In order to solve problem (4.3), one needs to handle the expectation functions in both its objective and constraints, and the SAA method is applied for a valid computational implementation of the expectation calculation. Let $\Xi := \{\xi_1, \cdots, \xi_N\}$

be the independent and identically distributed (i.i.d.) samples of random vector $\boldsymbol{\xi}$ with a, b chosen such that $Y(\xi) \in [a, b],\ \forall \xi \in \Xi$. Thus, the SAA of the SSD problem (4.3) is in the following form,

$$\begin{array}{ll} \min\limits_{x} & -\frac{1}{N}\sum_{i=1}^{N} x^T \xi_i \\ \text{s.t.} & x \in X_0, \\ & \frac{1}{N}\sum_{i=1}^{N}\left((\eta_j - x^T\xi_i)_+ - (\eta_j - Y(\xi_i))_+\right) \leq 0,\ j = 1, \cdots, N, \end{array} \tag{4.4}$$

where $\eta_j := Y(\xi_j)$. Let the $h_N(x, \eta) := \frac{1}{N}\sum_{i=1}^{N}\Big((\eta - x^T\xi_i)_+ - (\eta - Y(\xi_i))_+\Big)$, and the SAA of SSD constraints in problem (4.4) is satisfied if

$$\max_{j \in \{1, \cdots, N\}} h_N(x, \eta_j) \leq 0. \tag{4.5}$$

The constraint function (4.5) can be relaxed for the purpose of enlarged feasible set which may lead to better portfolio construction. A CVaR reformulation is firstly adopted and recall that the equivalence relation holds true in a continuum of CVaR. Before proceeding, be reminded that the definition of VaR of a random variable $\boldsymbol{\zeta}$ is

$$\text{VaR}_\beta(\boldsymbol{\zeta}) := \min_{\gamma \in \mathbb{R}}\{\gamma : \text{Prob}\{\boldsymbol{\zeta} \leq \gamma\} \geq \beta\},$$

where β is the VaR probability level. The definition of CVaR used in this thesis is adopted as in [78],

$$\text{CVaR}_\beta(\boldsymbol{\zeta}) := \min_{\gamma \in \mathbb{R}}(\gamma + \frac{1}{1-\beta}\mathbb{E}[(\boldsymbol{\zeta} - \gamma)_+]).$$

Then, if η_j is treated as a scenario of a random variable $\boldsymbol{\eta}$, it follows that

$$\text{CVaR}_\beta\big(h_N(x, \boldsymbol{\eta})\big) = \min_{\zeta \in \mathbb{R}}\big(\zeta + \frac{1}{1-\beta}\mathbb{E}_{P_\eta}[(h_N(x, \boldsymbol{\eta}) - \zeta)_+]\big), \tag{4.6}$$

where P_η denotes the discrete probability measure of random variable $\boldsymbol{\eta}$. It follows that problem (4.4) is approximated by

$$\begin{array}{ll} \min\limits_{x} & -\frac{1}{N}\sum_{i=1}^{N} x^T\xi_i \\ \text{s.t.} & x \in X_0, \\ & \text{CVaR}_\beta(h_N(x,\boldsymbol{\eta})) \leq 0. \end{array} \tag{4.7}$$

In addition, when $\boldsymbol{\eta}$ is uniformly distributed, i.e., $\text{Prob}\{\eta = Y(\xi_i)\} = \frac{1}{N}$, the constraint function of problem (4.7) is expressed as,

$$\begin{aligned} \text{CVaR}_\beta^N(h_N(x,\boldsymbol{\eta})) &= \min_{\zeta\in\mathbb{R}}\Big(\zeta + \frac{1}{1-\beta}\frac{1}{N}\sum_{j=1}^{N}(h_N(x,\eta_j)-\zeta)_+\Big) \\ &= \min_{\zeta\in\mathbb{R}}\left(\zeta + \frac{1}{1-\beta}\frac{1}{N}\sum_{j=1}^{N}\left(\frac{1}{N}\sum_{i=1}^{N}\Big((\eta_j - x^T\xi_i)_+ - (\eta_j - Y(\xi_i))_+\Big) - \zeta\right)_+\right). \end{aligned} \tag{4.8}$$

In turn, the *SAA-based CVaR-SSD relaxation problem* based on problem (4.4) is of the following form

$$\begin{array}{ll} \min\limits_{x} & -\frac{1}{N}\sum_{i=1}^{N} x^T\xi_i \\ \text{s.t.} & x \in X_0, \\ & \text{CVaR}_\beta^N(h_N(x,\boldsymbol{\eta})) \leq 0. \end{array} \tag{4.9}$$

Similar approximation scheme was recently considered by Anderson et. al. [2]. In the context of portfolio optimization, SSD constraints (4.5) are regarded as an extremely robust risk measure, while the CVaR reformulation in problem (4.9) can lead to a relaxation of SSD, achievable via particular choice of probability level β. More specifically, SSD constraints are relaxed in problem (4.9) by choosing the probability level $\beta < 1$. This can be seen from the fact that SSD requires all choices

of $\beta \in (0, 1]$ to be satisfied, and it is relaxed if some inequalities of the constraints of problem (4.4) with index $j \in \{1, \ldots, N\}$ can be violated. This in turn can be understood as an enlargement of the feasible set in the SSD problem (4.1) by omitting part of the support set of $\boldsymbol{\eta}$. In comparison with the full obedience of SSD, a less conservative solution that violates some of the inequalities may result in possibly better portfolio.

To demonstrate the effects of this relaxation, an illustrative example is presented as follows.

Example 4.1. *Consider the problem*

$$\begin{array}{ll} \min\limits_{z} & \mathbb{E}[G(z, \boldsymbol{\xi})] \\ \text{s.t.} & \mathbb{E}[(\eta - G(z, \boldsymbol{\xi}))_+] \leq \mathbb{E}[(\eta - Y(\boldsymbol{\xi}))_+], \ \forall \eta \in \mathbb{R}, \\ & z \in [1, 2], \end{array} \tag{4.10}$$

where $\boldsymbol{\xi} \sim \mathbb{U}[0, 1]$, $G(z, \boldsymbol{\xi}) = z\boldsymbol{\xi}$ *and*

$$Y(\boldsymbol{\xi}) := \begin{cases} \frac{i}{20}, & \boldsymbol{\xi} \in \big[0.05 \times i, 0.05 \times (i + 1)\big) \quad i = 0, \ldots, 19, \\ 1, & \boldsymbol{\xi} = 1. \end{cases} \tag{4.11}$$

Then for any $\eta < 0.1$ *and any* $z \in [1, 2]$, *the constraint* $\mathbb{E}[(\eta - G(z, \boldsymbol{\xi}))_+] \leq \mathbb{E}[(\eta - Y(\boldsymbol{\xi}))_+]$ *cannot be satisfied. Thus, problem (4.10) has no feasible solution. On the other hand, if* η *is treated as a random variable* $\boldsymbol{\eta}$ *uniformly distributed over* $[0, 1]$, *there is only* 10% *chance that* $Y(\boldsymbol{\xi})$ *cannot be "dominated" by* $G(z, \boldsymbol{\xi})$. *The relaxation of the proposed model can be seen analogously as if this "*10% *chance" is omitted. Thus,* $z = 2$ *becomes a feasible solution, as it obeys the relaxed CVaR constraint (4.6), and the corresponding expected loss* $-\mathbb{E}[G(2, \boldsymbol{\xi})] = -1$, *is better than the benchmark loss* $-\mathbb{E}[Y(\boldsymbol{\xi})] = -0.5$.

4.2 Convergence analysis

In this section, the convergence analysis is presented for both the optimal values and the behaviour of optimal solutions of SAA-based CVaR-SSD relaxation problem (4.9) as CVaR probability level $\beta \to 1$ and sample size $N \to \infty$. The convergence analysis between the SSD problem (4.3) and its SAA (4.4) is considered in this section. The convergence analysis of the SAA method of general stochastic optimization problem with SSD has been studied in several papers [43, 57, 100, 102]. More specifically, in [102], the exponential rate of convergence has been built under mild conditions. The convergence results were constructed for stationary points, which is of greater interests in nonconvex cases. In convex cases, stronger results can be achieve under similar conditions compared to that in the aforementioned references. The work flow of proofs in this section follows that in [102] applying to SSD problem (4.3) with the re-proposed condition more suitable for convex cases.

The SAA of SSD problem (4.4) (*SAA-SSD problem*) can be rewritten as

$$\begin{array}{ll} \min\limits_{x} & -\frac{1}{N}\sum_{i=1}^{N} x^T\xi_i \\ \text{s.t.} & x \in X_0, \\ & \frac{1}{N}\sum_{i=1}^{N}\left((\eta - x^T\xi_i)_+ - (\eta - Y(\xi_i))_+\right) \leq 0,\ \eta \in [a,b], \end{array} \tag{4.12}$$

and the convergence properties of optimal values and optimal solutions of SAA based problem (4.12) as $N \to \infty$ are considered.

Before proceeding to further theoretical results, the following assumption on CQ holds for the remaining of this chapter.

Assumption 4.1. Slater-type CQ *is satisfied in the SSD problem (4.3), i.e., there exists a point* $x_0 \in X_0$ *such that*

$$\sup_{\eta\in[a,b]} h(x_0,\eta) < 0.$$

Let $\mu \in \mathscr{C}_+^*([a,b])$ and define the Lagrange function of the SSD problem (4.3) as

$$L(x,\mu) = \mathbb{E}[-x^T\boldsymbol{\xi}] + \int_a^b h(x,\eta)\mu(d\eta),$$

and similarly for the SAA based problem (4.12)

$$L_N(x,\mu) = -\frac{1}{N}\sum_{i=1}^N x^T\xi_i + \int_a^b h_N(x,\eta)\mu(d\eta).$$

Lemma 4.1. *Suppose* $\mathbb{E}[|\boldsymbol{\xi}|] < \infty$, *then for all* $x \in X_0$ *and* $\eta \in [a,b]$, $\mathbb{E}[|x^T\boldsymbol{\xi}|] < \infty$, $h(x,\eta) < \infty$ *and* $\mathbb{E}[\|\nabla_x(x^T\boldsymbol{\xi})\|] < \infty$.

The above simple result ensures the properties of the objective and the constraint functions required for the convergence proofs. The following first-order necessary optimality conditions of problem (4.3) and problem (4.12) follow from [102, Section 2], [5, Theorem 5.107] and [22, Theorem 4.2].

Theorem 4.1 (first-order necessary conditions)**.** *Suppose* $\mathbb{E}[|\boldsymbol{\xi}|] < \infty$ *and let* $x^* \in X_0$ *be an optimal solution to the SSD problem (4.3). Then there exists* $\mu^* \in \mathscr{C}_+^*([a,b])$ *such that*

$$\begin{cases} 0 \in -\mathbb{E}[\boldsymbol{\xi}] + \int_a^b \partial_x h(x^*,\eta)\mu^*(d\eta) + \mathcal{N}_{X_0}(x^*), \\ h(x^*,\eta) \le 0, \forall \eta \in [a,b], \\ \int_a^b h(x^*,\eta)\mu^*(d\eta) = 0, \end{cases} \tag{4.13}$$

where $\partial_x h(\cdot,\eta)$ *denotes the Clarke subdifferential [20] of* h, *and*

$$\int_a^b \partial_x h(x,\eta)\mu(d\eta) = \left\{ \int_a^b \phi(\eta)\mu(d\eta) : \phi(\eta) \in \partial_x h(x,\eta) \quad \textit{and} \quad \phi(\eta) \quad \textit{is integrable} \right\}.$$

Similarly, the optimality conditions and first-order necessary conditions of the SAA problem (4.12) can be shown as follows.

Theorem 4.2 (first-order necessary conditions)**.** *Suppose* $\mathbb{E}[|\boldsymbol{\xi}|] < \infty$ *and let* $x_N \in X_0$ *be any optimal solution to the SAA-SSD problem (4.12). Then, w.p.1 problem (4.12) satisfies the Slater-type CQ and there exists* $\mu_N \in \mathscr{C}_+^*([a,b])$ *such that*

$$\begin{cases} 0 \in -\frac{1}{N}\sum_{i=1}^N \xi^i + \int_a^b \partial_x h_N(x_N,\eta)\mu_N(d\eta) + \mathcal{N}_{X_0}(x_N), \\ h_N(x_N,\eta) \le 0, \forall \eta \in [a,b], \\ \int_a^b h_N(x_N,\eta)\mu_N(d\eta) = 0. \end{cases} \tag{4.14}$$

A tuple $(x_N, \mu_N(\cdot))$ satisfying (4.14) is called a KKT pair of problem (4.12), where x_N is a *Clarke stationary point* and $\mu_N(\cdot)$ denotes the corresponding Lagrange multiplier.

Proposition 4.1. *Consider the SAA-SSD problem (4.12). Suppose* $\mathbb{E}[|\boldsymbol{\xi}|] < \infty$ *and Assumption 4.1 holds. Then, the sequence of the Langrange multipliers* $\{\mu_N\}$ *is bounded w.p.1.*

Proof. Note that the dual problem of (4.12) is

$$\max_{\mu \in \mathscr{C}_+^*([a,b])} \mathscr{T}_N(\mu) := \min_{x \in X_0} L_N(x,\mu). \tag{4.15}$$

It is then sufficient to show (4.15) satisfies the boundedness condition of upper level set in the weak topology w.p.1 when the sample size N is sufficiently large, i.e., there exists m_0 such that $\mathcal{W}(m_0, N) := \{\mu \in \mathscr{C}_+^*([a,b]) : \mathscr{T}_N \ge m_0\}$ is a bounded and nonempty set for N sufficiently large w.p.1.

Let

$$D_0 := \{x \in X_0, h(x,\eta) \le 0, \forall \eta \in [a,b]\}$$

be the feasible set of problem (4.3). Assume for the sake of a contradiction that the boundedness result does not hold. Then, for every $m_0 \in \mathbb{R}$ such that $\mathcal{W}(m_0, N)$ is nonempty for N sufficiently large w.p.1, there exists a sequence $\{m_k\}$ satisfies $m_k \geq m_0$, and a sequence (N_k, μ_k) such that $N_k \to \infty$, $\mu_k \in \mathcal{W}(m_k, N_k)$ for each k and $\|\mu_k\| \to \infty$. Since $\mu_k \in \mathcal{W}(m_k, N_k)$, then

$$L_{N_k}(x, \mu_k) \geq m_k \geq m_0, \forall x \in D_0.$$

Let $\hat{\mu}_k = \mu_k / \|\mu_k\|$ and since $\hat{\mu}_k$ is a Borel measure defined on compact set $[a, b]$, by Helly-Bray's theorem (see Theorems 9.2.1-9.2.3 and Remark 9.2.1 in [4]), it has a weakly convergent subsequence. Take a subsequence if necessary, assuming without loss of generality (wlog) that $\hat{\mu}_k \to \hat{\mu}$, which yields $\|\hat{\mu}\| = 1$. Dividing both sides of the above inequality by $\|\hat{\mu}_k\|$ and driving $k \to \infty$, and it is easy to observe that w.p.1 $\frac{1}{N_k} \sum_{i=1}^{N_k} x^T \xi^i \to \mathbb{E}_P[x^T \boldsymbol{\xi}]$, $h_N(x, \eta) \to h(x, \eta)$ uniformly w.r.t. $\eta \in [a, b]$ and then

$$\int_a^b h(x, \eta) \hat{\mu}(d\eta) \geq 0, \forall x \in D_0.$$

Note that since Slater-type CQ holds, there exists $x_0 \in D_0$ such that $h(x_0, \eta) < 0$ for all $\eta \in [a, b]$. Then it holds

$$\int_a^b h(x_0, \eta) \hat{\mu}(d\eta) \geq 0,$$

which implies $\|\hat{\mu}\| = 0$. This contradicts the fact that $\|\hat{\mu}\| = 1$. □

Sun and Xu [102] investigated the boundedness of Lagrange multipliers $\{\mu_N\}$ under nonzero abnormal multipliers constraint qualification (NNAMCQ) in more

general cases. Due to the convexity of problem (4.3), the easily checkable Slater-type CQ can be used to prove the following result.

Theorem 4.3. *Assume $\boldsymbol{\xi}$ is a continuous random variable, $\mathbb{E}[|\boldsymbol{\xi}|] < \infty$ and its support set Ξ is bounded. Then, sequence $\{(x_N, \mu_N)\}$ has cluster points and any cluster point of the sequence, denoted as (x^*, μ^*) is a KKT pair of problem (4.3) w.p.1. Moreover, for any $\alpha > 0$, there exist positive constants $C(\alpha)$ and $\beta(\alpha)$ independent of N such that*

$$\mathrm{Prob}\{\mathrm{d}(\mathrm{x_N}, \mathrm{X^*}) \geq \alpha\} \leq \mathrm{C}(\alpha)\mathrm{e}^{-\mathrm{N}\beta(\alpha)},$$

where X^* denotes the set of Clarke stationary points characterized by (4.13).

Remark 4.2. *The results follow from [102, Theorem 5 and Theorem 7]. With the benefit of convexity of problem (4.3) and linearity of $x^T\boldsymbol{\xi}$, together with Proposition 4.1, the conditions in [102, Theorem 7] can be simplified. Note also that the condition of the continuously differentiablility of $h(x, \eta)$ w.r.t. x for every $\eta \in [a, b]$ in [102, Theorem 5] can be weaken to the continuity of $h^o(x, \eta, u)$ w.r.t. (x, u), which is guaranteed by [102, Proposition 4], where $h^o(x, \eta, u)$ denotes the Clarke directional derivative with direction $u \in \mathbb{R}^n$, see [20].*

To clarify the notations, let v_N and X_N denote the optimal value and optimal solution sets of (4.4) respectively. v^* and X^* denote the optimal value and optimal solution set of problem (4.3) respectively. Similarly, let $v_N(\beta)$ and $X_N(\beta)$ be the optimal value and optimal solution sets of (4.9) with given β.

Theorem 4.4. *Assume $\boldsymbol{\xi}$ is a continuous random variable, $\mathbb{E}[|\boldsymbol{\xi}|] < \infty$ and its*

support set Ξ *is bounded. Then,*

$$\lim_{N\to\infty}\lim_{\beta\to 1} v_N(\beta) \to v^*, \tag{4.16}$$

and

$$\limsup_{N\to\infty}\limsup_{\beta\to 1} X_N(\beta) \subset X^*, \tag{4.17}$$

w.p.1.

Proof. By Assumption 4.1 , for N sufficiently large, problem (4.4) satisfies Slater-type CQ. Then by [2, Theorem 4], w.p.1,

$$v_N(\beta) \to v_N \text{ as } \beta \to 1 \text{ and } \limsup_{\beta\to 1} X_N(\beta) \subset X_N.$$

The statements (4.16) and (4.17) follow when combining Theorem 4.3. □

Remark 4.3. *The theoretical results mainly focus on the "approximation" part of the proposed model. The results on relaxation are mainly left in later chapters about numerical experiments where an application on portfolio selection will be presented to demonstrate the effectiveness of the proposed relaxation model.*

In a short summary, the convegence results show that the SAA approach to solve the proposed model (4.9) can be viewed as to solve the SSD problem (4.3) with $N \to \infty$ and probability level $\beta \to 1$.

Remark 4.4. *The general framework of two-stage stochastic model involves a response decision once a realization of random vector is revealed. In application described by solving the proposed portfolio selection problem, the underling mechanism*

of revenue generating is that all assets will be sold. More specifically, the second-stage decision is equal to the first-stage decision regardless of the realization of random return vector. Under this specific setting, no recourse decision is needed as responses.

4.3 Investment over finite horizon

In later chapter on numerical experiments, the concern will be the process of making portfolio decisions over a finite time investment horizon. The portfolio selection method based on the investment strategy proposed in this chapter can be viewed as to form a unit fraction of the entire investment horizon upon the choices of decision makers. For example, a rolling window investment strategy reviews the components of portfolio with, per say, daily basis. For every trading day, a portfolio is generated by solving the proposed problem (4.9), seen as the first-stage decision. Note that, although not explicitly formulated in the section of construction, the conventional assumption in the literature of portfolio selection holds that the portfolio can be bought at the market with the cost at the closing price of the previous trading day. At the end of the concerned trading day, the return of the portfolio is observed since the closing prices of components became available. Similar assumption at the construction stage of the portfolio holds at the selling stage where all the assets are assumed to be sold at their closing prices, and the changes in assets' prices can be treated as a realization of the random daily return vector. In this example, the daily return of the portfolio is realized by calculating the difference in capital before and after the concerned trading day. The strategy can be executed periodically through out the investment horizon.

Chapter 5

Two-stage stochastic Cournot-Nash game: oil market share

In this chapter, an application of two-stage stochastic C-N game of the world's crude oil market is presented. In particular, the random observation affecting the spot price of oil trading is observed only after the execution of first-stage decision, the extraction quantity of oil, while a second-stage decision, the supply quantity to the market, needs to be made as to respond to the realization of the random event. Moreover, the oil market is treated as an oligopolistic market in which the price is dominantly determined by the total supply to the market contributed by all the suppliers.

5.1 Two-stage stochastic Cournot-Nash game

The model proposed and studied in this section explores a real application in which the inherently "ordered" components over decision horizons are of great importance.

In particular, the decisions in the proposed stochastic strategy partially respond to the information that is available only at the time of decision making. The model is casted as a J-agent two-stage stochastic C-N equilibrium problem, which extends the classical deterministic C-N equilibrium problem (see [62]) to a more general stochastic setup.

Let $\boldsymbol{\xi} : \Omega \to \mathbb{R}^d$ be a random variable with compact support set $\Xi \subseteq \mathbb{R}^d$, and let $(\Xi, \mathcal{F}, P)$ denote the induced probability space. A two-stage *strategy* of agent $j \in \mathcal{J} := \{1, ..., J\}$ is denoted as

$$(x_j \in \mathbb{R}, \;\; y_j : \Xi \to \mathbb{R}, \mathcal{F}\text{-measurable}), \tag{5.1}$$

where x_j is a first-stage decision vector and $y_j \in \mathcal{Y}$ denotes a measurable second-stage *response function* with $\mathcal{Y}$ being the space of $\mathcal{F}$-measurable functions defined on Ξ. Let $\mathfrak{L}_n$ be the Lebesgue space of $\mathbb{R}^n$-valued functions with $\mathfrak{L}_n^\infty$ denotes the class of measurable essentially bounded functions. Following a similar treatment as in [12], the second-stage response function of random variable is required to be essentially bounded, i.e., $y_j \in \mathfrak{L}_1^\infty$. Collectively, the vector of strategies of all agents can be written as

$$(x \in \mathbb{R}^J, \;\; y : \Xi \to \mathbb{R}^J, \mathcal{F}\text{-measurable}). \tag{5.2}$$

A strategy $\left(x_j^*, y_j^*\right) \in \mathbb{R} \times \mathfrak{L}_1^\infty$ is said to be an C-N equilibrium if it solves the following

problem for all agents $j \in \mathcal{J}$.

$$\begin{aligned}
\max_{(x_j, y_j)} \quad & W_j^1(x_j, x_{-j}^*) + \mathbb{E}\big[W_j^2\big(\boldsymbol{\xi}, y_j(\boldsymbol{\xi}), y_{-j}^*(\boldsymbol{\xi})\big)\big], && \text{(objective function)} \\
\text{s.t.} \quad & x_j \in X_j, && \text{(first-stage constraints)} \\
& y_j(\xi) \in_{a.s.} Y_j, \quad g_j\big(\xi, x_j, y_j(\xi)\big) \leq_{a.s.} 0 && \text{(second-stage constraints)}
\end{aligned} \tag{5.3}$$

where

$$\begin{aligned}
x_{-j}^* &= (x_1^*, \ldots, x_{j-1}^*, x_{j+1}^*, \ldots, x_J^*), \\
y_{-j}^*(\xi) &= (y_1^*(\xi), \ldots, y_{j-1}^*(\xi), y_{j+1}^*(\xi), \ldots, y_J^*(\xi)),
\end{aligned}$$

with $y_j(\xi)$ denotes the value of response y_j to realization ξ, with

- $W_j^1 : \mathbb{R} \times \mathbb{R}^{J-1} \to \mathbb{R}$ a first-stage wealth function of agent j, concave and continuously differentiable w.r.t. x_j.
- $W_j^2 : \Xi \times \mathbb{R} \times \mathbb{R}^{J-1} \to \mathbb{R}$ a second-stage wealth function of agent j, concave, well-defined and finite.
- X_j, Y_j nonempty, closed and convex subsets of $\mathbb{R}$ and the second-stage constraints hold a.s..
- $g_j : \Xi \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ a continuously differentiable function w.r.t. $\big(x_j, y_j(\xi)\big)$ for a.e. $\xi \in \Xi$ and $\mathcal{F}$-measurable.

The model (5.3) is formulated under the assumption that all relevant uncertainty can be described by a random vector $\boldsymbol{\xi}$ with known distribution.

Remark 5.1. *The problem at concern needs to be solved by individual agent of the system, although the modelling relies on information that is not available to each*

agent, e.g., optimal decision of other agents. However, from the perspective of the entire system, the market[1] *"chooses"* $x \in \mathbb{R}^J$ *before a realization* $\xi \in \Xi$ *is revealed and later "selects"* $y(\xi) \in \mathbb{R}^J$ *with known realization, which solves the entire system.*

5.1.1 Stochastic commodity production and supply planning

The general application of commodity production and supply in an oligopolistic market is considered in this subsection, and it serves as a motivation as well as the practical economic problem of interest. A stochastic game is presented in such a way that the strategy of each agent in supply-side of the market can be described as the solution of the stochastic optimization problem (5.3).

The decision process follows that agent $j \in \mathcal{J}$ decides an optimal production quantity x_j of the commodity at the production stage, the first-stage decision vector. At the sales stage, each agent decides a supply quantity $y_j(\xi)$ to the market after a realization ξ is observed. It then follows that a total quantity $T(y(\xi)) := \sum_{j=1}^{J} y_j(\xi)$ is supplied to the oligopolistic market. Complication arises since the price of the commodity depends on the entire supply in the market, contributed by all J agent at the supply-side. Moreover, the price is not assumed to follow a deterministic inverse demand curve. Indeed, the market is treated under the condition that even with known quantity of total supply, the spot price is still subject to uncertainty. Recall that the study focuses on modelling oligopolistic markets behaviour which means that the price is dominantly affected by the supplied quantity in the market $T(y(\xi))$. Therefore, it follows that all the trades occur at the spot price $p : \Xi \times \mathbb{R} \to \mathbb{R}_+$,

[1] A "social planner" is commonly termed in literature of economics, and can be interpreted as one individual who oversees the system and assigns strategy to each agents.

determined by a stochastic inverse demand curve $p(\xi, T(y(\xi))$.

Further assumptions need to be restricted upon production and supply quantities for each agent, e.g., capability of production plant, logistic restriction, etc., and they are denoted as i.e., $x_j \in X_j$ and $y_j(\xi) \in_{a.s.} Y_j$. Note that the second-stage decision on supply need to hold for a.e. $\xi \in \Xi$. For example, it is natural to have non-negative requirements for both production and supply quantities, i.e., $X_j = \mathbb{R}_+$ and $Y_j = \mathbb{R}_+$. The relations between stage-wise decision variables x_j and y_j are captured by constraints $g_j\big(\xi, x_j, y_j(\xi)\big) \leq_{a.s.} 0$ in (5.3). For example, it is reasonable to require agent j's supply to the market cannot exceed that of his/her production quantity, i.e., $y_j(\xi) - x_j \leq_{a.s.} 0$. This can be interpreted as the fact that agents may have no stock to supply initially, or they need to preserve certain reserved quantities prior to each decision process. Nevertheless, every agent needs to formulate and solve a *two-stage stochastic programming problem with recourse* in the sense of achieving equilibria of a J-agents non-cooperative game of the market.

Problem (5.3) can also be viewed from the perspective of the intrinsic first-stage problem. As seen from the first-stage, agent $j \in \mathcal{J}$ wants to find a production quantity $x_j \geq 0$ that maximizes

$$W_j^1(x_j, x_{-j}^*) + \mathbb{E}\big[\Psi_j(x_j, x_{-j}^*, \boldsymbol{\xi})\big], \tag{5.4}$$

where,

$$\Psi_j(x_j, x_{-j}^*, \xi) = \sup_{y_j(\xi) \geq 0} \{W_j^2(\xi, y_j(\xi), y_{-j}^*(\xi)) \mid x_j \geq y_j(\xi), \text{ for a.e. } \xi \in \Xi\}. \tag{5.5}$$

Objective function in (5.4) can be seen as the expected profit of agent j's and problems (5.4)-(5.5) are formulated following the discussion in previous chapter. Recall

that the key feature of intrinsic first-stage problem compares to formulation (5.3) is the requirement on precise orders of decision execution, commonly known as the constraints of *nonanticipitivity*. In (5.4)-(5.5), the second-stage decisions are explicitly determined after the first-stage decision, provided for each x_j the second-stage problem is well-defined [93]. However, the study of optimality condition of (5.4)-(5.5), in the case of a general induced probability space[2] $(\Xi, \mathcal{F}, P)$, is very complicated since one needs to characterize the order of the decision process explicitly. For ease of analysis, assuming that there exists a multiplier $\lambda_j \in \mathfrak{L}_1^1$ corresponds to second-stage constraint. Therefore, the study of saddle-point condition of the Lagrangian formulation of problem (5.3) can be carried out. It is worth mentioning that the basic KKT condition of problem (5.3) (see [79]) introduces a second-stage multiplier $\tilde{\lambda}_j \in (\mathfrak{L}_1^\infty)^*$ for every $j \in \mathcal{J}$ which incorporates the two-stage decision making process. It has been shown by Rockafellar and Wets in [81] that any element of the dual space $(\mathfrak{L}_1^\infty)^*$ can be decomposed into a component of $\mathfrak{L}_1^1$ and a "singular" component, corresponding to the *multiplier of nonanticipativity*. The saddle-point condition is shown to be sufficient and "almost" necessary for optimality of problem (5.3), and the interested readers may refer to [79–82, 85] for more details.

The Lagrangian formulation of problem (5.3) associated with agent j is of the following form

$$L_j(x_j, x^*_{-j}, y_j, y^*_{-j}, \lambda_j) = L_j^1(x_j, x^*_{-j}) + \mathbb{E}\big[L_j^2\big(x_j, \boldsymbol{\xi}, y_j(\boldsymbol{\xi}), y^*_{-j}(\boldsymbol{\xi}), \lambda_j(\boldsymbol{\xi})\big)\big],$$

[2] In cases of finitely supported distribution, the equivalence between intrinsic first-stage problem and the original recourse problem can be established, and the optimality condition of the recourse problem can be applied, see for example [85].

where

$$L_j^1(x_j, x_{-j}^*) = W_j^1(x_j, x_{-j}^*),$$

$$L_j^2\big(x_j, \xi, y_j(\xi), y_{-j}^*(\xi), \lambda_j(\xi)\big) = W_j^2(\xi, y_j(\xi), y_{-j}^*(\xi)) + \lambda_j(\xi)(x_j - y_j(\xi)) \text{ for a.e. } \xi \in \Xi.$$

Remark 5.2. *The constraints $y_j(\xi) \leq_{a.s.} x_j$ can be interpreted as the situation under which the profit maximizing supply $y_j^*(\xi)$ of agent j is not necessarily equal to the total production quantity x_j. Recall the application in portfolio selection in previous chapter can be seen as a special case in which all the first-stage assets are sold after the random returns are realized. This feature of the proposed model differs from conventional requirement on production-clearing condition, i.e., all the produced goods are expected to supply to the market, and is only allow if one takes the dynamics of decisions into account.*

5.1.2 Construction of two-stage cost functions

The remaining of this section is devoted to specify and clarify the structures of wealth functions, W_j^1 and W_j^2, suitable for describing oligopolistic markets of homogeneous commodity.

- The production cost for j-th agent is assumed to be quadratic, i.e., for each $j \in \mathcal{J}$ the cost of producing x_j amount of production is $\frac{1}{2}c_j x_j^2 + a_j x_j$, for some $c_j > 0, a_j > 0$.

- The cost function of the supply or second-stage is linear and of stochastic nature, i.e., for each $j \in \mathcal{J}$ the cost of supplying $y_j(\xi)$ amount of commodity is $h_j(\xi)y_j(\xi)$ for a.e. $\xi \in \Xi$.

- A classic stochastic inverse demand curve is chosen, see for example [44], which takes the expression $p\big(\xi, T(y(\xi))\big) = p_0(\xi) - \gamma(\xi)T(y(\xi))$ for the spot price for a.e. $\xi \in \Xi$.

Remark 5.3. *The intuitive choice of linear cost function with an upper bound on first-stage decision in the production stage is not adopted. The linear cost model is commonly used by practitioners in business studies of oil industry but not widely accepted by researchers in economics, see [95]. Moreover, in the numerical experiments, it is found that the linear cost of the first-stage problem behave badly in the sense that the real market share cannot be reproduced.*

The model in turn requires two sets of parameters to be chosen, namely the deterministic parameters of the first-stage wealth model, c_j, a_j and stochastic parameters in the second-stage of the problem, $p_0(\xi), \gamma(\xi), h_j(\xi)$.

Remark 5.4. *Although the cost of the production, characterized by c_j and a_j, and the cost in supply, controlled by $h_j(\xi)$ are different for each agent, the spot price at which the trading occurs is common for all the agents.*

As will be described in later chapters on numerical experiments, the stochastic *benchmark price* excluding the effect of supply to the market $p_0 : \Xi \mapsto \mathbb{R}_+$ is estimated via statistical approaches based on real data. The supply discount $\gamma : \Xi \to \mathbb{R}_+$ acts as a market mechanism to adjust and reflect uncertainty in quantity in the market is also generated from historical data sets.

In order to respect the market mechanism of supply-demand relation, the following assumption is made through out this study.

Assumption 5.1. *There exists a* $\gamma_0 > 0$ *such that* $\gamma(\xi) \geq \gamma_0$ *for a.e.* $\xi \in \Xi$.

The assumption holds naturally for *common goods*, i.e., the price of common goods drops with larger quantity becomes available in the market provided the demand is always greater or equal to the supply, a condition always ensured under the oligopolistic market setting. Once the stochastic parameters are chosen to obey that of the historical observation, deterministic parameters are specified from an in-sample learning process dedicated to reduce the observation error to those of the known data sets. The detailed discussion is presented in later chapters.

Thus, agent j's stage-wise wealth functions are,

$$W_j^1(x_j, x_{-j}^*) = -\frac{1}{2}c_j x_j^2 - a_j x_j,$$

and

$$W_j^2(\xi, y_j(\xi), y_{-j}(\xi)) = \big(p_j(\xi) - \gamma(\xi)T(y(\xi))\big)y_j(\xi),$$

where the short-handed notation of the risk-adjusted spot price of agent j's is denoted by $p_j(\xi) := p_0(\xi) - h_j(\xi)$. Thus, for every agent $j \in \mathcal{J}$ the ordered decision problem reads

$$\begin{aligned} \max_{x_j} \quad & \mathbb{E}[\psi_j(x, \boldsymbol{\xi})] - \frac{1}{2}c_j x_j^2 - a_j x_j \\ \text{s.t.} \quad & 0 \leq x_j, \end{aligned} \tag{5.6}$$

where

$$\begin{aligned} \psi_j(\xi, x) = \quad \max_{y_j(\xi)} \quad & \Big(p_j(\xi) - \gamma(\xi)\Big(\sum_{i \neq j}^{J} y_i^*(\xi) + y_j(\xi)\Big)\Big)y_j(\xi) \\ \text{s.t.} \quad & 0 \leq y_j(\xi) \leq x_j, \text{ for a.e. } \xi \in \Xi. \end{aligned} \tag{5.7}$$

Note that the requirements in problem (5.7) hold almost surely in accordance with the a.s. constraints of the second-stage in problem (5.3). Similar to that of the intrinsic first-stage problems (5.4)-(5.5), problems (5.6)-(5.7) are also not easy to solve, especially in an SNEP with $J \geq 2$, see [13]. Further complication arises since the j-th agent's problem contains those of the other agents' strategies, not known at the decision horizon.

The approach adopted for this problem is to recast problems (5.6)-(5.7) of each agent as a stochastic equilibrium problem of the entire system of J agents. It follows that obtaining an equilibrium of the convex J-player game (5.6)-(5.7) is equivalent to finding solutions for all agents. It is well-known that stochastic equilibrium has been shown to be an effective approach to study and to solve two-stage multi-players stochastic game problems, see for instance [15, 70, 77, 85, 98]. In the remaining of this chapter, the saddle-point condition of the problem (5.6)-(5.7) is studied, rewritten in the form of two-stage stochastic problem (5.3). More specifically, assuming that for all $j \in \mathcal{J}$ there exists $\bar{\lambda}_j(\xi) \in \mathfrak{L}_1^1$ with $\bar{\lambda}(\xi) \geq_{a.s.} 0$ so that a strategy $\left(\bar{x}_j, \bar{y}_j\right) \in \mathbb{R}_+ \times \mathfrak{L}_+^\infty$ solves the following system.

$$
\begin{aligned}
& -c_j\bar{x}_j - a_j + \mathbb{E}\left[\bar{\lambda}_j(\boldsymbol{\xi})\right] \in \mathcal{N}_{[0,\infty)}(\bar{x}_j), \\
& p_j(\xi) - \gamma(\xi)\sum_{i\neq j}^{J}\bar{y}_i(\xi) - 2\gamma(\xi)\bar{y}_j(\xi) - \bar{\lambda}_j(\xi) \in_{a.s.} \mathcal{N}_{[0,\infty)}(\bar{y}_j(\xi)), && \text{(stationality)} \\
& \bar{x}_j \geq 0, \bar{y}_j(\xi) \geq_{a.s.} 0, \bar{x}_j - \bar{y}_j(\xi) \geq_{a.s.} 0, && \text{(feasibility)} \\
& \bar{\lambda}_j(\xi) \ \geq_{a.s.} \ 0, && \text{(dual feasibilty)} \\
& \bar{\lambda}_j(\xi) \perp_{a.s.} (\bar{x}_j - \bar{y}_j(\xi)). && \text{(complementarity)}
\end{aligned}
$$

In particular, *stationarity* is defined for the first-order necessary optimality condition

under the assertion $\partial_x \mathbb{E}\big[\psi_j(x, \boldsymbol{\xi})\big] \subseteq \mathbb{E}\big[\partial_x \psi_j(x, \boldsymbol{\xi})\big]$. The details about the above assertion are discussed recently in [7], and it is not hard to see that the above system is viewed as a weaker condition for optimality.

Rewritten in a compact form as an SVI, the optimal strategy-multiplier triplet $(x_j, y_j, \lambda_j) \in \mathbb{R}_+ \times \mathfrak{L}_+^\infty \times \mathfrak{L}_+^1$ must satisfy,

$$\begin{array}{lllll} 0 \leq & x_j & \perp & c_j x_j + a_j - \mathbb{E}\big[\lambda_j(\boldsymbol{\xi})\big] & \geq \quad 0, \\ 0 \leq_{a.s.} & y_j(\xi) & \perp_{a.s.} & -p_j(\xi) + \gamma(\xi) \sum\limits_{i \neq j}^{J} y_i(\xi) + 2\gamma(\xi) y_j(\xi) + \lambda_j(\xi) & \geq_{a.s.} 0, \\ 0 \leq_{a.s.} & \lambda_j(\xi) & \perp_{a.s.} & x_j - y_j(\xi) & \geq_{a.s.} 0. \end{array} \tag{5.8}$$

It then follows that since all agents in oligopolistic market act non-cooperatively, the collective equilibrium can be interpreted as that of the whole market. More specifically, let

$$x = (x_1, \ldots, x_J)^T$$

be the first-stage decision vectors of the system, and for a.e. $\xi \in \Xi$, second-stage decision vector

$$y(\xi) = \big(y_1(\xi), \ldots, y_J(\xi)\big)^T$$

and the corresponding multiplier vector

$$\lambda(\xi) = \big(\lambda_1(\xi), \ldots, \lambda_J(\xi)\big)^T$$

are denoted respectively. Analogously, parameter vectors can be written as

$$a = (a_1, \ldots, a_J)^T,$$

$$p(\xi) = \big(p_1(\xi), \ldots, p_J(\xi)\big)^T.$$

Then, the SVI for all agents can be treated as a two-stage SCP:

$$\begin{aligned} 0 \leq x \quad & \perp Cx - \mathbb{E}[\lambda(\boldsymbol{\xi})] + a \quad && \geq 0, \\ 0 \leq \begin{pmatrix} y(\xi) \\ \lambda(\xi) \end{pmatrix} \quad & \perp \begin{pmatrix} \Pi(\xi) & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} y(\xi) \\ \lambda(\xi) \end{pmatrix} + \begin{pmatrix} -p(\xi) \\ x \end{pmatrix} && \geq 0, \quad \text{for a.e. } \xi \in \Xi, \end{aligned} \tag{5.9}$$

where

$$C = \operatorname{diag}(c_1, c_2, ..., c_J),$$

$$\Pi(\xi) = \gamma(\xi)(ee^T + I).$$

It follows that for the whole system, a J-tuple of triplets

$$\left(x^*, y^*, \lambda^*\right) = \left((x_1^*, y_1^*, \lambda_1^*), \ldots, (x_J^*, y_J^*, \lambda_J^*)\right) \in \mathbb{R}^J \times \mathfrak{L}_J^\infty \times \mathfrak{L}_J^1$$

is called a solution of the two-stage SCP (5.9).

5.2 Structure of two-stage stochastic complementarity problem

In this section, the focus is on characterizing solutions of two-stage stochastic linear complementarity problem (5.9). From the derivation of first-order necessary optimality conditions of problem (5.6)-(5.7) and the monotonicity of problem (5.9), the following results can be derived on existence of solutions.

Proposition 5.1 (Theorem 2, [88]). *For any given fixed pair $(x, \xi) \in \mathbb{R}_+^J \times \Xi$, the second-stage problem (5.7) has a unique solution.*

Remark 5.5. *If the second-stage problem (5.7) is treated with given first-stage decision, one can write down its first order necessary optimality condition. The difficulty*

is that in the saddle-point characterization of the Lagrangian formulation of problems (5.6)-(5.7) the first order necessary condition cannot be written. However, the whole system (5.9) can be regarded as a necessary condition for achieving an equilibrium of problems (5.6)-(5.7).

Considering the entire two-stage decision process, represented by the two-stage stochastic linear complementarity problem (5.9), the following proposition holds.

Proposition 5.2. *The two-stage stochastic linear complementarity problem (5.9) has relatively complete recourse, i.e., for any $x \in \mathbb{R}^J_+$ and a.e. $\xi \in \Xi$ the second-stage problem of (5.9) is solvable.*

Proof. The coefficient matrix of the second-stage part of (5.9)

$$M(\xi) = \begin{pmatrix} \Pi(\xi) & I \\ -I & 0 \end{pmatrix}$$

is positive semidefinite for a.e. $\xi \in \Xi$. For any given $x \in \mathbb{R}^J_+$, it follows that there always exists a pair $\left(\hat{y}(\xi), \hat{\lambda}(\xi)\right) \in \mathbb{R}^J \times \mathbb{R}^J$, such that

$$\begin{pmatrix} \hat{y}(\xi) \\ \hat{\lambda}(\xi) \end{pmatrix} \geq 0, \quad \begin{pmatrix} \Pi(\xi) & I \\ -I & 0 \end{pmatrix} \begin{pmatrix} \hat{y}(\xi) \\ \hat{\lambda}(x) \end{pmatrix} + \begin{pmatrix} -p(\xi) \\ x \end{pmatrix} \geq 0, \qquad \text{for a.e. } \xi \in \Xi.$$

In details, the special choices of $\hat{y}(\xi) = 0$ and $\hat{\lambda}(\xi) = \max\{0, p(\xi)\}$ can always be taken, where the max function is taken component-wise. Then, the corresponding quadratic programming problem of the linear complementarity problem is feasible regardless of the choice of x. Thus, it follows from [74, Lemma 3.1.1, Theorem 3.1.2] that there exist at least one solution which solves the second-stage problem for any given pair $(x, \xi) \in \mathbb{R}^J_+ \times \Xi$. □

Although the second-stage problem (5.7) has a unique equilibrium for any given $(x, \xi) \in \mathbb{R}^J_+ \times \Xi$, the equilibrium system (5.9) may admit multiple solutions. This can be easily shown with an illustrative example.

Example 5.1. *Consider a duopoly game, i.e., $x = (x_1, x_2)^T \geq 0$, and $-p(\xi) \geq_{a.s.} 0$. If then follows that the corresponding second-stage part of complementarity system (5.9) reads*

$$0 \leq \begin{pmatrix} y_1(\xi) \\ y_2(\xi) \\ \lambda_1(\xi) \\ \lambda_2(\xi) \end{pmatrix} \perp \begin{pmatrix} 2\gamma(\xi) & \gamma(\xi) & 1 & 0 \\ \gamma(\xi) & 2\gamma(\xi) & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_1(\xi) \\ y_2(\xi) \\ \lambda_1(\xi) \\ \lambda_2(\xi) \end{pmatrix} + \begin{pmatrix} -p_1(\xi) \\ -p_2(\xi) \\ x_1 \\ x_2 \end{pmatrix} \geq 0, \quad \text{for a.e. } \xi \in \Xi. \tag{5.10}$$

Then, the solution set of (5.10) takes the form

$$\left\{ (0, 0, \tilde{\lambda}_1(\xi), \tilde{\lambda}_2(\xi)) : \tilde{\lambda}_1(\xi) = \begin{cases} 0, & x_1 > 0 \\ \lambda_1(\xi), & x_1 = 0 \end{cases}, \ \tilde{\lambda}_2(\xi) = \begin{cases} 0, & x_2 > 0 \\ \lambda_2(\xi), & x_2 = 0 \end{cases}, \quad \text{for a.e. } \xi \in \Xi \right\},$$

where $\lambda_1(\xi) \geq_{a.s.} 0, \lambda_2(\xi) \geq_{a.s.} 0$.

In Example 5.1, multiplier λ may admit multiple values when there exist some zero-valued components of x.

Technically, the multiple solutions of the second-stage problem will cause trouble when handling the two-stage stochastic complementarity system (5.9), both in computation and analysis, see [93] for more detailed discussion. To avoid this, the assumption ensuring the uniqueness of second-stage solution is usually made, see for instance [13, 15].

Remark 5.6. *In literatures in finance and economics, the terms "equilibrium price" or "shadow price" are used to denote multipliers corresponds to the constraints. In-*

terpreted as an "equilibrium price" associated with agents' production clearing, multiple values of λ would have ambiguous economical interpretations.

Motivated by these, a partial regularization modification is proposed to seek for one particular choice of equilibrium prices. For any $\epsilon > 0$, let

$$M^{\epsilon}(\xi) = \begin{pmatrix} \Pi(\xi) & I \\ -I & \epsilon I \end{pmatrix} \text{ and } q(x,\xi) = \begin{pmatrix} -p(\xi) \\ x \end{pmatrix}$$

and thus, the regularized second-stage SCP can be written as

$$0 \leq \begin{pmatrix} y(\xi) \\ \lambda(\xi) \end{pmatrix} \perp M^{\epsilon}(\xi) \begin{pmatrix} y(\xi) \\ \lambda(\xi) \end{pmatrix} + q(x,\xi) \geq 0, \quad \text{for a.e. } \xi \in \Xi. \tag{5.11}$$

Then, the corresponding regularized SCP of (5.9) is in the following form.

$$\begin{aligned} 0 \leq x \quad & \perp Cx - \mathbb{E}[\lambda(\boldsymbol{\xi})] + a \quad && \geq 0, \\ 0 \leq \begin{pmatrix} y(\xi) \\ \lambda(\xi) \end{pmatrix} \quad & \perp M^{\epsilon}(\xi) \begin{pmatrix} y(\xi) \\ \lambda(\xi) \end{pmatrix} + q(x,\xi) \quad && \geq 0, \quad \text{for a.e. } \xi \in \Xi. \end{aligned} \tag{5.12}$$

For a given pair $(x,\xi) \in \mathbb{R}^J_+ \times \Xi$, the second-stage problem of (5.9) and the regularized second-stage problem (5.11) are denoted by $\mathrm{LCP}(q(x,\xi), M(\xi))$ and $\mathrm{LCP}(q(x,\xi), M^{\epsilon}(\xi))$ respectively. Their solution functions are chosen from the respective solution sets and expressed by $z(q(x,\xi))$ and $z^{\epsilon}(q(x,\xi))$, detailed analysis can be seen in [16]. In the sequel, the expression ξ and x are omitted without causing any confusion, i.e., $\mathrm{LCP}(q,M) := \mathrm{LCP}(q(x,\xi), M(\xi))$ and $\mathrm{LCP}(q,M^{\epsilon}) := \mathrm{LCP}(q(x,\xi), M^{\epsilon}(\xi))$.

For clearer demonstration, recall that the illustrative Example 5.1, and consider its partially regularized approach. Thus, the second-stage of the regularized problem

takes the following form

$$0 \leq \begin{pmatrix} y_1(\xi) \\ y_2(\xi) \\ \lambda_1(\xi) \\ \lambda_2(\xi) \end{pmatrix} \perp \begin{pmatrix} 2\gamma(\xi) & \gamma(\xi) & 1 & 0 \\ \gamma(\xi) & 2\gamma(\xi) & 0 & 1 \\ -1 & 0 & \epsilon & 0 \\ 0 & -1 & 0 & \epsilon \end{pmatrix} \begin{pmatrix} y_1(\xi) \\ y_2(\xi) \\ \lambda_1(\xi) \\ \lambda_2(\xi) \end{pmatrix} + \begin{pmatrix} -p_1(\xi) \\ -p_2(\xi) \\ x_1 \\ x_2 \end{pmatrix} \geq 0, \quad \text{for a.e. } \xi \in \Xi. \tag{5.13}$$

Under the same condition as in Example 5.1, the unique solution of (5.13) can be obtained. In this example, the trivial solution $(\tilde{y}_1, \tilde{y}_2, \tilde{\lambda}_1, \tilde{\lambda}_2)^T = (0, 0, 0, 0)^T$ for a.e. $\xi \in \Xi$ is the unique solution. Due to the positive definiteness of C, it follows that the unique solution of the first-stage problem is $(x_1, x_2)^T = (0, 0)^T$. Then, for any $\epsilon > 0$, one particular solution of the original problem is obtained, the trivial solution in this example.

Remark 5.7. *The key feature of regularized method is that it promises the existence and uniqueness of solution due to the strongly monotone of regularized two-stage problem.*

In the remaining of this section, the relation between the first-stage and second-stage solutions are explored via the solution z^ϵ of LCP(q, M^ϵ). It will later help to analyze the structure of the second-stage solution.

Proposition 5.3. *For any fixed $\epsilon > 0$, the regularized problom (5.12) has a unique solution $(x^\epsilon, y^\epsilon, \lambda^\epsilon) \in \mathbb{R}^J \times \mathcal{Y} \times \mathcal{Y}$.*

Proof. The result can be obtained via a similar procedure as in [15, Proposition 2.1 (i)] and one needs to show that the condition in [15, Assumption 1] holds. Recall

that from Assumption 5.1, for a.e. $\xi \in \Xi$, it holds that

$$\begin{pmatrix} x \\ u(\xi) \\ v(\xi) \end{pmatrix}^T \begin{pmatrix} C & 0 & -I \\ 0 & \Pi(\xi) & I \\ I & -I & \epsilon I \end{pmatrix} \begin{pmatrix} x \\ u(\xi) \\ v(\xi) \end{pmatrix} \geq \tau(\|x\|^2 + \|u(\xi)\|^2 + \|v(\xi)\|^2),$$

where $\tau = 2\min\{\bar{c}, \gamma_0(J+1), \epsilon\}$ with $\bar{c}$ denoting the minimum diagonal element of C. $\square$

The following result gives the potential expressions of the second-stage decision, provided the first-stage decision is made and the observation of $\boldsymbol{\xi}$ is realized. This can be understood as if the agent has advanced in the decision horizon as to observe the second-stage execution of the market.

Theorem 5.1. *For any fixed $\epsilon > 0$, $x \geq 0$ and a.e. $\xi \in \Xi$, the j-th component of the solution of problem (5.11) $\left((y^\epsilon)_j, (\lambda^\epsilon)_j\right)$ is either $(0,0)$, or one of the following two forms:*

$$\begin{aligned} &\left(-\frac{\gamma(\xi)T^\epsilon - p_j(\xi)}{\gamma(\xi)}, \qquad 0\right), \\ &\left(-\frac{\epsilon(\gamma(\xi)T^\epsilon - p_j(\xi)) - x_j}{\epsilon\gamma(\xi)+1}, \quad \frac{\gamma(\xi)(T^\epsilon + x_j) - p_j(\xi)}{\epsilon\gamma(\xi)+1}\right) \end{aligned} \tag{5.14}$$

for $j \in \mathcal{J}$, where

$$T^\epsilon := \sum_{i=1}^{J} (y^\epsilon)_i = \frac{\gamma(\xi)\sum_{i\in\mathcal{I}_3} x_i + \epsilon\gamma(\xi)\sum_{i\in\mathcal{I}_2\cup\mathcal{I}_3} p_i(\xi) + \sum_{i\in\mathcal{I}_2} p_i(\xi)}{\big(\epsilon\gamma(\xi)(|\mathcal{I}_2| + |\mathcal{I}_3| + 1) + |\mathcal{I}_2| + 1\big)\gamma(\xi)} \tag{5.15}$$

with

$$\begin{aligned} \mathcal{I}_2 &= \{j \in \mathcal{J} : \gamma(\xi)T^\epsilon + (\lambda^\epsilon)_j - p_j(\xi) < 0, \quad (y^\epsilon)_j - x_j \leq 0\}, \\ \mathcal{I}_3 &= \{j \in \mathcal{J} : \gamma(\xi)T^\epsilon + (\lambda^\epsilon)_j - p_j(\xi) < 0, \quad (y^\epsilon)_j - x_j > 0\}, \end{aligned}$$

where $|\mathcal{I}_2|$ and $|\mathcal{I}_3|$ denote the cardinality of $\mathcal{I}_2$ and $\mathcal{I}_3$ respectively.

The detailed proof of the Theorem 5.1 is given in Appendix. For detailed discussion on the expression of the second-stage solution, the interested readers are referred to the Ph.D thesis of Mr. Jie Jiang [46], a collaborator of the author who partly contributed to the theoretical analysis in the study of oil market share. Note that the above theorem gives the forms of the unique solution of the second-stage regularized problem (5.11). However, it may not be used to aid numerical calculation since the partition of the index sets $\mathcal{I}$s is not known in advance. Nevertheless, it is suffice for the purposes of deriving structural properties of the solutions.

Due to the positive definiteness of M^ϵ and special structure of problem (5.11), the following Lipschitz continuous property can be seen, following [16, Corollary 2.1].

Lemma 5.1. *There exists $L > 0$ such that for any fixed $\epsilon \in (0, 1]$, it holds*

$$\|z^\epsilon\big(q(x_1, \xi)\big) - z^\epsilon\big(q(x_2, \xi)\big)\| \leq L\|x_1 - x_2\|, \quad \textit{for} \quad x_1, x_2 \in \mathbb{R}^J_+ \text{ and a.e. } \xi \in \Xi.$$

An interesting result derived from the proposed model is that an upper bound of the total supply to the market can be expressed for any given first-stage decision x and realization $\xi \in \Xi$.

Lemma 5.2. *For any fixed $\epsilon > 0$ and $(x, \xi) \in \mathbb{R}^J_+ \times \Xi$, T^ϵ has the following upper bound:*

$$T^\epsilon \leq \|x\|_1 + \left(\epsilon + \frac{1}{\gamma(\xi)}\right) \|p(\xi)\|_1.$$

Proof. The following derivation follows from (5.15) that

$$\begin{aligned} T^\epsilon =& \frac{\gamma(\xi)\sum_{i\in\mathcal{I}_3} x_i + \epsilon\gamma(\xi)\sum_{i\in\mathcal{I}_2\cup\mathcal{I}_3} p_i(\xi) + \sum_{i\in\mathcal{I}_2} p_i(\xi)}{(\epsilon\gamma(\xi)(|\mathcal{I}_2|+|\mathcal{I}_3|+1)+|\mathcal{I}_2|+1)\,\gamma(\xi)} \\ \leq& \frac{\gamma(\xi)\sum_{i=1}^J x_i + (\epsilon\gamma(\xi)+1)\sum_{i=1}^J |p_i(\xi)|}{(\epsilon\gamma(\xi)(|\mathcal{I}_2|+|\mathcal{I}_3|+1)+|\mathcal{I}_2|+1)\,\gamma(\xi)} \\ \leq& \frac{\gamma(\xi)\sum_{i=1}^J x_i + (\epsilon\gamma(\xi)+1)\sum_{i=1}^J |p_i(\xi)|}{\gamma(\xi)} \\ =& \|x\|_1 + \left(\epsilon + \frac{1}{\gamma(\xi)}\right)\|p_1(\xi)\|_1. \end{aligned}$$

□

As seen at the time before second-stage decisions are made, the convergence result of the second-stage LCP(q, M^ϵ) solutions can be established as $\epsilon \downarrow 0$ for any given pair $(x, \xi) \in \mathbb{R}^J \times \Xi$.

Proposition 5.4. *For any fixed $\epsilon > 0$ and $(x, \xi) \in \mathbb{R}^J_+ \times \Xi$, let $z^\epsilon(\xi) = (y^\epsilon(\xi), \lambda^\epsilon(\xi))$ denote the unique solution of the regularized problem LCP(q, M^ϵ). Then*

$$\lim_{\epsilon\downarrow 0} \|z^\epsilon(\xi) - \bar{z}(\xi)\| = 0,$$

where $\bar{z}(\xi) = (\bar{y}(\xi), \bar{\lambda}(\xi))$ denotes the unique least l_2-norm solution of the LCP(q, M). Moreover, the j-th component of the least l_2-norm solution of problem (5.9) has one of the following three forms:

$$\left\{(0,0), \left(-\frac{\gamma(\xi)\bar{T} - p_j(\xi)}{\gamma(\xi)}, 0\right), \left(x_j, -\gamma(\xi)(\bar{T}+x_j) + p_j(\xi)\right)\right\} \tag{5.16}$$

for $j \in \mathcal{J}$, where

$$\bar{T} := \lim_{\epsilon\downarrow 0} T^\epsilon = \sum_{i=1}^J \bar{y}_i.$$

Furthermore, for a.e. $\xi \in \Xi$ *there exists* $\bar{\kappa}(\xi) > 0$*, such that*

$$\|\lambda^{\epsilon}(\xi) - \bar{\lambda}(\xi)\| \leq \bar{\kappa}(\xi)\epsilon. \tag{5.17}$$

The details on the proof of Proposition 5.4 is given in Appendix.

5.3 Convergence analysis

In this section, the convergence of the unique solution of the regularized problem (5.12) to the solution set of the original problem as the regularized parameter ϵ decreases to zero is provided. The SAA method is used to solve the regularized problem for fixed choice of regularization parameter, similar approach can be seen in [14]. Combined with the results from the regularization approaches, the convergence property of the solution of the regularized SAA model as the number of samples goes to infinity is demonstrated.

5.3.1 Convergence of the regularized model

For any given first-stage decision, the convergence results for the second-stage have been shown in previous subsection. Thus, only the convergence properties of the first-stage decision vector needs to be considered in this part of the thesis, i.e., $x^{\epsilon} \in \mathbb{R}^{J}_{+}$ that solves problem (5.12) when the regularized parameter ϵ tends to zero. Then, the convergence property of the solution $(x^{\epsilon}, y^{\epsilon}, \lambda^{\epsilon})$ follows. From Proposition 5.3, for fixed $\epsilon > 0$ problem (5.12) admits a unique first-stage solution x^{ϵ}.

In the following, we consider the sequence of accumulation points of $\{x^{\epsilon}\}$ as $\epsilon \downarrow 0$.

For the existence of accumulation points, the following result about the boundedness of the first-stage solutions can be shown.

Proposition 5.5. *Suppose there exists $p_0 > 0$ such that for all $j \in \mathcal{J}$, $p_j(\xi) \leq_{a.s.} p_0$. Then, with $\epsilon \downarrow 0$, $\{x^\epsilon\}$ is bounded.*

Proof. From the condition on $p_j(\xi)$, there must exist a sufficiently large $\alpha > 0$ such that for any $j \in \mathcal{J}$

$$\gamma_0 \alpha - p_j(\xi) > 0, \quad \text{for a.e. } \xi \in \Xi.$$

Then it holds

$$-\frac{\gamma(\xi)(T^\epsilon + \alpha) - p_j(\xi)}{\epsilon\gamma(\xi) + 1} < 0, \quad \text{for a.e. } \xi \in \Xi.$$

Assume that $\{x^\epsilon\}$ is unbounded for the purpose of achieving a contradiction. Then, it follows that there exist some indices $j \in \mathcal{J}$, such that $(x^\epsilon)_j \geq \alpha$. Considering the j-th component of the first-stage complementarity relation,

$$0 \leq (x^\epsilon)_j \perp c_j(x^\epsilon)_j - \mathbb{E}\big[\big(\lambda^\epsilon(\boldsymbol{\xi})\big)_j\big] + a_j \geq 0,$$

which can be expressed, from (5.14), as

$$0 \leq (x^\epsilon)_j \perp c_j(x^\epsilon)_j + a_j \geq 0.$$

However, this complementarity relation cannot be obtained because $(x^\epsilon)_j > 0$ and $c_j(x^\epsilon)_j + a_j > 0$. This completes the proof. □

Remark 5.8. *Note that the conditions $p_j(\xi) \leq_{a.s.} p_0$ can be easily satisfied in many practical applications. For example, with finite data set of $p(\xi)$ with given $\xi \in \Xi$, it can always find an upper bound $p_0 := \max_j\{p_j(\xi)\}$.*

The next lemma concerns about the convergence results on the expected equilibrium price.

Lemma 5.3. *Suppose there exists a constant $p_0 > 0$ such that, for all $j \in \mathcal{J}$, $p_j(\xi) \leq_{a.s.} p_0$. Then, there exists a subsequence $\{\epsilon_k\}_{k=1}^{\infty}$ with $\epsilon_k \downarrow 0$ as $k \to \infty$ such that with $\lambda^k = \lambda^{\epsilon_k}$,*

$$\lim_{k\to\infty} \mathbb{E}[\lambda^k(\boldsymbol{\xi})] = \mathbb{E}[\bar{\lambda}(\boldsymbol{\xi})].$$

Proof. From the results of Proposition 5.5, there must exist an accumulation point of $\{x_\epsilon\}$ as $\epsilon \downarrow 0$, denoted by $\hat{x}$. Taking a subsequence $\{\epsilon_k\}_{k=1}^{\infty}$ with $\epsilon_k \downarrow 0$ as $k \to \infty$ if necessary. Deriving from Proposition 5.5 and inequality (5.17) that

$$\begin{aligned}
&\left\|\mathbb{E}[\lambda^k(\boldsymbol{\xi})] - \mathbb{E}[\bar{\lambda}(\boldsymbol{\xi})]\right\| \\
\leq &\left\|\mathbb{E}[\lambda^k(\boldsymbol{\xi})] - \mathbb{E}[\hat{\lambda}(\boldsymbol{\xi})]\right\| + \left\|\mathbb{E}[\hat{\lambda}(\boldsymbol{\xi})] - \mathbb{E}[\bar{\lambda}(\boldsymbol{\xi})]\right\| \\
\leq &\, \mathbb{E}[\bar{\kappa}(\boldsymbol{\xi})]\epsilon_k + \left\|\mathbb{E}[\hat{\lambda}(\boldsymbol{\xi})] - \mathbb{E}[\bar{\lambda}(\boldsymbol{\xi})]\right\|,
\end{aligned}$$

where for given $\xi \in \Xi$, $\hat{\lambda}(\xi)$ denotes the multiplier component of the least norm solution of problem (5.9) with $\hat{x}$. It follows from Lemma 5.1 that for a.e. $\xi \in \Xi$

$$\left\|\bar{\lambda}(\xi) - \hat{\lambda}(\xi)\right\| = 0.$$

Furthermore, the estimation

$$\begin{aligned}
\left\|\bar{\lambda}(\xi) - \hat{\lambda}(\xi)\right\| &\leq \|\bar{\lambda}(\xi)\| + \left\|\hat{\lambda}(\xi)\right\| \\
&\leq 4\sqrt{J}\big(\gamma(\xi)\|w\|_1 + \|p(\xi)\|_1\big), \quad \text{for a.e. } \xi \in \Xi,
\end{aligned}$$

can be derived where the last term comes from (5.16) with some $w > 0$. It then follows from the Lebesgue Dominated Convergence Theorem,

$$\left\|\mathbb{E}[\hat{\lambda}(\boldsymbol{\xi})] - \mathbb{E}[\bar{\lambda}(\boldsymbol{\xi})]\right\| = 0.$$

□

Theorem 5.2. *Any accumulation triplet of* $\{x^\epsilon, y^\epsilon, \lambda^\epsilon\}$ *as* $\epsilon \downarrow 0$ *is a solution of problem* (5.9).

Proof. The proof requires to verify that for any $\epsilon_k \downarrow 0$, the accumulation point $\hat{x}$ of subsequence $\{x^{\epsilon_k}\}$ is a first-stage solution of (5.9). Since x^{ϵ_k} is the first-stage solution of problem (5.12) for any $\epsilon_k > 0$, it has with $x^k = x^{\epsilon_k}$

$$0 \leq x^k \perp Cx^k - \mathbb{E}[\lambda^k(\boldsymbol{\xi})] + a \geq 0,$$

which, by using the 'min' NCP function (see, for example, [74]), can be rewritten as

$$\min\{x^k, Cx^k - \mathbb{E}[\lambda^k(\boldsymbol{\xi})] + a\} = 0.$$

By Lemma 5.3, it holds

$$0 = \lim_{k\to\infty} \min\{x^k, Cx^k - \mathbb{E}\big[\lambda^k(\xi)\big] + a\} = \min\{\hat{x}, C\hat{x} - \mathbb{E}[\bar{\lambda}(\boldsymbol{\xi})] + a\}$$

as $k \to \infty$. Thus we obtain that

$$\min\{\hat{x}, C\hat{x} - \mathbb{E}[\bar{\lambda}(\xi)] + a\} = 0.$$

The statement then follows from Proposition 5.4. □

5.3.2 Convergence of the regularized SAA model

In this subsection, the SAA scheme for solving the regularized problem (5.12) is studied and the focus is on the convergence of the regularized SAA approach. It is noteworthy that Chen, Sun and Xu considered a discrete approximation scheme in [15], which also leads to an approximation of the response variable in the second-stage problem.

Let $\xi_1, \xi_2, \ldots, \xi_N$ denote N independent identically distributed (i.i.d.) samples. Then, with slight abuse of notation, the following formulation of problem (5.12) with SAA can be obtained:

$$\begin{aligned} &0 \leq x \perp Cx - \frac{1}{N}\sum_{\ell=1}^{N} \lambda(\xi_\ell) + a \geq 0, \\ &0 \leq \begin{pmatrix} y(\xi_\ell) \\ \lambda(\xi_\ell) \end{pmatrix} \perp \begin{pmatrix} \Pi(\xi_\ell) & I \\ -I & \epsilon I \end{pmatrix} \begin{pmatrix} y(\xi_\ell) \\ \lambda(\xi_\ell) \end{pmatrix} + \begin{pmatrix} -p(\xi_\ell) \\ x \end{pmatrix} \geq 0,\ \ell = 1, \ldots, N. \end{aligned} \tag{5.18}$$

Or, the problem can be written collectively for all N samples

$$0 \leq \begin{pmatrix} x \\ v_1 \\ \vdots \\ v_N \end{pmatrix} \perp \begin{pmatrix} C & \frac{1}{N}B & \ldots & \frac{1}{N}B \\ -B^T & D_1^\epsilon & & \\ \vdots & & \ddots & \\ -B^T & & & D_N^\epsilon \end{pmatrix} \begin{pmatrix} x \\ v_1 \\ \vdots \\ v_N \end{pmatrix} + \begin{pmatrix} a \\ q_1 \\ \vdots \\ q_N \end{pmatrix} \geq 0, \tag{5.19}$$

where $C \in R^{J\times J}$, $B = \begin{pmatrix} 0 & -I \end{pmatrix} \in R^{J\times 2J}$, and for $\ell = 1, \ldots, N$, $D_\ell^\epsilon = \begin{pmatrix} \Pi(\xi_\ell) & I \\ -I & \epsilon I \end{pmatrix} \in R^{2J\times 2J}$, $v_\ell = (y(\xi_\ell),\, \lambda(\xi_\ell))^T, q_\ell = (-p(\xi_\ell), 0)^T$. Thus, (5.19) is treated as a large-scale deterministic linear complementarity problem:

$$0 \leq z \perp H^\epsilon z + \bar{q} \geq 0, \tag{5.20}$$

where $z = (x, v_1, ..., v_N)^T$, $\bar{q} = (a, q_1, \cdots, q_N)^T$, and H^ϵ denotes the coefficient matrix in (5.19).

The following assertion of existence and uniqueness of problem (5.18) follows from [74, Theorem 3.1.6].

Proposition 5.6. *For any fixed $\epsilon > 0$ and positive integer N, there exists a unique solution of problem (5.18).*

Recall the result of Lemma 5.1 and the following proposition can be shown in a similar way as in [14, Proposition 3.7].

Proposition 5.7. *Let $(y^\epsilon(\xi), \lambda^\epsilon(\xi))$ be the unique solution of the regularized second-stage problem (5.12) for any $(x, \xi) \in \mathbb{R}^J_+ \times \Xi$. Then,*

$$\frac{1}{N}\sum_{\ell=1}^{N} \lambda^\epsilon(\xi_\ell) \to \mathbb{E}[\lambda^\epsilon(\boldsymbol{\xi})]$$

with probability (w.p.) 1 as $N \to \infty$ uniformly on $\mathbb{B}(x, \delta) \cap \mathbb{R}^J_+$ for any $\delta > 0$,

Let x^ϵ_N denote the first J-components of the unique solution of problem of (5.18), and the following assertion holds.

Lemma 5.4. *Suppose there exists $p_0 > 0$ such that for all $j \in \mathcal{J}$, $p_j(\xi) \leq_{a.s.} p_0$. Then, with $\epsilon \downarrow 0$, $\{x^\epsilon_N\}$ is bounded.*

The proof is omitted since it can be shown analogously as in Proposition 5.5.

Theorem 5.3. *Suppose there exists $p_0 > 0$ such that for all $j \in \mathcal{J}$, $p_j(\xi) \leq_{a.s.} p_0$. Then, for any fixed $\epsilon > 0$, $x^\epsilon_N \to x^\epsilon$ w.p. 1 as $N \to \infty$.*

Proof. From Propositions 5.3, Proposition 5.6, and Lemma 5.4, for any fixed $\epsilon > 0$, both the regularized problem (5.12) and its SAA-regularized problem (5.18) have solutions and contained in some compact subset in $\mathbb{R}^J_+$. Known from Proposition 5.7 that

$$\frac{1}{N}\sum_{\ell=1}^{N} \lambda^\epsilon(\xi_\ell) \to \mathbb{E}[\lambda^\epsilon(\boldsymbol{\xi})]$$

as $N \to \infty$, uniformly w.r.t. x on any compact set. Then, $x_N^\epsilon \to x^\epsilon$ w.p. 1 as $N \to \infty$ by [92, Proposition 19]. $\square$

Combining Theorem 5.2 with Theorem 5.3, the following convergence result relates the original problem and the solution approach of the proposed model.

Theorem 5.4. *Suppose there exists $p_0 > 0$ such that for all $j \in \mathcal{J}$, $p_j(\xi) \leq_{a.s.} p_0$. Then,*

$$\limsup_{\epsilon \downarrow 0} \lim_{N \to \infty} x_N^\epsilon \subseteq S^*,$$

w.p. 1, where S^* denotes the optimal solution set of the first-stage problem of (5.9).

Chapter 6

Numerical Results on Applications

The practical applicability of the model is a great concern for researches in the field of applied mathematics. In this chapter, two models developed in previous chapters are put to test where the historical data are used for both in-sample demonstration and out-sample predictions.

In the application of portfolio selection, the SAA-based CVaR-SSD relaxation model is formulated against different portfolio construction methods in three stock exchange markets. The solutions of the model are then used directly for constructing assets allocation within a series of portfolios. The performances of the portfolios within an investment horizon will be demonstrated and analyzed, compared to both the benchmark portfolios and the solutions of different variations of the proposed SAA-based CVaR-SSD relaxation model. More specifically, one feature that is not obviously apparent in the theoretical formulation of the model is that it allows, to certain extend, flexibility in its formulation stage to be tuned to better perform in specific financial market.

The second application focuses more on the exploration of market mechanism

behind the observation of oil market share behaviour. Namely, one explanation of the interesting fact of stable market share through out the entire history of crude oil trading is approached from the aspects of the Cournot-Nash equilibrium, characterized by solutions of stochastic C-N games. The model is formulated obeying observed facts in crude oil extraction and trading together under mild conditions, widely accepted in both stochastic optimization literatures and economic studies of oil market. Through the numerical experiments attempting to reproduce historical oil market share, certain features in the modelling part of the two-stage stochastic game can be understood more clearly. Moreover, the insights of the numerical results further reflect the specialty of the game theory explanation of the world oil market.

For both applications, the particular computational methods used are firstly described. The data sets used for the experiments are explained in detail which is of research values on its own. One feature worth emphasising for the numerical experiments on general stochastic optimization problems is the method of sampling scenarios of random vectors. More specifically, history as seen at the present stage has already be determined while the realization in the future is not observable.

6.1 Portfolio selection via SAA-based CVaR-SSD relaxation problem

The holy grail of the problem of portfolio selection is, by definition, the construction of better portfolio over a given benchmark. In this section, the results of the empirical studies on SAA-based CVaR-SSD relaxation model are presented. The methodology of identifying advantageous portfolio construction strategies is to compare the resulting portfolios in terms of their performance, and in turn, to analyze

and compare the corresponding strategies of constructing them. All the portfolios under comparison are obtained by solving the corresponding portfolio optimization problems, e.g., index tracking, MV model, SSD model and SAA-based CVaR-SSD relaxation model. More specifically, index portfolios are drawn straightly from data sets as the perfect tracking of the corresponding index values, while MV portfolios are obtained by solving the standard non-shorting MV model, see problem (6.1.1).

For each data set, both index portfolio and MV portfolio are used as benchmarks for constructing and solving SSD problem and SAA-based CVaR-SSD relaxation problem. SSD portfolios are the solutions of SSD problems (4.3) with corresponding benchmarks, and CVaR-SSD portfolios are constructed by solving SAA-based CVaR-SSD relaxation problems (4.9) with both benchmarks and different choices of probability level β. As will be seen in later section, the probability level β may take a range of different values. Note that also lower values of β may be regarded as further relaxation of the SSD constraints, the best β, corresponding to the best CVaR-SSD portfolio, is dependent on particular data structure as well as market characteristics..

The performance of portfolio is quantified in terms of average daily returns, cumulated returns and risk control abilities, measured by standard deviation, Sharpe Ratio and Sortino Ratio[1]. The main finding from the empirical study is that the CVaR-SSD portfolio obtained by solving the proposed SAA-based CVaR-SSD relaxation problem with probability level $\beta < 1$ demonstrates comparable performance

[1] Sortino Ratio [96] is a modification of the Sharpe Ratio but penalizes only those returns falling below a required rate of return, while the Sharpe Ratio penalizes both upside and downside volatility equally. The ratio S is calculated as $S = \frac{R-T}{DR}$, where R is portfolio expected return, T is target rate of return, and DR is target semi-deviation.

w.r.t. SSD portfolio using the same benchmark. Moreover, with appropriate choice of probability level β, it outperforms SSD portfolio in most cases. A rule of thumb is to choose the value of β be less than but close to 1, while the "best fit" value depends on the data set as well as the choice of benchmarks.

6.1.1 Numerical implementation and data sets

The cutting plane algorithm:

The cutting plane method, proposed by Kelley [50], is used to solve SAA-based CVaR-SSD relaxation problem (4.9). By (4.6) and (4.9), let the problem be re-written as the following,

$$\begin{array}{ll} \min\limits_{x,\zeta} & -\frac{1}{N}\sum_{i=1}^{N} x^T\xi_i \\ \text{s.t.} & x \in X_0, \zeta \in \mathbb{R}, \\ & \zeta + \dfrac{1}{(1-\beta)N}\displaystyle\sum_{j=1}^{N}\left(\frac{1}{N}\sum_{i=1}^{N}((\eta_j - x^T\xi_i)_+ - (\eta_j - Y(\xi_i))_+) - \zeta\right)_+ \leq 0. \end{array}$$

For simpler notation, let

$$\Gamma_i(x,\zeta) := \left(\frac{1}{N}\sum_{j=1}^{N}\Big((\eta_j - x^T\xi_i)_+ - (\eta_j - Y(\xi_i))_+\Big) - \zeta\right)_+,$$

and denote

$$\Psi(x,\zeta) \;:=\; \zeta + \frac{1}{(1-\beta)N}\sum_{j=1}^{N}\Gamma_j(x,\zeta).$$

In addition, the following functions are defined for the usage of the cutting plane algorithm,

$$\rho(x,\zeta) \in \partial_x\Psi(x,\zeta), \quad \varrho(x,\zeta) \in \partial_\zeta\Psi(x,\zeta),$$

and

$$b_j(x,\zeta) \in \partial_x \Gamma_j(x,\zeta), \quad d_j(x,\zeta) \in \partial_\zeta \Gamma_j(x,\zeta).$$

More specifically, they can be computed as follows,

$$\rho(x,\zeta) = \frac{1}{(1-\beta)N}\sum_{j=1}^{N} b_j(x,\zeta), \qquad \varrho(x,\zeta) = 1 + \frac{1}{(1-\beta)N}\sum_{j=1}^{N} d_j(x,\zeta),$$

$$b_j(x,\zeta) = \begin{cases} 0, & \Gamma_j(x,\zeta) = 0, \\ \sum_{i=1}^{N} \gamma_{ij}, & others, \end{cases} \qquad \gamma_{ij}(x,\zeta) = \begin{cases} 0, & \eta_j - x^T\xi_i \le 0, \\ -\xi_j, & \eta_j - x^T\xi_i > 0, \end{cases}$$

$$d_j(x,\zeta) = \begin{cases} 0, & \Gamma_j(x,\zeta) = 0, \\ -1, & others. \end{cases}$$

Algorithm 1: Cutting Plane Method
Step 0. Set iteration $t = 0$, tolerance ϵ, $S_0 = \{(x,\zeta)|x \in X_0, \zeta \in \mathbb{R}\}$.
Step 1. Solve the linear programming problem

$$\begin{array}{ll} \min\limits_{(x,\zeta)} & -\frac{1}{N}\sum_{i=1}^{N} x^T\xi_i \\ \text{s.t.} & (x,\zeta) \in S_t. \end{array} \tag{6.1}$$

Let (x_t,ζ_t) denote the optimal solution of problem (6.1).
Step 2. If $\Psi(x_t,\zeta_t) \le \epsilon$, (x_t,ζ_t) is the optimal solution, stop; otherwise, go to Step 3.
Step 3. Set

$$S_{t+1} = S_t \cap \{(x,\zeta)|\Psi(x_t,\zeta_t) + (\rho(x_t,\zeta_t)^T, \varrho(x_t,\zeta_t))((x-x_t)^T, \zeta-\zeta_t)^T \le 0\}.$$

Step 4. t=t+1, go to Step 1.

Algorithm 1 is the standard cutting plane algorithm proposed in [50], where the details and convergence analysis of the algorithm are presented. All the numerical tests in this section are carried out in MATLAB R2014a installed on a Lenovo PC

with Windows 7 operating system and Intel Core i7 processor. IBM ILOG CPLEX Studio 12.4 solver is used for solving the subproblems (6.1) with cutting plane algorithm. The stoping criteria is set to be $\epsilon = 10^{-5}$ for all cases in this section of empirical studies.

Scopes of the experiments and data sets:

In this subsection, the specifications of the numerical experiments are explained and the description of data sets is stated. As per stated, the aim of portfolio selection is to construct portfolios by solving specially designed optimization problems subject to constraints, which reflect the underlining strategies of portfolio construction models. The analysis mainly investigates the following aspects of constructed portfolios.

- Validity of approximation results is tested empirically by comparing CVaR-SSD portfolios and SSD portfolios in-sample.
- Comparisons of the mean, standard deviation, Sharpe Ratio, Sortino Ratio of daily out-of-sample returns, and cumulated returns among SSD portfolios, CVaR-SSD portfolios and their corresponding benchmark portfolios.
- Sparsity results of CVaR-SSD portfolios, SSD portfolios and MV portfolios.

The data sets used are the historical data of three financial indices.

- *National Association of Securities Dealers Automated Quotations 100 index (NDX)*,
- *Standard and Poor's (S&P 500)*,
- *Financial Times Stock Exchange 100 Index (FTSE 100)*,

within the period $01/03/2016 - 30/09/2016$. These indices are treated as (benchmark) portfolios while all other portfolios are constructed based on real data of daily closing prices of the component stocks of the corresponding index. More specifically, as on 30/09/2016, NDX is calculated from the weighted sum of 100 largest non-financial companies based on market capitalization listed on NASDAQ. The daily closing prices of these stocks are publicly available from 01/03/2016 to 30/09/2016, and they form the pool of assets from which MV portfolios, SSD portfolios and CVaR-SSD portfolios are constructed from. Similarly, data sets of daily closing prices are obtained for all the 500 component stocks contributing to S&P 500 index, excepts for *Fortive*, whose was listed within the investigation period after spinning off from *Danaher* in July 2016, and *Westrock*, who was merged with *MeadWestvaco* before listing in *The New York Stock Exchange*. 100 assets listed in the *London Stock Exchange* contributing to FTSE 100 are available for portfolio constructions, except B share of *Royal Dutch Shell*. It is not included because of the lack of full pricing data due to different taxation issues than the A share of the company, which is also one of the 101 contributing stocks to the index. Besides the usage of indices as the choice of benchmark, constructions of the classical MV portfolios served as alternative choices.

For n risky securities, the classical non-shorting MV portfolio has minimal variance for a given expected return ρ and can be obtained by solving the following problem.

$$\begin{array}{ll} \min & x^T C x \\ \text{s.t.} & x^T \mu = \rho, \\ & x^T e = 1, \\ & x \geq 0, \end{array}$$

where e is an n-dimensional vector with all entries being one, random variable $\boldsymbol{r}_i$:

$\Omega \to \mathbb{R}$ is the return of the i^{th} security, $\mu_i = \mathbb{E}[\boldsymbol{r}_i]$ is its expected return and $\boldsymbol{r}(\xi) = (\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_n)^T$, $\mu = (\mu_1, \mu_2, \cdots, \mu_n)^T$, the covariance matrix of the returns $C = \mathbb{E}[(\boldsymbol{r} - \mu)(\boldsymbol{r} - \mu)^T]$. Note that C is an $n \times n$ positive semi-definite matrix. The MV portfolios are constructed and revised at daily basis with rolling window sampling strategy.

6.1.2 Methods and results of empirical studies

In-sample verification of convergence:

The first part of empirical studies is devoted to verify the theoretical analysis on the convergence of SAA-based CVaR-SSD relaxation problem to SSD problem as CVaR probability level $\beta \to 1$. Historical data within the period 01/03/2016−31/05/2016 is used as the in sample data set and the returns of the portfolios obtained by solving SSD problem (4.3) and proposed SAA-based CVaR-SSD relaxation problem (4.9) within this period are back-tested. The daily returns calculated from the difference between the closing prices of adjacent trading days are used to form the observation of random vector of return rates in-sample. Figure 6.1 shows the results in terms of returns in sample with NDX index as benchmark. CVaR-SSD portfolios obtained by solving SAA-based CVaR-SSD problems with different probability levels $\beta = 0.9, 0.8, 0.7$ are shown, and the SSD portfolio in sample returns are also plotted with the same choice of benchmark for comparison. It can be observed that as the probability level β tends to 1, the return plots of CVaR-SSD portfolios converge to that of the SSD portfolio, as demonstrated theoretically in Chapter 4. Note that In-sample tests are also carried out for both the S&P 500 and the FTSE 100 data sets and the same conclusion regarding the in-sample convergence in term of the

tendency of return plots are observed. Due to the similarity of the plots, only the in-sample test for the NDX data set is shown.

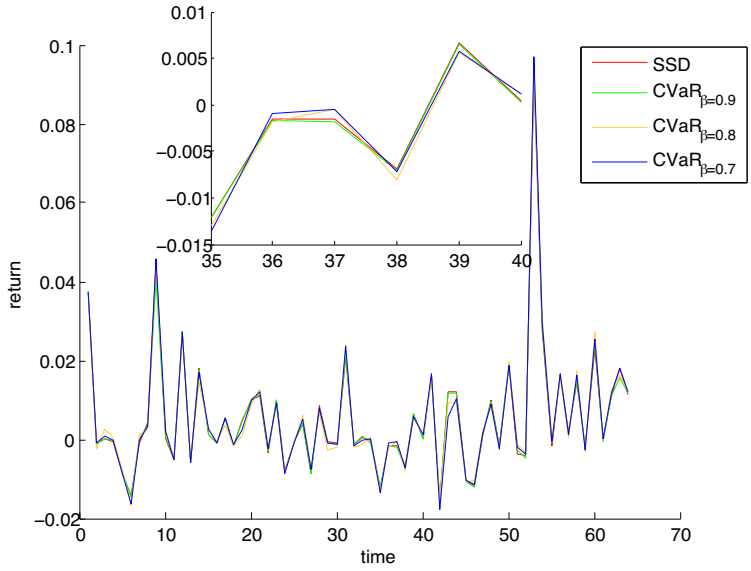

Figure 6.1: In-sample back testing with NDX data with NDX index as benchmark

Portfolio performance of out-of-sample tests:

The main numerical results are shown in this subsection, demonstrations of the results of out-of-sample testing for three sets of market data and with different choices of probability levels β for the SAA-based CVaR-SSD relaxation problem (4.9) are presented. At disposal, the three historical data sets cover the period 01/03/2016 – 30/09/2016. At any point of portfolio construction, the corresponding data set is divided into two categories. For a specific date of portfolio construction, the training part consists of all the available information prior to that date, while the remaining part of the data set is assumed unknown.

The comparisons of out-of-sample performance among portfolios are presented in daily basis. For all the portfolio constructions in this subsection, a "rolling one day" investment period is adapted. Portfolio optimization problems are consistent in the way that on any given date, a fixed size sampling window prior to the decision date is used as training data. More specifically, the daily closing prices of each of the component stocks and the corresponding benchmark portfolio are used within the training data as scenarios to construct current day portfolio. Based on these, different portfolio optimization problems are constructed and solved and their corresponding solutions are the decided portfolios for the trading day following the last day in the training sampling window. The investment horizon is from 01/06/2016 to 30/09/2016 and the portfolio is revised before every trading day, recall the detailed description in Chapter 4.

The indices are firstly used as benchmarks and the MV portfolios are standard to construct. For the first day of out-of-sample portfolio decisions on 01/06/2016, the training data covers the period $01/03/2016 - 31/05/2016$. Within the first period of training data sets, there are 64 trading days for NDX and S&P 500, while there are 62 trading days for FTSE 100. The returns of the portfolio obtained with data in this sampling window is evaluated using the closing prices of the component stocks on that day. Thus, the daily return of the first out-of-sample portfolio can be obtained. For the next trading day, the real closing prices on 01/06/2016 becomes known, and the real data within period $02/03/2016 - 01/06/2016$ is used as the training data set and the portfolio obtained with this data set is evaluated as the second out-of-sample portfolio. By analogy, with a daily rolling window sampling method, training data set will always have 64, for NDX and S&P 500, or 62, for FTSE 100,

scenarios, based on which the next-day portfolio is constructed and evaluated. Thus, for each portfolio optimization model, either 86 portfolio optimization problems or 87 portfolio optimization problems are solved and the performance of portfolios obtained are evaluated on the next working day's historical closing price. Correspondingly, each data set results in 86 or 87 out of sample daily portfolio decisions, covering the period $01/06/2016 - 30/09/2016$.

Concerning the benchmarks, within the training data set, the corresponding indices are available and the MV portfolios are standard to construct. For both benchmarks, the results are presented.

	mean	std	Sharpe Ratio	Sortino Ratio
Benchmark: index	0.0009	0.0088	0.1030	0.1389
SSD	0.0032	0.0118	0.2717	0.4501
$CVaR_{\beta=0.9}$	0.0033	0.0118	0.2784	0.4670
$CVaR_{\beta=0.8}$	0.0032	0.0117	0.2724	0.4486
$CVaR_{\beta=0.7}$	0.0033	0.0119	0.2810	0.4652
Benchmark: MV	0.0005	0.0078	0.0658	0.0841
SSD	0.0026	0.0102	0.2545	0.4129
$CVaR_{\beta=0.9}$	0.0027	0.0118	0.2696	0.4442
$CVaR_{\beta=0.8}$	0.0030	0.0117	0.2931	0.4916
$CVaR_{\beta=0.7}$	0.0030	0.0102	0.2943	0.5004

Table 6.1: NDX: average daily return, standard deviation, Sharpe Ratio, Sortino Ratio

Table 6.1 shows the average daily return rate, standard deviation, Sharpe Ratio and Sortino Ratio of the out-of-sample returns of 8 constructed portfolios as well as two benchmark portfolios using NDX data. With both benchmarks, SSD portfolios and CVaR-SSD portfolios perform better than the corresponding benchmarks with

higher mean, Sharpe Ratio and Sortino Ratio of returns, although with slightly less attractive standard deviation. Note that with NDX index as benchmark, solving SAA-based CVaR-SSD relaxation problem (4.9) with $\beta = 0.7, 0.9$ result in portfolios with slightly larger mean, Sharpe Ratio and Sortino Ratio than SSD portfolio. When $\beta = 0.8$, the performance of CVaR-SSD portfolio is comparable to that of SSD portfolios using the same comparison criteria. In the case of MV benchmark, the CVaR-SSD portfolios behave better than the SSD portfolio for all choices of probability levels for NDX data.

The out-of-sample compounded returns are computed over the period 01/06/2016 - 30/09/2016 and both SSD portfolios and CVaR-SSD portfolios perform better than the corresponding benchmark portfolios as shown in Figure 6.2 and Figure 6.3, respectively. In particular, both the index portfolios and MV portfolios have very small accumulative return, less than 8% for index and less than 5% for MV portfolio, while the CVaR-SSD portfolios perform consistently better though out the whole out of sample period, thus having a cumulated gain of over 30% for our choices of β with index benchmark. In addition, through out almost all the period 01/06/2016 - 30/09/2016, CVaR-SSD portfolios with $\beta = 0.9$ have the best compounded returns in the index benchmark case. For the benchmark being the MV portfolio returns, the portfolios with $\beta = 0.8, 0.7$ behave better than the SSD portfolio and the CVaR-SSD portfolio with $\beta = 0.9$ behaves similar to SSD portfolio in the cumulated return plots.

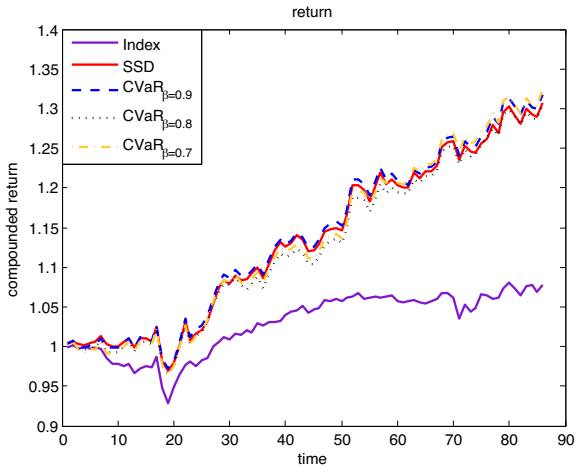

Figure 6.2: NDX: out-of-sample compounded daily returns (01/06/2016 - 30/9/2016), index returns as benchmark

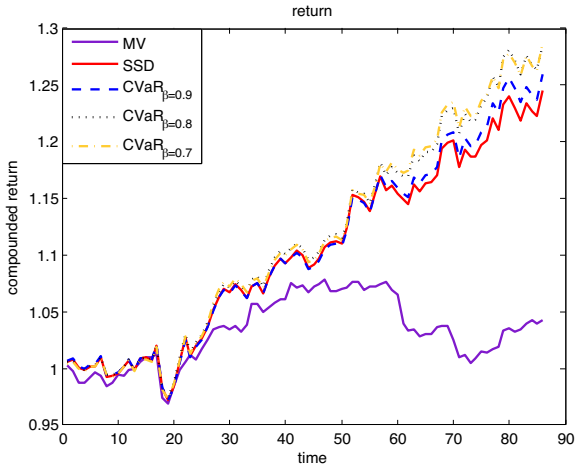

Figure 6.3: NDX: out-of-sample compounded daily returns (01/06/2016 - 30/9/2016), MV returns as benchmark

Table 6.2 shows the out-of-sample performance of comparison portfolios with S&P 500 data. Similar to the results of the NDX data, both SSD portfolios and

CVaR-SSD portfolios have better performance than the corresponding benchmarks. In the index benchmark case, CVaR-SSD portfolios with probability level $\beta = 0.9$ produce better results than that of other CVaR-SSD portfolios. Indeed, Figure 6.4 shows that CVaR-SSD portfolios with probability level $\beta = 0.9$ also have the best compounded return of 17.1% on the final day of test period, which is much better than the index return and maintaining advantageous over the other portfolios including the SSD portfolio. The plots for MV benchmark portfolios shown in Figure 6.5 are rather different in the sense that the MV portfolio, despite being the benchmark, behaves better than other portfolios in earlier stage of out-of-sample testing period. Both SSD portfolio and CVaR-SSD portfolios behave better than MV portfolio when long term observation is of consideration. Moreover, all CVaR-SSD portfolios with different values of β taken behave better than the SSD portfolios by the end of the out-of-sample testing period. It is observed that even the SSD portfolios seems to "fail" for relatively short investment horizon and only gain in a longer investment window. This is due to the structure of SSD constraints (respectively, the SAA-based CVaR-SSD constraints) takes account over a period of volatility of stock prices while remains not very sensitive to sudden and rapid changes in price fluctuation. In this particular example, on 23/06/2016, the sudden event occurred when people of the United Kingdom voted to leave EU, catching the financial markets by surprise, resulted in days of stock prices' free-fall.

	mean	std	Sharpe Ratio	Sortino Ratio
Benchmark: index	0.0004	0.0079	0.0534	0.0705
SSD	0.0017	0.0112	0.1490	0.2264
$CVaR_{\beta=0.9}$	0.0018	0.0111	0.1606	0.2444
$CVaR_{\beta=0.8}$	0.0016	0.0114	0.1442	0.2171
$CVaR_{\beta=0.7}$	0.0017	0.0115	0.1477	0.2241
Benchmark: MV	0.0003	0.0060	0.0421	0.0573
SSD	0.0009	0.0110	0.0802	0.1103
$CVaR_{\beta=0.9}$	0.0012	0.0110	0.1086	0.1517
$CVaR_{\beta=0.8}$	0.0013	0.0107	0.1196	0.1702
$CVaR_{\beta=0.7}$	0.0015	0.0110	0.1383	0.2001

Table 6.2: S&P 500: average daily return, standard deviation,Sharpe Ratio, Sortino Ratio

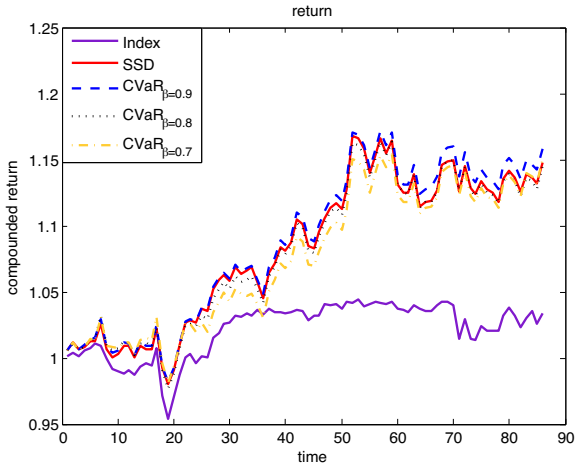

Figure 6.4: S&P 500: out-of-sample compounded daily returns (01/06/2016 - 30/9/2016), index returns as benchmark

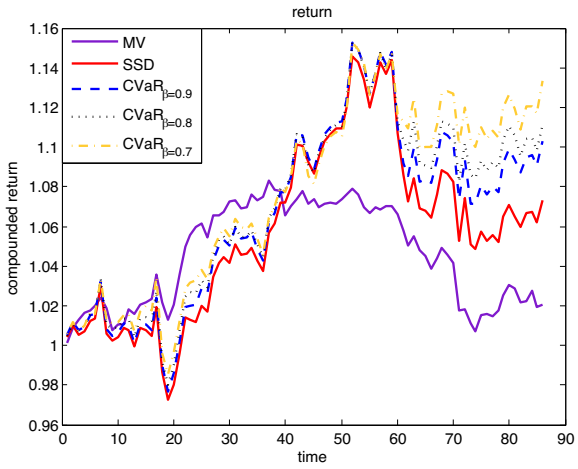

Figure 6.5: S&P 500: out-of-sample compounded daily returns (01/06/2016 - 30/9/2016), MV returns as benchmark

Similar behaviours can also be observed in the case of the FTSE 100 market data, shown in Table 6.3, Figure 6.6 and Figure 6.7. The SSD portfolio and CVaR-SSD portfolios with 3 different β have better out-of-sample performance than the benchmark portfolios. The best of such, the CVaR-SSD portfolio with confidence $\beta = 0.7$ has the best performance with 18.5% cumulated gain in the end of the index benchmark case. As for the case of MV portfolio returns as benchmark, the CVaR-SSD portfolios with $\beta = 0.8, 0.9$ behave comparable to the SSD portfolio and the relaxation leads to better portfolio when $\beta = 0.7$. This observation suggests that although with $\beta < 1$ a relaxation model of portfolio optimization problem can produce better portfolios than the conservative SSD portfolio. The "best" choice of β, or "level" of relaxation, is not universal across different data sets and benchmarks. From the results of empirical tests, it seems that the choice of β is subject to different data structure and benchmark choices.

The SSD portfolios and CVaR-SSD portfolios have been demonstrated to have an overall better performance than the corresponding benchmarks, especially the CVaR-SSD portfolios with probability levels β less but close to 1. CVaR-SSD portfolios with $\beta = 0.9$ is the best for NDX and S&P 500 data sets, and with $\beta = 0.7$, the portfolio outperforms the rest in the case of FTSE 100 data. With MV benchmarks, the portfolios corresponding to $\beta = 0.7$ behave the best over all three sets of data. To summarize, the CVaR-SSD portfolios always behave comparable to that of the SSD portfolios and, in most cases, much better with the same training sets.

	mean	std	Sharpe Ratio	Sortino Ratio
Benchmark: index	0.0012	0.0112	0.1080	0.1685
SSD	0.0017	0.0158	0.1094	0.1848
$CVaR_{\beta=0.9}$	0.0017	0.0155	0.1099	0.1836
$CVaR_{\beta=0.8}$	0.0020	0.0157	0.1254	0.2131
$CVaR_{\beta=0.7}$	0.0021	0.0164	0.1269	0.2185
Benchmark: MV	0.0018	0.0095	0.1901	0.3418
SSD	0.0023	0.0141	0.1606	0.2925
$CVaR_{\beta=0.9}$	0.0021	0.0134	0.1568	0.2747
$CVaR_{\beta=0.8}$	0.0021	0.0136	0.1578	0.2755
$CVaR_{\beta=0.7}$	0.0024	0.0141	0.1703	0.3058

Table 6.3: FTSE 100: average daily return, standard deviation, Sharpe Ratio, Sortino Ratio

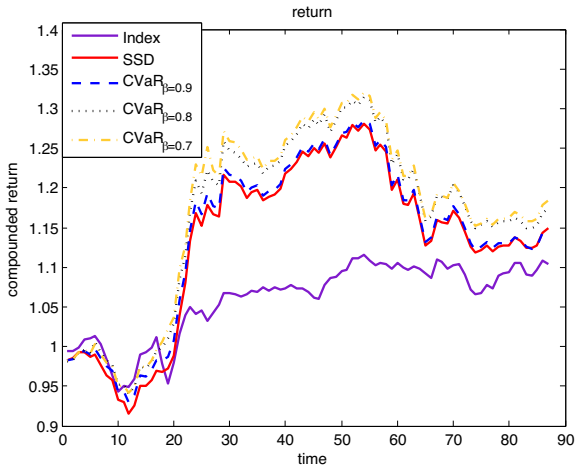

Figure 6.6: FTSE 100: out-of-sample compounded daily returns (01/06/2016 - 30/9/2016), index returns as benchmark

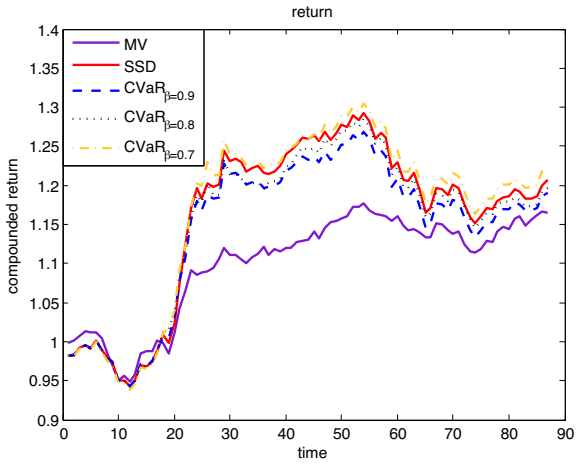

Figure 6.7: FTSE 100: out-of-sample compounded daily returns (01/06/2016 - 30/9/2016), MV returns as benchmark

Sparse portfolios:

Another observation of the constructed portfolios is the number of stocks in the composition of the efficient portfolios (Table 6.1.2). It is observed that the cardinalities of the portfolios have slight variations among different models within the same market. For example, considering the portfolios constructed by solving the SAA-based CVaR-SSD relaxation model with probability level $\beta = 0.9$ in the case of NDX data set with index as benchmark, 19 stocks are "selected" to construct the portfolios over the out-of-sample investment horizon of 86 trading days. On average, daily optimal portfolio has 4.6511 component stocks with the largest basket has 10 stocks. Baskets of similar size are traded for all SSD portfolios and CVaR-SSD portfolios with different probability levels β with both benchmarks in all three data sets, see Table 6.1.2.

Although MV portfolios can also achieve index-like performance with less components, desired under e.g., enhanced indexation setting [104]. Both SSD portfolios and CVaR-SSD portfolios contain much less number of stocks in their compositions, typically less than one tenth of the available stocks, while dominate the corresponding benchmarks in terms of performance.

NDX (100)	Index			MV		
	avg.	min.	max.	avg.	min.	max.
SSD	4.60	3	9	5.55	3	9
$CVaR_{\beta=0.9}$	4.65	3	10	5.53	2	9
$CVaR_{\beta=0.8}$	4.65	3	10	5.51	2	8
$CVaR_{\beta=0.7}$	4.64	3	9	5.50	3	8
FTSE (100)	Index			MV		
	avg.	min.	max.	avg.	min.	max.
SSD	4.05	2	9	5.16	2	9
$CVaR_{\beta=0.9}$	3.98	2	9	5.11	2	9
$CVaR_{\beta=0.8}$	3.90	2	8	5.01	2	8
$CVaR_{\beta=0.7}$	3.91	3	9	5.17	2	9
S&P (500)	Index			MV		
	avg.	min.	max.	avg.	min.	max.
SSD	5.87	3	10	6.62	4	11
$CVaR_{\beta=0.9}$	6.07	3	11	6.49	3	12
$CVaR_{\beta=0.8}$	6.07	3	12	6.70	4	12
$CVaR_{\beta=0.7}$	6.30	3	11	6.74	4	12

Table 6.4: Average, minimum and maximum of daily traded basket sizes of different models with both benchmarks in three data sets

Empirical study results of portfolio selection:

To conclude this section, it is found empirically that:

1. the portfolios obtained by solving SAA-based CVaR-SSD relaxation problem (4.9) converge to that obtained by solving problem (4.3) as $\beta \to 1$;

2. both SSD portfolios and CVaR-SSD portfolios have overall better performance than the corresponding benchmarks;

3. the performance of CVaR-SSD portfolios are always comparable to that of SSD portfolios, and when β is less but close to 1, the performance of CVaR-SSD portfolios is better than the more conservative SSD portfolios in most cases;

4. both SSD portfolios and CVaR-SSD portfolios are sparse portfolios.

6.2 Global oil market share

In this section, the numerical experiments of the second application are carried out firstly using randomly generated data to illustrate the effectiveness of the game model and its solution approach. Furthermore, the proposed two-stage stochastic Nash equilibrium problem is used to describe market share competition over the history of crude oil market. The results show that the model is capable of reproducing the actual oil market share based on in-sample data. Moreover, it is also shown that the model can make good out-of-sample predictions using historical data. All the tests are run in MATLAB 2016b on a personal computer with 32GB RAM and 8-core processor ($3.6 \times 8GHz$).

6.2.1 Progressive hedging method

Randomly generated problems are used for testing the regularized SAA approach to solve the two-stage stochastic Nash equilibrium problem. Recall that the model of interests takes the form of a scenario-based linear complementarity problem (5.18) or its equivalent expression (5.19) with sufficiently small ϵ. The solution process adopts the well-known methods of progressive hedging. The PHM is globally convergent and the convergence rate is linear for problem (5.19) with SAA approach.

The classical PHM to solve (5.19) is as follows

Algorithm 2: Progressive hedging method
Step 0. Given an initial point $x^0 \in \mathbb{R}^J$, let $x_\ell^0 = x^0 \in \mathbb{R}^J, v_\ell^0 \in \mathbb{R}^{2J}$ and $w_\ell^0 \in \mathbb{R}^J$, for $\ell = 1, \ldots, N$, such that $\frac{1}{N}\Sigma_{\ell=1}^N w_\ell^0 = 0$. Set the initial point $z^0 = (x^0, v_1^0, \ldots, v_N^0)^T$. Choose a step size $r > 0$. Set $k = 0$.
Step 1. If the point z^k satisfies the condition

$$\| \min(z^k, H^\epsilon z^k + \bar{q}) \| \leq 10^{-6},$$

output the solution z^k and terminate the algorithm; otherwise, go to Step 2.
Step 2. For $\ell = 1, \ldots, N$, find $(\hat{x}_\ell^k, \hat{v}_\ell^k)$ that solves linear complementarity problems

$$\begin{array}{l} 0 \leq x_\ell \perp C x_\ell + B v_\ell + a + w_\ell^k + r(x_\ell - x_\ell^k) \geq 0, \\ 0 \leq v_\ell \perp -B^T x_\ell + D_\ell^\epsilon v_\ell + q_\ell + r(v_\ell - v_\ell^k) \geq 0. \end{array} \tag{6.2}$$

Then let $\bar{x}^{k+1} = \frac{1}{N}\sum_{\ell=1}^N \hat{x}_\ell^k$, and for $\ell = 1, \ldots, N$, update

$$x_\ell^{k+1} = \bar{x}^{k+1}, v_\ell^{k+1} = \hat{v}_\ell^k, w_\ell^{k+1} = w_\ell^k + r(\hat{x}_\ell^k - x_\ell^{k+1}),$$

to get point $z^{k+1} = (\bar{x}^{k+1}, v_1^{k+1}, \ldots, v_N^{k+1})^T$.
Step 3. Set $k := k + 1$; go back to Step 1.

The PHM involves solving N independently sample-based LCP (6.2) at each iteration, Problem (6.2) is well-defined, since for $\ell = 1, \ldots, N$, the coefficient matrix

$$\begin{pmatrix} C + rI & B \\ -B^T & D_\ell^\epsilon + rI \end{pmatrix} \in \mathbb{R}^{3J \times 3J}$$

is positive definite for any $\epsilon > 0$. Thus, it has a unique solution for each $\ell = 1, \ldots, N$. For simplicity, denote (6.2) as

$$0 \leq z_\ell \perp \tilde{H}_\ell^\epsilon z_\ell + \tilde{q}_\ell \geq 0, \tag{6.3}$$

where $z_\ell = \begin{pmatrix} x_\ell \\ v_\ell \end{pmatrix}, \tilde{H}_\ell^\epsilon = \begin{pmatrix} C + rI & B \\ -B^T & D_\ell^\epsilon + rI \end{pmatrix}, \tilde{q}_\ell = \begin{pmatrix} a + w_\ell^k - r x_\ell^k \\ q_\ell - r v_\ell^k \end{pmatrix}$ with $\ell = 1, 2, \ldots, N$. Then, a large-scale LCP can be equivalently written for all N samples

and solve

$$0 \leq \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{pmatrix} \perp \begin{pmatrix} \tilde{H}_1^\epsilon & & & \\ & \tilde{H}_2^\epsilon & & \\ & & \ddots & \\ & & & \tilde{H}_N^\epsilon \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_N \end{pmatrix} + \begin{pmatrix} \tilde{q}_1 \\ \tilde{q}_2 \\ \vdots \\ \tilde{q}_N \end{pmatrix} \geq 0. \tag{6.4}$$

The structure of problem (6.4) enables us to use block computation to solve it, which can significantly improves the efficiency of the PHM. For example, suppose $N = mL$ with m, L being positive integer number. The equivalent m-block reformulation for (6.4) reads

$$0 \leq \begin{pmatrix} z_{i_1} \\ z_{i_2} \\ \vdots \\ z_{i_m} \end{pmatrix} \perp \begin{pmatrix} \tilde{H}_{i_1}^\epsilon & & & \\ & \tilde{H}_{i_2}^\epsilon & & \\ & & \ddots & \\ & & & \tilde{H}_{i_m}^\epsilon \end{pmatrix} \begin{pmatrix} z_{i_1} \\ z_{i_2} \\ \vdots \\ z_{i_m} \end{pmatrix} + \begin{pmatrix} \tilde{q}_{i_1} \\ \tilde{q}_{i_2} \\ \vdots \\ \tilde{q}_{i_m} \end{pmatrix} \geq 0, \tag{6.5}$$

where $i_j = iL + j, j = 1, \ldots, m, i = 0, \ldots, (m-1)L$.

Remark 6.1. *The main computation cost of the PHM is in* **Step 2** *due to the large sample size N despite the relative low cost in solving for one sample. Block implementation speeds up the computation by better exploring and rebalancing the computational load. In practice, the number of blocks are adjusted so that the overall computation time is minimized.*

To improve the efficiency of the PHM, the warm-start technique is also adopted, as suggested in [77] in choosing an initial point for subproblem (6.2). More specifically, the solution z^k of subproblem (6.2) at the k-th iteration is used as a starting point for the $(k+1)$-th iteration.

In the remaining of this subsection, the problem (6.3) for a given sample is analyzed in detail and the subscription $_\ell$ is omitted for ease of expression. In order to take advantage of the sparse structure of the subproblem (6.3), the smoothing Newton method proposed by Chen and Ye [18] is used. In what follows, a brief description of the smoothing Newton method is presented. It is well-known that solving (6.3) for a given sample is equivalent to solving the nonsmooth equation

$$F(z) = \min(z, \tilde{H}^\epsilon z + \tilde{q}) = 0. \tag{6.6}$$

The main idea of the smoothing Newton method is to use a smooth approximation function to approximate the nonsmooth function F and then solve the corresponding linear system. The smoothing is achieved by using the smooth Gariel-Moré approximation function $f : \mathbb{R}^{3J} \times \mathbb{R}_{++} \to \mathbb{R}^{3J}$ to approximate the nonsmooth function F. In numerical tests, the j-th component of f reads

$$f_j(z, \delta) = (z)_j - \frac{1}{2}\left(\sqrt{(\tilde{H}^\epsilon z + \tilde{q} - z)_j^2 + 4\delta^2} + (z - \tilde{H}^\epsilon z - \tilde{q})_j\right), \quad j = 1, \ldots, 3J,$$

where the corresponding j-th diagonal element of the Jacobian $\bar{D}(z_i)$ is

$$\bar{D}_{jj} = \frac{1}{2}\left(\frac{(z - \tilde{H}^\epsilon z - \tilde{q})_j}{\sqrt{(z - \tilde{H}^\epsilon z - \tilde{q})_j^2 + 4\delta^2}} + 1\right), \quad j = 1, \ldots, 3J.$$

Then, the smoothing Newton method for solving subproblem (6.6) requires to solve a linear equation to determine d^k at each iteration, namely

$$\nabla_z f(z^k, \delta_k) d^k + F(z^k) = 0, \tag{6.7}$$

where δ_k decreases to 0 according to the criterion in article [18]. To guarantee the well-defineness of the (6.7), the following result needs to be applied.

Theorem 6.1. *[32] For any diagonal matrix $\tilde{D} = diag(\tilde{D}_{jj}) \in \mathbb{R}^{J\times J}$ with $0 \leq \tilde{D}_{jj} \leq 1, j = 1, 2, \ldots, J$, the matrix $I - \tilde{D}(I - A)$ is nonsingular if and only if A is a P-matrix.* □

For any given sample, it is known that H^ϵ is positive definite and hence a P-matrix. Moreover, the $\bar{D}(z)$ is a diagonal matrix with its element on the interval $[0, 1]$ for any $(z, \delta) \in \mathbb{R}^{3J} \times \mathbb{R}_{++}$. Therefore, using Theorem 6.1, the Jacobian $\nabla_z f(z, \delta)$ is nonsingular for any $(z, \delta) \in \mathbb{R}^{3J} \times \mathbb{R}_{++}$. Thus, the linear equation (6.7) is well-defined.

Denoting the matrix $\bar{D}(z) = diag(\bar{D}_1(z), \bar{D}_2(z), \bar{D}_3(z))$, the Jacobian $\nabla_z f(z, \delta)$ at the point z is of the following structure

$$\begin{aligned} \nabla_z f(z,\delta) &= (I - \bar{D}(z)) + \bar{D}(z)\tilde{H}^\epsilon) \\ &\triangleq \begin{pmatrix} \Lambda_1(z) & 0 & -\bar{D}_1(z) \\ 0 & u_1(z)e^T + \Lambda_2(z) & \bar{D}_2(z) \\ \bar{D}_3(z) & -\bar{D}_3(z) & \Lambda_3(z) \end{pmatrix}, \end{aligned} \tag{6.8}$$

where $\Lambda_1(z) = \bar{D}_1(z)(C + (r-1)I) + I$, $\Lambda_2(z) = (\gamma(\xi) + (r-1))\bar{D}_2(z) + I$, $\Lambda_3(z) = (\epsilon + (r-1))\bar{D}_3(z) + I$ are all diagonal matrices, and $u_1(z) = \gamma(\xi)\bar{D}_2(z)e$. The sparse structure of (6.8) is used in its inverse computation. Specifically, $\nabla_z f(z, \delta)$ consists of only diagonal sub-matrix and the matrix $u_1(z)e^T + \Lambda_2(z)$, where the later is a sum of a diagonal sub-matrix and a rank-one matrix.

Noticing from (6.8), linear equation (6.7) is of the following form

$$\begin{pmatrix} \Lambda_1 & 0 & \Lambda_2 \\ 0 & u_1 u_2^T + \Lambda_3 & \Lambda_4 \\ \Lambda_5 & \Lambda_6 & \Lambda_7 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}, \tag{6.9}$$

where $\Lambda_i \in \mathbb{R}^J, i = 1, 2, \ldots, 7$ are diagonal matrices with Λ_1 and Λ_3 being nonsingular, $s_i \in \mathbb{R}^J, b_i \in \mathbb{R}^J, i = 1, 2, 3, u_1, u_2 \in \mathbb{R}^J$. For ease of notation, we use $\hat{\Lambda} = diag(1/\Lambda_{11}, \ldots, 1/\Lambda_{JJ})$ to represent the inverse of any invertible diagonal matrix $\Lambda = diag(\Lambda_{11}, \ldots, \Lambda_{JJ})$. Given the sparse structure of the coefficient matrix, (6.9) can be solved efficiently. More specifically, by the first two equations of (6.9), it follows

$$s_1 = \hat{\Lambda}_1(b_1 - \Lambda_2 s_3), \tag{6.10}$$

$$s_2 = (u_1 u_2^T + \Lambda_3)^{-1}(b_2 - \Lambda_4 s_3). \tag{6.11}$$

Directly substituting (6.10) and (6.11) into the third equation of (6.9), then

$$(\Lambda_7 - \Lambda_5\hat{\Lambda}_1\Lambda_2 - \Lambda_6(u_1 u_2^T + \Lambda_3)^{-1}\Lambda_4)s_3 = b_3 + const, \tag{6.12}$$

where $const = -(\Lambda_5\hat{\Lambda}_1 b_1 + \Lambda_6(u_1 u_2^T + \Lambda_3)^{-1} b_2)$. For computing the inverse matrix of $(u_1 u_2^T + \Lambda_3)$, the Sherman-Morrison formula is used in the following form.

$$(u_1 u_2^T + \Lambda_3)^{-1} = \hat{\Lambda}_3 - \frac{\hat{\Lambda}_3 u_1 u_2^T \hat{\Lambda}_3}{1 + u_2^T \hat{\Lambda}_3 u_1}. \tag{6.13}$$

Substituting (6.13) into (6.12), and it can be expressed

$$(\Lambda_0 + \alpha\tilde{u}_1\tilde{u}_2^T)s_3 = b_3 + const,$$

where $\alpha = 1/(1 + u_2^T\hat{\Lambda}_3 u_1), \Lambda_0 = \Lambda_7 - \Lambda_5\hat{\Lambda}_1\Lambda_2 - \Lambda_6\hat{\Lambda}_3\Lambda_4, \tilde{u}_1 = \Lambda_6\hat{\Lambda}_3 u_1, \tilde{u}_2 = \Lambda_4\hat{\Lambda}_3 u_2$. Then, if Λ_0 is nonsingular, using the Sherman-Morrison formula again, the solution of s_3 can be got immediately

$$s_3 = \left(\hat{\Lambda}_0 - \frac{\alpha\hat{\Lambda}_0\tilde{u}_1\tilde{u}_2^T\hat{\Lambda}_0}{1 + \alpha\tilde{u}_2^T\hat{\Lambda}_0\tilde{u}_1}\right)(b_3 + const). \tag{6.14}$$

Then, substituting the s_3 into (6.10) and (6.11), the solution of s_1 and s_2 are obtained, respectively.

From (6.14), one can see that the computation cost of s_3 is minimal, since only the inverse of several diagonal matrices are needed to be computed, namely, Λ_0, Λ_1, and Λ_3. Once s_3 is obtained, the calculation of s_1 and s_2 just needs to perform matrix-vector production. Therefore, the linear equation (6.9) can be solve efficiently.

Randomly generated problems:

For the first part of numerical test of this application, the problem of the form (5.19) are randomly generated. More specifically, a set of i.i.d. samples $\{\xi_\ell\}_{\ell=1}^N$ are randomly generated from a uniformly distribution over the interval $[0, 1]$. For $\ell = 1, \ldots, N$, set

$$p(\xi_\ell) = ((\xi_\ell)_1, (\xi_\ell)_2, \ldots, (\xi_\ell)_J)^T, \qquad \Pi(\xi_\ell) = \gamma(\xi_\ell)(ee^T + I) \triangleq (\xi_\ell)_1(ee^T + I).$$

Diagonal matrix $C \in \mathbb{R}^{J\times J}$ and $a \in \mathbb{R}^J$ are generated with its elements uniformly distributed over the interval $[1, 2]$. All the numerical results presented are the average value calculated from 10 independent runs.

To show the feasibility of the solution of the regularized problem compared to that of the original problem, the following residual value is constructed

$$\mathbf{Res} = \| \min(z, Hz + \bar{q})\|,$$

where H denotes the coefficient matrix (5.19) with $\epsilon = 0$.

Numerical results for $J = 10$ were listed in the Table 6.5 and 6.6. The first table is the solution obtained by PHM without block computation, while table 6.6

illustrates results with block implementation. For our generated experiments, it is found that $N = 50$ is the best block choice, measured by CPU time, for $J = 10$. The average number of iterations, the average cpu time, and the average value of **Res** were recorded in both tables. It is easy to see that the block implementation greatly reduces the cpu time. For the same value of ϵ, the number of iterations increases slightly when the sample size N increases. In cases where the sample size N is kept constant and the values of regularization parameter ϵ are chosen from $\epsilon = 10^{-3}$ to $\epsilon = 10^{-12}$, the iteration numbers are barely influenced as well as the cpu time. Furthermore, we observe the convergence of our regularization approach with decreasing values of ϵ, as have been proved in previous sections. Also notice that, the value of **Res** decreases when the ϵ diminishes from 10^{-3} to 10^{-12}. Numerically, it shows that the solution of the regularized problem is also that of the original problem when $\epsilon = 10^{-12}$.

Figure 6.8 illustrates the performance of the PHM measured by the number of iterations and cpu time for 10 players. It is also worth mentioning that although one might expect the problem to be more difficult to solve for a small ϵ, the numerical performance in our experiments remain roughly unaffected with decreasing values of ϵ.

Figure 6.9 demonstrates the convergence property of the first-stage solution x when the sample size gets large for the case $J = 10$. The convergence trend can be seen component-wisely as the solution x converges when the sample size N gets large.

N	$J(1+2N)$	Iter	CPU time/s	**Res**	Iter	CPU time/s	**Res**
			$\epsilon = 10^{-3}$			$\epsilon = 10^{-6}$	
10	210	146.30	0.26	4.42e-01	176.20	0.32	3.88e-04
50	1010	194.70	1.81	9.35e-01	197.40	1.83	9.32e-04
500	10010	208.70	26.72	3.00e+00	212.20	27.21	2.99e-03
2000	40010	222.60	154.97	5.93e+00	220.50	153.54	6.00e-03
5000	100010	224.70	623.53	9.49e+00	226.40	627.53	9.48e-03
			$\epsilon = 10^{-9}$			$\epsilon = 10^{-12}$	
10	210	152.70	0.27	1.08e-06	169.40	0.30	9.49e-07
50	1010	197.20	1.83	1.41e-06	194.40	1.80	9.75e-07
500	10010	212.70	27.21	3.21e-06	209.70	26.85	9.59e-07
2000	40010	220.30	153.34	6.16e-06	220.70	153.73	9.51e-07
5000	100010	226.70	628.89	9.60e-06	226.20	627.58	9.60e-07

Table 6.5: Numerical results for different regularization parameter ϵ and sample size N, $J = 10$ with individual sample.

N	$J(1+2N)$	Iter	CPU time/s	**Res**	Iter	CPU time/s	**Res**
			$\epsilon = 10^{-3}$			$\epsilon = 10^{-6}$	
10	210	162.60	0.21	4.04e-01	155.50	0.20	4.14e-04
50	1010	180.60	0.94	9.42e-01	184.00	0.93	9.31e-04
500	10010	203.30	10.67	2.97e+00	204.10	10.61	2.98e-03
2000	40010	213.80	42.38	5.89e+00	213.70	42.36	5.91e-03
5000	100010	218.20	115.15	9.36e+00	219.60	115.63	9.36e-03
			$\epsilon = 10^{-9}$			$\epsilon = 10^{-12}$	
10	210	143.20	0.19	1.08e-06	149.50	0.19	9.59e-07
50	1010	197.40	1.02	1.38e-06	191.10	0.97	9.58e-07
500	10010	205.10	10.58	3.19e-06	202.40	10.67	9.52e-07
2000	40010	214.10	42.46	6.07e-06	213.40	42.62	9.73e-07
5000	100010	219.00	115.35	9.50e-06	219.70	115.58	9.74e-07

Table 6.6: Numerical results for different regularization parameter ϵ and sample size N, $J = 10$ with block implementation.

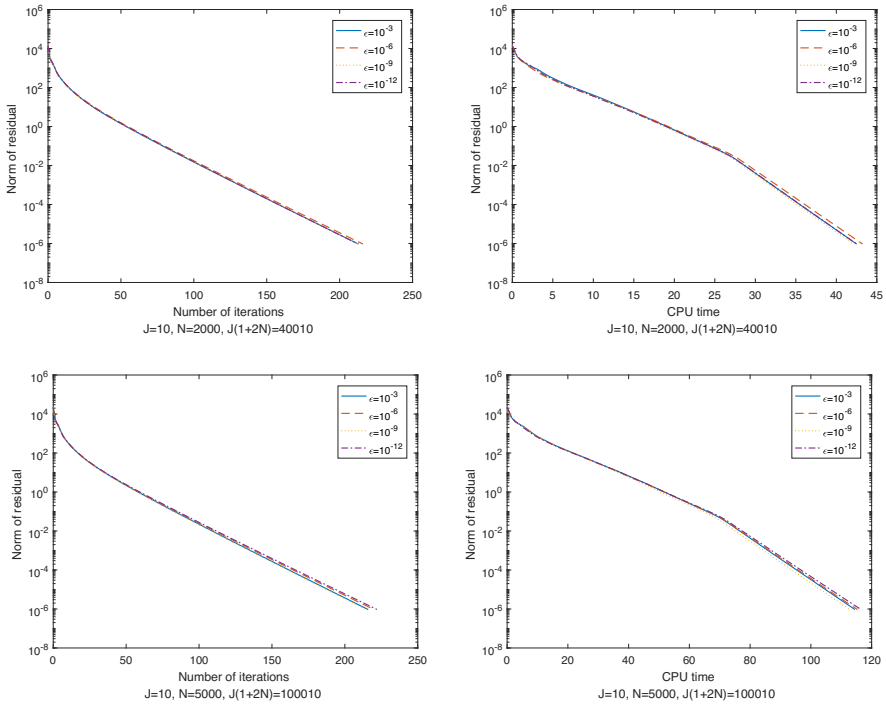

Figure 6.8: Numerical comparisons among different ϵ, $J = 10$.

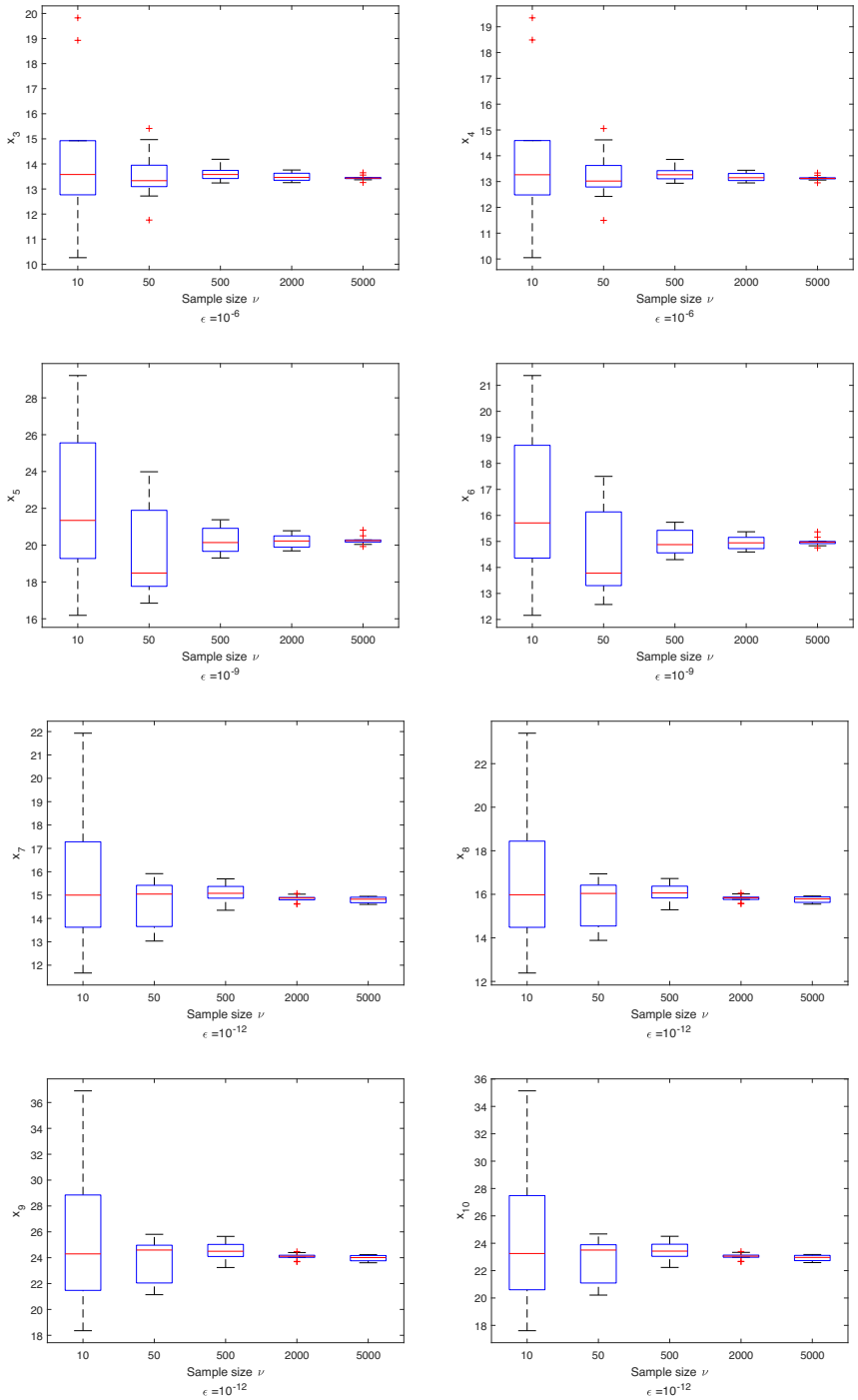

Figure 6.9: Convergence property of x with increasing N, $J = 10$.

6.2.2 Oil market share reproduction

In this subsection, the real application of interest: a two-stage stochastic Nash equilibrium model is tested in describing the crude oil market share. Namely, the strategies of crude oil exporting agents via solutions of reformulated SVI problem are investigated so as to recreate actual market shares. Based on historical data on crude oil market, in-sample back-tracking test is made to establish the effectiveness and validity of the model while explaining the market behaviour. Furthermore, the out-of-sample prediction capability of the proposed model is demonstrated when the results from the in-sample training data is used to specify model parameters. From the results of numerical tests, it is concluded that the proposed model is suitable to reproduce, predict and potentially capable to explain stable market shares of crude oil.

Facts on the crude oil market:

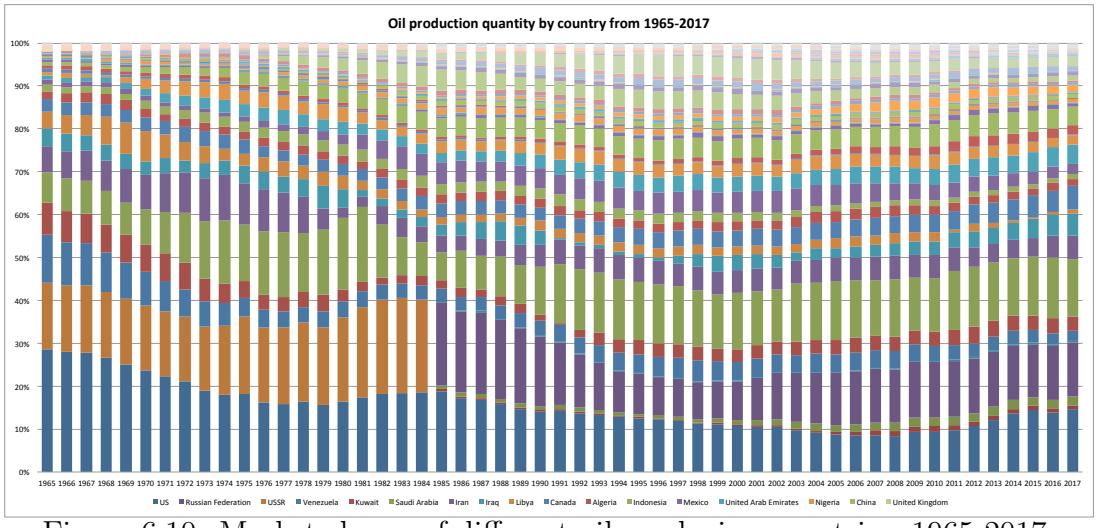

Figure 6.10: Market shares of different oil-producing countries, 1965-2017.

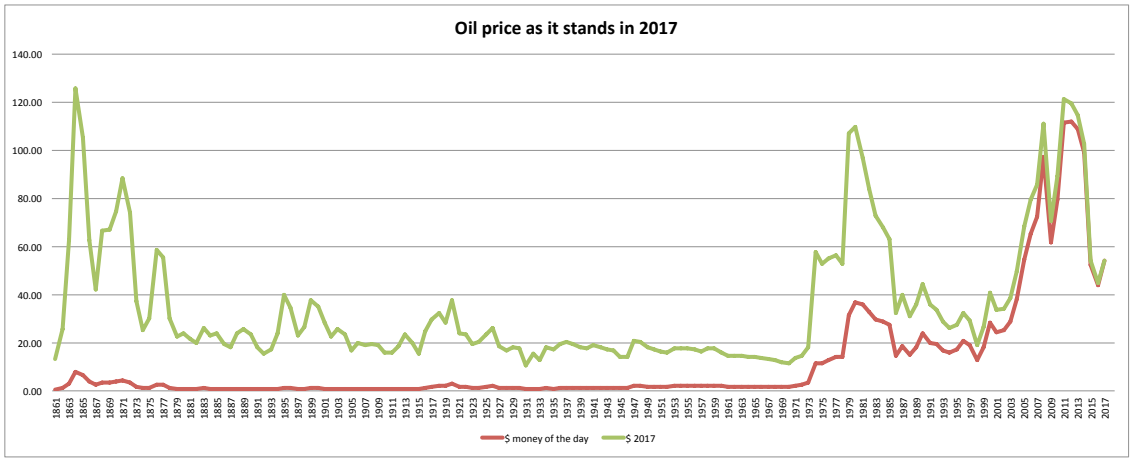

Figure 6.11: Inflation adjusted oil price in US $, 1861-2017.

Crude oil market is one of the most widely studied commodity market in the world. The market has long been described as of an 'intermediate between monopoly and perfect competition [42]", supported by extensive historical data and market observation. When the market encounters large sudden events, it took "oil shocks" [36, 37, 51]. Conceptually, consequences of supply and demand fluctuation, along with many other factors, are reflected most directly in dramatic changes in oil price. One interesting observation is that the market share behaves rather smoothly even during periods of oil shocks [51]. Majority of the world's crude oil is supplied by a few large oil exporting countries and they are viewed collectively as a finite number of large agents from which price-taking consumers purchase product at the same price [95]. Hence, it is no surprise that Cournot-Nash approaches have been adopted from earlier days of study in this field, see [90]. The study of oil shock gives rise to rich field of economic researches in oil market structure and as well as factors affecting oil price, see [6, 49, 52], etc. Note that the proposed model does not attempt to make explanation or future predictions on oil shocks but rather aims to make sense of

the stable characteristics of supply-side of oil market share: major agents acting non-cooperatively to achieve market equilibrium. The main focus is on the supply factor of oil, while treating other factors, e.g., demand, world economic situation, population, etc., as known information with uncertainty. The proposed model is found to be able to reproduce the market shares as well as to forecast future production plans of different oil-producing countries based on historical data.

In short, historical data are firstly used to determine model parameters as well as to approximate distribution for uncertainty in observations of the market. Model parameters are tested within a fixed sampling window so that the in-sample results obtained by solving our model matches that of training data set. It then follows that over a short decision horizon in the future, the trained model can be adopted to predict future production plans. More specifically, i.i.d. samples are taken using a fixed-size rolling window sampling method while the results are obtained by solving the proposed model with the adjusted set of parameters obtained by in-sample tests within the training window.

Data set, parameter selection and uncertainty description:

The data set used in this study are from

- *Statistical Review of World Energy*[2], yearly data from 1984, published every June by *bp. Inc*;
- *Oil Price Dynamics Report*[3], weekly by Federal Reserve Bank of New York since 1986;

[2] https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html

[3] https://www.newyorkfed.org/research/policy/oil_price_dynamics_report

- *U.S. Energy Information Administration*[4], weekly by United States Department of Energy.

Scenarios of the random vectors used in the numerical experiments are obtained from the empirical distributions generated by the data sets. More specifically, the empirical distributions of percentage price change, along with the contributions from demand changes, supply changes and the other effects are generated from the contents of oil dynamic report, consists of time series of weekly percentage change of Brent spot price and its corresponding components' contributions, i.e., those from demand, residual, and supply. The percentage change in price in $k+1$-th[5] week compared to that of the k-th week is denoted by $\triangle p^k$, while the contribution of demand and supply changes are denoted as $\triangle D^k$ and $\triangle S^k$ respectively. The contribution to the price change that does not correspond to demand and supply is denoted as residual $\triangle R^k$. Therefore, for any week within the oil price dynamics report, it holds that the historical Brent price change is *deterministicly* represented by

$$\triangle p^k = \triangle D^k + \triangle T^k + \triangle R^k. \tag{6.15}$$

Parameter determination:

In order to implement real market data into the proposed model (5.6)-(5.7), cost parameters c_j, a_j need to be determined correspond to production costs of j-th agent by analyzing training set. In practice, model parameters are adjusted via a brute force learning process in-sample. More specifically, if the training set consists only one trading day of information, the parameters are adjusted so that the solution,

[4] https://www.eia.gov

[5] k is used to denote the order in time, distinguished from that of i.i.d. samples

i.e., production quantities of J agents, matches the estimated historical observation. When more data become available, the parameters are adjusted so that the average of solutions is close to the real estimate. It is found from numerical implementation that the best fit solutions are obtained when values of c_j and a_j are taken to be inversely proportional to known market share within the sampling period. It is firstly learned that the common belief of "lower unit cost would lead to large sale" does not hold in the proposed game model, which also motivates the choice of quadratic production cost in the first-stage of the model.

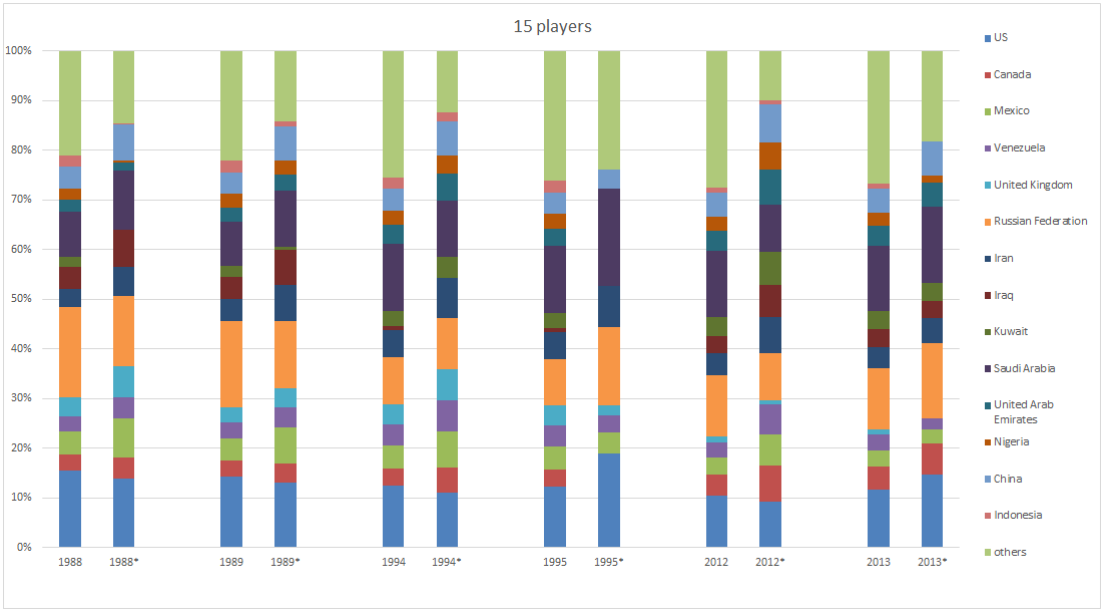

Figure 6.12: Oil market reproduction results with linear cost in the production over the chosen years.

Moreover, it is found in numerical experiments that the smaller the quadratic cost parameters c_j the greater market share the agent would have. This can be understood from the fact that the large oil producing agents are more flexible in adjusting

their production quantities as compared to smaller agents. A more mathematical interpretation is to think of c_j as related to the penalty parameters of the augmented Lagrangian formulation corresponding to production constraints of each agent: production cost is upper bounded with some unit production cost $\tilde{a}_j$. In this way, one can think of this observation been that the countries have greater reserve of oil suffer less when the production upper bound gets violated. Nevertheless, the quadratic cost is found to be more than adequate to reproduce historical observations.

Remark 6.2. *The linear cost function at the stage of production works badly especially in times where the oil price fluctuates violently. Conceptually, it was believed that the major oil producing agents control greater proportion of the entire supply due to their abilities to produce at low unit cost. It is not the case in the proposed game model, and it can be seen in Figure 6.12 that the low unit cost along may not be able to reproduce to certain level of satisfactory. A nature question to ask is that of the interpretation of the quadratic cost parameter c_j. One possible approach is to consider an additional constraint on the maximum production quantity. By reformulating such a constraint into an augmented Lagrangian function, the appearing of the quadratic term is related to the choice of penalty parameter of the aforementioned constraint on production cap. That is, the c_j can be interpreted as the "weight" of the penalization of extra quantity of production. In turn, it is concluded from both the analysis and the numerical results that the reason behind the stable global oil market share may be that the major producers are capable of producing "unscheduled" quantity without suffering too much of their profits. Hence, they have far greater advantages over other agents of the Nash game to adjust profit maximizing quantity*

of production x.

The scenarios of stochastic parameters are taken from empirical distribution of the sampling data set, guided by the analysis of [38]. In detail, i.i.d. samples are drawn to generate scenarios of market's supply discounting factor $\gamma(\xi)$ and j-th agent's adjusted price $p_j(\xi) = p_0(\xi) - h_j(\xi)$. Recall that the adopted expression for (scenario-based) inverse demand function of quantity offered to the market for a given scenario ξ^ℓ is in the following form

$$p^k(\xi_\ell) - \gamma^k(\xi_\ell)T^k(y_{\xi_\ell}).$$

In accordance with data structure, $p^k(\xi_\ell)$ corresponds to the adjusted predicted price from contributions other than supply, i.e.,

$$p^k(\xi_\ell) = p_0^k\big(1 + \triangle d^k(\xi_\ell) + \triangle r^k(\xi_\ell)\big),$$

where p_0^k denotes the known Brent spot price prior to that of the concerned time. Both $\triangle d^k(\xi_\ell)$ and $\triangle r^k(\xi_\ell)$ are random scenarios taken from empirical distributions of historical demand and residual distributions within the sampling set respectively. To estimate $\gamma^k(\xi)$, corresponding to k-week data, the data of spot quantity supplied to the market, T^k, is needed. In practice, it is very difficult to obtain reliable data on total supply to the market over short observation window. Rather, the trustworthy estimate on daily oil supply based on annual data is available along with the observation that a steady growth of about 1% per year over the last four decades, according to *Statistical Review of World Energy*. Therefore, realization of random instances of daily market supply $T^k(y_\xi)$ which are taken from a uniformly distributed interval between 99% and 101% of yearly based daily estimate, is used. Then, a set

of data of stochastic excessive supply discount factor $\{\gamma^k(\xi_\ell)\}$,

$$\gamma^k(\xi_\ell) = \frac{|p^k(\xi_\ell) - p_0^k|}{T^k(y_\xi)},$$

is generated, where p_0^k is known with certainty within the testing data set, and the $|\cdot|$ ensures that increase in quantity has a negative influence on price.

6.2.3 Performance in reproducing of the observed results

The results for numerical experiments on reproduced oil market share are presented. As shown in Fig 6.15, the in-sample results over the periods of oil shocks 2007, 2009 and 2014 are reproduced using solely from the available information within that year. For example, the in-sample experiment of 2009 uses all the available data obtained within 2009, e.g., daily price from 01/01/2009 to 31/12/2009. For the generation of random instances in-sample, the weekly price contribution covering the year 2009 is used to form empirical distributions for supply excluded spot prices $p(\xi_k)$, and discounting factor $\gamma(\xi_k)$, from which i.i.d. samples are taken to represent random scenarios. Hence, for in-sample experiments, the goal is to reproduce known market share with the adjustments of cost parameters a_j, c_j corresponding to j-th oil producing agent.

The out-of-sample tests cover the period from 2007 to 2017 with the *one year length equivalent of sampling window* in Figure 6.15. Note that one year sampling window is adopted since independent numerical results of selected years demonstrated that the market has a short memory in the sense that longer historical data would not provide extra information, at least in the setup of the proposed model and approaches. Prior to the date of interests, all the historical data are assumed to

be available while the rolling window sampling permits the usage of newly acquired data as soon as it becomes available. From the construction and interpretation of the two-stage model, it is reasonable to assume that cost of oil production remains unchanged over a short decision horizon. Furthermore, different lengths of the sampling window are tested for the tunings of parameters, especially during times of oil shocks for better in-sample fitting, and the best fit sampling window length is used in out-of-sample tests. For example, in order to obtain the market share of 2009, during times of the latest financial crisis, fixed length sampling windows are tested whose length varies from one year to five years.

Ignorance of price changes:

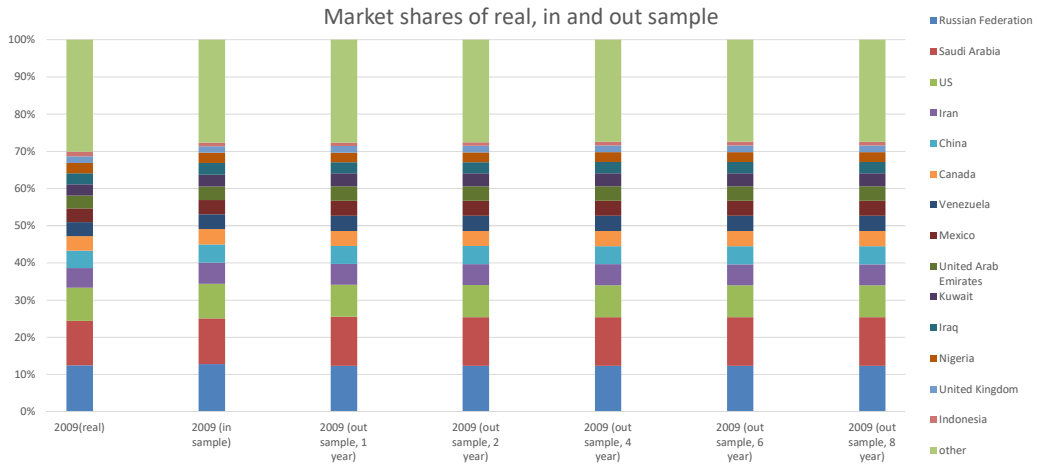

Figure 6.13: Out-of-sample results over 2009 with different length of historical data window.

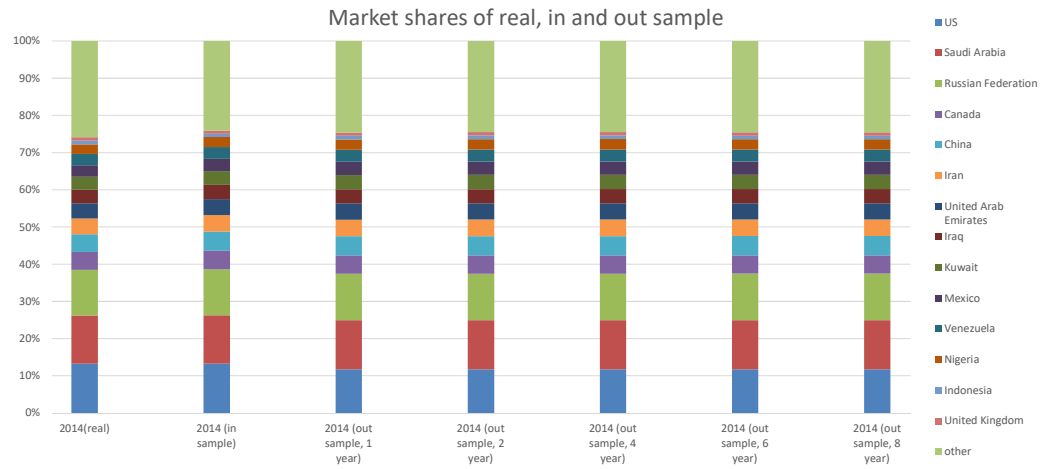

Figure 6.14: Out-of-sample results over 2014 with different length of historical data window.

Another interesting observation arises from the numerical tests is that the reproduction results vary very little with different lengths of memories of historical data. Numerical tests are carried out for the reproduction of years with great fluctuations in price. The different lengths of training data sets result into different empirical distributions, yet the reproduction quantity sees no significant differences. The results agree with the observations as seen in Figure 6.15 and Figure 6.11 that there is no clear connection between them.

If one wants to predict the production on 01/01/2009 using a sampling window of one year, one year sampling window from 01/01/2008 to 31/01/2008 is used to obtain empirical distributions from which scenarios of $\gamma(\xi_k)$, $p(\xi_k)$ are sampled. Taking N samples, the solutions method follows that described in previous subsection. For the result on 02/01/2019, the sampling window rolls forward using the data from 02/01/2008 to 01/01/2009 since the last day information becomes known, while the model remains unchanged. Therefore, for the market share prediction of any given

year, the daily market share results are obtained by taking the average of 250 runs, while the yearly estimate is formed in turn by taking the average on daily results. From Figure 6.15, it is observed that fairly good prediction quantities is achieved through out the testing period. However, it is worth mentioning that during the numerical experiments across that decision horizon the computational cost is much higher during periods over oil shocks compare to those of relative stable oil prices.

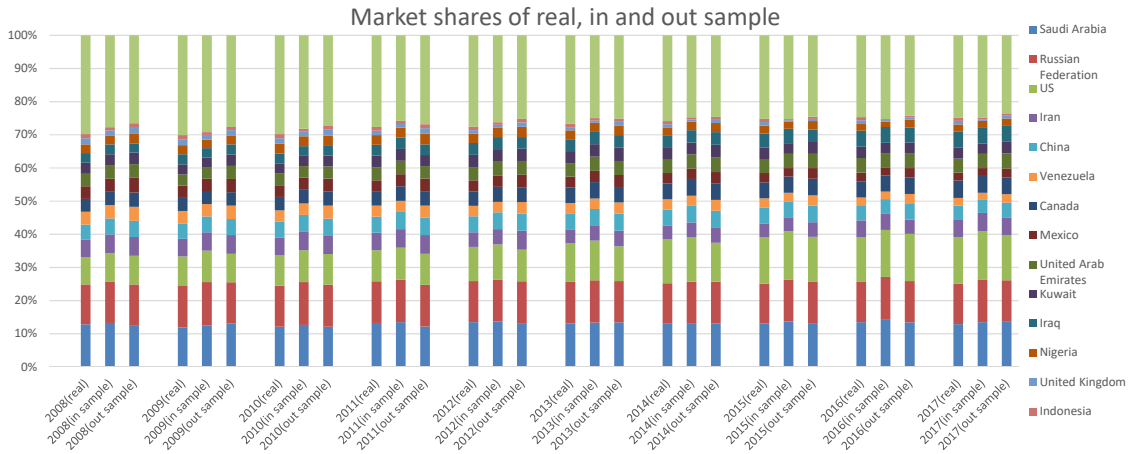

Figure 6.15: Real, in-sample and out-of-sample market shares results, 2008-2017.

Chapter 7

Concluding Remarks

7.1 Summary of the thesis contents

In this thesis, two models closely related to real world problems in the field of stochastic optimization are proposed and analyzed.

The first problem arises in the field of mathematical finance which searches for better asset allocation strategies. An effectively chosen portfolio is constructed by solving the proposed SAA-based CVaR-SSD relaxation model. The proposed model adopts and relaxes the well-known model of portfolio selection with SSD constraints. SAA method is firstly used to reduce the infinite number of SSD constraints to finite ones. A CVaR approximation of the SAA of SSD constraints enables the process of relaxation by specifically chosen probability level β. The proposed model can be regarded as a special problem in a class of problems with the level of SSD relaxation controlled by the values of β. The convergence analysis is demonstrated as the sample size N goes to infinity and the probability level β tends to 1. The choice of β which corresponds the best portfolio construction depends greatly on the structure of the data sets and thus the characteristics of the financial market. As a general guidance,

it should be chosen so that β is less but close to 1. Empirical tests are carried out on three sets of market data, and the model is of great promises for it provides consistently good results measured by different performance judges. The fact that the model is closely related to the portfolio selection model with SSD constraints is reflected in its overall outperforming of the benchmark portfolios.

Instead of the proposal of mathematical models to solve practical problems, the second application aims to explore the mechanism behind observation in economics. Rooted from literatures in the field of oil market analysis and classical game theoretical description of oligopolistic market, a two-stage stochastic Cournot-Nash game is formulated in attempt to provide an answer to the problem of stable oil market share regardless of the violent changes in oil price. The solution to the game is characterized by those of a two-stage SVI problem and studies are made to reveal its problem structure as well as properties of its solutions. Differ from the two-stage game, the nonuniquenss of solutions of two-stage SVI may cause trouble in the solution process not to mention the ambiguity it might bring when referring back to the economical interpretation of a particular solution. A regularization approach is developed to overcome this difficulty and theoretical properties of the solution of the regularized problem are derived. Furthermore, it is proved in the convergence analysis that the solution of the regularized problem is always one of the solutions of the original SVI. The numerical experiment is carried out after the adoption of the SAA method on the expectation calculation in the regularized SVI. The PHM method is used along with modification to better explore the special structure of the proposed SVI model. The validity of the game model is shown empirically by its ability to replicate historical observations in the oil market share. In-sample tests are ran to

best estimate parameters that characterizes the model, while stochastic parameters are drawn from formulated empirical distributions of historical data. Out-of-sample test is then performed with different scales of training sets over ten years period, and the reproduced results are in agreement of historical observations.

7.2 Future research directions

Future research directions will have two potential variations: better formulations of the models studied in this thesis; or extensions to other applications approachable by similar methodologies.

- The greatest efforts in presenting satisfactory results within this study are paid to the learning of parameters. For example, the CVaR probability level β in the application of portfolio selection is chosen with brute force by trails and errors. If one faces the problem of actual stock trading dynamics, it would be very difficult, if not entirely impossible, to decide the appropriate investment horizon beyond which the parameter needs to be reconsidered. Same problem arises in the application of oil market as well. The assumption that the cost parameter is revised with yearly basis may have worked in this particular example and reasonable following common sense but nevertheless an "educated" guess at its best. On the other hand, one thing that there is no shortage of is the flow of information. With more and more promising results appear from the researches of *machine learning*, one can try not to build the model based on grand assumptions but to adopt methods of learning to find a more realistic model for the formulation of the model.

- The techniques used in this thesis can be extended to other applications where similar mechanism may work as well. Indeed, as have seen in previous chapters, the theoretical analysis are often derived under more general framework than that of the problem of interests. In particular, the two-stage SVI problem arises from the two-stage SNEP but itself is a more general description of equilibrium system. One may use the technique of "reverse engineer" to work out the corresponding situations behind different variates of the two-stage SVI. For example, the cost function in the first-stage problem is set to be quadratic simply because the linear cost function with upper bound in production can never replicate the actual market share. It of course does not mean that the actual cost is of quadratic nature. This problem can also be related to the previous potential research direction of cooperating machine learning techniques a prior to the modelling process.

- Parallel computation and dynamic decisions can also be of great interests along with its potential for further researches. One bottleneck faced with numerical implementation of stochastic programming problems is the fact that in order to realize the situation under stochastic nature. Many scenarios need to be drawn which increases the computational load specially when high accuracy is required. In the calculation of second application in oil market, a block implementation is used under the PHM framework along with the smoothing technique for fast solving of subproblems. One can further explore the technique of parallel computation to speed up the entire load while the individual scenario is not so hard to solve. Also, if the data set is updated dynamically,

more scenarios can be drawn as the available data sets enlarge. The computation on these new scenarios may take equal amount of effort compare to the "old" scenarios but how to implement the newly coming results with the old ones while keeping the final solution dynamically consistent is also of great interests.

Bibliography

[1] G.J. Alexander and A.M. Baptista, A comparison of VaR and CVaR constraints on portfolio selection with the mean-variance model, *Manag. Sci.* **50** (2004), 1261-1273.

[2] E. Anderson, H. Xu and D. Zhang, Confidence levels for CVaR risk measures and minimax limits, manuscript, The University of Sydney, 2014.

[3] R.J. Aumann, Integrals of set-valued functions, *J. Math. Anal. Appl.,* **12** (1965), 1-12.

[4] K. B. Athreya and S. N. Lahiri, *Measure Theory and Probability Theory*, Springer Texts in Statistics, Springer, New York, 2006.

[5] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer series in operations research, Springer-Verlag, 2000.

[6] G. Bornstein, P. Krusell and S. Rebelo, Lags, costs, and shocks: An equilibrium model of the oil industry, *National Bureau of Economic Research*, 2017.

[7] J.V. Burke, X. Chen and H. Sun, Subdifferentiation and smoothing of non-smooth integral functionals, Department of applied mathematics, The Hong Kong Polytechnic University, 2017.

[8] N. A. Canakgoz and J. E. Beasley, Mixed-integer programming approaches for index tracking and en-hanced indexation, *European J. Oper. Res.*, **196** (2009), 384-399.

[9] X. Chen, Smoothing methods for nonsmooth, nonconvex minimization, *Math. Program.,* **134**:1 (2012), 71-99.

[10] X. Chen and M. Fukushima, A smoothing method for a mathematical program with P-matrix linear complementarity constraints, *Comput. Optim. Appl.,* **27**:3 (2004), 223-246.

[11] X. Chen and M. Fukushima, Expected residual minimization method for stochastic linear complementarity problems, *Math. Oper. Res.,* **30**:4 (2005), 1022-1038.

[12] X. Chen, T.K. Pong and R.J.B. Wets, Two-stage stochastic variational inequalities: an ERM-solution procedure, *Math. Program.,* **165** (2017), 71-111.

[13] X. Chen, A. Shapiro and H. Sun, Convergence analysis of sample average approximation of two-state stochastic generalized equations, *SIAM J. Optim.,* to appear.

[14] X. Chen, H. Sun and R.J.B. Wets, Regularized mathematical programs with stochastic equilibrium constraints: estimating structural demand models, *SIAM J. Optim.,* **25** (2015), 53-75.

[15] X. Chen, H. Sun and H. Xu, Discrete approximation of two-stage stochastic and distributionally robust linear complementarity problems, *Math. Program.*, (2018), to appear.

[16] X. Chen and S. Xiang, Newton iterations in implicit time-stepping scheme for differential linear complementarity systems, *Math. Program.,* **138** (2013), 579-606.

[17] X. Chen, R.J.B. Wets and Y. Zhang, Stochastic variational inequalities: residual minimization smoothing sample average approximations, *SIAM J. Optim.,* **22** (2012), 649-673.

[18] X. Chen and Y. Ye, On homotopy-smoothing methods for box-constrained variational inequalities, *SIAM J. Control. Optim.,* **37** (1999), 589-616.

[19] X. Chen and Y. Ye, On smoothing methods for the P_0 matrix linear complementarity problem, *SIAM J. Optim.,* **11** (2000), 341-363.

[20] F.H. Clarke, *Optimization and nonsmooth analysis*, SIAM, 1990.

[21] G.B. Dantzig, Linear programming under uncertainty, *Managemnet Sci.,* **1** (1955), 197-206.

[22] D. Dentcheva and A. Ruszczyński, Optimization with stochastic dominance constraints, *SIAM J. Optim.*, **14** (2003), 548-566.

[23] D. Dentcheva and A. Ruszczyński, Semi-infinite probabilistic constraints: optimality and convexification, *Optimization*, **53** (2004), 583-601.

[24] D. Dentcheva and A. Ruszczyński, Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints, *Math. Program.*, **99** (2004), 329-350.

[25] D. Dentcheva and A. Ruszczyński, Portfolio optimization with stochastic dominance constraints, *J. Bank. Financ.*, **30** (2006), 433-451.

[26] D. Dentcheva and A. Ruszczyński, Inverse cutting plane methods for optimization problems with second-order stochastic dominance constraints, *Optimization*, **59** (2010), 323-338.

[27] D. Dentcheva and A. Ruszczyński, Risk-averse portfolio optimization via stochastic dominance constraints, in *Handbook of Quantitative Finance and Risk Management* (eds. C. Lee, A. Lee and J. Lee), Springer, NewYork, (2015), 2281-2302.

[28] C. Fábián, G. Mitra and D. Roman, Processing second-order stochastic dominance models using cutting-plane representations, *Math. Program.*, **130** (2011), 33-57.

[29] F. Facchinei and J.S. Pang, *Finite-dimensional variational inequalities and complementarity problems*, Springer Science & Business Media, 2007.

[30] A. Fischer, A Newton-type method for positive-semidefinite linear complementarity problems, *J. Optim. Theory. Appl.,* **86**:3 (1995). 585-608.

[31] S.A. Gabriel and J.D. Fuller, A Benders decomposition method for solving stochastic complementarity problems with an application in energy, *Comput. Econ.,* **35** (2010), 301-329.

[32] S.A. Gabriel and J.J. More, Smoothing of mixed complementarity problems, in *Complementarity and Variational Problems: State of the Art, M.C. Ferris and J.S. Pang, eds.,* SIAM, (1997), 105-116.

[33] T.S. Genc, S.S. Reynolds and S. Sen, Dynamic oligopolistic games under uncertainty: A stochastic programming approach, *J. Econ. Dyn. Control.,* **31** (2007), 55-80.

[34] S.J. Grossman, Nash equilibrium and the industrial organization of markets with large fixed costs, *Econometrica*, (1981), 1149-1172.

[35] M. Gugat, A parametric review on the Mangasarian-Fromovitz constraint qualification, *Math. Program.*, **85** (1999), 643-653.

[36] J.D. Hamilton, This is what happened to the oil price-macroeconomy relationship, *J. Monet. Econ.,* **38** (1996), 215-220.

[37] J.D. Hamilton, What is an oil shock?, *J. Econom.,* **113**:2 (2003), 363-398.

[38] J.D. Hamilton, Historical oil shocks, *National Bureau of Economic Research*, 2011.

[39] P.T. Harker and J.S. Pang, Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications, *Math. Program.,* **48** (1990), 161-200.

[40] B.F. Hobbs and J.S. Pang, Nash-Cournot equilibria in electric power markets with piecewise linear demand functions and joint constraints, *Oper. Res.,* **55** (2007), 113-127.

[41] T. Homem-de-Mello and S. Mehrota, A cutting surface method for uncertain linear programs with polyhedral stochastic dominance constraints, *SIAM J. Optim.*, **20** (2009), 1250-1273.

[42] H. Hotelling, The economics of exhaustible resources, *J. Polit. Econ.,* **39**:2 (1931), 137-175.

[43] J. Hu, T. Homen-De-Mello and S. Mehrotra, Sample average approximation of stochastic dominance constrained programs, *Math. Program.*, **133** (2012), 171-201.

[44] J. Huang, M. Leng and M. Parlar, Demand functions in decision modeling: A comprehensive survey and research directions, *Decis. Sci.,* **44**:3 (2013), 557-609.

[45] B. Jadamba, F. Raciti, Variational inequality approach to stochastic Nash equilibrium problems with an application to Cournot oligopoly, *J. Optim. Theory. Appl.,* **165**:3 (2015), 1050-1070.

[46] J. Jiang, Quantification and convergence analysis of two-stage stochastic variational inequality problems, Ph.D thesis, The Hong Kong Polytechnic University, submitted (2019).

[47] A. Jofré, R.T. Rockafellar and R.J.B. Wets, Variational inequalities and economic equilibrium, *Math. Oper. Res.,* **32**:1 (2007), 32-50.

[48] P. Jorion, *Value at Risk: The New Benchmark for Controlling Market Risk*, McGraw-Hill Inc., US, 2006.

[49] L. Juvenal and I. Petrella, Speculation in the oil market, *J. Appl. Econom.,* **30** (2015), 621-649.

[50] J.E. Kelley, The cutting-plane method for solving convex programs, *SIAM J. Appl. Math.*, **8** (1960), 703-712.

[51] L. Kilian, Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market, *Am. Econ. Rev.,* **99**:3 (2009), 1053-1069.

[52] L. Kilian and D.P. Murphy, The role of inventories and speculative trading in the global market for crude oil, *J. Appl. Econom.,* **29** (2014), 454-478.

[53] A.J. King and R.T. Rockafellar, Asymptotic theory for solutions in statistical estimation and stochastic programming, *Math. Oper. Res.,* **18**:1 (1993), 148-162.

[54] P. Kolm, R. Tütüncü and F. Fabozzi, 60 Years of portfolio optimization: Practical challenges and current trends, *European J. Oper. Res.*, **234** (2014), 356-371.

[55] H. Konno and H. Yamazaki, Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market, *Manag. Sci.*, **37** (1991), 519-531.

[56] H. Konno, H. Shirakawa and H. Yamazaki, A mean-absolute deviation-skewness portfolio optimization model, *Ann. Oper. Res.*, **45.1** (1993), 205-220.

[57] Y. Liu and H. Xu, Stability and sensitivity analysis of stochastic programs with second order dominance constraints, *Math. Program.*, **142** (2013), 435-460.

[58] H. Markowitz, Portfolio Selection, *J. Finance*, (1952), 77-91.

[59] H. Markowitz, *Portfolio Selection*, John Wiley & Sons, New York, 1959.

[60] C.R. McConnell, S.L. Brue and S.M. Flynn, *Economics: Principles, problems, and policies*, Boston McGraw-Hill/Irwin, 2009.

[61] K. Mosler and M. Scarsini, (eds.), Stochastic Orders and Decision under Risk, *Institute of Mathematical Statistics*, Hayward, CA, 1991.

[62] F.H. Murphy, H.D. Sherali and A.L. Soyster, A mathematical programming approach for determining oligopolistic market equilibrium, *Math. Program.,* **24** (1982), 92-105.

[63] R.B. Myerson, Nash equilibrium and the history of economic theory, *Journal of Economic Literature*, **37**:3 (1999), 1067-1082.

[64] A. Nemirovski, A. Juditsky, G. Lan and A. Shapiro, Robust stochastic approximation approach to stochastic programming, *SIAM J. Optim.,* **19**:4 (2009), 1574-1609.

[65] W. Ogryczak and A. Ruszczyński, From stochastic dominance to mean-risk models: Semideviations as risk measures, *Euro. J. Oper. Res.*, **116** (1999), 33-50.

[66] W. Ogryczak and A. Ruszczyński, Dual stochastic dominance and related mean-risk models, *SIAM J. Optim.*, **13** (2002), 60-78.

[67] J. Outrata, M.C. Ferris, M. Červinka and M. Outrata, On Cournot-Nash-Walras equilibria and their computation, *Set-Valued Var. Anal.*, **24** (2016), 387-402.

[68] D. Ralph and H. Xu, Convergence of stationary points of sample average two-stage stochastic programs: A generalized equation approach, *Math. Oper. Res.,*, **36**:3 (2011), 568-592.

[69] J.S. Pang, G. Scutari, D.P. Palomar and F. Facchinei, Design of cognitive radio systems under temperature-interference constraints: A variational inequality approach, *IEEE Trans. Signal Process.,* **58**:6 (2010), 3251-3271.

[70] J.S. Pang, S. Sen and U.V. Shanbhag, Two-stage non-cooperative games with risk-averse players, *Math. Program.,* **165**:1 (2017), 235-290.

[71] J.S. Pang, C.L. Su, Y.C. Lee, A constructive approach to estimating pure characteristics demand models with pricing, *Oper. Res.,*, **63** (2015), 639-659.

[72] A. Philpott, M.C. Ferris and R.J.B. Wets, Equilibrium, uncertainty and risk in hydro-thermal electricity systems, *Math. Program.,* **157** (2016), 483-513.

[73] U. Ravat and U.V. Shanbhag, On the characterization of solution sets of smooth and nonsmooth convex stochastic Nash games, *SIAM J. Optim.,* **21**:3 (2011), 1168-1199.

[74] C.W. Richard, J.S. Pang and S.E. Richard, *The linear complementarity problem*, Academic Press, 1992.

[75] R.T. Rockafellar, *Convex analysis*, Princeton university press, 2015.

[76] R.T. Rockafellar, Integrals which are convex functionals, *Pac. J. Math.,* **24**:3 (1968), 525-539.

[77] R.T. Rockafellar and J. Sun, Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging, *Math. Program.,* (2018), to appear.

[78] R.T. Rockafellar and S. Uryasev, Optimization of conditional value-at-risk, *J. Risk*, **2** (2000), 21-41.

[79] R.T. Rockafellar and R.J.B. Wets, Stochastic convex programming: Kuhn-Tucker conditions, *J. Math. Econ.,* **2** (1975), 349-370.

[80] R.T. Rockafellar and R.J.B. Wets, Stochastic convex programming: Basic duality, *Pac. J. Math.,* **62** (1976), 173-195.

[81] R.T. Rockafellar and R.J.B. Wets, Stochastic convex programming: Singular multipliers and extended duality singular multipliers and duality, *Pac. J. Math.,* **62** (1976), 173-195.

[82] R.T. Rockafellar and R.J.B. Wets, Stochastic convex programming: Relatively complete recourse and induced feasibility, *SIAM. J. Control. Optim.,* **14** (1976), 574-589.

[83] R.T. Rockafellar and R.J.B. Wets, Scenarios and policy aggregation in optimization under uncertainty, *Math. Oper. Res.,* **16** (1991), 119-147.

[84] R.T. Rockafellar and R.J.B. Wets, *Variational analysis*, Springer Science & Business Media, 2009.

[85] R.T. Rockafellar and R.J.B. Wets, Stochastic variational inequalities: single-stage to multistage, *Math. Program.,* **165** (2017), 331-360.

[86] D. Roman, K. Darby-Dowman and G. Mitra, Portfolio construction based on stochastic dominance and target return distributions, *Math. Program.*, **108** (2000), 541-569.

[87] D. Roman, G. Mitra and V. Zverovich, Enhanced indexation based on second-order stochastic dominance, *European J. Oper. Res.*, **228** (2013), 273-281.

[88] B.J. Rosen, Existence and uniqueness of equilibrium points for concave n-person games, *Econometrica*, (1965), 520-534.

[89] G. Rudolf and A. Ruszczyński, Optimization problems with second order stochastic dominance constraints: Duality, compact formulations, and cut generation methods, *SIAM J. Optim.*, **19** (2008), 1326-1343.

[90] S.W. Salant, Exhaustible resources and industrial structure: A Nash-Cournot approach to the world oil market, *J. Polit. Econ.,* **84** (1976), 1079-1093.

[91] A. Shapiro, On uniqueness of Lagrange multipliers in optimization problems subject to cone constraints, *SIAM J. Optim.,* **7**:2 (1997), 508-518.

[92] A. Shapiro, Monte Carlo sampling methods, *Handbooks in operations research and management science*, **10** (2003), 353-425.

[93] A. Shapiro, D. Dentcheva and A. Ruszczyński, *Lectures on stochastic programming: modeling and theory*, SIAM, 2009.

[94] W.E. Sharpe, The Sharpe Ratio, *J. Portfolio Manage.*, **21** (1994), 49-58.

[95] J.L. Smith, World oil: market or mayhem?, *J. Econ. Perspect.,* **23** (2009), 145-164.

[96] F.A. Sortino and L.N. Price, Performance measurement in a downside risk framework, *J. Invest.*, **3** (2009), 59-64.

[97] J. Sun, L.-Z. Liao and B. Rodrigues, Quadratic two-stage stochastic optimization with coherent measures of risk, *Math. Program.,* **168**:1-2 (2018), 599-613.

[98] H. Sun, C.L. Su and X. Chen, SAA-regularized methods for multiproduct price optimization under the pure characteristics demand model, *Math. Program.,* **165** (2017), 361-389.

[99] G.A. Whitmore and M.C. Findlay (eds.), *Stochastic Dominance: An Approach to Decision-Making Under Risk*, D.C.Heath, Lexington, MA. 1978.

[100] H. Sun, H. Xu and Y. Wang, A smoothing penalized sample average approximation method for stochastic programs with second order stochastic dominance constraints, *Asia-Pac. J. Oper. Res.*, **30** (2013).

[101] H. Sun, H. Xu, R. Meskarian and Y. Wang, Exact penalization, level function method and modified cutting-plane method for stochastic programs with second order stochastic dominance constraints, *SIAM J. Optim.*, **23** (2013), 602-631.

[102] H. Sun and H. Xu, Convergence analysis of stationary points in sample average approximation of stochastic programs with second order stochastic dominance constraints, *Math. Program.*, **143** (2014), 31-59.

[103] H. Xu, J.S. Pang, F. Ordónez and M. Dessouky, Complementarity models for traffic equilibrium with ridesharing, *Transport. Res. B-meth.,* **81** (2015), 161-182.

[104] F. Xu, M. Wang, Y.H. Dai and D. Xu, A sparse enhanced indexation model with chance and cardinality constraints, *J. Glob. Optim.*, (2017), 1-21.

[105] H. Xu and D. Zhang, Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications, *Math. Program.,* **119** (2009), 371-401.

[106] J. Yao, I. Adler and S.S. Oren, Modeling and computing two-settlement oligopolistic equilibrium in a congested electricity network, *Oper. Res.,* **56** (2008), 34-47.

Appendices

Proof of Theorem 5.1. The proof was mainly contributed by Mr. Jie Jiang and is included for the completeness of the project. More details can be found in the Ph.D thesis [46].

Proof. By direct computation, it holds that

$$\left(M^{\epsilon}(\xi)\begin{pmatrix} y^{\epsilon} \\ \lambda^{\epsilon} \end{pmatrix} + q(x,\xi)\right)_j = \begin{cases} \gamma(\xi)(y^{\epsilon})_j + \gamma(\xi)T^{\epsilon} + (\lambda^{\epsilon})_j - p_j(\xi), \ j = 1,\ldots,J; \\ x_{j-J} - (y^{\epsilon})_{j-J} + \epsilon(\lambda^{\epsilon})_{j-J}, \ j = J+1,\ldots,2J. \end{cases}$$

Then, problem (5.11) can be rewritten as below:

$$\begin{cases} 0 \le (y^{\epsilon})_j \perp \gamma(\xi)(y^{\epsilon})_j + \gamma(\xi)T^{\epsilon} + (\lambda^{\epsilon})_j - p_j(\xi) \ge 0, \\ 0 \le (\lambda^{\epsilon})_j \perp x_j - (y^{\epsilon})_j + \epsilon(\lambda^{\epsilon})_j \ge 0, \end{cases} \tag{1}$$

for $j \in \mathcal{J}$. From the first complementarity condition in (1), it has $(y^{\epsilon})_j$ is of the following:

$$(y^{\epsilon})_j = \begin{cases} -\dfrac{\gamma(\xi)T^{\epsilon} + (\lambda^{\epsilon})_j - p_j(\xi)}{\gamma(\xi)}, & \gamma(\xi)T^{\epsilon} + (\lambda^{\epsilon})_j - p_j(\xi) < 0; \\ 0, & \gamma(\xi)T^{\epsilon} + (\lambda^{\epsilon})_j - p_j(\xi) \ge 0 \end{cases} \tag{2}$$

for $j \in \mathcal{J}$. Similarly, it can be derived that

$$(\lambda^\epsilon)_j = \begin{cases} \dfrac{(y^\epsilon)_j - x_j}{\epsilon}, & (y^\epsilon)_j - x_j > 0; \\ 0, & (y^\epsilon)_j - x_j \leq 0 \end{cases} \tag{3}$$

for $j \in \mathcal{J}$. Note that $(y^\epsilon)_j = 0$ implies $(y^\epsilon)_j = 0 \leq x_j$, and we have $(\lambda^\epsilon)_j = 0$. Then, based on (2) and (3), for all three cases:

$$\begin{cases} (y^\epsilon)_j = 0, \ (\lambda^\epsilon)_j = 0 & \text{for } j \in \mathcal{I}_1; \\ (y^\epsilon)_j = -\dfrac{\gamma(\xi)T^\epsilon + (\lambda^\epsilon)_j - p_j(\xi)}{\gamma(\xi)}, \ (\lambda^\epsilon)_j = 0 & \text{for } j \in \mathcal{I}_2; \\ (y^\epsilon)_j = -\dfrac{\gamma(\xi)T^\epsilon + (\lambda^\epsilon)_j - p_j(\xi)}{\gamma(\xi)}, \ (\lambda^\epsilon)_j = \dfrac{(y^\epsilon)_j - x_j}{\epsilon} & \text{for } j \in \mathcal{I}_3, \end{cases}$$

where

$$\mathcal{I}_1 := \{j \in \mathcal{J} : \gamma(\xi)T^\epsilon + (\lambda^\epsilon)_j - p_j(\xi) \geq 0, (y^\epsilon)_j - x_j \leq 0\},$$

$$\mathcal{I}_2 := \{j \in \mathcal{J} : \gamma(\xi)T^\epsilon + (\lambda^\epsilon)_j - p_j(\xi) < 0, (y^\epsilon)_j - x_j \leq 0\},$$

$$\mathcal{I}_3 := \{j \in \mathcal{J} : \gamma(\xi)T^\epsilon + (\lambda^\epsilon)_j - p_j(\xi) < 0, (y^\epsilon)_j - x_j > 0\}.$$

It follows that,

$$((y^\epsilon)_j, (\lambda^\epsilon)_j) = \begin{cases} (0,0), & j \in \mathcal{I}_1; \\ \left(-\dfrac{\gamma(\xi)T^\epsilon - p_j(\xi)}{\gamma(\xi)}, 0\right), & j \in \mathcal{I}_2; \\ \left(-\dfrac{\epsilon\gamma(\xi)T^\epsilon - x_j - \epsilon p_j(\xi)}{\epsilon\gamma(\xi)+1}, -\dfrac{\gamma(\xi)(T^\epsilon + x_j) - p_j(\xi)}{\epsilon\gamma(\xi)+1}\right), & j \in \mathcal{I}_3, \end{cases}$$

which verifies (5.14). For the remaining of the proof, let $j \in \mathcal{I}_2$, and

$$-\gamma(\xi)(y^\epsilon)_j = \gamma(\xi)T^\epsilon - p_j(\xi)$$

and thus

$$-\gamma(\xi)\sum_{i\in\mathcal{I}_2}(y^\epsilon)_i = |\mathcal{I}_2|\gamma(\xi)T^\epsilon - \sum_{i\in\mathcal{I}_2} p_i(\xi). \tag{4}$$

Analogously, it can be derived from

$$(y^\epsilon)_j = -\frac{\gamma(\xi)T^\epsilon + (\lambda^\epsilon)_j - p_j(\xi)}{\gamma(\xi)} \text{ and } (\lambda^\epsilon)_j = \frac{(y^\epsilon)_j - x_j}{\epsilon} \text{ for } j \in \mathcal{I}_3$$

that

$$-\gamma(\xi)\sum_{i\in\mathcal{I}_3}(y^\epsilon)_i = |\mathcal{I}_3|\gamma(\xi)T^\epsilon + \frac{1}{\epsilon}\sum_{i\in\mathcal{I}_3}(y^\epsilon)_i - \frac{1}{\epsilon}\sum_{i\in\mathcal{I}_3}x_i - \sum_{i\in\mathcal{I}_3}p_i(\xi). \tag{5}$$

Combining that of (4) and (5), it is obtained

$$-\gamma(\xi)T^\epsilon = (|\mathcal{I}_2| + |\mathcal{I}_3|)\gamma(\xi)T^\epsilon + \frac{1}{\epsilon}\sum_{i\in\mathcal{I}_3}(y^\epsilon)_i - \frac{1}{\epsilon}\sum_{i\in\mathcal{I}_3}x_i - \sum_{i\in\mathcal{I}_2\cup\mathcal{I}_3}p_i(\xi).$$

Therefore,

$$\frac{1}{\epsilon}\sum_{i\in\mathcal{I}_3}(y^\epsilon)_i = -(|\mathcal{I}_2| + |\mathcal{I}_3| + 1)\gamma(\xi)T^\epsilon + \frac{1}{\epsilon}\sum_{i\in\mathcal{I}_3}x_i + \sum_{i\in\mathcal{I}_2\cup\mathcal{I}_3}p_i(\xi). \tag{6}$$

Substituting (6) into (5), and

$$\begin{aligned}
|\mathcal{I}_3|\gamma(\xi)T^\epsilon =& -\left(\gamma(\xi) + \frac{1}{\epsilon}\right)\sum_{i\in\mathcal{I}_3}(y^\epsilon)_i + \frac{1}{\epsilon}\sum_{i\in\mathcal{I}_3}x_i + \sum_{i\in\mathcal{I}_3}p_i(\xi) \\
=& -(\epsilon\gamma(\xi) + 1)\left(-(|\mathcal{I}_2| + |\mathcal{I}_3| + 1)\gamma(\xi)T^\epsilon + \frac{1}{\epsilon}\sum_{i\in\mathcal{I}_3}x_i + \sum_{i\in\mathcal{I}_2\cup\mathcal{I}_3}p(\xi)_i\right) \\
&+ \frac{1}{\epsilon}\sum_{i\in\mathcal{I}_3}x_i + \sum_{i\in\mathcal{I}_3}p_i(\xi) \\
=& (\epsilon\gamma(\xi) + 1)(|\mathcal{I}_2| + |\mathcal{I}_3| + 1)\gamma(\xi)T^\epsilon - \gamma(\xi)\sum_{i\in\mathcal{I}_3}x_i - \epsilon\gamma(\xi)\sum_{i\in\mathcal{I}_2\cup\mathcal{I}_3}p_i(\xi) - \sum_{i\in\mathcal{I}_2}p_i(\xi).
\end{aligned}$$

Then,

$$(\epsilon\gamma(\xi)(|\mathcal{I}_2| + |\mathcal{I}_3| + 1) + |\mathcal{I}_2| + 1)\,\gamma(\xi)T^\epsilon = \gamma(\xi)\sum_{i\in\mathcal{I}_3}x_i + \epsilon\gamma(\xi)\sum_{i\in\mathcal{I}_2\cup\mathcal{I}_3}p_i(\xi) + \sum_{i\in\mathcal{I}_2}p_i(\xi),$$

that is,

$$T^{\epsilon} = \frac{\gamma(\xi)\sum_{i\in\mathcal{I}_3} x_i + \epsilon\gamma(\xi)\sum_{i\in\mathcal{I}_2\cup\mathcal{I}_3} p_i(\xi) + \sum_{i\in\mathcal{I}_2} p_i(\xi)}{(\epsilon\gamma(\xi)(|\mathcal{I}_2| + |\mathcal{I}_3| + 1) + |\mathcal{I}_2| + 1)\,\gamma(\xi)}.$$

This completes the proof. □

Proof of Proposition 5.4.

Proof. Let $\hat{z} = (\hat{y}, \hat{\lambda})$ be any solution of LCP(q, M) and

$$\begin{aligned} 0 &\geq (z^\epsilon - \hat{z})^T (M^\epsilon z^\epsilon + q - (M\hat{z} + q)) \\ &= (z^\epsilon - \hat{z})^T (M^\epsilon z^\epsilon - M\hat{z}) \\ &= (z^\epsilon - \hat{z})^T M (z^\epsilon - \hat{z}) + (z^\epsilon - \hat{z})^T \begin{pmatrix} 0 \\ \epsilon\lambda^\epsilon \end{pmatrix} \\ &\geq (z^\epsilon - \hat{z})^T \begin{pmatrix} 0 \\ \epsilon\lambda^\epsilon \end{pmatrix} \\ &= \epsilon(\lambda^\epsilon - \hat{\lambda})^T \lambda^\epsilon, \end{aligned}$$

where the second inequality follows from the positive semidefiniteness of M. Then,

$$\|\lambda^\epsilon\|^2 \leq \hat{\lambda}^T \lambda^\epsilon \leq \|\hat{\lambda}\| \|\lambda^\epsilon\|,$$

which implies the boundedness of λ^ϵ,

$$\|\lambda^\epsilon\| \leq \|\hat{\lambda}\|. \tag{7}$$

It follows from (7) that any accumulation point of $\{\lambda^\epsilon\}$ as $\epsilon \downarrow 0$ is the least l_2-norm solution. Since M is positive semidefinite, we know from [74, Theorem 5.6.2] that there is a unique least l_2-norm solution. On the other hand, we know from Proposition 5.1, for any fixed (x, ξ), $\hat{y}$ is unique. Therefore, the limit of z^ϵ exists as $\epsilon \downarrow 0$ and converges to the least l_2-norm solution of LCP(q, M).

Due to the existence of limit for z^ϵ as $\epsilon \downarrow 0$, (5.16) can be derived directly from (5.14). In what follows, we focus on deriving the expression (5.17). To this end, for

each $j \in \mathcal{J}$, three cases are discussed:

$$\gamma(\xi)T^{\epsilon} + (\lambda^{\epsilon})_j - p_j(\xi) \geq 0,\ (y^{\epsilon})_j - x_j \leq 0, \tag{8}$$

$$\gamma(\xi)T^{\epsilon} + (\lambda^{\epsilon})_j - p_j(\xi) < 0,\ (y^{\epsilon})_j - x_j \leq 0, \tag{9}$$

$$\gamma(\xi)T^{\epsilon} + (\lambda^{\epsilon})_i - p_j(\xi) < 0,\ (y^{\epsilon})_j - x_j > 0. \tag{10}$$

Case 1: If there exists a sequence $\{\epsilon_k\}_{k=1}^{\infty}$ converging to 0 such that (8) holds,

$$\lim_{k\to\infty} \left((y^{\epsilon_k})_j, (\lambda^{\epsilon_k})_j\right) = (0,0).$$

Thus, $\left|(\lambda^{\epsilon_k})_j - \bar{\lambda}_j\right| = 0$.

Case 2: If there exists a sequence $\{\epsilon_k\}_{k=1}^{\infty}$ converging to 0 such that (9) holds, an estimation holds

$$\begin{aligned}\lim_{k\to\infty} \left((y^{\epsilon_k})_j, (\lambda^{\epsilon_k})_j\right) &= \lim_{k\to\infty} \left(-\frac{\gamma(\xi)T^{\epsilon_k} - p_j(\xi)}{\gamma(\xi)}, 0\right) \\ &= \left(-\frac{\gamma(\xi)\lim_{k\to\infty} T^{\epsilon_k} - p_j(\xi)}{\gamma(\xi)}, 0\right) \\ &= \left(-\frac{\gamma(\xi)\bar{T} - p_j(\xi)}{\gamma(\xi)}, 0\right).\end{aligned}$$

Thus, $\left|(\lambda^{\epsilon_k})_j - \bar{\lambda}_j\right| = 0$.

Case 3: If there exists a sequence $\{\epsilon_k\}_{k=1}^{\infty}$ converging to 0 such that (10) holds,

$$\begin{aligned}\lim_{k\to\infty} \left((y^{\epsilon_k})_j, (\lambda^{\epsilon_k})_j\right) &= \lim_{k\to\infty} \left(-\frac{\epsilon_k\gamma(\xi)T^{\epsilon_k} - x_j - \epsilon_k p_j(\xi)}{\epsilon_k\gamma(\xi) + 1}, -\frac{\gamma(\xi)(T^{\epsilon_k} + x_j) - p_j(\xi)}{\epsilon_k\gamma(\xi) + 1}\right) \\ &= \left(x_j, -\gamma(\xi)(\bar{T} + x_j) + p_j(\xi)\right).\end{aligned}$$

Thus,

$$
\begin{aligned}
&\left|(\lambda^{\epsilon_k})_j - \bar{\lambda}_j\right| \\
&= \left|(\lambda^{\epsilon_k})_j + \gamma(\xi)(\bar{T} + x_j) - p_j(\xi)\right| \\
&= \left|-\frac{\gamma(\xi)(T^{\epsilon_k} + x_j) - p_j(\xi)}{\epsilon_k\gamma(\xi) + 1} + \gamma(\xi)(\bar{T} + x_j) - p_j(\xi)\right| \\
&= \frac{\left|-\gamma(\xi)(T^{\epsilon_k} + x_j) + p_j(\xi) + \gamma(\xi)(\bar{T} + x_j) - p_j(\xi) + \epsilon_k\gamma(\xi)(\gamma(\xi)(\bar{T} + x_j) - p_j(\xi))\right|}{\epsilon_k\gamma(\xi) + 1} \\
&\leq \frac{\gamma(\xi)\left|T^{\epsilon_k} - \bar{T}\right| + \left|\gamma(\xi)(\gamma(\xi)(\bar{T} + x_j) - p_j(\xi))\right|\epsilon_k}{\epsilon_k\gamma(\xi) + 1}.
\end{aligned}
$$

Collectively, it is known from **Case 1**, **Case 2** and **Case 3** that

$$(y^{\epsilon_k})_j - \bar{y}_j = 0, \tag{11}$$

$$(y^{\epsilon_k})_j - \bar{y}_j = -(T^{\epsilon_k} - \bar{T}), \tag{12}$$

$$(y^{\epsilon_k})_j - \bar{y}_j = \frac{-\gamma(\xi)T^{\epsilon_k} + p_j(\xi) - \gamma(\xi)x_j}{\epsilon_k\gamma(\xi) + 1} \cdot \epsilon_k. \tag{13}$$

Furthermore, it has that $T^{\epsilon_k} - \bar{T} \geq 0$ always holds. For the purpose of arriving at a contradiction, assume that $T^{\epsilon_k} - \bar{T} < 0$. Then (12) implies that

$$y^{\epsilon_k})_j - \bar{y}_j > 0.$$

Moreover, (11) and (13) induce

$$
\begin{aligned}
(y^{\epsilon_k})_j - \bar{y}_j &= 0, \\
(y^{\epsilon_k})_j - \bar{y}_j &\geq x_j - \bar{y}_j \geq 0,
\end{aligned}
$$

respectively. Clearly, it arrives at $T^{\epsilon_k} - \bar{T} \geq 0$, which contradicts our assumption.

In addition,

$$T^{\epsilon_k} - \bar{T} \leq \frac{-\gamma(\xi)T^{\epsilon_k} + \|p(\xi)\|_1 + \gamma(\xi)\|x\|_1}{\epsilon_k\gamma(\xi) + 1} \cdot \epsilon_k$$
$$\leq \left(\|p(\xi)\|_1 + \gamma(\xi)\|x\|_1\right)\epsilon_k.$$

Then, it follows that

$$\begin{aligned}
&\left|(\lambda^{\epsilon_k})_j - \bar{\lambda}_j\right| \\
&\leq \frac{\gamma(\xi)\left|T^{\epsilon_k} - \bar{T}\right| + \left|\gamma(\xi)(\gamma(\xi)(\bar{T} + x_j) - p_j(\xi))\right|\epsilon_k}{\epsilon_k\gamma(\xi) + 1} \\
&\leq \frac{\gamma(\xi)\left(\|p(\xi)\|_1 + \gamma(\xi)\|x\|_1\right) + \left|\gamma(\xi)(\gamma(\xi)(\bar{T} + x_j) - p_j(\xi))\right|}{\epsilon_k\gamma(\xi) + 1} \cdot \epsilon_k \\
&\leq \left(\gamma(\xi)\left(\|p(\xi)\|_1 + \gamma(\xi)\|x\|_1\right) + \gamma(\xi)^2\left(\|x\|_1 + \frac{\|p(\xi)\|_1}{\gamma(\xi)} + \|x\|_1\right) + \gamma(\xi)\|p(\xi)\|_1\right)\epsilon_k \\
&\leq 3\left(\gamma(\xi)^2\|x\|_1 + \gamma(\xi)\|p(\xi)\|_1\right)\epsilon_k,
\end{aligned}$$

where the third inequality follows Lemma 5.2 and the continuity of T^ϵ that

$$\bar{T} \leq \|x\|_1 + \frac{\|p(\xi)\|_1}{\gamma(\xi)}.$$

To summarize, for each $j \in \mathcal{J}$, it always holds

$$\left|(\lambda^{\epsilon})_j - \bar{\lambda}_j\right| \leq 3\left(\gamma(\xi)^2\|x\|_1 + \gamma(\xi)\|p(\xi)\|_1\right)\epsilon.$$

Then, according to the definition of l_2-norm, for any given $x \in \mathbb{R}^J_+$ one can compute

$$\bar{\kappa}(\xi) := 3\sqrt{J}\left(\gamma(\xi)^2\|x\|_1 + \gamma(\xi)\|p(\xi)\|_1\right).$$

□