Journal of Mathematical Research with Applications
Jan., 2021, Vol. 41, No. 1, pp. 99–110
DOI:10.3770/j.issn:2095-2651.2021.01.010
Http://jmre.dlut.edu.cn

Wasserstein Distributionally Robust Option Pricing

Wei LIU[1,2], Li YANG[1], Bo YU[1,*]
1. *School of Mathematical Sciences, Dalian University of Technology, Liaoning* 116024*, P. R. China;*
2. *School of Applied Mathematics, Beijing Normal University, Zhuhai, Guangdong* 519087*, P. R. China;*

Abstract In this paper, the option pricing problem is formulated as a distributionally robust optimization problem, which seeks to minimize the worst case replication error for a given distributional uncertainty set (DUS) of the random underlying asset returns. The DUS is defined as a Wasserstein ball centred the empirical distribution of the underlying asset returns. It is proved that the proposed model can be reformulated as a computational tractable linear programming problem. Finally, the results of the empirical tests are presented to show the significance of the proposed approach.

Keywords option pricing; Wasserstein distance; distributionally robust optimization

MR(2020) Subject Classification 90C15; 90C22; 90C25; 90C47

1. Introduction

The problem of pricing derivative securities has been one of the most well studied problems in finance and mathematical finance. The most well-known approach for pricing options is the Black-Scholes-Merton model, introduced by [1] and [2]. As the future market price of the underlying asset is an uncertain quantity, the Black-Scholes-Merton pricing model includes a key assumption, that the underlying asset returns follow a log-normal distribution with known volatility. There is sample empirical evidence suggesting that the strong assumption of the underlying asset price following a stationary geometric Brownian motion does not hold. Attempts have been made to model the volatility of the underlying asset as a stochastic quantity [3–6].

The option pricing problem is typically modeled as optimization problem under uncertainty [7]. To date, optimization under uncertainty has been addressed by several complementary modeling paradigms that differ mainly in the representation of uncertainty. For instance, stochastic programming assumes that the uncertainty is governed by a known probability distribution and seeks to minimize a probability functional such as the expected cost or a quantile of the cost distribution [8]. Contrary to stochastic programming, robust optimization ignores all distributional information and aims to minimize the worst-case cost under all possible uncertainty realizations [9, 10]. While stochastic programs may rely on distributional information that is not

Received January 20, 2020; Accepted October 24, 2020
Supported by the National Natural Science Foundation of China (Grant Nos. 11571061; 11401075; 11971092) and the Fundamental Research Funds for the Central Universities (Grant No. DUT17RC(4)38).
* Corresponding author
E-mail address: liuwei@bnuz.edu.cn (Wei LIU); yangli96@dlut.edu.cn (Li YANG); yubo@dlut.edu.cn (Bo YU)

available or hard to acquire in practice, robust optimization models may adopt an overly pessimistic view of the uncertainty and thereby promote over-conservative decisions. The emerging field of distributionally robust optimization aims to bridge the gap between the conservatism of robust optimization and the specificity of stochastic programming: it seeks to minimize a worst-case probability functional (e.g., the worst-case expectation), where the worst case is taken with respect to a distributional uncertainty set (DUS), that is, a family of distributions consistent with the given prior information on the uncertainty. The vast majority of the existing literature focuses on DUSs characterized through moment and support information [11–18]. For such kind of DUSs, the DRO model can be reformulated as a second-order cone program (SOCP) or semidefinite program (SDP) and efficient algorithms can be developed accordingly. However, DUSs can also be constructed via distance measures in the space of probability distributions. More general and important probability metric used to define the metric-based DUSs are the Wasserstein metric [19–24] and the ϕ-divergence [11, 24–30]. The Prohorov metric [31], the goodness-of-fit [32] and the likelihood function [33] were also used to define the DUSs. To the best of our knowledge, distributionally robust optimization paradigm has not been established for the option pricing problem. In this paper we aim to close this gap by adopting distributionally robust optimization based on a new Wasserstein DUS due to its attractive measure concentration properties. In particular, we propose to model the underlying price dynamics using Wasserstein DUS. We then utilize the ϵ-arbitrage approach [34] where one seeks a self-financing dynamic portfolio strategy that most closely approximates the payoff of an option. This choice of the l_1 norm to measure the error in replication when matched with Wasserstein DUS results in a computational tractable problem. In addition, we adapt our approach to capture the phenomenon of "implied volatility smile" that characterizes the classical Black-Scholes-Merton model. The reason behind the implied volatility smile may be the different levels of risk aversion of an option writer for different strikes.

The paper is structured as follows. In Section 2, we utilize the ϵ-arbitrage approach to price a European call option to obtain the corresponding stochastic optimization problem. In Section 3, the computational tractable reformulation of the distributionall robust option pricing problem based on Wasserstein DUS can be derived. Section 4 contains computational results and Section 5 includes our conclusions.

2. Option pricing stochastic optimization problem

An option is a contract that gives its owner the right to trade in a fixed number of shares of the specified underlying securities at a fixed price at any time on or before a given date. The act of making this transaction is referred to as exercising the option. The fixed price is termed the strike price, and the given date, the expiration date. A call option gives the right to buy the shares; a put option gives the right to sell the shares [4]. An option is associated with a payoff function. The payoff function determines the value of the option after the realization of random returns of the underlying securities. For instance, a European Call options payoff is given by $f(S_T, K) = [S_T - K]_+$, where $[x]_+ := \max\{x, 0\}$, S_T denotes the price of the underlying security at the time of expiry T, and K denotes the strike price. The option pricing problem refers to the problem of calculating the value of an option before the realization of the random returns. The idea of the ϵ-arbitrage approach [34] is to find a replicating portfolio that consists of the underlying stock $\mathbb{S}$ and a risk-free asset $\mathbb{B}$ so that the value of this portfolio at the time of exercise matches the payoff of the option as closely as possible. The replication error is given by $|f(S_T, K) - V_T|$, where V_T is the value of the portfolio at the time of exercise T. In a distributionally robust optimization setting, our goal is to find a portfolio that minimizes the worst case expected replication error (denoted by ϵ), between the portfolio wealth and the option payoff, over all possible probability distributions of the underlying stock returns that fall into a predetermined DUS. The price of the option would thus be the initial value of this replicating portfolio. We now consider a discrete model of the underlying stock price movements where the price of the stock changes at discrete points of time. Denote the return from period $[t, t+1)$ by $\tilde{r}_t^S$. Suppose that the random returns variables $\{\tilde{r}_1^S, \tilde{r}_2^S, \ldots, \tilde{r}_T^S\}$ be identical and independent random variables. In the context of a European call option, the associated optimization problem can be represented as follows:

$$\begin{aligned} \min_{u_t^S, u_t^B, v_t} \quad & \left|[S_T - K]_+ - \left(u_T^S + u_T^B\right)\right| \\ \text{s.t.} \quad & u_t^S = (1 + \tilde{r}_{t-1}^S)\left(u_{t-1}^S + v_{t-1}\right), \quad \forall t = 1, \ldots, T, \\ & u_t^B = (1 + r_{t-1}^B)\left(u_{t-1}^B - v_{t-1}\right), \quad \forall t = 1, \ldots, T, \end{aligned} \tag{2.1}$$

where u_t^S is the amount invested in the underlying stock, u_t^B is the amount invested in the risk-less asset, and v_t is the amount traded from the underlying stock to the risk-less asset during the period $[t, t+1)$. From optimization perspective, we seek to minimize the replication error and obtain the price of the option would then be given by $u_0^S + u_0^B$. By introducing the following variable transformations: $x_t^S = u_t^S/\tilde{\xi}_t^S$, $x_t^B = u_t^B/\xi_t^B$, $y_t = v_t/\tilde{\xi}_t^S$, where $\tilde{\xi}_t^S := \prod_{i=0}^{t-1}\left(1 + \tilde{r}_i^S\right)$, is the cumulative return up to time t, and $\xi_t^B := \prod_{i=0}^{t-1}(1 + r_i^B)$, problem (2.1) can be rewritten as

$$\begin{aligned} \min_{x_t^S, x_t^B, y_t} \quad & |[S_0\tilde{\xi}_T^S - K]_+ - (\tilde{\xi}_T^S x_T^S + \xi_T^B x_T^B)| \\ \text{s.t.} \quad & x_t^S = x_{t-1}^S + y_{t-1}, && \forall t = 1, \ldots, T, \\ & x_t^B = x_{t-1}^B - y_{t-1}\frac{\tilde{\xi}_{t-1}^S}{\xi_{t-1}^B}, && \forall t = 1, \ldots, T, \end{aligned}$$

where $\tilde{\xi}_0^S := 1$ and $\xi_0^B := 1$. Substituting all intermediate x_t^S, x_t^B, we obtain the following formulation:

$$\min_{x_0^S, x_0^B, \boldsymbol{y}} \left|[S_0\tilde{\xi}_T^S - K]_+ - \left(x_0^S + \sum_{t=1}^{T} y_{t-1}\right)\tilde{\xi}_T^S - x_0^B\xi_T^B + \sum_{t=0}^{T-1} y_t \frac{\xi_T^B}{\xi_t^B}\tilde{\xi}_t^S\right|. \tag{2.2}$$

By denoting $\boldsymbol{a} := (0, \ldots, 0, S_0)' \in \mathbb{R}^T$, $\boldsymbol{b}(x_0^S, \boldsymbol{y}) := (y_1\frac{\xi_T^B}{\xi_1^B}, \ldots, y_{t-1}\frac{\xi_T^B}{\xi_{t-1}^B}, -x_0^S - \sum_{t=1}^{T} y_{t-1})' \in \mathbb{R}^T$, $c(x_0^B, \boldsymbol{y}) := -x_0^B\xi_T^B + y_0\frac{\xi_T^B}{\xi_0^B} \in \mathbb{R}$, $\tilde{\boldsymbol{\xi}}^S := (\tilde{\xi}_1^S, \tilde{\xi}_2^S, \ldots, \tilde{\xi}_T^S)' \in \mathbb{R}^T$, problem (2.2) can be simplified as

$$\min_{x_0^S, x_0^B, \boldsymbol{y}} |[\boldsymbol{a}'\tilde{\boldsymbol{\xi}}^S - K]_+ + \boldsymbol{b}(x_0^S, \boldsymbol{y})'\tilde{\boldsymbol{\xi}}^S + c(x_0^B, \boldsymbol{y})|. \tag{2.3}$$

For ease of notations, $\tilde{\boldsymbol{\xi}}^S$ is replaced by $\tilde{\boldsymbol{\xi}}$ in what follows. Suppose that the random cumulative returns $\tilde{\boldsymbol{\xi}}$ of the underlying stock enjoys the probability distribution $\mathbb{Q}$. By taking the expectation of the replication error function in problem (2.3), we obtain the following optimization problem:

$$\min_{x_0^S, x_0^B, \boldsymbol{y}} \mathbb{E}^{\mathbb{Q}}[|[\boldsymbol{a}'\tilde{\boldsymbol{\xi}} - K]_+ + \boldsymbol{b}(x_0^S, \boldsymbol{y})'\tilde{\boldsymbol{\xi}} + c(x_0^B, \boldsymbol{y})|], \tag{2.4}$$

which is a stochastic optimization problem. In data-driven setting, we are given the historical stock price information, i.e., the sample set $\{\hat{\boldsymbol{\xi}}^1, \hat{\boldsymbol{\xi}}^2, \ldots, \hat{\boldsymbol{\xi}}^N\}$, which are typically assumed to be drawn IID from the probability distribution $\mathbb{Q}$. Using the Sample Average Approximation (SAA) approach, problem (2.4) can be approximated as

$$\min_{x_0^S, x_0^B, \boldsymbol{y}} \frac{1}{N} \sum_{i=1}^{N} |[\boldsymbol{a}'\hat{\boldsymbol{\xi}}^i - K]_+ + \boldsymbol{b}(x_0^S, \boldsymbol{y})'\hat{\boldsymbol{\xi}}^i + c(x_0^B, \boldsymbol{y})|, \tag{2.5}$$

which is a convex optimization problem. We prove the following proposition.

Proposition 2.1 *The optimization problem (2.5) is equivalent to the following linear optimization problem*

$$\begin{array}{lll}
\min\limits_{x_0^S, x_0^B, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\epsilon}} & \frac{1}{N}\sum_{i=1}^N \epsilon_i & \\
\text{s.t.} & z_i + \boldsymbol{b}(x_0^S, \boldsymbol{y})'\hat{\boldsymbol{\xi}}^i + c(x_0^B, \boldsymbol{y}) \le \epsilon_i, & \forall i \le N, \\
& -z_i - \boldsymbol{b}(x_0^S, \boldsymbol{y})'\hat{\boldsymbol{\xi}}^i - c(x_0^B, \boldsymbol{y}) \le \epsilon_i, & \forall i \le N, \\
& \boldsymbol{a}'\hat{\boldsymbol{\xi}}^i - K \ge z_i, & \forall i \le N, \\
& \boldsymbol{z} \ge 0. &
\end{array} \tag{2.6}$$

Proof It is easy to derive problem (2.6) by introducing the epigraphical auxiliary variables $\boldsymbol{\epsilon}$ and $\boldsymbol{z} \ge 0$ for problem (2.5). □

3. Wasserstein distributionally robust options price model

Problem (2.4) is a stochastic optimization problem that is dependent on the probability distribution $\mathbb{Q}$ of the random cumulative returns $\tilde{\boldsymbol{\xi}}$ of the underlying stock. In practice, however, $\mathbb{Q}$ is only indirectly observable through the historical stock price information, i.e., the sample set $\{\hat{\boldsymbol{\xi}}^1, \hat{\boldsymbol{\xi}}^2, \ldots, \hat{\boldsymbol{\xi}}^N\}$. Thus, the probability distribution $\mathbb{Q}$ is itself uncertain, which motivates us to address problem (2.4) from a distributionally robust optimization perspective. This means that we use the sample set $\{\hat{\boldsymbol{\xi}}^1, \hat{\boldsymbol{\xi}}^2, \ldots, \hat{\boldsymbol{\xi}}^N\}$ to construct a DUS $\mathcal{P}$, that is, a family of distributions that contains the unknown distribution $\mathbb{Q}$ with high confidence. Then we solve the distributionally robust optimization problem which minimizes the worst-case expected replication error function. In this paper we propose to use the Wasserstein metric to construct $\mathcal{P}$ as a ball in the space of probability distributions.

Definition 3.1 *The Wasserstein metric* $d_W: \mathcal{M}(\Xi) \times \mathcal{M}(\Xi) \to \mathbb{R}_+$ *is defined via*

$$d_W(\mathbb{Q}, \mathbb{P}) := \inf\left\{ \int_{\Xi^2} \|\boldsymbol{\xi} - \boldsymbol{\xi}'\| \Pi(\mathrm{d}\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\xi}') : \begin{array}{l} \Pi \textit{ is a joint distribution of } \boldsymbol{\xi} \textit{ and } \boldsymbol{\xi}' \\ \textit{with marginals } \mathbb{Q} \textit{ and } \mathbb{P}, \textit{ respectively} \end{array} \right\},$$

where $\Xi \subseteq \mathbb{R}^m$ *is the support set of* $\boldsymbol{\xi}$ *and* $\|\cdot\|$ *represents an arbitrary norm on* Ξ.

We denote by $\mathbb{B}_\rho(\hat{\mathbb{P}}_N) := \{\mathbb{Q} : d_W(\mathbb{Q}, \hat{\mathbb{P}}_N) \le \rho\}$ the ball of radius ρ centered at $\hat{\mathbb{P}}_N$ with respect to the Wasserstein metric in what follows. The center of the Wasserstein ball is the empirical distribution $\hat{\mathbb{P}}_N = \frac{1}{N}\sum_{i=1}^N \delta_{\hat{\boldsymbol{\xi}}^i}$, where $\delta_{\hat{\boldsymbol{\xi}}^i}$ denotes the Dirac function at $\hat{\boldsymbol{\xi}}^i$. Thus, the following distributionally robust optimization problem based the Wasserstein ball can be derived:

$$\min_{x_0^S, x_0^B, \boldsymbol{y}} \max_{\mathbb{Q}\in\mathbb{B}_\rho(\hat{\mathbb{P}}_N)} \mathbb{E}^{\mathbb{Q}}[|[\boldsymbol{a}'\tilde{\boldsymbol{\xi}} - K]_+ + \boldsymbol{b}(x_0^S, \boldsymbol{y})'\tilde{\boldsymbol{\xi}} + c(x_0^B, \boldsymbol{y})|]. \tag{3.1}$$

The following theorem presents a tractable reformulation of the distributionally robust optimization problem (3.1) and thus constitutes the main result of this paper.

Theorem 3.2 *Let $\Xi = \{\boldsymbol{\xi} : C\boldsymbol{\xi} \le \boldsymbol{d}\}$. Then problem (3.1) is equivalent to the following convex optimization problem:*

$$\begin{array}{lll}
\min & \epsilon & \\
\text{s.t.} & \lambda\rho + \frac{1}{N}\sum_{i=1}^N s_i \le \epsilon, & \\
& \mu\rho + \frac{1}{N}\sum_{i=1}^N t_i \le \epsilon, & \\
& (\boldsymbol{d} - C\hat{\boldsymbol{\xi}}^i)'\boldsymbol{\gamma}_{i1} - K + c(x_0^B, \boldsymbol{y}) + (\boldsymbol{a} + \boldsymbol{b}(x_0^S, \boldsymbol{y}))'\hat{\boldsymbol{\xi}}^i \le s_i, & \forall i \le N, \\
& (\boldsymbol{d} - C\hat{\boldsymbol{\xi}}^i)'\boldsymbol{\gamma}_{i2} + c(x_0^B, \boldsymbol{y}) + \boldsymbol{b}(x_0^S, \boldsymbol{y})'\hat{\boldsymbol{\xi}}^i \le s_i, & \forall i \le N, \\
& K\alpha + \boldsymbol{d}'\boldsymbol{\beta} - c(x_0^B, \boldsymbol{y}) + \boldsymbol{z}_i'\hat{\boldsymbol{\xi}}^i \le t_i, & \forall i \le N, \\
& \alpha\boldsymbol{a} + \boldsymbol{b}(x_0^S, \boldsymbol{y}) + \boldsymbol{z}_i + C'\boldsymbol{\beta} = 0, & \forall i \le N, \\
& \|\boldsymbol{a} + \boldsymbol{b}(x_0^S, \boldsymbol{y}) - C'\boldsymbol{\gamma}_{i1}\|_* \le \lambda, & \forall i \le N, \\
& \|\boldsymbol{b}(x_0^S, \boldsymbol{y}) - C'\boldsymbol{\gamma}_{i2}\|_* \le \lambda, & \forall i \le N, \\
& \boldsymbol{\gamma}_{i1} \ge 0, \boldsymbol{\gamma}_{i2} \ge 0, & \forall i \le N, \\
& \lambda \ge 0, \mu \ge 0, 0 \le \alpha \le 1, \boldsymbol{\beta} \ge 0, &
\end{array} \tag{3.2}$$

where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$.

Proof Introducing the epigraphical auxiliary variable ϵ, problem (3.1) can be formulated as

$$\begin{array}{lll}
\min\limits_{x_0^S, x_0^B, \boldsymbol{y}, \epsilon} & \epsilon & \\
\text{s.t.} & \mathbb{E}^{\mathbb{Q}}[[\boldsymbol{a}'\tilde{\boldsymbol{\xi}} - K]_+ + \boldsymbol{b}(x_0^S, \boldsymbol{y})'\tilde{\boldsymbol{\xi}} + c(x_0^B, \boldsymbol{y})] \le \epsilon, & \forall \mathbb{Q} \in \mathbb{B}_\tau(\hat{\mathbb{P}}_N), \\
& \mathbb{E}^{\mathbb{Q}}[-[\boldsymbol{a}'\tilde{\boldsymbol{\xi}} - K]_+ - \boldsymbol{b}(x_0^S, \boldsymbol{y})'\tilde{\boldsymbol{\xi}} - c(x_0^B, \boldsymbol{y})] \le \epsilon, & \forall \mathbb{Q} \in \mathbb{B}_\tau(\hat{\mathbb{P}}_N),
\end{array}$$

which is rewritten further as

$$\begin{array}{ll}
\min\limits_{x_0^S, x_0^B, \boldsymbol{y}, \epsilon} & \epsilon \\
\text{s.t.} & \mathbb{E}^{\mathbb{Q}}[\max\{(\boldsymbol{a} + \boldsymbol{b}(x_0^S, \boldsymbol{y}))'\tilde{\boldsymbol{\xi}} - K + c(x_0^B, \boldsymbol{y}), \boldsymbol{b}(x_0^S, \boldsymbol{y})'\tilde{\boldsymbol{\xi}} + c(x_0^B, \boldsymbol{y})\}] \le \epsilon, \\
& \qquad \forall \mathbb{Q} \in \mathbb{B}_\tau(\hat{\mathbb{P}}_N), \\
& \mathbb{E}^{\mathbb{Q}}[\min\{(-\boldsymbol{a} - \boldsymbol{b}(x_0^S, \boldsymbol{y}))'\tilde{\boldsymbol{\xi}} + K - c(x_0^B, \boldsymbol{y}), -\boldsymbol{b}(x_0^S, \boldsymbol{y})'\tilde{\boldsymbol{\xi}} - c(x_0^B, \boldsymbol{y})\}] \le \epsilon, \\
& \qquad \forall \mathbb{Q} \in \mathbb{B}_\tau(\hat{\mathbb{P}}_N).
\end{array} \tag{3.3}$$

It is easy to derive that problem (3.3) is equivalent to the following optimization problem:

$$\begin{aligned}
\min_{x_0^S, x_0^B, \boldsymbol{y}, \epsilon} \quad & \epsilon \\
\text{s.t.} \quad & \max_{\mathbb{Q}\in\mathbb{B}_\rho(\hat{\mathbb{P}}_N)} \mathbb{E}^{\mathbb{Q}}[\max\{(\boldsymbol{a}+\boldsymbol{b}(x_0^S,\boldsymbol{y}))'\tilde{\boldsymbol{\xi}} - K + c(x_0^B,\boldsymbol{y}), \boldsymbol{b}(x_0^S,\boldsymbol{y})'\tilde{\boldsymbol{\xi}} + c(x_0^B,\boldsymbol{y})\}] \le \epsilon, \\
& \max_{\mathbb{Q}\in\mathbb{B}_\rho(\hat{\mathbb{P}}_N)} \mathbb{E}^{\mathbb{Q}}[\min\{(-\boldsymbol{a}-\boldsymbol{b}(x_0^S,\boldsymbol{y}))'\tilde{\boldsymbol{\xi}} + K - c(x_0^B,\boldsymbol{y}), -\boldsymbol{b}(x_0^S,\boldsymbol{y})'\tilde{\boldsymbol{\xi}} - c(x_0^B,\boldsymbol{y})\}] \le \epsilon.
\end{aligned} \tag{3.4}$$

For ease of notation, we suppress the dependence on the decision variable x_0^S, x_0^B and $\boldsymbol{y}$ in the constraints of problem (3.4), and denote $l_1(\tilde{\boldsymbol{\xi}}) = (\boldsymbol{a}+\boldsymbol{b}(x_0^S,\boldsymbol{y}))'\tilde{\boldsymbol{\xi}} - K + c(x_0^B,\boldsymbol{y})$, $l_2(\tilde{\boldsymbol{\xi}}) = \boldsymbol{b}(x_0^S,\boldsymbol{y})'\tilde{\boldsymbol{\xi}} + c(x_0^B,\boldsymbol{y})$, $L_{\max}(\tilde{\boldsymbol{\xi}}) = \max_{i=1,2}\{l_i(\tilde{\boldsymbol{\xi}})\}$ and $L_{\min}(\tilde{\boldsymbol{\xi}}) = \min_{i=1,2}\{-l_i(\tilde{\boldsymbol{\xi}})\}$. Thus problem (3.4) can be represented as follows:

$$\begin{aligned}
\min \quad & \epsilon \\
\text{s.t.} \quad & \max_{\mathbb{Q}\in\mathbb{B}_\rho(\hat{\mathbb{P}}_N)} \mathbb{E}^{\mathbb{Q}}[L_{\max}(\tilde{\boldsymbol{\xi}})] \le \epsilon, \\
& \max_{\mathbb{Q}\in\mathbb{B}_\rho(\hat{\mathbb{P}}_N)} \mathbb{E}^{\mathbb{Q}}[L_{\min}(\tilde{\boldsymbol{\xi}})] \le \epsilon.
\end{aligned} \tag{3.5}$$

Firstly, we consider the following worst-case expectation problem in the left-hand side of the first constraint of problem (3.5), that is,

$$\max_{\mathbb{Q}\in\mathbb{B}_\rho(\hat{\mathbb{P}}_N)} \{\mathbb{E}^{\mathbb{Q}}[L_{\max}(\tilde{\boldsymbol{\xi}})]\}. \tag{3.6}$$

By using Definition 3.1, problem (3.6) can be rewritten as

$$\begin{aligned}
&\max_{\mathbb{Q}\in\mathbb{B}_\rho(\hat{\mathbb{P}}_N)} \{\mathbb{E}^{\mathbb{Q}}[L_{\max}(\tilde{\boldsymbol{\xi}})]\} \\
&= \left\{ \begin{array}{ll} \max\limits_{\Pi,\mathbb{Q}} & \mathbb{E}^{\mathbb{Q}}[L_{\max}(\tilde{\boldsymbol{\xi}})] \\ \text{s.t.} & \int_{\Xi^2} \|\boldsymbol{\xi}-\boldsymbol{\xi}'\| \Pi(\mathrm{d}\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\xi}') \le \rho \\ & \left\{ \begin{array}{l} \Pi \text{ is a joint distribution of } \boldsymbol{\xi} \text{ and } \boldsymbol{\xi}' \\ \text{with marginals } \mathbb{Q} \text{ and } \hat{\mathbb{P}}_N, \text{ respectively} \end{array} \right. \end{array} \right. \\
&= \left\{ \begin{array}{ll} \max\limits_{\Pi,\mathbb{Q}} & \int_{\Xi} L_{\max}(\boldsymbol{\xi}) \mathbb{Q}(\mathrm{d}\boldsymbol{\xi}) \\ \text{s.t.} & \int_{\Xi^2} \|\boldsymbol{\xi}-\boldsymbol{\xi}'\| \Pi(\mathrm{d}\boldsymbol{\xi}, \mathrm{d}\boldsymbol{\xi}') \le \rho \\ & \left\{ \begin{array}{l} \Pi \text{ is a joint distribution of } \boldsymbol{\xi} \text{ and } \boldsymbol{\xi}' \\ \text{with marginals } \mathbb{Q} \text{ and } \hat{\mathbb{P}}_N, \text{ respectively} \end{array} \right. \end{array} \right. \\
&= \left\{ \begin{array}{ll} \max\limits_{\mathbb{Q}_i\in\mathcal{M}(\Xi)} & \frac{1}{N}\sum_{i=1}^N \int_{\Xi} L_{\max}(\boldsymbol{\xi}) \mathbb{Q}_i(\mathrm{d}\boldsymbol{\xi}) \\ \text{s.t.} & \frac{1}{N}\sum_{i=1}^N \int_{\Xi} \|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}^i\| \mathbb{Q}_i(\mathrm{d}\boldsymbol{\xi}) \le \rho, \\ & \mathbb{Q}_i \in \mathcal{M}(\Xi), \forall i \le N, \end{array} \right.
\end{aligned} \tag{3.7}$$

where the last equality follows from the law of total probability. The Lagrangian function of problem (3.7) can be obtained $\mathcal{L}(\mathbb{Q}_1,\dots,\mathbb{Q}_N;\lambda) = \lambda\rho + \frac{1}{N}\sum_{i=1}^N \int_{\Xi}(L_{\max}(\boldsymbol{\xi}) - \lambda\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}^i\|)\mathbb{Q}_i(\mathrm{d}\boldsymbol{\xi})$.

Using a standard duality argument, we obtain

$$
\begin{aligned}
&\max_{\mathbb{Q}_1,\ldots,\mathbb{Q}_N\in\mathcal{M}(\Xi)}\min_{\lambda\geq 0}\mathcal{L}(\mathbb{Q}_1,\ldots,\mathbb{Q}_N;\lambda)\\
&=\min_{\lambda\geq 0}\max_{\mathbb{Q}_i\in\mathcal{M}(\Xi)}\lambda\rho+\frac{1}{N}\sum_{i=1}^{N}\int_{\Xi}(L_{\max}(\boldsymbol{\xi})-\lambda\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}^i\|)\mathbb{Q}_i(\mathrm{d}\boldsymbol{\xi})\\
&=\begin{cases}\inf\limits_{\lambda,s_i} & \lambda\rho+\frac{1}{N}\sum_{i=1}^{N}s_i\\ \text{s.t.} & \sup\limits_{\boldsymbol{\xi}\in\Xi}\{L_{\max}(\boldsymbol{\xi})-\lambda\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}^i\|\}\leq s_i,\quad\forall i\leq N,\\ & \lambda\geq 0.\end{cases}\\
&=\begin{cases}\inf\limits_{\lambda,s_i} & \lambda\rho+\frac{1}{N}\sum_{i=1}^{N}s_i\\ \text{s.t.} & \sup\limits_{\boldsymbol{\xi}\in\Xi}\{l_1(\boldsymbol{\xi})-\max\limits_{\|\boldsymbol{z}_{i1}\|_*\leq\lambda}\boldsymbol{z}_{i1}'(\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}^i)\}\leq s_i,\quad\forall i\leq N,\\ & \sup\limits_{\boldsymbol{\xi}\in\Xi}\{l_2(\boldsymbol{\xi})-\max\limits_{\|\boldsymbol{z}_{i2}\|_*\leq\lambda}\boldsymbol{z}_{i2}'(\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}^i)\}\leq s_i,\quad\forall i\leq N,\\ & \lambda\geq 0.\end{cases}\\
&=\begin{cases}\inf\limits_{\lambda,s_i} & \lambda\rho+\frac{1}{N}\sum_{i=1}^{N}s_i\\ \text{s.t.} & \min\limits_{\|\boldsymbol{z}_{i1}\|_*\leq\lambda}\sup\limits_{\boldsymbol{\xi}\in\Xi}\{l_1(\boldsymbol{\xi})-\boldsymbol{z}_{i1}'(\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}^i)\}\leq s_i\quad\forall i\leq N,\\ & \min\limits_{\|\boldsymbol{z}_{i2}\|_*\leq\lambda}\sup\limits_{\boldsymbol{\xi}\in\Xi}\{l_2(\boldsymbol{\xi})-\boldsymbol{z}_{i2}'(\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}^i)\}\leq s_i\quad\forall i\leq N,\\ & \lambda\geq 0,\end{cases}
\end{aligned}
\tag{3.8}
$$

where the second equality follows from the fact that $\mathcal{M}(\Xi)$ contains all the Dirac distributions supported on $\mathcal{M}(\Xi)$ and the third equality follows from the definition of the dual norm. For ease of notation, we suppress the dependence on the decision variable α_0^S, α_0^B and $\boldsymbol{y}$ in the functions $\boldsymbol{b}(x_0^S,\boldsymbol{y}))$ and $c(x_0^B,\boldsymbol{y})$, then denote $l_1(\boldsymbol{\xi})=\boldsymbol{a}+\boldsymbol{b}'\boldsymbol{\xi}-K+c$ and $l_2(\boldsymbol{\xi})=\boldsymbol{b}'\boldsymbol{\xi}+c$ in what follows. Thus problem (3.8) can be simplified as follows:

$$
\begin{aligned}
&\begin{cases}\inf\limits_{\lambda,s_i} & \lambda\rho+\frac{1}{N}\sum_{i=1}^{N}s_i\\ \text{s.t.} & \min\limits_{\|\boldsymbol{z}_{i1}\|_*\leq\lambda}\sup\limits_{\boldsymbol{\xi}\in\Xi}\{(\boldsymbol{a}+\boldsymbol{b}-\boldsymbol{z}_{i1})'\boldsymbol{\xi}-K+c+\boldsymbol{z}_{i1}'\hat{\boldsymbol{\xi}}^i\}\leq s_i,\quad\forall i\leq N,\\ & \min\limits_{\|\boldsymbol{z}_{i2}\|_*\leq\lambda}\sup\limits_{\boldsymbol{\xi}\in\Xi}\{(\boldsymbol{b}-\boldsymbol{z}_{i2})'\boldsymbol{\xi}+c+\boldsymbol{z}_{i2}'\hat{\boldsymbol{\xi}}^i\}\leq s_i,\quad\forall i\leq N,\\ & \lambda\geq 0.\end{cases}\\
&=\begin{cases}\inf\limits_{\lambda,s_i} & \lambda\rho+\frac{1}{N}\sum_{i=1}^{N}s_i\\ \text{s.t.} & \sup\limits_{\boldsymbol{\xi}\in\Xi}\{(\boldsymbol{a}+\boldsymbol{b}-\boldsymbol{z}_{i1})'\boldsymbol{\xi}-K+c+\boldsymbol{z}_{i1}'\hat{\boldsymbol{\xi}}^i\}\leq s_i,\quad\forall i\leq N,\\ & \sup\limits_{\boldsymbol{\xi}\in\Xi}\{(\boldsymbol{b}-\boldsymbol{z}_{i2})'\boldsymbol{\xi}+c+\boldsymbol{z}_{i2}'\hat{\boldsymbol{\xi}}^i\}\leq s_i,\quad\forall i\leq N,\\ & \lambda\geq 0,\|\boldsymbol{z}_{i1}\|_*\leq\lambda,\|\boldsymbol{z}_{i2}\|_*\leq\lambda,\quad\forall i\leq N.\end{cases}
\end{aligned}
\tag{3.9}
$$

Due to the fact that the maximization problems of the left-hand side of the two groups of constraints in problem (3.9) are the linear optimization problems, where $\Xi=\{\boldsymbol{\xi}:C\boldsymbol{\xi}\leq\boldsymbol{d}\}$, by using strong linear programming duality, we obtain

$$
\begin{aligned}
\inf_{\lambda,\boldsymbol{s}} \quad & \lambda\rho + \frac{1}{N}\sum_{i=1}^{N} s_i \\
\text{s.t.} \quad & (\boldsymbol{d} - C\hat{\boldsymbol{\xi}}^i)'\boldsymbol{\gamma}_{i1} - K + c + (\boldsymbol{a}+\boldsymbol{b})'\hat{\boldsymbol{\xi}}^i \le s_i, && \forall i \le N, \\
& (\boldsymbol{d} - C\hat{\boldsymbol{\xi}}^i)'\boldsymbol{\gamma}_{i2} + c + \boldsymbol{b}'\hat{\boldsymbol{\xi}}^i \le s_i, && \forall i \le N, \\
& \|\boldsymbol{a}+\boldsymbol{b} - C'\boldsymbol{\gamma}_{i1}\|_* \le \lambda, \|\boldsymbol{b} - C'\boldsymbol{\gamma}_{i2}\|_* \le \lambda, && \forall i \le N, \\
& \lambda \ge 0, \boldsymbol{\gamma}_{i1} \ge 0, \boldsymbol{\gamma}_{i2} \ge 0, && \forall i \le N.
\end{aligned} \tag{3.10}
$$

Secondly, we consider the following worst-case expectation problem in the left-hand side of the first constraint of problem (3.5), that is,

$$
\max_{\mathbb{Q}\in\mathbb{B}_\rho(\hat{\mathbb{P}}_N)} \{\mathbb{E}^{\mathbb{Q}}[L_{\min}(\widetilde{\boldsymbol{\xi}})]\}. \tag{3.11}
$$

Similarly, problem (3.11) can be reformulated as follows:

$$
\begin{aligned}
\inf_{\alpha,\boldsymbol{\beta},} \quad & \mu\rho + \tfrac{1}{N}\textstyle\sum_{i=1}^{N} t_i \\
\text{s.t.} \quad & K\alpha + \boldsymbol{d}'\boldsymbol{\beta} - c + \boldsymbol{z}_i'\hat{\boldsymbol{\xi}}^i \le t_i, && \forall i \le N, \\
& \alpha\boldsymbol{a} + \boldsymbol{b} + \boldsymbol{z}_i + C'\boldsymbol{\beta} = 0, && \forall i \le N, \\
& \mu \ge 0, 0 \le \alpha \le 1, \boldsymbol{\beta} \ge 0.
\end{aligned} \tag{3.12}
$$

By combining problem (3.10) and problem (3.12), problem (3.5) can be reformulated as follows:

$$
\begin{aligned}
\inf \quad & \epsilon \\
\text{s.t.} \quad & \lambda\rho + \tfrac{1}{N}\sum_{i=1}^{N} s_i \le \epsilon, \\
& \mu\rho + \tfrac{1}{N}\sum_{i=1}^{N} t_i \le \epsilon, \\
& (\boldsymbol{d} - C\hat{\boldsymbol{\xi}}^i)'\boldsymbol{\gamma}_{i1} - K + c + (\boldsymbol{a}+\boldsymbol{b})'\hat{\boldsymbol{\xi}}^i \le s_i, && \forall i \le N, \\
& (\boldsymbol{d} - C\hat{\boldsymbol{\xi}}^i)'\boldsymbol{\gamma}_{i2} + c + \boldsymbol{b}'\hat{\boldsymbol{\xi}}^i \le s_i, && \forall i \le N, \\
& K\alpha + \boldsymbol{d}'\boldsymbol{\beta} - c + \boldsymbol{z}_i'\hat{\boldsymbol{\xi}}^i \le t_i, && \forall i \le N, \\
& \alpha\boldsymbol{a} + \boldsymbol{b} + \boldsymbol{z}_i + C'\boldsymbol{\beta} = 0, && \forall i \le N, \\
& \|\boldsymbol{a}+\boldsymbol{b} - C'\boldsymbol{\gamma}_{i1}\|_* \le \lambda, && \forall i \le N, \\
& \|\boldsymbol{b} - C'\boldsymbol{\gamma}_{i2}\|_* \le \lambda, && \forall i \le N, \\
& \boldsymbol{\gamma}_{i1} \ge 0, \boldsymbol{\gamma}_{i2} \ge 0, && \forall i \le N, \\
& \lambda \ge 0, \mu \ge 0, 0 \le \alpha \le 1, \boldsymbol{\beta} \ge 0,
\end{aligned}
$$

that is,

$$
\begin{aligned}
\min \quad & \epsilon \\
\text{s.t.} \quad & \lambda\rho + \frac{1}{N}\sum_{i=1}^{N} s_i \le \epsilon, \\
& \mu\rho + \frac{1}{N}\sum_{i=1}^{N} t_i \le \epsilon,
\end{aligned}
$$

$$
\begin{aligned}
&(\boldsymbol{d} - C\hat{\boldsymbol{\xi}}^i)'\boldsymbol{\gamma}_{i1} - K + c(x_0^B, \boldsymbol{y}) + (\boldsymbol{a} + \boldsymbol{b}(x_0^S, \boldsymbol{y}))'\hat{\boldsymbol{\xi}}^i \le s_i, \quad && \forall i \le N,\\
&(\boldsymbol{d} - C\hat{\boldsymbol{\xi}}^i)'\boldsymbol{\gamma}_{i2} + c(x_0^B, \boldsymbol{y}) + \boldsymbol{b}(x_0^S, \boldsymbol{y})'\hat{\boldsymbol{\xi}}^i \le s_i, && \forall i \le N,\\
&K\alpha + \boldsymbol{d}'\boldsymbol{\beta} - c(x_0^B, \boldsymbol{y}) + \boldsymbol{z}_i'\hat{\boldsymbol{\xi}}^i \le t_i, && \forall i \le N,\\
&\alpha\boldsymbol{a} + \boldsymbol{b}(x_0^S, \boldsymbol{y}) + \boldsymbol{z}_i + C'\boldsymbol{\beta} = 0, && \forall i \le N,\\
&\|\boldsymbol{a} + \boldsymbol{b}(x_0^S, \boldsymbol{y}) - C'\boldsymbol{\gamma}_{i1}\|_* \le \lambda, && \forall i \le N,\\
&\|\boldsymbol{b}(x_0^S, \boldsymbol{y}) - C'\boldsymbol{\gamma}_{i2}\|_* \le \lambda, && \forall i \le N,\\
&\boldsymbol{\gamma}_{i1} \ge 0, \boldsymbol{\gamma}_{i2} \ge 0, && \forall i \le N,\\
&\lambda \ge 0, \mu \ge 0, 0 \le \alpha \le 1, \boldsymbol{\beta} \ge 0,
\end{aligned}
$$

thus the conclusion holds. □

Remark 3.3 When $\|\cdot\|$ is taken as 1-norm and ∞-norm, $\|\cdot\|_*$ is ∞-norm and 1-norm, respectively. Thus problem (3.2) can be reduced to the linear optimization problem. When $\|\cdot\|$ is taken as 2-norm, problem (3.2) can be reduced to an SOCP problem [35], which can be solved efficiently by CVX software package [36].

4. Numerical results

We now present the power of Wasserstein distributionally robust option pricing model in three empirical experiments.

• In the first experiment, we aim to price SSE 50 ETF 42 days European call options with spot price ¥3.0210 for various strikes in the range ¥2.7 − ¥3.4.

• In the second experiment, we aim to price SSE 50 ETF 35 days European call options with spot price ¥2.9870 for various strikes in the range ¥2.7 − ¥3.4.

• In the last experiment, we aim to price SSE 50 ETF 29 days European call options with spot price ¥2.9810 for various strikes in the range ¥2.7 − ¥3.4

No.	K/S	ρ	Mkt price	Model price	Erro
1	0.894	0.038	0.3216	0.3220	0.0004
2	0.927	0.032	0.2266	0.2267	0.0001
3	0.960	0.032	0.1462	0.1450	-0.0012
4	0.993	0.036	0.0845	0.0839	-0.0006
5	1.026	0.042	0.0451	0.0440	-0.0011
6	1.059	0.060	0.0233	0.0225	-0.0008
7	1.092	0.076	0.0120	0.0117	-0.0003
8	1.125	0.087	0.0068	0.0074	0.0006

Table 1 SSE 50 ETF 42 days European call options

All optimization problems are implemented in MATLAB®2014b via the modeling language CVX [36] and solved with the second-order cone programming solver SDPT3. All experiments are run on a PC (Inter®Core$^{\text{TM}}$i5-4590, 3.30GHz, 4.00GB). The three experiment results are displayed in Tables 1–3, respectively. The experiment results show that our approach produces prices that are close to those observed in the options market.

No.	K/S	ρ	Mkt price	Model price	Erro
1	0.904	0.0010	0.3088	0.3072	-0.0016
2	0.937	0.0005	0.2131	0.2133	0.0002
3	0.971	0.0004	0.1281	0.1272	-0.0009
4	1.004	0.0005	0.0649	0.0636	-0.0013
5	1.038	0.0011	0.0283	0.0278	-0.0005
6	1.071	0.0015	0.0115	0.0122	0.0007
7	1.105	0.0023	0.0053	0.0061	0.0008
8	1.138	0.0033	0.0029	0.0025	-0.0004

Table 2 SSE 50 ETF 35 days European call options

No.	K/S	ρ	Mkt price	Model price	Erro
1	0.906	0.0062	0.2930	0.2925	-0.0005
2	0.939	0.0034	0.1960	0.1970	0.0010
3	0.973	0.0025	0.1081	0.1071	-0.0010
4	1.006	0.0045	0.0464	0.0454	-0.0010
5	1.040	0.0072	0.0169	0.0164	-0.0005
6	1.073	0.0088	0.0063	0.0064	0.0001
7	1.107	0.0150	0.0031	0.0039	0.0008
8	1.141	0.0251	0.0019	0.0020	0.0001

Table 3 SSE 50 ETF 29 days European call options

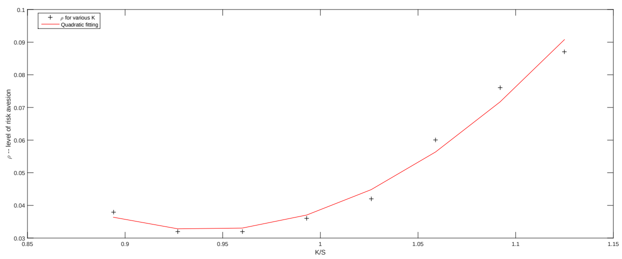

Figure 1 ρ vs K/S -SSE 50 ETF 42 days European call options

In addition, for each of the experiments, we fit a quadratic function to these values of "K/S" and "ρ" in Tables 1–3, and observe that the quadratic dependence of the ρ's on "K/S", that is, capture the phenomenon of "implied volatility smile" that characterizes the classical Black-Scholes-Merton model (see Figures 1–3).

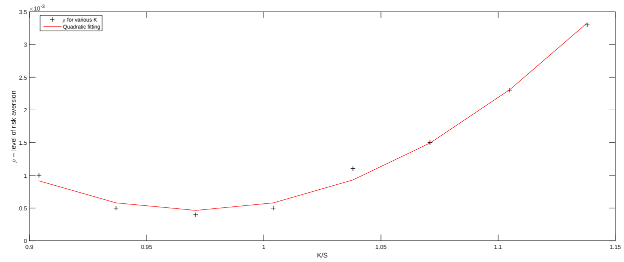

Figure 2 ρ vs K/S -SSE 50 ETF 35 days European call options

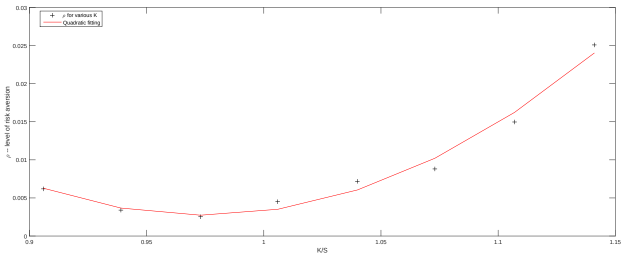

Figure 3 ρ vs K/S -SSE 50 ETF 29 days European call options

5. Conclusion

In this paper, the distributionally robust optimization is applied to the option pricing problem, which seeks to minimize the worst case replication error for the Wasserstein DUS. We obtain the corresponding computational tractable reformulation. The experiment results show that our approach produces prices that are close to those observed in the options market and capture the phenomenon of "implied volatility smile" that characterizes the classical Black-Scholes-Merton model. Applying our approach to price Asian, Lookback, American and Index options will be our future work.

Acknowledgements We thank the referees for their time and comments, and the Teacher Research Capacity Promotion Program of Beijing Normal University Zhuhai.

References

[1] F. BLACK, M. SCHOLES. *The pricing of options and corporate liabilities*. J. Polit. Econ., 1973, **81**(3): 637–654.

[2] R. C. MERTON. *An intertemporal capital asset pricing model*. Econometrica, 1973, **41**(5): 867–887.

[3] R. C. MERTON. *Option pricing when underlying stock returns are discontinuous*. J. Financial Econ., 1976, **3**: 125–144.

[4] J. C. COX, S. A. ROSS. *The valuation of options for alternative stochastic processes*. J. Financial Econ., 1976, **3**: 145–166.

[5] J. C. HULL, A. WHITE. *The pricing of options on assets with stochastic volatilities*. J. Finance, 1987, **42**(2): 281–300.

[6] D. B. MADAN, P. CARR, E. C. CHANG. *The variance gamma process and option pricing*. Review of Finance, 1998, **2**(1): 79–105.

[7] C. BANDI, D. BERTSIMAS. *Robust option pricing*. European J. Oper. Res., 2014, **239**(3): 842–853.

[8] A. SHAPIRO, D. DENTCHEVA, A. RUSZCZYŃSKI. *Lectures on Stochastic Programming: Modeling and Theory*. MPS-SIAM series on optimization (SIAM, Philadelphia, USA), 2nd edition, 2014.

[9] A. BEN-TAL, A. NEMIROVSKI. *Robust convex optimization*. Math. Oper. Res., 1998, **23**(4): 769–805.

[10] D. BERTSIMAS, D. B. BROWN, C. CARAMANIS. *Theory and applications of robust optimization*. SIAM Rev., 2011, **53**(3): 464–501.

[11] L. E. GHAOUI, M. OKS, F. OUSTRY. *Worst-case value-at-risk and robust portfolio optimization: a conic programming approach*. Oper. Res., 2003, **51**(4): 543–556.

[12] I. POPESCU. *Robust mean-covariance solutions for stochastic optimization*. Oper. Res., 2007, **55**(1): 98–112.

[13] E. DELAGE, Yinyu YE. *Distributionally robust optimization under moment uncertainty with application to data-driven problems*. Oper. Res., 2010, **58**(3): 595–612.

[14] J. GOH, M. SIM. *Distributionally robust optimization and its tractable approximations*. Oper. Res., 2010, **58**(4): 902–917.

[15] S. ZYMLER, D. KUHN, B. RUSTEM. *Distributionally robust joint chance constraints with second-order moment information*. Math. Program., Ser. A, 2013, **137**(1-2): 167–198.

[16] W. WIESEMANN, D. KUHN, M. SIM. *Distributionally robust convex optimization*. Oper. Res., 2014, **62**(6): 1358–1376.

[17] Zhilin KANG, Xun LI, Zhongfei LI, et al. *Data-driven robust mean-CVaR portfolio selection under distribution ambiguity*. Quant. Finance, 2019, **19**(1): 105–121.

[18] Aifan LING, Jie SUN, Naihua XIU, et al. *Robust two-stage stochastic linear optimization with risk aversion*. European J. Oper. Res., 2017, **256**(1): 215–229.

[19] S. SHAFIEEZADEH-ABADEH, P. M. ESFAHANI, D. KUHN. *Distributionally robust logistic regression*. Neural Information Processing Systems, 2015, 1576–1584.

[20] P. MOHAJERIN ESFAHANI, D. KUHN. *Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations*. Math. Program., Ser. A, 2018, **171**(1-2): 115–166.

[21] Chaoyue ZHAO, Yongpei GUAN. *Data-driven risk-averse stochastic optimization with Wasserstein metric*. Oper. Res. Lett., 2018, **46**(2): 262–267.

[22] R. GAO, A. J. KLEYWEGT. *Distributionally robust stochastic optimization with Wasserstein distance*. 2016, arXiv:1604.02199v2 [math.OC].

[23] D. WOZABAL. *Robustifying convex risk measures for linear portfolios: a nonparametric approach*. Oper. Res., 2014, **62**(6): 1302–1315.

[24] A. SHAPIRO. *Distributionally robust stochastic programming*. SIAM J. Optim., 2017, **27**(4): 2258–2275.

[25] Zhaolin HU, L. J. HONG. *Kullback-Leibler divergence constrained distributionally robust optimization*. Optimization Online http://www.optimization-online.org/DBHTML/2012/11/3677.html.

[26] G. BAYRAKSAN, D. K. LOVE *Data-driven stochastic programming using phi-divergences*. TutORials in Operations Research, 2015, 1–19.

[27] Ruiwei JIANG, Yongpei GUAN. *Data-driven chance constrained stochastic program*. Math. Program., Ser. A, 2016, **158**(1-2): 291–327.

[28] G. C. CALAFIORE. *Ambiguous risk measures and optimal robust portfolios*. SIAM J. Optim., 2007, **18**(3): 853–877.

[29] A. BEN-TAL, D. DEN HERTOG, A. DE WAEGENAERE, et al. *Robust solutions of optimization problems affected by uncertain probabilities*. Management Science, 2013, **59**(2): 341–357.

[30] K. POSTEK, D. DEN HERTOG, B. MELENBERG. *Computationally tractable counterparts of distributionally robust constraints on risk measures*. SIAM Rev., 2016, **58**(4): 603–650.

[31] E. ERDOGAN, G. IYENGAR. *Ambiguous chance constrained problems and robust optimization*. Math. Program., 2006, **107**(1): 37–61.

[32] D. BERTSIMAS, V. GUPTA, N. KALLUS. *Robust sample average approximation*. Math. Program., 2018, **171**(1): 217–282.

[33] Zizhuo WANG, P. W. GLYNN, Yinyu YE. *Likelihood robust optimization for data-driven problems*. Comput. Manag. Sci., **13**(2): 241–261.

[34] D. BERTSIMAS, L. KOGAN, A. W. LO. *Hedging derivative securities and incomplete markets: an ϵ-arbitrage approach*. Oper. Res., 2001, **49**(3): 372–397.

[35] M. S. LOBO, L. VANDENBERGHE, S. BOYD, et al. *Applications of second-order cone programming*. Linear Algebra Appl., 1998, **284**(1): 193–228.

[36] M. GRANT, S. BOYD. *CVX: Matlab software for disciplined convex programming*. 2010, version 2.0 beta. http://cvxr.com/cvx.