Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 18, No. 4, pp. 679–693, 2022

DISTRIBUTIONALLY ROBUST PORTFOLIO SELECTION WITH TRANSACTION COSTS*

Jingnan Chen, Lei Sun and Ning Zhang†

Abstract: We propose a portfolio selection model that is robust to the distribution of the asset return. Mathematically, the objective is to maximize the worst-case expected utility over a set of distributions with an ℓ_1 regularization incorporated to account for transaction costs. Under a piecewise-linear concave utility function, the distributionally robust optimization model is equivalent to a semidefinite program that can be solved effectively using the Alternating Direction Method of Multipliers. To evaluate the performance of the robust portfolio, we compute portfolio metrics including the average return, the volatility, and the worst-case optimized certainty equivalent (OCE) risk. Out-of-sample empirical results show that the robust portfolio has superior performance with regard to the above portfolio metrics.

Key words: *portfolio selection, distributionally robust optimization, transaction costs, ℓ_1 regularization, worst-case optimized certainty equivalence*

Mathematics Subject Classification: *90C22, 91G10, 90C25*

1 Introduction

Markowitz's seminal work in 1952 [22] raises investors' awareness of asset return and risk being two indispensable components of investment decision making. Since the introduction of the Markowitz's mean-variance portfolio optimization model that selects a portfolio in a frictionless market with the minimal variance risk, industrial practitioners and academic researchers have noticed that the Markowitz portfolios do not usually perform well in the real market, primarily due to their lacking robustness to model parameters and assuming zero market friction.

The implementation of Markowitz's portfolio selection model requires estimations of the asset return vector and covariance matrix. Both theoretical analyses and computational results have shown that slight changes in parameter estimations can lead to significantly different portfolio compositions [4, 7]. According to Fabozzi et al. [10], treating point estimates of the expected returns and the covariance matrix as error-free in portfolio selection might not necessarily correspond to prudent investor behavior. Investors are thus more comfortable with portolios that would perform well under different scenarios, thereby attaining

*The research was partially supported by the National Natural Science Foundation of China under grants 11801023, 72171012, 11971106, the Beihang University Research Fund Program under project ZG216S1983, Guangdong Basic and Applied Basic Research Foundation 2022A1515010088, and the Start-up Research Fund of DGUT 211135046.

†Corresponding author.

some protection from inevitable estimation inaccuracies. Robust optimization approaches have then been proposed to reduce the sensitivity of the optimal portfolio weights to the input parameter uncertainties and have thus enhanced the real-world applicability of the classical mean-variance model [17, 28, 25]. Additionally, the traditional mean-variance portfolio selection framework does not account for the market friction while transaction costs are a key instance of such friction that can affect portfolios' net performances in practice.

In this article, we propose the following portfolio selection model featuring distributional robustness and incorporating transaction costs:

$$\begin{array}{ll} \max_y \inf\limits_{\mathbb{P}\in\mathbb{F}} & \mathbb{E}_{\mathbb{P}}\left[U(y,\xi)\right] - \rho\|y-\tilde{y}\|_1 \\ \text{s.t.} & e^\top y = 1, \end{array} \tag{1.1}$$

where $y \in \Re^m$ is the vector of asset holdings in the portfolio with y_i denoting the portion of the total capital invested in asset i, $\tilde{y} \in \Re^m$ denotes the starting portfolio, $\xi \in \Re^m$ is the asset return vector which is an m-dimensional random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $U(\cdot,\cdot)$ is the utility function, $\mathbb{F}$ is a set of distributions with certain properties, and ρ is a tunable penalty parameter. We assume that a whole unit of capital has to be invested and hence $e^\top y = 1$, where e is an all-one vector. Linear constraints $Dy \geq d$ can be further added to represent practical investment restrictions. For instance, we may control the maximal or minimal allowable positions of some assets, i.e., $y^l \leq y \leq y^u$. Note that $-\sum_{i:y_i<0} y_i = (\|y\|_1 - 1)/2$ since $e^\top y = 1$. Therefore, if we set $\tilde{y} = 0$, adding the ℓ_1 regularization can also control the short positions.

Our objective is to maximize the expected utility over a set of distributions with the same first and second order moments. We do not require the complete knowledge of the asset return distribution, which is normally hard to obtain in practice, while the information of the mean return and the covariance matrix is more accessible [24]. It is noted in [5] that "for large investors whose principal cost is a fixed bid-ask spread, transaction costs are effectively proportional to the gross market value of the selected portfolio, and thus to the ℓ_1 penalty term." Although the ℓ_1 term is not the real transaction cost, yet controlling the value of the ℓ_1 term helps manage the trading volume and thus the trading cost. For ease of expression, we assume that the same spread applies to the universe of assets being considered, similar as in [5], yet our model and solution method can easily be extended to the general case of asset-specific bid-ask spreads. The ℓ_1 penalty also acts as a measure of portfolio leverage [20] and controls the total amount of short positions in the portfolio [8]. According to [5], adding an ℓ_1 penalty in the context of Markowitz portfolio optimization stabilizes the optimization model and promotes solution sparsity.

The distributionally robust framework addresses the important concern faced by investors regarding the ambiguity in the knowledge of the asset return distribution. Considering the piecewise-linear utility function in [23] and [29] that approximates the exponential utility function given a risk-aversion parameter, we show that model (1.1) can be converted to an equivalent semidefinite programming problem (SDP). We adopt the 2-block Alternating Direction Method of Multipliers (ADMM) to find the optimal portfolio and prove the convergence of the algorithm. We further compute the worst-case optimized uncertainty equivalent (OCE), the negative of which is shown to be a convex risk measure, within the framework of [2]. Performance of model (1.1) is evaluated through out-of-sample empirical tests, which indicate enhanced portfolio stability, especially during the 07-09 crisis period of high market volatility, and improved portfolio risk-return performance led by incorporating the robustness and adding the ℓ_1 penalty.

The remaining of the paper proceeds as follows. In Section 2, we reformulate model (1.1) into a 2-block convex conic program. Moreover, we introduce the notion of the worst-case

OCE risk and the computational framework. In Section 3, we present details of solving the reformulated model by ADMM. Section 4 reports empirical results and Section 5 concludes the paper.

The following notations will be used throughout this article. Let $\mathcal{S}^m$ be the linear space of all $m \times m$ real symmetric matrices and $\mathcal{S}^m_+$ be the cone of all positive semidefinite matrices. I and $\mathcal{I}$ denote the identity matrix of appropriate dimensions and the identity operator, respectively. For any vector $x \in \Re^n$ and matrix $X \in \Re^{m\times m}$, denote $\|x\|$ as the Euclidean norm and $\|X\|$ as the Frobenius norm.

2 Model Formulation and Risk Measure

In this section, we first reformulate the distributionally robust portfolio selection model (1.1) to a tractable semidefinite program. As the exact distribution of the random asset return is unknown, we introduce the worst-case OCE risk, i.e., a convex risk measure under ambiguous distributions proposed in [23], to gauge the portfolio risk exposure.

To begin with, for notational convenience we set

$$h(y) := \inf_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}\left[U(y,\xi)\right] - \rho\|y-\tilde{y}\|_1$$

and

$$\mathcal{Y} := \{y \in \Re^m \,|\, e^\top y = 1\}.$$

As mentioned in the introduction, the utility function takes the following form:

$$U(y,\xi) = \min_{k\in\{1,\ldots,K\}} \left\{\alpha_k \xi^\top y + \beta_k\right\}, \tag{2.1}$$

which approximates the exponential utility function $(1 - exp(-\alpha x))/\alpha$ given a risk-aversion parameter α. We aim to find the worst-case expected utility $\inf_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}\left[U(y,\xi)\right]$ over the distribution set

$$\mathbb{F} := \left\{\mathbb{P} \,|\, \mathbb{P}(\xi \in \Re^m) = 1,\, \mathbb{E}_{\mathbb{P}}[\xi] = \mu,\, \mathbb{E}_{\mathbb{P}}[\xi\xi^\top] = \Sigma\right\},$$

which is a family of distributions of the random return vector ξ with known mean vector μ and covariance matrix Σ. In practice, investors usually do not have the complete information of the multivariate distribution $\mathbb{P}$ to compute the exact expected utility while the first and second order moment information is more accessible (see [24]). Let $g(y) := \inf_{\mathbb{P}\in\mathbb{F}} \mathbb{E}_{\mathbb{P}}\left[U(y,\xi)\right]$ be the worst-case expected utility of portfolio y. Then a portfolio that maximizes such worst-case utility is obtained by solving

$$\bar{y} = \arg\max_{y\in\mathcal{Y}}\ g(y).$$

Transaction costs are important in determining the net value of a portfolio and hence need to be incorporated when making investment decisions. For institutional investors, broker commissions can often be omitted due to their high trading volumes and an ℓ_1 penalty term suffices to capture the transaction costs [5]. In our portfolio selection model (1.1), we add the ℓ_1 penalty of the trading volume to the worst-case expected utility $g(y)$ to control the transaction costs.

Assumption 2.1. The mean vector μ and covariance matrix Σ of the random asset return ξ are finite and satisfy $\Sigma - \mu\mu^\top \succ 0$.

With Assumption 2.1 and Theorem 2.1 in [3], for any given $y \in \mathcal{Y}$, $h(y)$ is the optimal value of the following semidefinite program:

$$\begin{aligned} h(y) = \sup_{R,r,r_0} \quad & \langle R,\, \Sigma \rangle + \mu^\top r + r_0 - \rho \|y - \widetilde{y}\|_1 \\ \text{s.t.} \quad & \begin{pmatrix} -R & \frac{-r+\alpha_k y}{2} \\ \frac{(-r+\alpha_k y)^\top}{2} & \beta_k - r_0 \end{pmatrix} \succeq 0,\ k = 1, \ldots, K. \end{aligned} \tag{2.2}$$

Consequently, model (1.1) can be equivalently written as

$$\begin{aligned} \max_{y} \sup_{R,r,r_0} \quad & \langle R,\, \Sigma \rangle + \mu^\top r + r_0 - \rho \|y - \widetilde{y}\|_1 \\ \text{s.t.} \quad & \begin{pmatrix} -R & \frac{-r+\alpha_k y}{2} \\ \frac{(-r+\alpha_k y)^\top}{2} & \beta_k - r_0 \end{pmatrix} \succeq 0,\ k = 1, \ldots, K, \\ & e^\top y = 1. \end{aligned} \tag{2.3}$$

2.1 OCE risk

In this section, we introduce the worst-case OCE risk proposed by Natarajan et al. [23], which is a convex risk measure within the framework of OCE developed by Ben-Tal and Teboulle [1, 2]. The general expression of OCE is given as follows:

$$\bar{S}_U(x) := \sup_{v \in \mathcal{R}} \{ v + \mathbb{E}_{\mathbb{P}_x}[U(x - v)] \}, \tag{2.4}$$

where $U(\cdot)$ is a normalized concave utility function and x is a random variable with probability distribution $\mathbb{P}_x$. Suppose that an investor expects to earn a future uncertain income of x and he has a choice to consume part of it now. If he consumes v, then $x - v$ is left for later consumption and the resulting present value becomes $v + \mathbb{E}_x[U(x - v)]$. Thus, the OCE (2.4) measures the sure present value of a future uncertain income, accounting for the optimal allocation between the present and the future consumptions.

In this paper, we only consider the case where the utility function takes a piecewise-linear form. For a fixed portfolio y, define the portfolio payoff $x := \xi^\top y$ with the first order moment $\mathbb{E}_{\mathbb{P}_x}[x] = \mu^\top y$ and the second order moment $\mathbb{E}_{\mathbb{P}_x}[xx^\top] = y^\top \Sigma y$, denoted by μ_x and σ_x^2, respectively. According to [23], the worst-case OCE is defined as

$$S_U(x) = S_U(\xi^\top y) := \sup_{v \in \mathcal{R}} \left\{ v + \inf_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}} \left[\min_{k \in \{1,2,\ldots,K\}} \alpha_k(\xi^\top y - v) + \beta_k \right] \right\}, \tag{2.5}$$

and $\rho_U(x) := -S_U(x)$ is the corresponding risk measure.

We follow the computational framework in [23] to obtain the OCE risk of a portfolio. Specifically, we first compute the worst-case expected utility

$$\inf_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}} \left[\min_{k \in \{1,2,\ldots,K\}} \alpha_k \xi^\top y + \beta_k \right],$$

which can be reformulated as

$$\inf_{\mathbb{P}_x \in \mathbb{F}_x} \mathbb{E}_{\mathbb{P}_x} \left[\min_{k \in \{1,2,\ldots,K\}} \alpha_k x + \beta_k \right], \tag{2.6}$$

due to Proposition 1 in [24]. Specifically, we can further rewrite problem (2.6) as the following convex optimization problem:

$$\begin{aligned} \min_{\mathbb{P}} \quad & \int \Big(\min_{k\in\{1,2,\ldots,K\}} \alpha_k x + \beta_k \Big) d\mathbb{P}_x(x) \\ \text{s.t.} \quad & \int d\mathbb{P}_x(x) = 1, \\ & \int x d\mathbb{P}_x(x) = \mu_x, \\ & \int x^2 d\mathbb{P}_x(x) = \sigma_x. \end{aligned}$$

Therefore, from the duality theory in infinite-dimensional convex optimization problem (see, e.g., [26, 19]), we can obtain the optimal solution of model (2.6) via solving the dual optimization problem:

$$\begin{aligned} \sup_{m_0,m_1,M} \quad & m_0 + \mu_x m_1 + M(\mu_x^2 + \sigma_x^2) \\ \text{s.t.} \quad & m_0 + m_1 x + M x^2 \leq \min_{k\in\{1,2,\ldots,K\}}\{\alpha_k x + \beta_k\}, \quad \forall x \in \mathcal{R}, \end{aligned} \tag{2.7}$$

where m_0, m_1, and M are dual variables. With the worst-case expected utility obtained by solving the second order cone program (2.7), finding the worst-case OCE in (2.5) becomes a maximization problem with a single variable v. For detailed computational procedures of the worst-case OCE, refer to [23].

3 Solution Method: ADMM

In this section, we aim to solve model (2.3) to obtain the optimal robust and sparse portfolio where the sparsity is owing to the ℓ_1 regularization. We first recall some established results that will be used in the subsequent analyses.

3.1 Preliminaries

Let $\mathcal{Z}$ be a convex set and we define the indicator function

$$\mathbb{I}_{\mathcal{Z}}(z) := \begin{cases} 0, & \text{if } z \in \mathcal{Z}, \\ +\infty, & \text{otherwise.} \end{cases} \tag{3.1}$$

For a given closed proper convex function $f : \mathcal{X} \to \Re \cup \{+\infty\}$, the Moreau-Yosida regularization of f at $x \in \mathcal{X}$ is defined by

$$\Psi_f(x) = \min_{x'\in\mathcal{X}} \left\{ f(x') + \frac{1}{2}\|x' - x\|^2 \right\}. \tag{3.2}$$

The minimization problem in (3.2) has a unique optimal solution $\text{Prox}_f(x)$ called the proximal point of x associated with f and Prox_f is called the proximal mapping.

If $f_1(x) = \mathbb{I}_{\mathcal{Z}}(x)$, then the corresponding proximal point of x is

$$\text{Prox}_{f_1}(x) = \Pi_{\mathcal{Z}}(x),$$

where $\Pi_{\mathcal{Z}}(\cdot)$ is the metric projection operator over $\mathcal{Z}$. If $f_2(x) = \rho\|x\|_1 = \rho\sum_{i=1}^m |x_i|$, then the corresponding proximal point of x is

$$\text{Prox}_{f_2}(x) = \text{sgn}(x) \cdot \max(0, |x| - \rho), \tag{3.3}$$

where sgn$(\cdot)$ is the sign function.

Now we briefly review the 2-block ADMM. Consider the following 2-block convex composite optimization problem:

$$\begin{array}{cl} \min\limits_{x\in\mathcal{X},y\in\mathcal{Y}} & p(x)+f(x)+q(y)+g(y) \\ \text{s.t.} & \mathcal{A}x+\mathcal{B}y=c, \end{array} \tag{3.4}$$

where $\mathcal{A}:\mathcal{X}\to\mathcal{Z}$ and $\mathcal{B}:\mathcal{Y}\to\mathcal{Z}$ are two linear operators between finite dimensional Euclidean spaces $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$. $f:\mathcal{X}\to\Re$ and $g:\mathcal{Y}\to\Re$ are two continuously differentiable convex functions; $p:\mathcal{X}\to\Re\cup\{+\infty\}$ and $q:\mathcal{Y}\to\Re\cup\{+\infty\}$ are two closed proper convex (not necessarily smooth) functions.

Let $\sigma>0$ be a given parameter. The augmented Lagrangian function of (3.4) is defined by

$$\mathcal{L}_\sigma(x,y;z):=p(x)+f(x)+q(y)+g(y)+\langle z,\mathcal{A}x+\mathcal{B}y-c\rangle+\frac{\sigma}{2}\|\mathcal{A}x+\mathcal{B}y-c\|^2.$$

The 2-block ADMM iterative scheme can be described as follows:

$$\left\{\begin{array}{l} x^{k+1}\in\ \operatorname{argmin}\left\{\mathcal{L}_\sigma(x,y^k;z^k)+\frac{1}{2}\|x-x^k\|^2_{\mathcal{S}}\,|\,x\in\mathcal{X}\right\}, \\ y^{k+1}\in\ \operatorname{argmin}\left\{\mathcal{L}_\sigma(x^{k+1},y;z^k)+\frac{1}{2}\|y-y^k\|^2_{\mathcal{T}}\,|\,y\in\mathcal{Y}\right\}, \\ z^{k+1}=z^k+\tau\sigma(\mathcal{A}x^{k+1}+\mathcal{B}y^{k+1}-c). \end{array}\right. \tag{3.5}$$

If the proximal terms $\mathcal{S}=0$ and $\mathcal{T}=0$, the iterative scheme (3.5) is the classical ADMM developed in[13] and [14]. In [12] and [15], the authors establish the global convergence of the classical ADMM with any $\tau\in(0,(1+\sqrt{5})/2)$ when the adjoint operator $\mathcal{A}^*$ of $\mathcal{A}$ is surjective and $\mathcal{B}$ is an identity operator. A more general and easy-to-use convergence theorem is developed in [11]. To know more about the convergent analysis of 2-block ADMM, refer to [11, 15, 16, 18, 21, 30, 31], and references therein.

3.2 ADMM for solving model (2.3)

For notational convenience, denote

$$A:=\begin{pmatrix}\Sigma & \mu\\ \mu^\top & 1\end{pmatrix},\ X:=\begin{pmatrix}R & r/2\\ r^\top/2 & r_0\end{pmatrix},\ C_k:=\begin{pmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \beta_k\end{pmatrix},\ \mathcal{P}_k y:=\frac{\alpha_k}{2}\begin{pmatrix}\mathbf{0} & y\\ y^\top & 0\end{pmatrix},$$

for $k\in\{1,\ldots,K\}$. Introducing an auxiliary variable $Z:=(Z_1,\ldots,Z_k)\in\mathcal{S}^{m+1}\times\cdots\times\mathcal{S}^{m+1}$, we can reformulate model (2.3) into the following standard form of 2-block convex composite optimization program:

$$\begin{array}{cl} \min_{X,y,z,Z} & -\langle A,\,X\rangle+\rho\|z\|_1+\sum_{k=1}^K\mathbb{I}_{\mathcal{S}_+^{m+1}}(Z_k) \\ \text{s.t.} & e^\top y-1=0, \\ & X-\mathcal{P}_k y+Z_k=C_k,\ k=1,\ldots,K, \\ & y-\tilde{y}-z=0. \end{array} \tag{3.6}$$

The Lagrangian function associated with (3.6) is

$$\begin{aligned}\mathcal{L}(X,y,z,Z;x,\Gamma,\gamma)=&-\langle A,\,X\rangle+\rho\|z\|_1+\textstyle\sum_{k=1}^K\mathbb{I}_{\mathcal{S}_+^{m+1}}(Z_k)\\ &+\textstyle\sum_{k=1}^K\langle\Gamma_k,C_k-X+\mathcal{P}_k y-Z_k\rangle+\langle\gamma,1-e^\top y\rangle+\langle x,y-\tilde{y}-z\rangle,\end{aligned}$$

and the corresponding Karush-Kuhn-Tucker (KKT) system is given by

$$\begin{cases} e^\top y - 1 = 0,\ y - \tilde{y} - z = 0,\ X - \mathcal{P}_k y + Z_k = C_k,\ k = 1,\dots,K, \\ A + \sum_{k=1}^K \Gamma_k = 0,\ \sum_{k=1}^K \mathcal{P}^*\Gamma_k - \gamma e + x = 0, \\ z - \mathrm{Prox}_{\rho\|\cdot\|_1}(z - x) = 0,\ Z_k - \Pi_{\mathcal{S}_+^{m+1}}(\Gamma_k + Z_k) = 0,\ k = 1,\dots,K. \end{cases} \tag{3.7}$$

Let $\sigma > 0$ be a given parameter and the augmented Lagrangian function is

$$\begin{aligned} \mathcal{L}_\sigma(X, y, z, Z; x, \Gamma, \gamma) = \ & \mathcal{L}(X, y, z, Z; x, \Gamma, \gamma) \\ & + \tfrac{\sigma}{2}(\|y - \tilde{y} - z\|^2 + \textstyle\sum_{k=1}^K \|C_k - X + \mathcal{P}_k y - Z_k\|^2 + \|1 - e^\top y\|^2). \end{aligned} \tag{3.8}$$

The iteration scheme of ADMM for (3.6) is described in Algorithm 1. According to Schur complement, H in the linear system is positive definite. Next we show the global convergence of Algorithm 1. Without loss of generality, assume that the solution set of KKT system (3.7) is nonempty. The global convergence results of Algorithm 1 can be directly obtained from Theorem B.1 in [11]. We present the proof here for the sake of completeness.

Theorem 3.1. *Let $\{(X^t, y^t, z^t, Z^t, \gamma^t, \Gamma^t, x^t)\}$ be the sequence generated by the ADMM in Algorithm 1. The sequence $\{(X^t, y^t, z^t, Z^t)\}$ converges to an optimal solution of problem (3.6) and the sequence $\{(\gamma^t, \Gamma^t, x^t)\}$ converges to its dual solution.*

Proof. We first reformulate (3.6) into the framework of (3.4). Specifically, set

$$p(X, y) := 0,\ f(X, y) := -\langle A, X\rangle,\ q(Z, z) := \rho\|z\|_1 + \sum_{k=1}^K \mathbb{I}_{\mathcal{S}_+^{m+1}}(Z_k),\ g(Z, z) := 0,$$

Define linear operators $\mathcal{A}_1 : \mathcal{S}^{m+1} \times \Re^m \to \Re$, $\mathcal{A}_2 : \mathcal{S}^{m+1} \times \Re^m \to \mathcal{S}_K^{m+1}$, and $\mathcal{A}_3 : \mathcal{S}^{m+1} \times \Re^m \to \Re$ as follows:

$$\mathcal{A}_1 := (0,\, e^\top),\ \mathcal{A}_2 := \begin{pmatrix} \mathcal{I} & -\mathcal{P}_1 \\ \vdots & \vdots \\ \mathcal{I} & -\mathcal{P}_K \end{pmatrix},\ \text{and } \mathcal{A}_3 := (0,\, \mathcal{I}).$$

Then, we can set

$$\mathcal{A} := \begin{pmatrix} \mathcal{A}_1 \\ \mathcal{A}_2 \\ \mathcal{A}_3 \end{pmatrix},\ \mathcal{B} := \begin{pmatrix} 0 & 0 & \dots & 0 \\ \mathcal{I} & 0 & \dots & 0 \\ & \ddots & \ddots & \vdots \\ & & \mathcal{I} & 0 \\ & & & -\mathcal{I} \end{pmatrix},\ \text{and } c := \begin{pmatrix} 1 \\ C_1 \\ \vdots \\ C_k \\ \tilde{y} \end{pmatrix}.$$

It is easy to check that $\mathcal{B}^*\mathcal{B} \succ 0$.

$$\mathcal{A}^*\mathcal{A} = \mathcal{A}_1^*\mathcal{A}_1 + \mathcal{A}_2^*\mathcal{A}_2 + \mathcal{A}_3^*\mathcal{A}_3 = \mathcal{A}_2^*\mathcal{A}_2 + \begin{pmatrix} 0 & 0 \\ 0 & ee^\top + I \end{pmatrix} \succeq 0.$$

For any given $0 \neq \Delta := (D; d) \in \mathcal{S}^{m+1} \times \Re^m$, we have

$$\langle \mathcal{A}^*\mathcal{A}\Delta,\, \Delta\rangle = \langle \mathcal{A}_2^*\mathcal{A}_2\Delta,\, \Delta\rangle + d^\top(ee^\top + I)d.$$

It can be easily seen that $\langle \mathcal{A}^*\mathcal{A}\Delta,\, \Delta\rangle > 0$ when $d \neq 0$. When $d = 0$ and $D \neq 0$, $\langle \mathcal{A}^*\mathcal{A}\Delta,\, \Delta\rangle = K\langle D, D\rangle > 0$. Thus, we have $\mathcal{A}^*\mathcal{A} \succ 0$ and hence the conditions in Theorem B.1 of [11] are satisfied. Therefore, the conclusion holds. □

Algorithm 1 : ADMM for solving (3.6)

Let $\sigma > 0$ and $\tau \in (0, (1+\sqrt{5})/2)$ be given parameters. Choose an initial point $(X^0, y^0, z^0, Z^0, \gamma^0, \Gamma^0, x^0)$ and perform the following steps for $t = 0, 1, \ldots$ until that the error tolerance level is met.

Step 1. Compute $(X^{t+1}, y^{t+1}) = \arg\min_{X,y} \mathcal{L}_\sigma(X, y, z^t, Z^t; x^t, \Gamma^t, \gamma^t)$. Specifically,

$$
\begin{aligned}
R^{t+1} &= \tfrac{1}{K\sigma}\left(\textstyle\sum_{k=1}^K (\Gamma_k^t)_{11} + \Sigma\right) - \textstyle\sum_{k=1}^K (Z_k^t)_{11}, \\
r_0^{t+1} &= \tfrac{1}{K\sigma}\left(\textstyle\sum_{k=1}^K (\Gamma_k^t)_{22} + 1\right) - \textstyle\sum_{k=1}^K \left((Z_k^t)_{22} + \beta_k\right),
\end{aligned}
$$

where $(\Gamma_k^t)_{11}, (Z_k^t)_{11} \in \mathcal{S}^m, (\Gamma_k^t)_{12}, (Z_k^t)_{12} \in \Re^m,\ (\Gamma_k^t)_{22}, (Z_k^t)_{22} \in \Re$, and

$$
\Gamma_k^t = \begin{pmatrix} (\Gamma_k^t)_{11} & (\Gamma_k^t)_{12} \\ (\Gamma_k^t)_{12}^T & (\Gamma_k^t)_{22} \end{pmatrix}, \quad Z_k^t = \begin{pmatrix} (Z_k^t)_{11} & (Z_k^t)_{12} \\ (Z_k^t)_{12}^T & (Z_k^t)_{22} \end{pmatrix}.
$$

Moreover, $\zeta := [r^{t+1}; y^{t+1}]$ is the solution to the following linear equation:

$$
H\zeta = rhs, \tag{3.9}
$$

where

$$
\begin{aligned}
H &:= \sigma \begin{pmatrix} \frac{1}{2}KI & -\frac{1}{2}\sum_{k=1}^K \alpha_k I \\ -\frac{1}{2}\sum_{k=1}^K \alpha_k I & \frac{1}{2}\sum_{k=1}^K \alpha_k^2 I + ee^\top + I \end{pmatrix}, \\
rhs &:= \begin{pmatrix} \sum_{k=1}^K (\Gamma_k^t - \sigma Z_k^t)_{12} + \mu \\ \sum_{k=1}^K \alpha_k(-\Gamma_k^t + \sigma Z_k^t)_{12} + e\gamma^t + x^t \end{pmatrix}.
\end{aligned}
$$

Step 2. Compute $(z^{t+1}, Z^{t+1}) := \arg\min_{z,Z} \mathcal{L}_\sigma(X^{t+1}, y^{t+1}, z, Z; x^t, \Gamma^t, \gamma^t)$. Specifically,

$$
\begin{aligned}
z^{t+1} &= \arg\min_z \rho\|z\|_1 + \tfrac{\sigma}{2}\|z - (y^{t+1} - \widetilde{y} - \tfrac{1}{\sigma}x^t)\|^2 \\
&= \operatorname{sgn}(y^{t+1} - \widetilde{y} - \tfrac{1}{\sigma}x^t)\max\left\{0, |y^{t+1} - \widetilde{y} - \tfrac{1}{\sigma}x^t| - \tfrac{\rho}{\sigma}e\right\},
\end{aligned}
$$

and for each $k \in \{1, \ldots, K\}$,

$$
Z_k^{t+1} = \Pi_{\mathcal{S}_+^{m+1}}\left(\mathcal{P}_k y^{t+1} + C_k - X^{t+1} + \frac{1}{\sigma}\Gamma_k^t\right).
$$

Step 3. Compute

$$
\begin{aligned}
\gamma^{t+1} &= \gamma^t + \tau\sigma(1 - e^\top y^{t+1}), \\
\Gamma_k^{t+1} &= \Gamma_k^t + \tau\sigma(C_k - X^{t+1} + \mathcal{P}_k y^{t+1} - Z_k^{t+1}),\ k = 1, \ldots, K, \\
x^{t+1} &= x^t + \tau\sigma(z^{t+1} - y^{t+1} + \widetilde{y}).
\end{aligned}
$$

Based on the relative residuals of the KKT system (3.7), we measure the quality of a solution (X, y, z, Z) to (3.6) obtained from Algorithm 1 by

$$
\eta := \max\{\eta_{P_1}, \eta_{P_2}, \eta_{P_3}, \eta_{D_1}, \eta_{D_2}, \eta_{C_1}, \eta_{C_2}\},
$$

where

$$
\begin{aligned}
&\eta_{P_1} = \|e^\top y - 1\|,\ \eta_{P_2} = \max_{k\in\{1,\dots,K\}} \left\{\frac{X-\mathcal{B}_k y+Z_k-C_k}{1+\|C_k\|}\right\},\ \eta_{P_3} = \frac{\|y-\widetilde{y}-z\|}{1+\|y\|+\|\widetilde{y}\|+\|z\|},\\
&\eta_{D_1} = \frac{\|A+\sum_{k=1}^K \Gamma_k\|}{1+\|A\|},\ \eta_{D_2} = \frac{\|\sum_{k=1}^K \mathcal{B}^*\Gamma_k - e\gamma - x\|}{1+\|m\gamma\|},\\
&\eta_{C_1} = \frac{\|z-\text{prox}_{\rho\|\cdot\|_1}(z-x)\|}{1+\|z\|},\ \eta_{C_2} = \max_{k\in\{1\dots,K\}} \left\{\frac{Z_k - \Pi_{\mathcal{S}_+^{m+1}}(\Gamma_k+Z_k)}{1+\|Z_k\|+\|\Gamma_k\|}\right\}.
\end{aligned}
$$

For a given error tolerance ε, Algorithm 1 terminates when $\eta < \varepsilon$. We set $\varepsilon = 10^{-5}$ for tests in section 4.

4 Empirical Results

We present two examples that evaluate the performance of robust and sparse portfolios through out-of-sample empirical tests using monthly data of 48 industry portfolios. We consider the five-piece linear utility function in Uichanco ([29]) (i.e., $U(y,\xi) = \min_{k\in\{1,2,\dots,5\}} \{\alpha_k \xi^\top y + \beta_k\}$) that approximates the exponential utility function $(1 - exp(-\alpha\xi^\top y))/\alpha$. For the risk-neutral investor, $\alpha = 0$. When α is large, portfolios with more risk exposures become more highly penalized. Chapter 2 of [10] indicates that the upper bound on the risk aversion is usually 4 for most portfolio allocation decisions. Hence, we consider $\alpha = 1$ in our examples. Values of coefficients α_k and β_k, $k = 1, 2, \dots, 5$, are given in Table 1.

k	1	2	3	4	5
α_k	86.4266	26.0312	7.8404	2.3615	0.7113
β_k	284.7210	55.2174	7.9214	0.2509	-0.0791

Table 1: Parameters of the five-piece utility function ($\alpha = 1$).

Following the computational procedures in Section 2.1, we compute the worst-case OCE risk as follows:

$$
S_U(x) = \begin{cases} -\mu_x - 0.8695 + 4.9868\sigma_x,\ \sigma_x \geq 0.3706 \\ -\mu_x + 0.0361 + 6.7303\sigma_x^2,\ otherwise, \end{cases} \tag{4.1}
$$

where x denotes $\xi^\top y$ as in Section 2.1.

Example 1. This example shows merits of incorporating the ℓ_1 regularization in portfolio selection using the same rolling-sample method as in [5] and [9]. Specifically, given a penalty parameter ρ, we construct the first portfolio at the beginning of January 1995 using a data sample of past 60 months from January 1990 to December 1994. The mean vector μ and covariance matrix Σ in model (2.3) are chosen as sample mean and sample covariance, respectively. The optimal portfolio is then held for three months till the end of March 1995 and the portfolio's out-of-sample monthly returns are recorded. We repeat the procedure and obtain a new portfolio at the beginning of April 1995 by solving (2.3) with μ and Σ updated using data from April 1990 to March 1995. When updating our portfolio every

Details of the dataset can be found on Kennith French's website `http://www.mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

quarter, we can either keep the same value of ρ or adjust the value of ρ to keep the number of selected assets (i.e., number of active positions) unchanged. We continue the exercise till the end of 2009 and with a series of out-of-sample monthly returns we calculate the average return, its standard deviation, and worst-case OCE risk.

To test effects of the ℓ_1 regularization on portfolio selection, we conduct the above exercise under $\rho = 0$ and record the number of active positions for each portfolio. It turns out that all of the portfolios contain a total of 48 active positions. Then we construct three portfolio strategies requiring 35, 24, and 11 active positions by adjusting the value of ρ each time the portfolio is updated. For each of these four strategies, we compute the aforementioned three quantities over 5 non-overlapping 3-year evaluation periods. It can be seen from Table 2 that except for the period 2001-2003, portfolio strategies developed with the regularization term added have significantly better performances in terms of the average return, the standard deviation, and the worst-case OCE risk than the one obtained in the absence of the ℓ_1 regularization. While in the period of 2001-2003, regularized portfolios have smaller volatility and worst-case OCE risk compared to the unregularized portfolio. For better visualization we also draw two bar charts of these strategies' out-of-sample average return (Figure 1) and worst-case OCE risk (Figure 2) respectively, which clearly illustrates the performance improvements led by incorporating the ℓ_1 penalty in portfolio selection.

Evaluation period	NSA	Mean	Stdev	Worst-Case OCE risk
01/1995-12/1997	48	0.3638	2.3284	10.3778
	35	1.3112	1.3833	4.7175
	24	1.3164	1.2855	4.2246
	11	1.5711	1.2700	3.9155
01/1998-12/2000	48	0.2256	4.7734	22.7089
	35	0.8007	1.6806	6.7106
	24	0.7041	1.7403	7.1049
	11	0.3183	2.4115	10.8379
01/2001-12/2003	48	-0.031	4.1276	19.7450
	35	-0.3663	2.8079	13.4992
	24	-0.1592	2.9346	13.9240
	11	-0.0531	3.2793	15.5368
01/2004-12/2006	48	0.1234	2.4321	11.1355
	35	0.8470	1.4329	5.4291
	24	0.8240	1.4639	5.6067
	11	0.8827	1.2652	4.5571
01/2007-12/2009	48	-0.9536	4.9047	24.5429
	35	-0.1030	2.3437	10.9211
	24	-0.3781	2.0502	9.7325
	11	-0.2110	3.1032	14.8165

Table 2: Out-of-sample performances of portfolios with different numbers of selected assets (abbreviated as NSA).

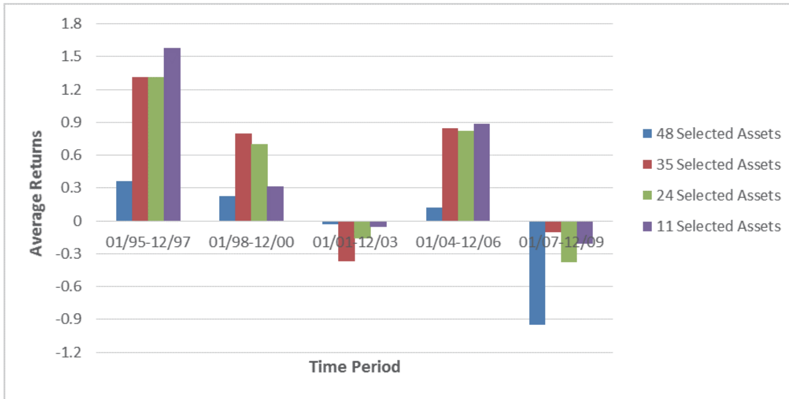

Figure 1: Out-of-sample average returns of portfolios with different numbers of selected assets.

Example 2. We compare out-of-sample performances of the robust and sparse portfolio and the equally weighted portfolio. We continue to apply the rolling-sample procedure described in Example 1 but with an evaluation period of three months instead of three years. This time we fix the value of ρ to be 1 and compute a strategy's return, standard deviation, and worst-case OCE risk over each evaluation period. Investors may also chosse the value for ρ using the cross-validation as in [6]. Column 1 of Table 3 provides a list of 12 evaluation periods starting from January 2007 and ending at December 2009. According to Table 3, the robust and sparse portfolio has better performances than the equally weighted portfolio. Specifically, in 9 out of 12 testing quarters the robust and sparse portfolios have higher average returns and lower OCE risk than the equally weighted portfolios. In addition, during the crisis period from July 2007 to March 2009 when there was very high market volatility, portfolios obtained under the robust and sparse optimization model outperform the equally weighted portfolio significantly by having higher portfolio return and 50% less worst-case OCE risk, indicating that the robust and sparse portfolios are better immunized against the market fluctuations. Table 3 also shows that when the market increases sharply, the equally weighted strategy has better performance compared to the robust strategy, which is more conservative in compensation for robustness.

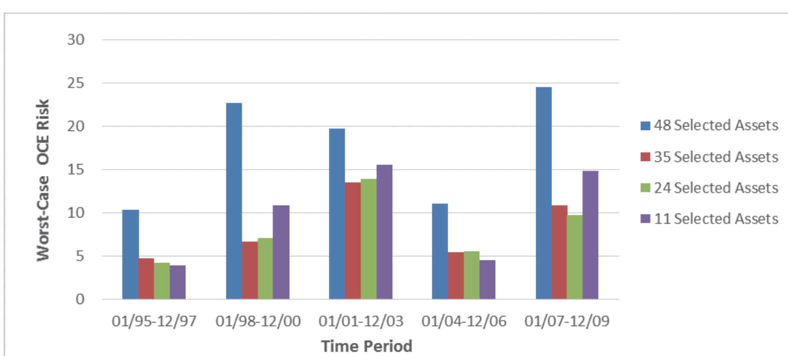

Figure 2: Out-of-sample OCE risk of portfolios with different numbers of selected assets.

Evaluation period	Methods	Mean	Stdev	Worst-Case OCE risk
01/2007-03/2007	eq_w	1.2051	1.6970	6.3880
	robust	-0.7783	1.7139	8.4557
04/2007-06/2007	eq_w	2.5430	3.2359	12.7243
	robust	2.8698	3.9249	15.8334
07/2007-09/2007	eq_w	0.4381	3.5137	16.2145
	robust	1.0657	0.4111	0.1149
10/2007-12/2007	eq_w	-0.6221	3.7980	18.6925
	robust	-0.5919	1.8181	8.7889
01/2008-03/2008	eq_w	-2.6726	2.2720	13.1331
	robust	-0.1944	1.4088	6.3503
04/2008-06/2008	eq_w	-0.3271	7.2287	35.5057
	robust	2.0854	3.4811	14.4046
07/2008-09/2008	eq_w	-3.2451	6.9158	36.8633
	robust	0.0591	4.0150	19.0934
10/2008-12/2008	eq_w	-8.7858	12.4995	70.2488
	robust	-0.6947	3.5479	17.5179
01/2009-03/2009	eq_w	-3.4637	10.8853	56.8770
	robust	-0.9231	3.0207	15.1172
04/2009-06/2009	eq_w	7.3654	9.1552	37.4203
	robust	-6.2543	16.0266	85.3062
07/2009-09/2009	eq_w	6.2669	3.1745	8.6942
	robust	2.0360	1.2681	3.4183
10/2009-12/2009	eq_w	2.1378	5.3834	23.8386
	robust	6.3474	5.4908	20.1646

Table 3: Out-of-sample performances of the equally weighted strategy and the robust and sparse strategy ($\rho = 1$).

5 Conclusion

We introduce a new model for portfolio selection that incorporates robustness to reduce the model sensitivity to parameter estimations. Transaction costs are considered and modeled by an ℓ_1 regularization term penalizing high trading volumes. Under the objective of maximizing a piecewise-linear concave utility function over a set of distributions sharing the same first and second order moment information, our distributionally robust model is equivalent to a semidefinite program that can be solved effectively using the 2-block alternating direction method of multipliers. We show the global convergence of the solution algorithm. In addition, we compute the worst-case OCE to measure the performance of the robust and sparse portfolio and demonstrate merits of incorporating the robustness and the ℓ_1 penalty into a utility-maximization model for portfolio selection, including enhancing portfolio stability and improving risk-return performance through out-of-sample empirical tests.

References

[1] A. Ben-Tal and M. Teboulle, Expected utility, penalty functions and duality in stochastic nonlinear programming, *Manage. Sci.* 32 (1986) 1445–1466.

[2] A. Ben-Tal and M. Teboulle, An old-new concept of convex risk measures: the optimized certainty equivalent, *Math. Finance* 17 (2007) 449–476.

[3] D. Bertsimas, X. Doan, K. Natarajan and C. Teo, Models for minimax stochastic linear optimization problems with risk aversion. *Math. Oper. Res.* 35 (2010) 580–602.

[4] M. Best and R. Grauer, On the sensitivity of mean-variance efficient portfolios to changes in asset means: some analytical and computational results, *Rev. Financ. Stud.* 4 (1991) 315–342.

[5] J. Brodie, I. Daubechies, C. De Mol, D. Giannone and I. Loris, Sparse and stable Markowitz portfolios, *Proceedings of the National Academy of Sciences*, 106 (2009) 12267–12272.

[6] A. Chinco, A. Clark-Joseph and Y. Mao, Sparse signals in the cross-section of returns, *J. Financ.* 74 (2017) 449-492.

[7] V. Chopra and W. Ziemba, The effects of errors in means, variances, and covariances on optimal portfolio choice, *J. Portf. Manage.* 19 (1993) 6–11.

[8] V. DeMiguel, L. Garlappi, F. Nogales and R. Uppal, A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms. *Manage. Sci.* 55 (2009) 798–812.

[9] V. DeMiguel , L. Garlappi and R. Uppal, Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy? *Rev.of Financ. Stud.* 22 (2009) 1915–1953.

[10] F. Fabozzi, P. Kolm, D. Pachamanova and S. Focardi, *Robust Portfolio Optimization and Management*, John Wiley & Sons, 2007.

[11] M. Fazel, T. Pong, D. Sun and P. Tseng, Hankel matrix rank minimization with applications to system identification and realization, *SIAM J. Matrix Anal. Appl.* 34 (2013) 946–977.

[12] M. Fortin and R. Glowinski, Chapter III on decomposition-coordination methods using an augmented Lagrangian. *Studies in Mathematics and Its Applications* 15 (1983) 97–146.

[13] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput. Math. Appl.* 2 (1976) 17–40 .

[14] R. Glowinski and A. Marroco, Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires,*Analyse Numérique* 9 (1975) 41–76.

[15] R. Glowinski, *Lectures on Numerical Methods for Non-linear Variational Problems*, Springer Science & Business Media, 2008.

[16] R. Glowinski, On alternating direction methods of multipliers: a historical perspective, in: *Modeling, Simulation and Optimization for Science and Technology*, Springer, 2014, pp.59–82.

[17] D. Goldfarb and G. Iyengar, Robust portfolio selection problems. *Math. Oper. Res.* 28 (2003) 1–38.

[18] B. He, H. Yang and S. Wang, Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities, *J. Optim. Theory Appl.* 106 (2000) 337–356.

[19] K. Isii, On the sharpness of Chebyshev-type inequalities, *Ann. Inst. Statist. Math.* 14 (1962) 185–197.

[20] B. Jacobs and K. Levy, Leverage aversion and portfolio optimality, *Financ. Anal. J.* 68 (2012) 89–94.

[21] Y. Liu , X. Yuan, S. Zeng and J. Zhang, Partial error bound conditions and the linear convergence rate of the alternating direction method of multipliers, *SIAM J. Numer. Anal.*, 56 (2018) 2095-2123.

[22] H. Markowitz, Portfolio selection. *J. Financ.* 7 (1952) 77–91.

[23] K. Natarajan, M. Sim and J. Uichanco, Tractable robust expected utility and risk models for portfolio optimization, *Math. Finance* 20 (2010) 695–731.

[24] I. Popescu, Robust mean-covariance solutions for stochastic optimization. *Oper. Res.* 55 (2007) 98–112.

[25] P. Qian, Z. Wang and Z. Wen, A composite risk measure framework for decision making under uncertainty, *J. Oper. Res. Soc. China* 7 (2019) 43–68.

[26] R. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[27] W. Rogosinsky, Moments of non-negative mass, *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 245 (1958) 1-27.

[28] R. Tütüncü and M. Koenig, Robust asset allocation, *Ann. Oper. Res.* 132 (2004) 157–187.

[29] J. Uichanco, *Ambiguous risk measures and piecewise linear utility models in portfolio management*, Doctoral Dissertation, National University of Singapore, 2007.

[30] J. Yang and Y. Zhang, Local Linear Convergence of an ADMM-Type Splitting Framework for Equality Constrained Optimization, *J. Oper. Res. Soc. China* 9 (2021) 307-319.

[31] X. Yuan, S. Zeng and J. Zhang, Discerning the linear convergence of admm for structured convex optimization through the lens of variational analysis, *J. Mach. Learn. Res.* 21 (2020) -1-75.

Manuscript received 20 September 2021
revised 6 January 2022
accepted for publication 21 April 2022

Jingnan Chen
School of Economics and Management
Beihang University, Beijing 100191, China
E-mail address: jchen@buaa.edu.cn

Lei Sun
School of Economics and Management
Beihang University Beijing 100191, China
E-mail address: l1249107884@163.com

Ning Zhang
School of Computer Science and Technology
Dongguan University of Technology, Dongguan 523808, China
E-mail address: zhangning@dgut.edu.cn