A Markowitz Portfolio Approach to Options Trading

Licheng Zhao and Daniel P. Palomar, *Fellow, IEEE*

***Abstract*—In this paper, we study the problem of option portfolio design under the Markowitz mean-variance framework. We extend the common practice of a pure-stock portfolio and include options in the design. The options returns are modeled statistically with first- and second-order moments, enriching the conventional delta-gamma approximation. The naive mean-variance formulation allows for a zero-risk design that, in a practical scenario with parameter estimation errors, is totally misleading and leads to bad results. This zero-risk fallacy can be circumvented with a more realistic robust formulation. Transaction cost is also considered in the formulation for a proper practical design. We propose an efficient BSUM-M-based algorithm to solve the optimization problem. The proposed algorithm can perform as well as the off-the-shelf solvers but with a much lower computational time—up to one order of magnitude lower. Numerical results based on real data are conducted and the performance is presented in terms of Sharpe ratio, cumulative profit and loss, drawdown, overall return over turnover, value at risk, expected shortfall, and certainty equivalent.**

***Index Terms*—Option portfolio, transaction cost, robustness, BSUM-M, Sharpe ratio.**

I. Introduction

PORTFOLIO design has attracted great attention from researchers ever since Markowitz introduced the mean-variance portfolio optimization framework in 1952 [1] (for which he got the Nobel price in 1990). This framework plays a fundamental role in modern portfolio theory by using a statistical modeling in the portfolio formulation. It aims at achieving a trade-off between expected return and risk (measured by portfolio variance). This framework is well-known for its flexibility: if an investor is willing to take a risk, more weight is given to the expected return; otherwise, more weight is placed on risk.

In the open literature, most works merely considered stocks in the portfolio design. The reason for this is very straightforward: stock data is readily available online and seems relatively easy to understand and manipulate. One can easily estimate the expected return and covariance matrix of a certain number of stocks (albeit usually with a dubious quality of estimates). However, the applicability of the Markowitz framework is not limited to stocks. It would be desirable to extend it to include derivatives. For the sake of concreteness, we focus on a specific kind of derivative named "options" (vanilla options, to be exact), but the approach can be straightforwardly employed with other derivatives. Traditionally, derivatives are regarded as hedging instruments, but, in this paper, we will reveal their potential for investment under the Markowitz framework.

Manuscript received October 3, 2017; revised February 13, 2018 and June 7, 2018; accepted June 18, 2018. Date of publication June 22, 2018; date of current version July 5, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mats Bengtsson. This work was supported by the Hong Kong RGC 16208917 Research Grant. *(Corresponding author: Licheng Zhao.)*

The authors are with the Hong Kong University of Science and Technology, Hong Kong (e-mail: lzhaoai@connect.ust.hk; palomar@ust.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2018.2849733

A. Preliminaries on Options

Admittedly, derivatives, especially options, are more complicated than stocks, see [2]–[4] for popular textbooks on derivatives. A standard (vanilla) option contract consists of the following parameters: option price, the underlying asset (mostly stocks), expiration date, and strike price. A call (put) option gives the option holder the right, rather than obligation, to buy (sell) the underlying asset by the expiration date for the strike price. American options can be exercised at any time before expiration, while European options can only be exercised on the expiration date. Most of the trading options on exchanges are American style.

The price of an option is associated with the following factors: current price of the underlying asset S_0, strike price K, risk-free interest rate r_{free}, time to expiration T, and (underlying) volatility σ. One popular approach to evaluating a European call or put option is the Black-Scholes-Merton formula [5], [6] (for which another Nobel price was awarded in 1997):

$$\begin{aligned}\text{Call Price} &= S_0 N(d_1) - Ke^{-rT} N(d_2)\\ \text{Put Price} &= Ke^{-rT} N(-d_2) - S_0 N(-d_1),\end{aligned} \tag{1}$$

where

$$\begin{aligned}d_1 &= \frac{\log(S_0/K) + \left(r_{\text{free}} + \sigma^2/2\right)T}{\sigma\sqrt{T}},\\ d_2 &= \frac{\log(S_0/K) + \left(r_{\text{free}} - \sigma^2/2\right)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T},\end{aligned} \tag{2}$$

and $N(x)$ is the cumulative distribution function for standard Gaussian distribution. Let us look at a toy example to gain some insight into how much an option is worth.

Example 1: Suppose the current price of the underlying stock is \$105, the strike price is \$100, the risk-free interest rate is 2% per annum, and the volatility is 20% per annum. The option is European style and expires in 6 months. Thus, by applying the Black-Scholes-Merton formula, we obtain the option price as follows:

$$\begin{cases}\text{Call Price} = & \$9.24\\ \text{Put Price} = & \$3.24.\end{cases} \tag{3}$$

There are several reasons why one may want to include options in the portfolio design. The first reason is the comparatively higher return, which can be seen from the aforementioned example. We already calculated the call option price: \$9.24. Suppose the stock price increases by \$1 on the next trading day, and thus the call option price moves up to \$9.93 according to the Black-Scholes-Merton formula. The return in stock is only $1/105 = 0.95\%$, while the return in the option is $(9.93 - 9.24)/9.24 = 7.47\%$, which is much more attractive. We should note that the comparatively higher return of options comes at the cost of higher tail and kurtosis risk because options prices are known to be heavily-tailed and not log-normal distributed in practice.[1] Options are a convenient method to make directional bets on stocks without too much exposure. The second reason is the convenience of taking a short position. When an investor believes the market will go down, he has no other choice than to short sell if he only trades stocks. During the short selling period, he has to maintain a margin account so that the deposit is not lower than the minimum requirement. On the other hand, if options are considered in the trading, he can conveniently buy a few put options instead, and no extra money is required. While many advantages can be identified, we still need to point out that options are not a panacea. The option market often experiences a larger percentage change than the stock market, so frequent adjustment or rebalancing of the option position is a must. Another weakness is concerned with the expiration date. If the market takes an unexpected large move opposite to the investor's belief and the expiration date is drawing near, the loss will be inevitable.

B. Related Works

So far, there have been two prevailing philosophies for option portfolio design. The first one is based on the single period portfolio optimization framework [7]–[9]. The investors delicately design an option portfolio at the current time in the hope of maximizing the expected return or minimizing the value at risk (VaR) on the expiration date [8]. This philosophy works well with European options because they cannot be executed until expiration. One weakness of this philosophy is that it fails to consider the trading of option contracts: options can be bought and sold as well as executed. Moreover, this philosophy is faced with the difficulty of long-term return estimation. In order to make the estimation result reliable, it is recommended that only near-expiration options are chosen. The final concern relates to the risk management. The option market is more volatile in terms of percentage change and it is very risky to conduct a one-shot investment without further adjustments. Any adverse market move before the expiration date could result in a big loss for the investors.

In view of all the drawbacks, we would prefer to design a dynamic option portfolio that is subject to daily adjustment, i.e., actively trading the options rather than passively executing them. This is exactly the second prevailing philosophy, which is based on the delta-gamma approximation of the function of the option price (not necessarily the Black-Scholes-Merton formula) [9]–[11]. This approximation is by nature a first-order Taylor expansion in time difference and second-order in stock price difference and can be rederived from the stochastic differential equation perspective, as we will elaborate in later sections. We will approximate the option price difference up to the second order and ignore any higher order statistics. The biggest advantage of adopting this philosophy is the flexibility. We can include virtually all the vanilla options in the portfolio, whether the lifespan is short or long and the style is European or American. Besides, we can conduct daily adjustments very conveniently. The approximation is renewed on a daily basis, so the portfolio is dynamically updated. The possible weakness of this philosophy is also obvious: we need extra information on the partial derivatives and we need to update them frequently to ensure a valid approximation. Fortunately, some powerful terminals, e.g., the Bloomberg terminal, offer such statistical information. Some software libraries can also compute options-related quantities such as QuantLib [12]. On top of that, the designed portfolio strongly depends on the current-moment data, so the investment decision could be myopic.

C. Contribution

The major contributions of this paper are:

1) We derive the expressions of mean and variance for a portfolio with mixed stocks and options. To the best of our knowledge, we are the first to apply the Markowitz mean-variance framework in option portfolio design by means of exploiting both first- and second-order statistics of option returns using stochastic differential equations. Conventionally, the delta-gamma approximation focuses on the expected return of options, which corresponds to the mean term of the proposed optimization problem. Thus, we enrich the traditional delta-gamma approximation by further considering second-order statistics, i.e., the variance of option return. It is also worth noticing that modern portfolio selection practice involves more sophisticated measures of risk, e.g., VaR or CVaR (Conditional VaR). We will look into these more realistic approaches in our future work.
2) We identify a weakness in the variance term of the option portfolio in the form of a zero-risk subspace, which in theory seems good but in practice crumbles due to estimation errors in the parameters (we call this phenomenon the zero-risk fallacy). We then introduce different kinds of robustness to fix this problem. Eventually, we present a unified formulation including all the proposed modifications. We additionally propose a Black-Litterman model with specific views derived from the stock-options relationships.
3) We additionally consider transaction costs in the formulation.
4) We propose an efficient BSUM-M-based algorithm to solve the portfolio design problem. It is especially useful when the off-the-shelf solvers are not available on some online financial programming platforms. According to synthetic simulation results, the proposed algorithm can achieve as good solutions as MOSEK (an off-the-shelf solver) and the computational time is within 0.3 seconds when the problem size is smaller than 500, around one half or one order of magnitude faster, depending on the choice of parameters.
5) We demonstrate via real-data numerical simulations the superior performance of our proposed stock-option portfolio compared to the pure-stock portfolio design (in some cases achieving a Sharpe ratio of 3.60 compared to the 0.73 of the benchmark).

D. Organization and Notation

The rest of the paper is organized as follows. In Section II, we present the price model and derive the problem formulation. In Section III, we improve the original naive formulation by introducing different kinds of robustness so as to fix the

[1] We ignore the heavy-tail issue and higher-order risk (risk higher than the second order) in this paper and leave them for future work.

zero-risk fallacy. In Section IV, we provide an efficient algorithm to solve the robust formulation, which serves as an alternative to the off-the-shelf solvers. Finally, Section V presents numerical results, and the conclusions are drawn in Section VI.

The following notation is adopted. Boldface upper-case letters represent matrices, boldface lower-case letters denote column vectors, and standard lower-case or upper-case letters stand for scalars. $\mathbb{R}$ denotes the real field. $\odot$ stands for the Hadamard product. $[x]_+ = \max(x, 0)$. $\|\cdot\|_p$ denotes the ℓ_p-norm of a vector. $\nabla(\cdot)$ represents the gradient of a multivariate function. $\mathbf{1}$ stands for the all-one vector and $\mathbf{I}$ stands for the identity matrix. $\mathbf{X}^T$, $\mathrm{Tr}(\mathbf{X})$, $\mathrm{rank}(\mathbf{X})$, and $\lambda_{\max}(\mathbf{X})$ denote the transpose, trace, rank, and the largest eigenvalue of $\mathbf{X}$, respectively. $\sigma_i(\mathbf{X})$ is the ith largest singular value of $\mathbf{X}$. $\mathrm{Diag}(\mathbf{x})$ is a diagonal matrix with $\mathbf{x}$ filling its principal diagonal. Block diagonal matrix

$$\begin{bmatrix} \mathbf{x}_1 & & & \\ & \mathbf{x}_2 & & \\ & & \ddots & \\ & & & \mathbf{x}_I \end{bmatrix}$$

is compactly rewritten as $\mathrm{Blkdiag}(\{\mathbf{x}_i\}_{i=1}^I)$. $\mathbf{X} \succeq \mathbf{0}$ means $\mathbf{X}$ is positive semidefinite. $\|\mathbf{X}\|_{\sigma p}$ and $\|\mathbf{X}\|_{a,b}$ denotes the matrix Schatten p-norm and $\ell_{a,b}$-norm of $\mathbf{X}$, respectively.

$$\mathrm{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}.$$

II. Price Modeling and Problem Statement

A. Price Modeling

We denote the stock price at time t as S_t and make the following assumption on the stock price process.

Assumption 1 ([2]): The stock price S_t satisfies the following geometric Brownian motion:

$$dS_t = \mu S_t dt + \sigma S_t dz_t, \tag{4}$$

where μ and σ are given parameters standing for the mean and volatility of the percentage change of S_t, and z_t is the Wiener process.

We denote the price of a particular derivative (could be options, futures, etc.) at time t as F_t. We impose the following assumption on F_t.

Assumption 2: Let F_t be a function of time and the price of its underlying, i.e., $F_t = F_t(S_t, t)$ where $0 \le t \le T$ and T is the time from now to the expiration date.

According to Itô's lemma [2, Sec. 14.6], the differential of F_t is

$$dF_t = \left(\frac{\partial F_t}{\partial S_t}\mu S_t + \frac{\partial F_t}{\partial t} + \frac{1}{2}\frac{\partial^2 F_t}{\partial S_t^2}\sigma^2 S_t^2\right) dt + \frac{\partial F_t}{\partial S_t}\sigma S_t dz_t. \tag{5}$$

In practice, we can hardly expect to obtain stock data on a continuous time basis, so we modify the stochastic differential equations to their discrete counterparts. The notations in (4) and (5) are changed as follows: Δt in place of dt, Δz_t in place of dz_t, ΔS_t in place of dS_t, and ΔF_t in place of dF_t. Now that Δt is no longer arbitrarily small, the expressions of ΔS_t and ΔF_t are merely an approximation. We assume the following approximation is valid.

Assumption 3: In discrete time, the underlying stock price approximately satisfies a geometric Brownian motion

$$\Delta S_t \simeq \mu S_t \Delta t + \sigma S_t \Delta z_t, \tag{6}$$

with the percentage drift μ and volatility σ staying constant for a short-term period Δt. The difference of F_t is assumed to be validly approximated by its continuous differential counterpart, i.e.,

$$\Delta F_t \simeq \left(\frac{\partial F_t}{\partial S_t}\mu S_t + \frac{\partial F_t}{\partial t} + \frac{1}{2}\frac{\partial^2 F_t}{\partial S_t^2}\sigma^2 S_t^2\right)\Delta t + \frac{\partial F_t}{\partial S_t}\sigma S_t \Delta z_t. \tag{7}$$

Higher order statistics (Greeks) are not considered in this paper.

We can estimate the value of μ and σ from historical stock prices. Note that (7) could be the price change of any derivative. In this paper, we specify the derivatives as vanilla call and put options for the sake of concreteness, whose prices are denoted as C_t and P_t, respectively. The expressions of ΔC_t and ΔP_t can be readily obtained from (7):

$$\Delta C_t \simeq \left(\Delta_{C,t}\mu S_t + \Theta_{C,t} + \frac{1}{2}\Gamma_{C,t}\sigma^2 S_t^2\right)\Delta t + \Delta_{C,t}\sigma S_t \Delta z_t \tag{8}$$

and

$$\Delta P_t \simeq \left(\Delta_{P,t}\mu S_t + \Theta_{P,t} + \frac{1}{2}\Gamma_{P,t}\sigma^2 S_t^2\right)\Delta t + \Delta_{P,t}\sigma S_t \Delta z_t, \tag{9}$$

where $\Delta_{C,t} = \frac{\partial C_t}{\partial S_t}$, $\Delta_{P,t} = \frac{\partial P_t}{\partial S_t}$, $\Theta_{C,t} = \frac{\partial C_t}{\partial t}$, $\Theta_{P,t} = \frac{\partial P_t}{\partial t}$, $\Gamma_{C,t} = \frac{\partial^2 C_t}{\partial S_t^2}$, and $\Gamma_{P,t} = \frac{\partial^2 P_t}{\partial S_t^2}$. Note that the first term of ΔC_t or ΔP_t is exactly the well-known delta-gamma approximation. Although the Black-Scholes-Merton formula only applies to European options, the aforementioned analysis can be extended to American options as well as any other derivative that follows the form $F_t = F_t(S_t, t)$.

Remark 2: In continuous time, we can easily see in (5) that the source of the variance of the option is the underlying stock because they share the same stochastic source dz_t. In discrete time, for any $0 < t < T$, the option value F_t is a deterministic function of time index t and its underlying stock price S_t based on the evaluation of future pay-out. Now for $t+1$, we do not know S_{t+1} yet. If we knew it, then F_{t+1} would be deterministic and $F_{t+1} - F_t \triangleq \Delta F_t$ would be deterministic as well. However, S_{t+1} is only characterized statistically with some mean and variance. This translates into the stochastic nature of $S_{t+1} - S_t \triangleq \Delta S_t$. We can see from eq. (6) and (7) that ΔS_t and ΔF_t share the same stochastic source Δz_t. This indicates that the source of the variance of the option is the underlying stock. Moreover, this source is consistent with the Black-Scholes world. We can rederive the Black-Scholes-Merton formula with Assumptions 1 and 2, Itô's lemma, and the boundary conditions for vanilla call or put options. Technical details can be found in [2, Chap. 15.6].

Remark 3: Now let us look into these partial derivatives. These partial derivatives are named Greeks in the financial industry, namely, Δ, Θ, and Γ. They are used to measure the sensitivity of the price of options to a change in the underlying stock price. To simplify notation, we omit the subscript "t" for the moment. Δ is defined as the rate of change of the option price with respect to the underlying stock price. Θ is known as the time decay parameter. It is the rate of change of the option value with respect to the passage of time. The definition of Γ is the rate of change of the option' s Δ with respect to the underlying stock price. If Γ is small, Δ changes mildly; if Γ is large, Δ is sensitive to the underlying stock price. To sum up, each of these Greeks measures a different dimension to the risk in an option position. Interested readers can refer to [2, Chap. 19] for more details.

B. Warm-up: Rederivation of Markowitz Mean-Variance Framework for Stocks

Suppose we construct a portfolio consisting of I stocks, with the proportion of the total budget B allocated to the ith stock w_i (note that by definition $\sum_{i=1}^{I} w_i = 1$). The value of the portfolio is then $\Pi = \sum_{i=1}^{I} Bw_i = B$. For simplicity of notation, we drop the time subscript "t" for the moment. Now we study the percentage change of Π:

$$\begin{aligned}\frac{\Delta\Pi}{\Pi} &= \frac{\sum_{i=1}^{I} Bw_i \frac{\Delta S_i}{S_i}}{B} \overset{(a)}{\simeq} \sum_{i=1}^{I} w_i \frac{\mu_i S_i \Delta t + \sigma_i S_i \Delta z_i}{S_i} \\ &= \underbrace{\sum_{i=1}^{I} w_i \mu_i \Delta t}_{\text{deterministic}} + \underbrace{\sum_{i=1}^{I} w_i \sigma_i \Delta z_i}_{\text{stochastic}} \end{aligned} \tag{10}$$

where (a) follows from (6). Recall that z_i's are Wiener processes with $\mathsf{E}[\Delta z_i] = 0$, $\mathsf{E}[\Delta z_i \Delta z_j] = \rho_{ij}\Delta t$ with ρ_{ij} being correlation coefficients. Thus,

$$\mathsf{E}\left[\frac{\Delta\Pi}{\Pi}\right] \simeq \sum_{i=1}^{I} w_i \mu_i \Delta t \triangleq \mathbf{w}^T \boldsymbol{\mu} \times \Delta t, \tag{11}$$

$$\begin{aligned}\mathsf{Var}\left[\frac{\Delta\Pi}{\Pi}\right] &\simeq \mathsf{E}\left[\left(\sum_{i=1}^{I} w_i \sigma_i \Delta z_i\right)^2\right] \triangleq \mathsf{E}\left[\left(\mathbf{w}^T \mathrm{Diag}(\boldsymbol{\sigma})\Delta\mathbf{z}\right)^2\right] \\ &= \mathbf{w}^T \mathrm{Diag}(\boldsymbol{\sigma})\, \mathsf{E}\left[\Delta\mathbf{z}\Delta\mathbf{z}^T\right] \mathrm{Diag}(\boldsymbol{\sigma})\,\mathbf{w} \\ &= \mathbf{w}^T \mathrm{Diag}(\boldsymbol{\sigma}) \begin{bmatrix} \rho_{11} & \cdots & \rho_{1I} \\ \vdots & \ddots & \vdots \\ \rho_{I1} & \cdots & \rho_{II} \end{bmatrix} \mathrm{Diag}(\boldsymbol{\sigma})\,\mathbf{w} \times \Delta t \\ &\triangleq \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \times \Delta t, \end{aligned} \tag{12}$$

where $\boldsymbol{\mu}$ is the expected stock return and $\boldsymbol{\Sigma}$ is the covariance of the stock returns. In order to achieve a tradeoff between the portfolio expected return ($\mathsf{E}[\frac{\Delta\Pi}{\Pi}]$) and risk ($\mathsf{Var}[\frac{\Delta\Pi}{\Pi}]$), we eventually obtain the Markowitz mean-variance optimization problem [1] [13, Sec. 5.1.1].

C. Extension to Portfolio with Mixed Stocks and Options

Now we construct a portfolio of mixed stocks and options and we only consider vanilla call and put options. We include I stocks in the portfolio and for each stock, e.g., the ith stock, we consider M_i call options and N_i put options. We redefine the normalized portfolio vector $\mathbf{w}$ as

$$\mathbf{w} = \left[\, \mathbf{w}_1^T \,\cdots\, \mathbf{w}_i^T \,\cdots\, \mathbf{w}_I^T \,\right]^T, \tag{13}$$

where

$$\mathbf{w}_i = \Bigg[\underbrace{w_{S,i}}_{\text{stock}} \; \underbrace{w_{C,1i} \,\cdots\, w_{C,M_i i}}_{\text{call options}} \; \underbrace{w_{P,1i} \,\cdots\, w_{P,N_i i}}_{\text{put options}} \Bigg]^T. \tag{14}$$

The expression of the percentage change of Π is (time subscript "t" is dropped for simplicity of notation)

$$\begin{aligned}\frac{\Delta\Pi}{\Pi} &= \frac{\sum_{i=1}^{I}\left(Bw_{S,i}\frac{\Delta S_i}{S_i} + \sum_{m=1}^{M_i} Bw_{C,mi}\frac{\Delta C_{mi}}{C_{mi}} + \sum_{n=1}^{N_i} Bw_{P,ni}\frac{\Delta P_{ni}}{P_{ni}}\right)}{B} \\ &\overset{(a)}{\simeq} \sum_{i=1}^{I}\Bigg[\frac{w_{S,i}}{S_i}\left(\mu_i S_i \Delta t + \sigma_i S_i \Delta z_i\right) \\ &+ \sum_{m=1}^{M_i}\frac{w_{C,mi}}{C_{mi}}\Delta_{C,mi}\sigma_i S_i \Delta z_i + \sum_{n=1}^{N_i}\frac{w_{P,ni}}{P_{ni}}\Delta_{P,ni}\sigma_i S_i \Delta z_i \\ &+ \sum_{m=1}^{M_i}\frac{w_{C,mi}}{C_{mi}}\left(\Delta_{C,mi}\mu_i S_i + \Theta_{C,mi} + \frac{1}{2}\Gamma_{C,mi}\sigma_i^2 S_i^2\right)\Delta t \\ &+ \sum_{n=1}^{N_i}\frac{w_{P,ni}}{P_{ni}}\left(\Delta_{P,ni}\mu_i S_i + \Theta_{P,ni} + \frac{1}{2}\Gamma_{P,ni}\sigma_i^2 S_i^2\right)\Delta t\Bigg] \end{aligned} \tag{15}$$

where (a) follows from the expressions of ΔC_{mi} and ΔP_{ni} in (8) and (9) and $\Delta_{C,mi}$, $\Delta_{P,ni}$, $\Theta_{C,mi}$, $\Theta_{P,ni}$, $\Gamma_{C,mi}$, and $\Gamma_{P,ni}$ follow the same definition as in (8) and (9) up to a subscript difference. Thus,

$$\begin{aligned}\mathsf{E}\left[\frac{\Delta\Pi}{\Pi}\right] &\simeq \sum_{i=1}^{I}\Bigg[w_{S,i}\mu_i\Delta t \\ &+ \sum_{m=1}^{M_i}\frac{w_{C,mi}}{C_{mi}}\left(\Delta_{C,mi}\mu_i S_i + \Theta_{C,mi} + \frac{1}{2}\Gamma_{C,mi}\sigma_i^2 S_i^2\right)\Delta t \\ &+ \sum_{n=1}^{N_i}\frac{w_{P,ni}}{P_{ni}}\left(\Delta_{P,ni}\mu_i S_i + \Theta_{P,ni} + \frac{1}{2}\Gamma_{P,ni}\sigma_i^2 S_i^2\right)\Delta t\Bigg] \\ &\triangleq \mathbf{w}^T\mathbf{u} \times \Delta t, \end{aligned} \tag{16}$$

where

$$\mathbf{u} = \left[\, \mathbf{u}_1^T \,\cdots\, \mathbf{u}_i^T \,\cdots\, \mathbf{u}_I^T \,\right]^T, \tag{17}$$

and

$$\begin{aligned}\mathbf{u}_i = \Bigg[&\mu_i, \frac{1}{C_{1i}}\left(\Delta_{C,1i}\mu_i S_i + \Theta_{C,1i} + \frac{1}{2}\Gamma_{C,1i}\sigma_i^2 S_i^2\right), \ldots, \\ &\frac{1}{C_{M_i i}}\left(\Delta_{C,M_i i}\mu_i S_i + \Theta_{C,M_i i} + \frac{1}{2}\Gamma_{C,M_i i}\sigma_i^2 S_i^2\right), \\ &\frac{1}{P_{1i}}\left(\Delta_{P,1i}\mu_i S_i + \Theta_{P,1i} + \frac{1}{2}\Gamma_{P,1i}\sigma_i^2 S_i^2\right), \ldots, \\ &\frac{1}{P_{N_i i}}\left(\Delta_{P,N_i i}\mu_i S_i + \Theta_{P,N_i i} + \frac{1}{2}\Gamma_{P,N_i i}\sigma_i^2 S_i^2\right)\Bigg]^T; \end{aligned} \tag{18}$$

$$\begin{aligned}\mathsf{Var}\left[\frac{\Delta\Pi}{\Pi}\right] &\simeq \mathsf{Var}\left[\sum_{i=1}^{I}\left(w_{S,i}+\sum_{m=1}^{M_i}\frac{w_{C,mi}}{C_{mi}}\Delta_{C,mi}S_i\right.\right.\\ &\qquad\left.\left.+\sum_{n=1}^{N_i}\frac{w_{P,ni}}{P_{ni}}\Delta_{P,ni}S_i\right)\sigma_i\Delta z_i\right]\\ &\overset{(a)}{=} \mathsf{Var}\left[\mathbf{w}^T\mathbf{V}\mathrm{Diag}\left(\boldsymbol{\sigma}\right)\Delta\mathbf{z}\right]\\ &= \mathbf{w}^T\mathbf{V}\mathrm{Diag}\left(\boldsymbol{\sigma}\right)\mathsf{E}\left[\Delta\mathbf{z}\Delta\mathbf{z}^T\right]\mathrm{Diag}\left(\boldsymbol{\sigma}\right)\mathbf{V}^T\mathbf{w}\\ &\overset{(b)}{=} \mathbf{w}^T\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^T\mathbf{w}\times\Delta t, \qquad (19)\end{aligned}$$

where (a) follows from defining $\mathbf{V} = \mathrm{Blkdiag}(\{\mathbf{v}_i\}_{i=1}^{I})$ and $\mathbf{v}_i = [1\ \frac{\Delta_{C,1i}S_i}{C_{1i}}\ \cdots\ \frac{\Delta_{C,M_i i}S_i}{C_{M_i i}}\ \frac{\Delta_{P,1i}S_i}{P_{1i}}\ \cdots\ \frac{\Delta_{P,N_i i}S_i}{P_{N_i i}}]^T$ and (b) from the definition of $\boldsymbol{\Sigma}$ as in (12).

Remark 4 (No risk-free arbitrage): We denote $N \triangleq I + \sum_{i=1}^{I}(M_i + N_i)$. Note that the length of $\mathbf{w}$ is N, the size of $\mathbf{V}$ is $N \times I$, and the size of $\boldsymbol{\Sigma}$ is $I \times I$. This means that the covariance matrix $\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^T$ is highly rank-deficient and has a nontrivial null space. This result makes sense because the prices of options are perfectly determined from those of their underlying stocks. Under the assumption of no riskless arbitrage opportunities (the Black-Scholes world), any stock-option portfolio in that null subspace (i.e., with a zero risk) shall achieve a zero excess return. Furthermore, any stock-option portfolio with a nonzero risk can achieve a nonzero excess return but in theory the same return can be achieved with only stocks. The only hope to construct a stock-option portfolio with a better performance than only stocks is that the original assumption of no riskless arbitrage does not hold. In addition, in a practical case, one has to deal with the errors of parameter estimation. Admittedly, in the absence of estimation error, one can find an arbitrage portfolio, just like what [14] conjectured. With estimation error, this naive risk measurement can be disastrous since the claimed zero risk fails to materialize in practice. We name this phenomenon "zero-risk fallacy" and it will be properly addressed in Section III. Paper [15] suggests that there may exist an efficient asset subset with which we can achieve the same performance.

Remark 5 (Investment opportunities): In the Black-Scholes world, it is assumed that no riskless arbitrage opportunities exist and the true values of the parameters are known to all the financial market participants and are applied by everyone in the evaluation of different securities. In the practical financial world, the assumption of no riskless arbitrage does not seem to hold. There exist undervalued and overpriced assets. Some people can do better in recognizing the mispricing than others due to a better knowledge, so we can have some hope to construct a stock-option portfolio with a better performance than only stocks. The true statistical parameters are hidden to everyone. Every financial market participant is trying to make a better mean-variance estimation so as to take advantage of others' mispricings.

D. Transaction Cost Concerns

Trading of stocks or options incurs transaction costs. Take the commission rule of Interactive Brokers[2] for example. The transaction cost for stocks is the minimum of USD 0.005 per share and 0.5% of the trade value if this number is larger than USD 1.00; otherwise the cost is USD 1.00. The transaction cost caused by trading options of fewer than 10,000 monthly contracts (one contract represents 100 shares of stocks) is

- USD 0.70 per contract if the premium (option price) is no less than USD 0.10,
- USD 0.50 per contract if the premium is between USD 0.05 and 0.10, and
- USD 0.25 per contract if the premium is smaller than USD 0.05

if this number is larger than USD 1.00. For simplicity, we assume our trading volume is always large enough to exceed USD 1.00, whether stocks or options. Also, the trading volume of options is always smaller than 10,000 monthly contracts. In this case, the transaction cost of trading w_S dollars of stocks at price S is

$$\text{Transaction Cost} = 0.5\% \times w_S / \max(1, S), \qquad (20)$$

while the transaction cost of trading w_C (w_P) dollars of call (put) options at price C (P) is

$$\text{Transaction Cost} = \begin{cases}\eta(C) \times w_C / C \text{ call}\\ \eta(P) \times w_P / P \text{ put}\end{cases}, \qquad (21)$$

where

$$\eta(C \text{ or } P) = \begin{cases} 0.70/100 & C \text{ or } P \geq 0.1\\ 0.50/100 & 0.05 \leq C \text{ or } P < 0.1\,.\\ 0.25/100 & C \text{ or } P < 0.05\end{cases}$$

To sum up, the expression of transaction cost can be compactly written as

$$B\left\|(\mathbf{w} - \mathbf{w}_0) \odot \mathbf{q}\right\|_1, \qquad (22)$$

where $\mathbf{w}$ is the target portfolio, $\mathbf{w}_0$ is the current portfolio, and

$$\mathbf{q} = \left[\,\mathbf{q}_1^T\ \cdots\ \mathbf{q}_i^T\ \cdots\ \mathbf{q}_I^T\,\right]^T, \qquad (23)$$

with

$$\mathbf{q}_i = \left[\frac{0.5\%}{\max(1,S_i)}\ \frac{\eta(C_{1i})}{C_{1i}}\ \cdots\ \frac{\eta(C_{M_i i})}{C_{M_i i}}\ \frac{\eta(P_{1i})}{P_{1i}}\ \cdots\ \frac{\eta(P_{N_i i})}{P_{N_i i}}\right]^T. \qquad (24)$$

The transaction cost penalty coincides with the LASSO estimation technique [16], [17].

E. Problem Formulation

The ideal investment portfolio has the following characteristics: 1) high (expected) return, 2) low risk, and 3) low turnover. The first two characteristics are self-explanatory. As for low turnover, the true motivation is to lower the transaction cost caused by rebalancing. Since we can already model transaction cost, we can directly minimize this quantity. In order to design a desirable portfolio, we want to achieve a tradeoff between high expected return ($\mathsf{E}[\frac{\Delta\Pi}{\Pi}]$), low risk (measured by $\mathsf{Var}[\frac{\Delta\Pi}{\Pi}]$), and low transaction cost ($B\|(\mathbf{w}-\mathbf{w}_0)\odot\mathbf{q}\|_1$). Thus we naturally formulate the optimization problem as

$$\begin{aligned}\underset{\mathbf{w}}{\mathsf{minimize}}\quad & -\mathbf{w}^T\mathbf{u} + \lambda\mathbf{w}^T\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^T\mathbf{w}\\ & + \xi B\left\|(\mathbf{w}-\mathbf{w}_0)\odot\mathbf{q}\right\|_1\\ \mathsf{subject\ to}\quad & \mathbf{1}^T\mathbf{w} = 1\\ & \mathbf{w} \geq \mathbf{0}, \qquad (25)\end{aligned}$$

where λ and ξ are positive regularization parameters. The scaling factor Δt that appears in (16) and (19) is removed in $\mathsf{E}[\frac{\Delta\Pi}{\Pi}]$

[2]https://www.interactivebrokers.com

and $\mathsf{Var}[\frac{\Delta\Pi}{\Pi}]$ and it is absorbed in the parameter ξ. We additionally impose the long-only constraint because 1) short selling of stocks and options requires an extra margin deposit [2, Sec. 10.7] which we do not want to consider (although it could be allowed in the formulation); 2) short selling of options can be very risky, especially for risk-averse investors [8].

III. Robust Risk Measurement

In Section II-C, we derived the expression of $\mathsf{Var}[\frac{\Delta\Pi}{\Pi}]$, which is regarded as a measure of risk: $\mathsf{Var}[\frac{\Delta\Pi}{\Pi}] \propto \mathbf{w}^T\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^T\mathbf{w}$. This quantity could be zero even if $\mathbf{w} \neq \mathbf{0}$ because $\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^T$ is rank deficient. This zero-risk phenomenon cannot be realized in practice since we are investing in risky assets, and thus we name this phenomenon risk-free fallacy. We see two reasons to account for the risk-free fallacy. The first reason is that, assuming the covariance $\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^T$ can be perfectly estimated, the portfolio is still exposed to a higher order risk despite that it is risk-free to a lower order. There do exist methods for hedging a higher order risk, but such methods often introduce new heavier tail risks that are even harder to hedge. Including higher order moments enriches the risk measurement so that the overall risk will never achieve zero. The second reason is that the covariance $\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^T$ contains the estimation error due to the estimation error in $\mathbf{V}$ and $\boldsymbol{\Sigma}$. Higher order statistics are not easy to obtain in practice, so an alternative is to work on the covariance to suppress the effect of estimation error. In the following, we propose introducing different types of robustness so as to force the risk term never to become zero.

A. Stochastic Robustness

One way to deal with the zero-risk fallacy is to make the quadratic matrix full rank, and the most straightforward way is to introduce stochastic robustness, i.e., model the parameters as random variables around the noisy estimates. We borrow the idea from [18], a wireless communications application, where the authors acknowledged the imperfectness of channel state information (as opposed to naively assuming the estimates were perfect) and modeled the parameters statistically. In their modeling scheme, the channel parameter consists of a deterministic component equal to the estimate and a zero-mean stochastic component modeling the estimation error.

Let's start from the definition of covariance. We define the asset return as $\mathbf{r} = [\mathbf{r}_1^T, \ldots, \mathbf{r}_I^T]$ with

$$\mathbf{r}_i^T = [\underbrace{r_{S,i}}_{\text{stock return}}, \underbrace{r_{C,1i}, \ldots, r_{C,M_i i}}_{\text{returns of call options}}, \underbrace{r_{P,1i}, \ldots, r_{P,N_i i}}_{\text{returns of put options}}]. \tag{26}$$

For stock return, $r_S = \frac{\Delta S}{S}$; for options return, $r_C = \frac{\Delta C}{C}$ and $r_P = \frac{\Delta P}{P}$. The expressions can be found in (15). Applying the law of total covariances, we get

$$\begin{aligned}\mathsf{Cov}[\mathbf{r}] &= \mathsf{Cov}[\mathsf{E}[\mathbf{r}|\boldsymbol{\Sigma},\mathbf{V}]] + \mathsf{E}[\mathsf{Cov}[\mathbf{r}|\boldsymbol{\Sigma},\mathbf{V}]]\\ &\simeq \mathsf{Cov}[\mathbf{u}] + \mathsf{E}\left[\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^T\right],\end{aligned} \tag{27}$$

where the expected return $\mathbf{u}$ comes from (17) and the approximation results from (16) and (19). Following the logic of stochastic robustness, we can use the same modeling on the three parameters $\mathbf{u}$, $\mathbf{V}$, and $\boldsymbol{\Sigma}$. We model $\mathbf{u} = \bar{\mathbf{u}} + \mathbf{m}$ where $\bar{\mathbf{u}}$ is the estimation of $\mathbf{u}$ (the μ's come from sample mean and the Greeks are obtained from the Bloomberg terminal) and $\mathbf{m}$ is elementwisely independent and identically distributed with $\mathsf{E}[\mathbf{m}] = \mathbf{0}$ and $\mathsf{Cov}[\mathbf{m}] = \mathbf{D}_\mathbf{m}$, a diagonal matrix. One way to construct $\mathbf{D}_\mathbf{m}$ is to impose variance on μ, Δ, Θ, and Γ independently and then figure out the overall variance for each r_S, r_C, and r_P so as to form $\mathbf{D}_\mathbf{m}$'s principal diagonal. Recall that $\mathbf{V} = \mathrm{Blkdiag}(\{\mathbf{v}_i\}_{i=1}^I)$, so we can assume $\forall i$, $\mathbf{v}_i = \bar{\mathbf{v}}_i + \mathbf{n}_i$ where $\bar{\mathbf{v}}_i$ is deterministic and $\mathbf{n}_i$ is stochastic with

$$\begin{cases}\mathsf{E}[\mathbf{n}_i] = \mathbf{0}\\ \mathsf{E}\left[\mathbf{n}_i\mathbf{n}_j^T\right] = \mathbf{0},\ i \neq j\\ \mathsf{E}\left[\mathbf{n}_i\mathbf{n}_i^T\right] = \mathbf{R}_i \succeq \mathbf{0}.\end{cases} \tag{28}$$

Thus,

$$\begin{aligned}\mathbf{V} &= \mathrm{Blkdiag}\left(\{\mathbf{v}_i\}_{i=1}^I\right)\\ &= \mathrm{Blkdiag}\left(\{\bar{\mathbf{v}}_i\}_{i=1}^I\right) + \mathrm{Blkdiag}\left(\{\mathbf{n}_i\}_{i=1}^I\right)\\ &\triangleq \bar{\mathbf{V}} + \mathbf{N}.\end{aligned} \tag{29}$$

We similarly model $\boldsymbol{\Sigma}$: $\boldsymbol{\Sigma} = \bar{\boldsymbol{\Sigma}} + \boldsymbol{\Xi}$ with $\mathsf{E}[\boldsymbol{\Xi}] = \mathbf{0}$. We additionally assume $\mathbf{u}$, $\mathbf{V}$, and $\boldsymbol{\Sigma}$ are statistically independent. In the first stage of derivation, we merely rewrite the covariance matrix in terms of the stochastic model just described and obtain

$$\begin{aligned}\mathsf{Var}\left[\mathbf{w}^T\mathbf{r}\right] &= \mathbf{w}^T\mathsf{Cov}[\mathbf{r}]\,\mathbf{w} = \mathbf{w}^T\left(\mathsf{Cov}[\mathbf{u}] + \mathsf{E}\left[\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^T\right]\right)\mathbf{w}\\ &= \mathbf{w}^T\mathbf{D}_\mathbf{m}\mathbf{w} + \mathbf{w}^T\mathsf{E}\left[\left(\bar{\mathbf{V}}+\mathbf{N}\right)\left(\bar{\boldsymbol{\Sigma}}+\boldsymbol{\Xi}\right)\left(\bar{\mathbf{V}}^T+\mathbf{N}^T\right)\right]\mathbf{w}\\ &= \mathbf{w}^T\mathbf{D}_\mathbf{m}\mathbf{w} + \mathbf{w}^T\mathsf{E}\left[\left(\bar{\mathbf{V}}+\mathbf{N}\right)\bar{\boldsymbol{\Sigma}}\left(\bar{\mathbf{V}}^T+\mathbf{N}^T\right)\right]\mathbf{w}\\ &= \mathbf{w}^T\mathbf{D}_\mathbf{m}\mathbf{w} + \mathbf{w}^T\mathsf{E}\left[\bar{\mathbf{V}}\bar{\boldsymbol{\Sigma}}\bar{\mathbf{V}}^T\right]\mathbf{w} + 2\mathbf{w}^T\mathsf{E}\left[\bar{\mathbf{V}}\bar{\boldsymbol{\Sigma}}\mathbf{N}^T\right]\mathbf{w}\\ &\quad + \mathbf{w}^T\mathsf{E}\left[\mathbf{N}\bar{\boldsymbol{\Sigma}}\mathbf{N}^T\right]\mathbf{w}.\end{aligned} \tag{30}$$

We notice that 1) $\mathbf{w}^T\mathsf{E}[\bar{\mathbf{V}}\bar{\boldsymbol{\Sigma}}\bar{\mathbf{V}}^T]\mathbf{w} = \mathbf{w}^T\bar{\mathbf{V}}\bar{\boldsymbol{\Sigma}}\bar{\mathbf{V}}^T\mathbf{w}$ and 2) $\mathbf{w}^T\mathsf{E}[\bar{\mathbf{V}}\bar{\boldsymbol{\Sigma}}\mathbf{N}^T]\mathbf{w} = 0$. Therefore,

$$\mathsf{Var}\left[\mathbf{w}^T\mathbf{r}\right] = \mathbf{w}^T\mathbf{D}_\mathbf{m}\mathbf{w} + \mathbf{w}^T\bar{\mathbf{V}}\bar{\boldsymbol{\Sigma}}\bar{\mathbf{V}}^T\mathbf{w} + \mathbf{w}^T\mathsf{E}\left[\mathbf{N}\bar{\boldsymbol{\Sigma}}\mathbf{N}^T\right]\mathbf{w}. \tag{31}$$

The second stage of derivation is reflected in the following lemma.

Lemma 6: Suppose $\mathbf{N} = \mathrm{Blkdiag}(\{\mathbf{n}_i\}_{i=1}^I)$ satisfies (28). Then,

$$\mathsf{E}\left[\mathbf{N}\bar{\boldsymbol{\Sigma}}\mathbf{N}^T\right] = \mathrm{Blkdiag}\left(\{\bar{\Sigma}_{ii}\mathbf{R}_i\}\right) \triangleq \mathbf{D}_\mathbf{n}. \tag{32}$$

Proof: The proof is straightforward and is omitted due to space restrictions. ■

With stochastic robustness, we modify the original ideal and naive risk term $\mathbf{w}^T\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^T\mathbf{w}$ to a more realistic and meaningful one: $\mathbf{w}^T(\bar{\mathbf{V}}\bar{\boldsymbol{\Sigma}}\bar{\mathbf{V}}^T + \mathbf{D})\mathbf{w}$ where $\mathbf{D} = \mathbf{D}_\mathbf{m} + \mathbf{D}_\mathbf{n}$. $\mathbf{D}_\mathbf{m}$ is a diagonal matrix and $\mathbf{D}_\mathbf{n}$ is a block diagonal matrix, so $\mathbf{D}$ is block diagonal. $\bar{\mathbf{V}}\bar{\boldsymbol{\Sigma}}\bar{\mathbf{V}}^T + \mathbf{D}$ is a full rank covariance matrix. This modeling scheme indeed avoids the zero-risk fallacy, but it seems to have no effect on the parameter $\boldsymbol{\Sigma}$ except for a notational difference. This is because $\mathbf{w}^T\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^T\mathbf{w}$ is linear in $\boldsymbol{\Sigma}$. By taking the expectation, we can only get its deterministic component.

Remark 7: The initial risk measurement is wrongly estimated in practice and suffers from zero-risk fallacy due to lack of higher order statistics and estimation error in parameters. After we introduce stochastic robustness, the new risk measurement overcomes the zero-risk fallacy and is observed to enjoy better performance. The introduced robustness helps to suppress the effect of estimation error and, in a sense, make a slightly

TABLE I
SUMMARY OF MODIFIED RISK EXPRESSIONS

Category of Robustness		Expression	No.	Remark
Stochastic Only		$\mathbf{w}^T\left(\bar{\mathbf{V}}\bar{\boldsymbol{\Sigma}}\bar{\mathbf{V}}^T+\mathbf{D}\right)\mathbf{w}$	I	Convex in $\mathbf{w}$.
Stochastic & Worst-Case	Matrix Norm (Schatten Norm) Uncertainty	$\mathbf{w}^T\left[\bar{\mathbf{V}}\left(\bar{\boldsymbol{\Sigma}}+\varepsilon\mathbf{I}\right)\bar{\mathbf{V}}^T+\mathbf{D}\right]\mathbf{w}$	II	Convex in $\mathbf{w}$.
	Matrix Norm ($\ell_{a,b}$-norm) Uncertainty	$\mathbf{w}^T\left(\bar{\mathbf{V}}\bar{\boldsymbol{\Sigma}}\bar{\mathbf{V}}^T+\mathbf{D}\right)\mathbf{w}$ $+\varepsilon\left\Vert\bar{\mathbf{V}}^T\mathbf{w}\right\Vert_a\left\Vert\bar{\mathbf{V}}^T\mathbf{w}\right\Vert_b$	III	1. May not be convex in general. 2. Convex in $\mathbf{w}$ if a=b. 3. Boil down to Expression II if a=b=2.
	Elementwise Uncertainty	$\mathbf{w}^T\left(\bar{\mathbf{V}}\bar{\boldsymbol{\Sigma}}\bar{\mathbf{V}}^T+\mathbf{D}\right)\mathbf{w}$ $+\left\vert\bar{\mathbf{V}}^T\mathbf{w}\right\vert^T\mathbf{E}\left\vert\bar{\mathbf{V}}^T\mathbf{w}\right\vert,$ $E_{ij}=\varepsilon_{ij}$	IV	1. $\vert\cdot\vert$ means taking absolute value elementwisely. 2. May not be convex in general. 3. Convex in $\mathbf{w}$ if $\mathbf{E}$=$\varepsilon\mathbf{1}\mathbf{1}^T$.

better prediction on risk than before so that we are able to take advantage of others' mispricings.

B. One Step Further: Worst-Case Robustness

Since stochastic robustness provides no protection from the inaccurate estimation of $\boldsymbol{\Sigma}$, we consider further imposing worst-case robustness so as to take into account the uncertainty of $\boldsymbol{\Sigma}$. We assume $\boldsymbol{\Sigma}$ has a nominal value $\bar{\boldsymbol{\Sigma}}$ and lies within an uncertainty set $\mathcal{U}_{\boldsymbol{\Sigma}}$, which we specify in the following.

1) Matrix Norm (Schatten p-Norm) Uncertainty: We define the Schatten p-norm as

$$\|\mathbf{A}\|_{\sigma p}=\left(\sum_{i=1}^{\operatorname{rank}(\mathbf{A})}\left[\sigma_i\left(\mathbf{A}\right)\right]^p\right)^{\frac{1}{p}}. \tag{33}$$

We introduce the following lemma to show the impact of Schatten p-norm uncertainty.

Lemma 8: When $\mathcal{U}_{\boldsymbol{\Sigma}}=\{\boldsymbol{\Sigma}|\|\boldsymbol{\Sigma}-\bar{\boldsymbol{\Sigma}}\|_{\sigma p}\leq\varepsilon\}$,

$$\max_{\boldsymbol{\Sigma}\in\mathcal{U}_{\boldsymbol{\Sigma}}}\left(\mathbf{w}^T\bar{\mathbf{V}}\boldsymbol{\Sigma}\bar{\mathbf{V}}^T\mathbf{w}\right)=\mathbf{w}^T\bar{\mathbf{V}}\left(\bar{\boldsymbol{\Sigma}}+\varepsilon\mathbf{I}\right)\bar{\mathbf{V}}^T\mathbf{w}. \tag{34}$$

Proof: The proof is straightforward and is omitted due to space restrictions. ■

2) Matrix Norm ($\ell_{a,b}$-norm) Uncertainty: We define the matrix $\ell_{a,b}$-norm ($a\geq1$ and $b\geq1$) as

$$\|\mathbf{A}\|_{a,b}=\left(\sum_j\left(\sum_i|a_{ij}|^a\right)^{\frac{b}{a}}\right)^{\frac{1}{b}}=\left(\sum_j\|\mathbf{A}\mathbf{e}_i\|_a^b\right)^{\frac{1}{b}}. \tag{35}$$

Note that when $a=b$, we have $\|\mathbf{A}\|_{a,a}=\|\operatorname{vec}(\mathbf{A})\|_a$. Now we present the following lemma.

Lemma 9: When $\mathcal{U}_{\boldsymbol{\Sigma}}=\{\boldsymbol{\Sigma}|\|\boldsymbol{\Sigma}-\bar{\boldsymbol{\Sigma}}\|_{a,b}\leq\varepsilon\}$,

$$\begin{aligned}&\max_{\boldsymbol{\Sigma}\in\mathcal{U}_{\boldsymbol{\Sigma}}}\left(\mathbf{w}^T\bar{\mathbf{V}}\boldsymbol{\Sigma}\bar{\mathbf{V}}^T\mathbf{w}\right)\\&=\mathbf{w}^T\bar{\mathbf{V}}\bar{\boldsymbol{\Sigma}}\bar{\mathbf{V}}^T\mathbf{w}+\varepsilon\left\|\bar{\mathbf{V}}^T\mathbf{w}\right\|_{\frac{a}{a-1}}\left\|\bar{\mathbf{V}}^T\mathbf{w}\right\|_{\frac{b}{b-1}}.\end{aligned} \tag{36}$$

Proof: The proof is straightforward and is omitted due to space restrictions. ■

Remark 10: It can be observed that 1) if $a=b$, the expression $\|\bar{\mathbf{V}}^T\mathbf{w}\|_{\frac{a}{a-1}}\|\bar{\mathbf{V}}^T\mathbf{w}\|_{\frac{b}{b-1}}=\|\bar{\mathbf{V}}^T\mathbf{w}\|_{\frac{a}{a-1}}^2$ is convex in $\mathbf{w}$; if $a=b=2$, $\|\bar{\mathbf{V}}^T\mathbf{w}\|_{\frac{a}{a-1}}\|\bar{\mathbf{V}}^T\mathbf{w}\|_{\frac{b}{b-1}}=\|\bar{\mathbf{V}}^T\mathbf{w}\|_2^2=\mathbf{w}^T\bar{\mathbf{V}}\bar{\mathbf{V}}^T\mathbf{w}$, which boils down to the Schatten p-norm case.

Remark 11: Now that $a\geq1$ and $b\geq1$, the ranges of $\frac{a}{a-1}$ and $\frac{b}{b-1}$ are also $[1,+\infty)$. With a slight abuse of notation, we replace $\|\bar{\mathbf{V}}^T\mathbf{w}\|_{\frac{a}{a-1}}$ with $\|\bar{\mathbf{V}}^T\mathbf{w}\|_a$ and $\|\bar{\mathbf{V}}^T\mathbf{w}\|_{\frac{b}{b-1}}$ with $\|\bar{\mathbf{V}}^T\mathbf{w}\|_b$ for simple notation when presenting the formulation.

3) Elementwise Uncertainty: We introduce the following lemma to show the impact of elementwise uncertainty.

Lemma 12: When $\mathcal{U}_{\boldsymbol{\Sigma}}=\{\boldsymbol{\Sigma}||\Sigma_{ij}-\bar{\Sigma}_{ij}|\leq\varepsilon_{ij},\ \forall i,j\}$ where $\varepsilon_{ij}=\varepsilon_{ji}$,

$$\max_{\boldsymbol{\Sigma}\in\mathcal{U}_{\boldsymbol{\Sigma}}}\left(\mathbf{w}^T\bar{\mathbf{V}}\boldsymbol{\Sigma}\bar{\mathbf{V}}^T\mathbf{w}\right)=\mathbf{w}^T\bar{\mathbf{V}}\bar{\boldsymbol{\Sigma}}\bar{\mathbf{V}}^T\mathbf{w}+\sum_{i,j}\varepsilon_{ij}\left|\mathbf{w}_i^T\bar{\mathbf{v}}_i\right|\left|\mathbf{w}_j^T\bar{\mathbf{v}}_j\right| \tag{37}$$

(Recall that $\bar{\mathbf{V}}=\operatorname{Blkdiag}(\{\bar{\mathbf{v}}_i\})$).

Proof: The proof is straightforward and is omitted due to space restrictions. ■

C. Summary of Modified Formulations

We summarize the aforementioned modified risk expressions in Table I. We may rewrite some of the expressions for the sake of clarity.

We notice that Expressions III and IV are not convex in general. For simplicity and convenience, we only focus on the scenarios where the modified problems are convex. For Expression III, we set $a=b$ and additionally confine a to 1, 2, and $+\infty$ because these values are most commonly used. For Expression IV, we simply let $\varepsilon_{ij}=\varepsilon$ and thus $\mathbf{E}=\varepsilon\mathbf{1}\mathbf{1}^T$.

It is easy to see that Expressions I, II, and III can be unified into Expression III: when $\varepsilon=0$, Expression III becomes Expression I; when $a=b=2$, Expression III becomes Expression II. Now that in Expression IV we set $\mathbf{E}=\varepsilon\mathbf{1}\mathbf{1}^T$, we obtain

$$\begin{aligned}&\mathbf{w}^T\left(\bar{\mathbf{V}}\bar{\boldsymbol{\Sigma}}\bar{\mathbf{V}}^T+\mathbf{D}\right)\mathbf{w}+\left|\bar{\mathbf{V}}^T\mathbf{w}\right|^T\mathbf{E}\left|\bar{\mathbf{V}}^T\mathbf{w}\right|\\&=\mathbf{w}^T\left(\bar{\mathbf{V}}\bar{\boldsymbol{\Sigma}}\bar{\mathbf{V}}^T+\mathbf{D}\right)\mathbf{w}+\varepsilon\left|\bar{\mathbf{V}}^T\mathbf{w}\right|^T\mathbf{1}\mathbf{1}^T\left|\bar{\mathbf{V}}^T\mathbf{w}\right|\\&=\mathbf{w}^T\left(\bar{\mathbf{V}}\bar{\boldsymbol{\Sigma}}\bar{\mathbf{V}}^T+\mathbf{D}\right)\mathbf{w}+\varepsilon\left(\mathbf{1}^T\left|\bar{\mathbf{V}}^T\mathbf{w}\right|\right)^2\\&=\mathbf{w}^T\left(\bar{\mathbf{V}}\bar{\boldsymbol{\Sigma}}\bar{\mathbf{V}}^T+\mathbf{D}\right)\mathbf{w}+\varepsilon\left\|\bar{\mathbf{V}}^T\mathbf{w}\right\|_1^2,\end{aligned} \tag{38}$$

which means, when $a=b=1$, Expression III becomes Expression IV.

D. A Unified Formulation

To this end, we have been able to unify the four expressions in Table I into one if the aforementioned parameter settings are applied. The unified problem formulation is formally presented as follows:

$$\begin{aligned}
\underset{\mathbf{w}}{\text{minimize}} \quad & -\mathbf{w}^T\mathbf{u} + \lambda\left[\mathbf{w}^T\mathbf{A}\mathbf{w} + \varepsilon\left\|\bar{\mathbf{V}}^T\mathbf{w}\right\|_a^2\right] \\
& + \xi B\left\|(\mathbf{w}-\mathbf{w}_0)\odot\mathbf{q}\right\|_1 \\
\text{subject to} \quad & \mathbf{1}^T\mathbf{w} = 1 \\
& \mathbf{w} \geq \mathbf{0}, \qquad (39)
\end{aligned}$$

where $\mathbf{A} = \bar{\mathbf{V}}\bar{\mathbf{\Sigma}}\bar{\mathbf{V}}^T + \mathbf{D}$ and $a = 1, 2$, or $+\infty$. Note that if one seeks to maximize the Sharpe Ratio of the portfolio, the robustness discussion will stay the same.

E. Black-Litterman Extension

We consider incorporating the idea of the Black-Litterman portfolio [19], which involves financial views. The view vector $\boldsymbol{\nu}$ is modeled on the random return $\mathbf{r}$ [20]:

$$\boldsymbol{\nu} = \mathbf{P}\mathbf{r} + \mathbf{e}, \qquad (40)$$

where $\mathbf{P}$ is a view-based parameter and $\mathbf{e}$ measures error. Thus, the expected return $\mathbf{u}$ and robustified covariance $\mathbf{A}$ ($\mathbf{A} = \bar{\mathbf{V}}\bar{\mathbf{\Sigma}}\bar{\mathbf{V}}^T + \mathbf{D}$, as is defined in Sec. III-D) are modified as

$$\mathbf{u}_{\text{BL}} = \mathbf{u} + \mathbf{A}\mathbf{P}^T\left(\mathbf{P}\mathbf{A}\mathbf{P}^T + \mathbf{\Omega}\right)^{-1}(\boldsymbol{\nu} - \mathbf{P}\mathbf{u}) \qquad (41)$$

and

$$\mathbf{A}_{\text{BL}} = \mathbf{A} - \mathbf{A}\mathbf{P}^T\left(\mathbf{P}\mathbf{A}\mathbf{P}^T + \mathbf{\Omega}\right)^{-1}\mathbf{P}\mathbf{A}. \qquad (42)$$

The view-based parameter $\mathbf{P}$ is decided by relating the stocks with the corresponding options. Taking one stock S and one option C, for example, we have

$$[-S\Delta\ C]\begin{bmatrix} r_S \\ r_C \end{bmatrix} \simeq \Theta + \frac{1}{2}\Gamma\sigma^2 S^2, \qquad (43)$$

inferring from (6) and (8) ($r_S = \Delta S/S$ and $r_C = \Delta C/C$). In this case, we obtain $\mathbf{p} = [-S\Delta\ C]$ (now $\mathbf{P}$ is a row-vector) and $\nu = \Theta + \frac{1}{2}\Gamma\sigma^2 S^2$ (now $\boldsymbol{\nu}$ is a scalar). The parameter $\mathbf{\Omega}$ can be chosen as a scaled identity.

IV. Efficient Algorithms for Portfolio Optimization: An Alternative to Solvers

With problem (39) being convex, we could conveniently call an off-the-shelf solver, e.g., MOSEK [21], SeDuMi [22], or SDPT3 [23], to obtain a global optimal solution. However, it can be tedious to rewrite the problem formulation in the correct format of the solver. For that purpose, one can conveniently use a "modeling framework" to do this tedious reformulation like cvx [24], which is available for Matlab, R, Python, and Julia. This convenience comes at the cost of a higher computational cost.

While the off-the-shelf solvers can be applied, they may not support all the simulation platforms. There are many widely used online financial programming platforms, e.g., Worldquant, Quantopian, JoinQuant, Ricequant, and Uquant, and not all of them may support appropriate solvers. In the following we develop a specialized algorithm simple to implement that does not require any off-the-shelf solver. Not only such an algorithm is convenient but it is even faster than solvers (according to the simulations about one half or one order of magnitude faster).

From the constraint of (39) we find there exists a coupling linear constraint $\mathbf{1}^T\mathbf{w} = 1$, which could be tackled with the popular approach of alternating direction method of multipliers (ADMM) [25]. In this paper, we consider an advanced version of ADMM named BSUM-M [26] for algorithm design. Let us take a look at this general method.

A. BSUM-M Overview

Consider the following general convex problem:

$$\begin{aligned}
\underset{\{\mathbf{x}_k\}_{k=1}^K}{\text{minimize}} \quad & f(\{\mathbf{x}_k\}) + \sum_{k=1}^K g_k(\mathbf{x}_k) \\
\text{subject to} \quad & \sum_{k=1}^K \mathbf{H}_k\mathbf{x}_k = \mathbf{h} \\
& \mathbf{x}_k \in \mathcal{X}_k,\ k = 1, 2, \ldots, K \qquad (44)
\end{aligned}$$

where f is a smooth convex function, g_k is a nonsmooth convex function, and $\mathcal{X}_k$ is a convex set for any k. Define

$$\mathbf{x}_{-k}^{(l)} \triangleq \left(\mathbf{x}_1^{(l)}, \ldots, \mathbf{x}_{k-1}^{(l)}, \mathbf{x}_{k+1}^{(l-1)}, \ldots, \mathbf{x}_K^{(l-1)}\right). \qquad (45)$$

The BSUM-M algorithm is briefly described in the following table.

At each iteration $l \geq 1$:

$$\begin{cases}
\mathbf{y}^{(l+1)} = \mathbf{y}^{(l)} + \alpha^{(l)}\left(\mathbf{h} - \sum_{k=1}^K \mathbf{H}_k\mathbf{x}_k^{(l)}\right) \\
\mathbf{x}_k^{(l+1)} = \arg\min_{\mathbf{x}_k\in\mathcal{X}_k} u_k\left(\mathbf{x}_k; \mathbf{x}_k^{(l)}, \mathbf{x}_{-k}^{(l+1)}\right) \\
\qquad\qquad - \left[\mathbf{y}^{(l+1)}\right]^T\mathbf{H}_k\mathbf{x}_k + g_k(\mathbf{x}_k)
\end{cases} \qquad (46)$$

where $\alpha^{(l)} > 0$ is the step size for dual update and u_k is an upper bound of $f(\{\mathbf{x}_k\}) + \frac{\rho}{2}\|\mathbf{h} - \sum_{k=1}^K \mathbf{H}_k\mathbf{x}_k\|_2^2$ at a given iterate $(\mathbf{x}_k^{(l)}, \mathbf{x}_{-k}^{(l+1)})$.

Convergence: The main result of convergence of BSUM-M is elaborated in [26, Theorem 2.1]. Every limit point of $\{\{\mathbf{x}_k^{(l)}\}_{k=1}^K, \mathbf{y}^{(l)}\}$ is a primal and dual optimal solution. The upper bound function u_k must satisfy a few conditions [26, Assumption B], so that the convergence criteria are satisfied. The conditions are given as follows. We denote

$$\mathbf{H}_{-k}\mathbf{x}_{-k}^{(l)} = \sum_{k'=1}^{k-1}\mathbf{H}_{k'}\mathbf{x}_{k'}^{(l)} + \sum_{k'=k+1}^{K}\mathbf{H}_{k'}\mathbf{x}_{k'}^{(l-1)} \qquad (47)$$

and u_k must satisfy

1) $u_k(\mathbf{x}_k^{(l)}; \mathbf{x}_k^{(l)}, \mathbf{x}_{-k}^{(l+1)}) = f(\mathbf{x}_k^{(l)}, \mathbf{x}_{-k}^{(l+1)}) + \frac{\rho}{2}\|\mathbf{h} - \mathbf{H}_k\mathbf{x}_k^{(l)} - \mathbf{H}_{-k}\mathbf{x}_{-k}^{(l)}\|_2^2$, $\forall\mathbf{x}_k^{(l)}$ being feasible, $\forall k$,

2) $u_k(\mathbf{x}_k; \mathbf{x}_k^{(l)}, \mathbf{x}_{-k}^{(l+1)}) \geq f(\mathbf{x}_k, \mathbf{x}_{-k}^{(l+1)}) + \frac{\rho}{2}\|\mathbf{h} - \mathbf{H}_k\mathbf{x}_k - \mathbf{H}_{-k}\mathbf{x}_{-k}^{(l)}\|_2^2$, $\forall\mathbf{x}_k, \mathbf{x}_k^{(l)}, \mathbf{x}_{-k}^{(l+1)}$ being feasible, $\forall k$,

3) $\nabla_{\mathbf{x}_k} u_k(\mathbf{x}_k; \mathbf{x}_k^{(l)}, \mathbf{x}_{-k}^{(l+1)})|_{\mathbf{x}_k=\mathbf{x}_k^{(l)}} = \nabla_{\mathbf{x}_k}(f(\mathbf{x}_k, \mathbf{x}_{-k}^{(l+1)}) + \frac{\rho}{2}\|\mathbf{h} - \mathbf{H}_k\mathbf{x}_k - \mathbf{H}_{-k}\mathbf{x}_{-k}^{(l)}\|_2^2)|_{\mathbf{x}_k=\mathbf{x}_k^{(l)}}$, $\forall\mathbf{x}_k^{(l)}, \mathbf{x}_{-k}^{(l+1)}$ being feasible, $\forall k$,

4) $u_k(\mathbf{x}_k; \mathbf{x}_k^{(l)}, \mathbf{x}_{-k}^{(l+1)})$ is continuous in $\mathbf{x}_k$ and $(\mathbf{x}_k^{(l)}, \mathbf{x}_{-k}^{(l+1)})$ and also strongly convex in $\mathbf{x}_k$,

5) $u_k(\mathbf{x}_k; \mathbf{x}_k^{(l)}, \mathbf{x}_{-k}^{(l+1)})$ has a Lipchitz continuous gradient.

Interested readers may refer to [26] for more technical details.

B. Implementation of BSUM-M

We find the convergence conditions in [26, Theorem 2.1] are either readily satisfied or can be satisfied via doing a change of variables, so we are guaranteed to obtain the optimal solution. We observe in (39) that a could be 1, 2, or $+\infty$, so we look into each case by case in the following. We start with $a = 2$, which is the simplest case.

1) Case I, $a = 2$: When $a = 2$, the objective of (39) can be rewritten as

$$-\mathbf{w}^T\mathbf{u} + \mathbf{w}^T\mathbf{B}\mathbf{w} + \xi B \left\|(\mathbf{w} - \mathbf{w}_0) \odot \mathbf{q}\right\|_1, \tag{48}$$

where

$$\mathbf{B} = \lambda\mathbf{A} + \lambda\varepsilon\bar{\mathbf{V}}\bar{\mathbf{V}}^T. \tag{49}$$

Here $f(\mathbf{w}) = -\mathbf{w}^T\mathbf{u} + \mathbf{w}^T\mathbf{B}\mathbf{w}$ and $u_{\mathbf{w}}(\mathbf{w}; \mathbf{w}^{(l)})$ should be an upper bound of $f(\mathbf{w}) + \frac{\rho}{2}(1 - \mathbf{1}^T\mathbf{w})^2$. We introduce the following lemma to derive its upper bound function.

Lemma 13 ([27, Lemma 1]): The quadratic function $\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{p}^T\mathbf{x} + \text{const}$ ($\mathbf{P}$ is real symmetric) is upper bounded at $\mathbf{x}_0$ by

$$\lambda_{\max}(\mathbf{P})\,\mathbf{x}^T\mathbf{x} + \mathbf{r}^T\mathbf{x} + \text{const}', \tag{50}$$

where $\mathbf{r} = 2\mathbf{P}\mathbf{x}_0 - 2\lambda_{\max}(\mathbf{P})\mathbf{x}_0 + \mathbf{p}$, and this upper bound function satisfies the aforementioned convergence conditions.

By applying Lemma 13, we obtain

$$\begin{aligned} &f(\mathbf{w}) + \frac{\rho}{2}\left(1 - \mathbf{1}^T\mathbf{w}\right)^2 \\ &= \mathbf{w}^T\left(\mathbf{B} + \frac{\rho}{2}\mathbf{1}\mathbf{1}^T\right)\mathbf{w} - (\rho\mathbf{1} + \mathbf{u})^T\mathbf{w} + \frac{\rho}{2} \\ &\le \lambda_u\mathbf{w}^T\mathbf{w} + \mathbf{b}^T\mathbf{w} + \text{const}, \end{aligned} \tag{51}$$

where $\lambda_u = \lambda_{\max}(\mathbf{M}) > 0$, $\mathbf{b} = 2\mathbf{M}\mathbf{w}^{(l)} - 2\lambda_u\mathbf{w}^{(l)} - \rho\mathbf{1} - \mathbf{u}$, and $\mathbf{M} = \mathbf{B} + \frac{\rho}{2}\mathbf{1}\mathbf{1}^T$. So we set $u_{\mathbf{w}}(\mathbf{w}; \mathbf{w}^{(l)}) = \lambda_u\mathbf{w}^T\mathbf{w} + \mathbf{b}^T\mathbf{w} + \text{const}$.

Therefore, the update steps of BSUM-M are

$$\nu^{(l+1)} = \nu^{(l)} + \alpha^{(l)}\left(1 - \mathbf{1}^T\mathbf{w}^{(l)}\right) \tag{52}$$

and

$$\begin{aligned} \mathbf{w}^{(l+1)} = \; &\arg\min_{\mathbf{w}\ge\mathbf{0}} \lambda_u\mathbf{w}^T\mathbf{w} + \mathbf{b}^T\mathbf{w} \\ &- \nu^{(l+1)}\mathbf{1}^T\mathbf{w} + \xi B\left\|(\mathbf{w} - \mathbf{w}_0) \odot \mathbf{q}\right\|_1, \end{aligned} \tag{53}$$

i.e., $\forall k$,

$$\begin{aligned} w_k^{(l+1)} = \; &\arg\min_{w_k\ge 0} \lambda_u w_k^2 + \left(b_k - \nu^{(l+1)}\right)w_k \\ &+ \xi B q_k\left|w_k - w_{0,k}\right|. \end{aligned} \tag{54}$$

We introduce the following lemma to give the closed form solution of $w_k^{(l+1)}$.

Lemma 14: There holds

$$\begin{aligned} &\arg\min_{x\ge 0} \alpha x^2 + \beta x + \gamma\left|x - x_0\right| \\ &= \begin{cases} \left[\frac{\gamma-\beta}{2\alpha}\right]_+ & x_0 \ge \frac{\gamma-\beta}{2\alpha} \\ -\frac{\gamma+\beta}{2\alpha} & x_0 \le -\frac{\gamma+\beta}{2\alpha} \\ x_0 & -\frac{\gamma+\beta}{2\alpha} < x_0 < \frac{\gamma-\beta}{2\alpha} \end{cases} \\ &\triangleq \left[\left[x_0\right]_{-\frac{\gamma+\beta}{2\alpha}}^{\frac{\gamma-\beta}{2\alpha}}\right]_+, \end{aligned} \tag{55}$$

where $\alpha > 0$, $\gamma > 0$, and $x_0 \ge 0$.

Proof: The proof is straightforward and is omitted due to space restrictions. ■

Thus, we can compactly express $w_k^{(l+1)}$ as

$$w_k^{(l+1)} = \left[\left[w_{0,k}\right]_{-\frac{\xi B q_k + b_k - \nu^{(l+1)}}{2\lambda_u}}^{\frac{\xi B q_k - b_k + \nu^{(l+1)}}{2\lambda_u}}\right]_+. \tag{56}$$

2) Case II, $a = 1$: Now we look into a slightly more involved case. When $a = 1$, we rewrite (39) as

$$\begin{aligned} \underset{\mathbf{w},\mathbf{t},\mathbf{s}_1,\mathbf{s}_2}{\text{minimize}} \quad & -\mathbf{w}^T\mathbf{u} + \mathbf{w}^T(\lambda\mathbf{A})\mathbf{w} + \lambda\varepsilon\left(\mathbf{1}^T\mathbf{t}\right)^2 \\ & + \xi B\left\|(\mathbf{w} - \mathbf{w}_0) \odot \mathbf{q}\right\|_1 \\ \text{subject to} \quad & \bar{\mathbf{V}}^T\mathbf{w} + \mathbf{s}_1 = \mathbf{t},\ \mathbf{s}_1 \ge \mathbf{0} \\ & \bar{\mathbf{V}}^T\mathbf{w} - \mathbf{s}_2 = -\mathbf{t},\ \mathbf{s}_2 \ge \mathbf{0} \\ & \mathbf{1}^T\mathbf{w} = 1 \\ & \mathbf{w} \ge \mathbf{0}. \end{aligned} \tag{57}$$

Here $f(\mathbf{w}, \mathbf{t}, \mathbf{s}_1, \mathbf{s}_2) = -\mathbf{w}^T\mathbf{u} + \mathbf{w}^T(\lambda\mathbf{A})\mathbf{w} + \lambda\varepsilon(\mathbf{1}^T\mathbf{t})^2$ and we need to design three upper bound functions: $u_{\mathbf{w}}$, $u_{\mathbf{t}}$, and $u_{\mathbf{s}}$ (i.e., $u_{(\mathbf{s}_1,\mathbf{s}_2)}$) for BSUM-M implementation. We introduce the following lemma to obtain the upper bound functions.

Lemma 15: The upper bound functions for $f(\mathbf{w}, \mathbf{t}, \mathbf{s}_1, \mathbf{s}_2) + \frac{\rho}{2}[\|\bar{\mathbf{V}}^T\mathbf{w} + \mathbf{s}_1 - \mathbf{t}\|_2^2 + \|\bar{\mathbf{V}}^T\mathbf{w} - \mathbf{s}_2 + \mathbf{t}\|_2^2 + (1 - \mathbf{1}^T\mathbf{w})^2]$ are given as follows:

$$\begin{cases} u_{\mathbf{w}}\left(\mathbf{w}; \mathbf{w}^{(l)}, \mathbf{t}^{(l)}, \mathbf{s}_1^{(l)}, \mathbf{s}_2^{(l)}\right) \\ \qquad = \lambda_{\max}(\mathbf{M}) \cdot \mathbf{w}^T\mathbf{w} + \mathbf{b}^T\mathbf{w} + \text{const} \\ u_{\mathbf{t}}\left(\mathbf{t}; \mathbf{w}^{(l+1)}, \mathbf{t}^{(l)}, \mathbf{s}_1^{(l)}, \mathbf{s}_2^{(l)}\right) \\ \qquad = \mathbf{t}^T\left(\rho\mathbf{I} + \lambda\varepsilon\mathbf{1}\mathbf{1}^T\right)\mathbf{t} - \rho\left(\mathbf{s}_1^{(l)} + \mathbf{s}_2^{(l)}\right)^T\mathbf{t} + \text{const} \\ u_{\mathbf{s}}\left(\mathbf{s}_1, \mathbf{s}_2; \mathbf{w}^{(l+1)}, \mathbf{t}^{(l+1)}, \mathbf{s}_1^{(l)}, \mathbf{s}_2^{(l)}\right) \\ \qquad = \frac{\rho}{2}\mathbf{s}_1^T\mathbf{s}_1 + \rho\left(\bar{\mathbf{V}}^T\mathbf{w}^{(l+1)} - \mathbf{t}^{(l+1)}\right)^T\mathbf{s}_1 + \frac{\rho}{2}\mathbf{s}_2^T\mathbf{s}_2 \\ \qquad\quad - \rho\left(\bar{\mathbf{V}}^T\mathbf{w}^{(l+1)} + \mathbf{t}^{(l+1)}\right)^T\mathbf{s}_2 + \text{const}, \end{cases} \tag{58}$$

where

$$\mathbf{M} = \lambda\mathbf{A} + \rho\bar{\mathbf{V}}\bar{\mathbf{V}}^T + \frac{\rho}{2}\mathbf{1}\mathbf{1}^T \tag{59}$$

and

$$\mathbf{b} = 2\mathbf{M}\mathbf{w}^{(l)} - 2\lambda_u\mathbf{w}^{(l)} + \rho\bar{\mathbf{V}}\left(\mathbf{s}_1^{(l)} - \mathbf{s}_2^{(l)}\right) - \rho\mathbf{1} - \mathbf{u}. \tag{60}$$

All of them satisfy the aforementioned convergence conditions.

Proof: The proof is straightforward and is omitted due to space restrictions. ■

Therefore, the update steps of BSUM-M are

$$\begin{cases} \mathbf{z}_1^{(l+1)} = \mathbf{z}_1^{(l)} + \alpha^{(l)} \left(\mathbf{t}^{(l)} - \bar{\mathbf{V}}^T \mathbf{w}^{(l)} - \mathbf{s}_1^{(l)} \right) \\ \mathbf{z}_2^{(l+1)} = \mathbf{z}_2^{(l)} + \alpha^{(l)} \left(-\mathbf{t}^{(l)} - \bar{\mathbf{V}}^T \mathbf{w}^{(l)} + \mathbf{s}_2^{(l)} \right) \\ \nu^{(l+1)} = \nu^{(l)} + \alpha^{(l)} \left(1 - \mathbf{1}^T \mathbf{w}^{(l)} \right), \end{cases} \tag{61}$$

the update of $\mathbf{w}$ follows (56),

$$\begin{aligned} \mathbf{t}^{(l+1)} &= \arg\min \mathbf{t}^T \left(\rho \mathbf{I} + \lambda\varepsilon \mathbf{1}\mathbf{1}^T \right) \mathbf{t} - \rho \left(\mathbf{s}_1^{(l)} + \mathbf{s}_2^{(l)} \right)^T \mathbf{t} \\ &\quad + \left[\mathbf{z}_1^{(l+1)} - \mathbf{z}_2^{(l+1)} \right]^T \mathbf{t} \\ &= \frac{1}{2} \left(\rho \mathbf{I} + \lambda\varepsilon \mathbf{1}\mathbf{1}^T \right)^{-1} \left[\rho \left(\mathbf{s}_1^{(l)} + \mathbf{s}_2^{(l)} \right) - \mathbf{z}_1^{(l+1)} + \mathbf{z}_2^{(l+1)} \right] \\ &= \frac{1}{2} \left(\rho^{-1} \mathbf{I} - \frac{\lambda\varepsilon}{\rho^2 + \rho\lambda\varepsilon \mathbf{1}^T \mathbf{1}} \mathbf{1}\mathbf{1}^T \right) \cdot \\ &\quad \left[\rho \left(\mathbf{s}_1^{(l)} + \mathbf{s}_2^{(l)} \right) - \mathbf{z}_1^{(l+1)} + \mathbf{z}_2^{(l+1)} \right], \end{aligned} \tag{62}$$

and

$$\begin{aligned} &\left(\mathbf{s}_1^{(l+1)}, \mathbf{s}_2^{(l+1)} \right) \\ &= \arg \min_{\mathbf{s}_1, \mathbf{s}_2 \geq \mathbf{0}} \frac{\rho}{2} \mathbf{s}_1^T \mathbf{s}_1 + \rho \left(\bar{\mathbf{V}}^T \mathbf{w}^{(l+1)} - \mathbf{t}^{(l+1)} \right)^T \mathbf{s}_1 \\ &\quad + \frac{\rho}{2} \mathbf{s}_2^T \mathbf{s}_2 - \rho \left(\bar{\mathbf{V}}^T \mathbf{w}^{(l+1)} + \mathbf{t}^{(l+1)} \right)^T \mathbf{s}_2 \\ &\quad - \left[\mathbf{z}_1^{(l+1)} \right]^T \mathbf{s}_1 + \left[\mathbf{z}_2^{(l+1)} \right]^T \mathbf{s}_2 \\ &= \left(\left[-\bar{\mathbf{V}}^T \mathbf{w}^{(l+1)} + \mathbf{t}^{(l+1)} + \mathbf{z}_1^{(l+1)} / \rho \right]_+ , \right. \\ &\left. \left[\bar{\mathbf{V}}^T \mathbf{w}^{(l+1)} + \mathbf{t}^{(l+1)} - \mathbf{z}_2^{(l+1)} / \rho \right]_+ \right). \end{aligned} \tag{63}$$

3) Case III, $a = +\infty$: When $a = +\infty$, problem (39) is reformulated as

$$\begin{aligned} \underset{\mathbf{w}, t, \mathbf{s}_1, \mathbf{s}_2}{\text{minimize}} \quad & -\mathbf{w}^T \mathbf{u} + \mathbf{w}^T (\lambda \mathbf{A}) \mathbf{w} + \lambda\varepsilon t^2 \\ & + \xi B \left\| (\mathbf{w} - \mathbf{w}_0) \odot \mathbf{q} \right\|_1 \\ \text{subject to} \quad & \bar{\mathbf{V}}^T \mathbf{w} + \mathbf{s}_1 = t\mathbf{1}, \ \mathbf{s}_1 \geq \mathbf{0} \\ & \bar{\mathbf{V}}^T \mathbf{w} - \mathbf{s}_2 = -t\mathbf{1}, \ \mathbf{s}_2 \geq \mathbf{0} \\ & \mathbf{1}^T \mathbf{w} = 1 \\ & \mathbf{w} \geq \mathbf{0}, \end{aligned} \tag{64}$$

which is very similar to (64), so we present the update steps of BSUM-M directly as follows:

$$\begin{cases} \mathbf{z}_1^{(l+1)} = \mathbf{z}_1^{(l)} + \alpha^{(l)} \left(t^{(l)} \mathbf{1} - \bar{\mathbf{V}}^T \mathbf{w}^{(l)} - \mathbf{s}_1^{(l)} \right) \\ \mathbf{z}_2^{(l+1)} = \mathbf{z}_2^{(l)} + \alpha^{(l)} \left(-t^{(l)} \mathbf{1} - \bar{\mathbf{V}}^T \mathbf{w}^{(l)} + \mathbf{s}_2^{(l)} \right) \\ \nu^{(l+1)} = \nu^{(l)} + \alpha^{(l)} \left(1 - \mathbf{1}^T \mathbf{w}^{(l)} \right), \end{cases} \tag{65}$$

Algorithm 1: BSUM-M-based Portfolio Optimization Algorithm, $a = 2$.

Require: Initialization: $\nu^{(0)}$, $\mathbf{w}^{(0)}$, $l = 0$, ρ;
1: $\mathbf{B} = \lambda \mathbf{A} + \lambda\varepsilon \bar{\mathbf{V}} \bar{\mathbf{V}}^T$;
2: $\mathbf{M} = \mathbf{B} + \frac{\rho}{2} \mathbf{1}\mathbf{1}^T$;
3: $\lambda_u = \lambda_{\max}(\mathbf{M})$;
4: **repeat**
5: $\nu^{(l+1)} = \nu^{(l)} + \alpha^{(l)} (1 - \mathbf{1}^T \mathbf{w}^{(l)})$;
6: $\mathbf{b} = 2\mathbf{M}\mathbf{w}^{(l)} - 2\lambda_u \mathbf{w}^{(l)} - \rho \mathbf{1} - \mathbf{u}$;
7: Compute $w_k^{(l+1)}$ using (56), $\forall k$;
8: $l = l + 1$;
9: **until** convergence

the update of $\mathbf{w}$ follows (56),

$$\begin{aligned} & t^{(l+1)} \\ &= \arg\min \left(\rho \mathbf{1}^T \mathbf{1} + \lambda\varepsilon \right) t^2 - \rho \left(\mathbf{s}_1^{(l)} + \mathbf{s}_2^{(l)} \right)^T \mathbf{1} t \\ &\quad + \left[\mathbf{z}_1^{(l+1)} - \mathbf{z}_2^{(l+1)} \right]^T \mathbf{1} t \\ &= \frac{1}{2} \left(\rho \mathbf{1}^T \mathbf{1} + \lambda\varepsilon \right)^{-1} \left[\rho \left(\mathbf{s}_1^{(l)} + \mathbf{s}_2^{(l)} \right) - \mathbf{z}_1^{(l+1)} + \mathbf{z}_2^{(l+1)} \right]^T \mathbf{1}, \end{aligned} \tag{66}$$

and

$$\begin{aligned} &\left(\mathbf{s}_1^{(l+1)}, \mathbf{s}_2^{(l+1)} \right) \\ &= \arg \min_{\mathbf{s}_1, \mathbf{s}_2 \geq \mathbf{0}} \frac{\rho}{2} \mathbf{s}_1^T \mathbf{s}_1 + \rho \left(\bar{\mathbf{V}}^T \mathbf{w}^{(l+1)} - t^{(l+1)} \mathbf{1} \right)^T \mathbf{s}_1 \\ &\quad + \frac{\rho}{2} \mathbf{s}_2^T \mathbf{s}_2 - \rho \left(\bar{\mathbf{V}}^T \mathbf{w}^{(l+1)} + t^{(l+1)} \mathbf{1} \right)^T \mathbf{s}_2 \\ &\quad - \left[\mathbf{z}_1^{(l+1)} \right]^T \mathbf{s}_1 + \left[\mathbf{z}_2^{(l+1)} \right]^T \mathbf{s}_2 \\ &= \left(\left[-\bar{\mathbf{V}}^T \mathbf{w}^{(l+1)} + t^{(l+1)} \mathbf{1} + \mathbf{z}_1^{(l+1)} / \rho \right]_+ , \right. \\ &\left. \left[\bar{\mathbf{V}}^T \mathbf{w}^{(l+1)} + t^{(l+1)} \mathbf{1} - \mathbf{z}_2^{(l+1)} / \rho \right]_+ \right). \end{aligned} \tag{67}$$

We summarize the efficient algorithms for portfolio optimization in Algorithms 1 and 2. Some steps may be reorganized for code efficiency.

C. Computational Complexity

We present an analysis on the computational complexity of Algorithms 1 and 2 in this subsection and later in section V we will show numerical results. The analysis is done on a per-iteration basis. First we study Algorithm 1. Each iteration consists of three steps: updating ν, $\mathbf{b}$, and $\mathbf{w}$ sequentially. Recall that the length of $\mathbf{w}$ is $I + \sum_{i=1}^{I} (M_i + N_i) = N$. The most costly step in the update of ν is the summation of $\mathbf{w}^{(l)}$, of complexity $\mathcal{O}(N)$. The most costly step in the update of $\mathbf{b}$ is the matrix multiplication $\mathbf{M}\mathbf{w}^{(l)}$ ($\mathbf{M} \in \mathbb{R}^{N \times N}$), of complexity $\mathcal{O}(N^2)$. The computation of $\mathbf{w}^{(l+1)}$ is by nature elementwise. According to Lemma 14, updating one element of $\mathbf{w}^{(l+1)}$ needs

Algorithm 2: BSUM-M-based Portfolio Optimization Algorithm, $a = 1$ or $+\infty$.

Require: Initialization: $\nu^{(0)}$, $\mathbf{w}^{(0)}$, $l = 0$, ρ;
$\mathbf{M} = \lambda\mathbf{A} + \rho\bar{\mathbf{V}}\bar{\mathbf{V}}^T + \frac{\rho}{2}\mathbf{1}\mathbf{1}^T$;
$\lambda_u = (\mathbf{M})$;
repeat
  Dual update: if $a = 1$, cf. (61); if $a = +\infty$, cf. (65);
  $\mathbf{b} = 2\mathbf{M}\mathbf{w}^{(l)} - 2\lambda_u\mathbf{w}^{(l)} + \rho\bar{\mathbf{V}}(\mathbf{s}_1^{(l)} - \mathbf{s}_2^{(l)}) - \rho\mathbf{1} - \mathbf{u}$;
  Compute $w_k^{(l+1)}$ using (56), $\forall k$;
  Update $\mathbf{t}$ or t: if $a = 1$, cf. (62); if $a = +\infty$, cf. (66);
  Update $\mathbf{s}_1, \mathbf{s}_2$: if $a = 1$, cf. (63); if $a = +\infty$, cf. (67);
  $l = l + 1$;
until convergence

$\mathcal{O}(1)$, thus in total $\mathcal{O}(N)$. Therefore, the overall computational complexity of Algorithm 1 per iteration is $\mathcal{O}(N^2)$, neglecting lower-order terms.

Next we look into Algorithm 2. Each iteration consists of five steps: updating $(\mathbf{z}_1, \mathbf{z}_2, \nu)$, $\mathbf{b}$, $\mathbf{w}$, $\mathbf{t}$ or t, and $(\mathbf{s}_1, \mathbf{s}_2)$ sequentially. The most costly step in the update of $(\mathbf{z}_1, \mathbf{z}_2)$ is the matrix multiplication $\bar{\mathbf{V}}^T\mathbf{w}^{(l)}$ ($\bar{\mathbf{V}} \in \mathbb{R}^{N\times I}$) whether $a = 1$ or $+\infty$, of complexity $\mathcal{O}(NI)$. The complexity of updating ν is $\mathcal{O}(N)$. The complexity of updating $\mathbf{b}$ is $\mathcal{O}(N^2 + NI)$ due to $\mathbf{M}\mathbf{w}^{(l)}$ and $\bar{\mathbf{V}}(\mathbf{s}_1^{(l)} - \mathbf{s}_2^{(l)})$. The computation of $\mathbf{w}^{(l+1)}$ is still $\mathcal{O}(N)$. When $a = 1$, the update of $\mathbf{t}$ requires $\mathcal{O}(I^2)$ (an $I \times I$ matrix multiplies an $I \times 1$ vector); and when $a = +\infty$, updating t only requires $\mathcal{O}(I)$ (summation of a length-I vector). Finally, the update of $(\mathbf{s}_1, \mathbf{s}_2)$ requires $\mathcal{O}(NI)$ as well, due to $\bar{\mathbf{V}}^T\mathbf{w}^{(l+1)}$, whether $a = 1$ or $+\infty$. In our problem, $N = I + \sum_{i=1}^{I}(M_i + N_i) \gg I$. Therefore, the overall computational complexity of Algorithm 2 per iteration is $\mathcal{O}(N^2)$, neglecting non-dominant terms.

V. Numerical Simulations

In this section, we present numerical results on both synthetic experiments (algorithmic performance) and real-data experiments (financial performance). All simulations are performed on a PC with a 3.20 GHz i5-4570 CPU and 8 GB RAM. The off-the-shelf solver is specified as MOSEK built in the CVX toolbox or MOSEK directly. If the proposed algorithm (Algorithm 1 or 2) is applied, we set by default $\rho = 1$, $\alpha^{(l)} = \rho/\sqrt{l}$, and the termination condition is $\|\mathbf{w}^{(l+1)} - \mathbf{w}^{(l)}\|_2/\|\mathbf{w}^{(l)}\|_2 < 10^{-8}$ or the number of iterations reaches 5×10^4 unless otherwise specified.

A. Synthetic Experiments

We study the algorithmic performance in this subsection. We compare the performance of different methods to solve (39). The experiment settings are as follows. We set $\lambda = 1$, $\xi = 1$, and $\varepsilon = 0.01$. We randomly generate $\mathbf{u}$ and $\bar{\mathbf{V}}$ according to their length or size. We also randomly generate $\mathbf{A}$ as positive semidefinite, $\mathbf{q}$ as elementwise positive, and $\mathbf{w}_0$ as a feasible point on the simplex. When we run Algorithm 1 or 2, we randomly initialize the dual variables: ν, $\mathbf{z}_1$, and $\mathbf{z}_2$ if applicable; we randomly initialize the primal variables as well, but we additionally make them primal feasible: $\mathbf{w}$, $\mathbf{t}$ or t, $\mathbf{s}_1$, and $\mathbf{s}_2$ if applicable. We regard the built-in Matlab solver fmincon as a benchmark.

1) Single Realization Scenario: We set the number of stocks to 10 ($I = 10$) and the combined number of stocks and options to 100 ($N = 100$ and the length of $\mathbf{w}$ is also 100) and run all the computational methods once. We present the convergence property of Algorithm 1 or 2 in Fig. 1. It can be observed that although the objective does not decrease monotonically in all iterations, the first ten iterations seem to be monotonic. The converged objective always equals the optimal value computed by CVX, and is always lower than the optimal value computed by fmincon. It indicates that the solution computed by fmincon is not optimal and thus fmincon does not give reliable results when solving (39). We should use the solver MOSEK (whether called via CVX or directly) or the proposed algorithm instead.

Next, we take a look at the computational time of different methods. The result is presented in Table II. It can be seen that all four methods take very little time to finish the computation, among which the proposed algorithm and MOSEK are the fastest, almost one order of magnitude faster than the slowest ones. However, one realization is not sufficient to draw a conclusion, so we move on to the next experiment.

2) Multiple Realization Scenario: Now we set $I = N/10$ and vary N in $\{50, 100, 150, \ldots, 450, 500\}$. For a given tuple of N and I, we run each method 100 times and the reported performance is averaged over the 100 instances. We already know from the previous experiment that the optimized objective of fmincon is always higher than that of MOSEK and the proposed algorithm, so we do not have to reproduce this result. Our focus is shifted to the difference in solutions between the proposed algorithm and MOSEK. It is observed that the gap between the converged objective of the proposed algorithm and the optimal value of MOSEK (whether called via CVX or directly) is negligible, so we study the distance between the solutions instead. We call MOSEK via CVX. Solution distance is defined as follows:

$$\text{Solution Distance} = \left\|\mathbf{w}^\star_{\text{MOSEK}} - \mathbf{w}^\star_{\text{Proposed}}\right\|_\infty . \tag{68}$$

Apart from solution distance, we are also concerned with the computational time of different methods.

In the simulation that follows, we will present the results of the aforementioned two performance metrics. First we present the solution distance between the proposed algorithm and MOSEK in Fig. 2. The solution distance is generally low for all problem sizes, smaller than 3×10^{-3}. When $a = 1$ or 2, the solution distance stays steady; when $a = +\infty$, the solution distance displays a seemingly decreasing trend when $N \leq 350$, and after that, fluctuates in the range $[4, 6] \times 10^{-5}$. Next, let us take a look at computational time, which is displayed in Fig. 3. Among the four methods, fmincon is always the slowest. MOSEK called via CVX is the second slowest method because it spends extra time on problem reformulation before calling the solver. The proposed algorithm seems to be the fastest method, even faster than MOSEK when called directly. When $a = 1$ or $+\infty$, the proposed method is about half an order of magnitude faster than MOSEK (called directly), and when $a = 2$, the proposed method is about one order of magnitude faster than MOSEK (called directly). Taking a global view, we can find that both MOSEK and the proposed method can provide a solution within 1 second, which is efficient enough for financial applications, so both methods are recommended.

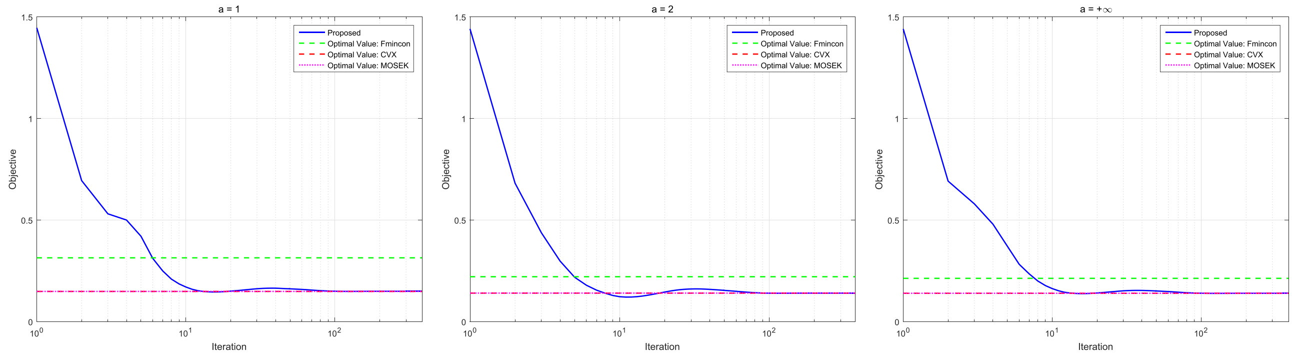


Fig. 1. Convergence plot: objective versus iteration, $N = 100$, $I = 10$; $a = 1$ (left), $a = 2$ (middle), and $a = +\infty$ (right).

TABLE II
COMPUTATIONAL TIME (SEC) OF DIFFERENT METHODS (ONE REALIZATION), $N = 100$, $I = 10$

	$a=1$	$a=2$	$a=+\infty$
Fmincon	0.2557	0.3547	0.2804
CVX	0.2226	0.1629	0.2320
MOSEK	0.0363	0.0353	0.0331
Proposed Algorithm	0.0284	0.0263	0.0306

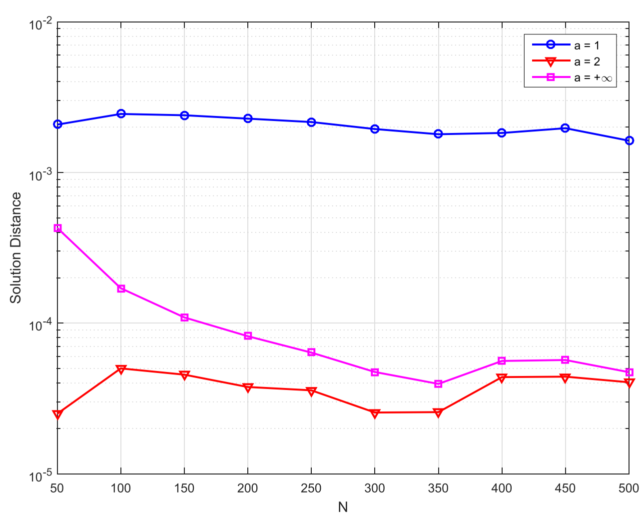


Fig. 2. Solution distance versus problem size N.

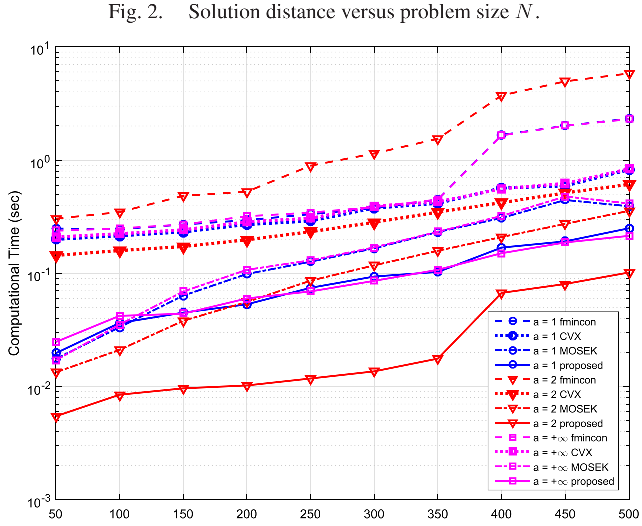


Fig. 3. Computational time versus problem size N.

B. Real-Data Experiments

We move on to real data experiments. We obtain stock and option data from the Bloomberg terminal. The trading period goes from Nov. 18, 2014 to Jan. 20, 2017. We choose 6 stocks as underlying assets, namely AAPL, BRKB, FB, MSFT, WMT, and XOM; we also choose 128 liquid options (strike price close to stock price) for portfolio design, namely AAPL: 14 calls, 14 puts; BRKB: 11 calls, 11 puts; FB: 18 calls, 18 puts; MSFT: 7 calls, 7 puts; WMT: 6 calls, 6 puts; XOM: 8 calls, 8 puts. Note that one stock can be the underlying of many options because there are many different strike prices and expiration dates to choose from. The total budget B is 1 dollar. We conduct daily rebalancing in the simulation and on each trading day, the investment is fixed to be $B = 1$. In order to evaluate the performance of trading, we adopt the following performance metrics:

$$1.\ \text{Cumulative PnL}\,(t) = \sum_{i=1}^{t}\Big[\text{PnL}_i - \underbrace{B\,\|(\mathbf{w}_i - \mathbf{w}_{i-1})\odot\mathbf{q}\|_1}_{\text{transaction cost}}\Big]$$

where $\text{PnL}_i = B\mathbf{w}_{i-1}^T\mathbf{r}_i$ and $\mathbf{r}_i$ denotes the return on the ith day,

$$2.\ \text{Sharpe Ratio (annual)} = \frac{\text{mean}\,(\{\text{PnL}_i\})}{\text{std}\,(\{\text{PnL}_i\})}\times\sqrt{252},$$

3. Drawdown (t) (in percentage)

$$= \frac{\min\left(\text{PnL}_t - \max_{i=1,\ldots,t}\text{PnL}_i,\ 0\right)}{\max_{i=1,\ldots,t}\text{PnL}_i},$$

$$4.\ \text{ROT} = \frac{\text{Cumulative PnL}\,(T)}{\sum_{i=1}^{T} B\,\|\mathbf{w}_i - \mathbf{w}_{i-1}\|_1}/10^{-4}. \quad (69)$$

We will adopt three additional metrics on portfolio return. We define portfolio return at time t as $x_t = \text{PnL}_t/$ Cumulative PnL $(t-1)$. These three metrics are

$$5.\text{VaR}_\alpha\,(t) = \min\left\{x_{t,0}\,\middle|\,\Pr\,(x_t \le x_{t,0}) \ge \alpha\right\},\ \alpha = 0.99,$$

$$6.\ \text{ES (Expected Shortfall, a.k.a CVaR)} = \mathsf{E}\left[x_t\,\middle|\,x_t \ge \text{VaR}_\alpha(t)\right],$$

$$7.\ \text{CE (Certainty Equivalent)} = U^{-1}\left(\mathsf{E}\,[U\,(x_t)]\right),$$

$$U\,(a) = \log\,(1+a)\,,\ U^{-1}\,(a) = e^a - 1. \quad (70)$$

We choose the Sharpe ratio as our major performance metric, i.e., we want this quantity as large as possible. The other metrics are minor metrics; we present them merely as a reference. As for the tuning parameter $(\lambda, \varepsilon, \xi)$, we choose them via cross validation. We use data from the last 20 trading days to form the training and validation set–80 percent for training and 20 percent for validation. The estimation of stock return mean and covariance matrix is based on the training set. The covariance matrix is simply computed as the sample covariance matrix, but note that many other more robust and better methods exists [28], [29]. We select the pair $(\lambda, \varepsilon, \xi)$ achieving the highest Sharpe ratio. Since we only include 6 stocks, a training period of $20 \times 0.8 = 16$ days[3] is enough since $T/I = 16/6 \simeq 3$.

1) Including Options is Worthwhile: First we show that including options in the portfolio design is worthwhile. We compare the performance of including and excluding the options. The benchmark is constructing a portfolio with stocks only (the aforementioned 6 stocks), either allowing or forbidding short selling. The formulation is the traditional mean-variance trade-off with transaction cost penalty[4]:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & -\mathbf{w}^T\mathbf{u} + \lambda \mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} \\ & + \xi B \left\|(\mathbf{w} - \mathbf{w}_0) \odot \mathbf{q}\right\|_1 \\ \text{subject to} \quad & \begin{cases} \mathbf{1}^T\mathbf{w} \le 1, \ \mathbf{w} \ge \mathbf{0}. & \text{(long only)} \\ \text{or} & \\ \|\mathbf{w}\|_1 \le 1. & \text{(short selling allowed)} \end{cases} \end{aligned} \tag{71}$$

What we propose is including different call and put options, which is solving (25). We present the result in Fig. 4 and Table III. As can be seen in the figure and table, the portfolio containing both stocks and options is much more profitable than that with pure stocks. The cumulative PnL reaches close to 8. The Sharpe ratio is around twice as much as trading stocks (long only) and ten times as much as trading stocks (short selling allowed). This is mainly because 1) implementing a risk-hedging trading strategy, investors can expect to obtain a higher return using options than only stocks and 2) the transaction cost of trading the same amount (of shares) of options and stocks is almost the same. However, in terms of drawdown, including options improves the performance slightly but not as much as one would desire. This could result from the aforementioned "zero risk" fallacy in risk modeling. Additionally, we can observe that when short selling is allowed in stock trading, the cumulative PnL is even lower. This indicates the potential risk of short selling. For risk-averse investors, imposing a long-only constraint is preferred. We can also observe that including options may cause a larger potential loss judging from the metrics of VaR and ES, although a much larger CE is achieved meanwhile.

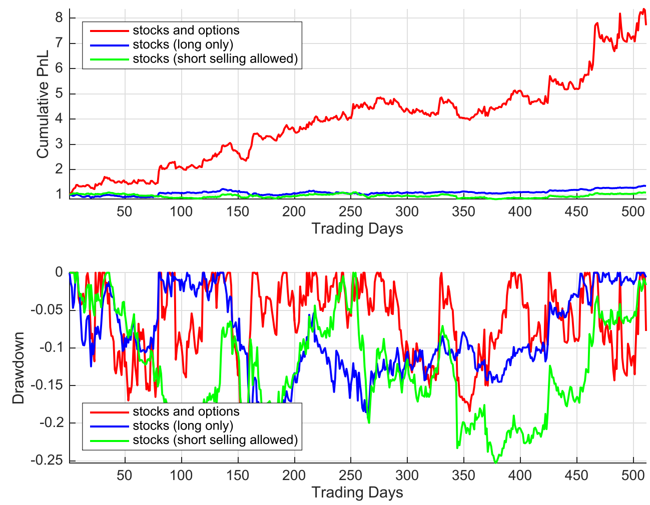


Fig. 4. Performance evaluation (cumulative PnL and drawdown versus trading days) of stock-only and stock plus options.

TABLE III
PERFORMANCE EVALUATION (SHARPE RATIO, ROT, VAR, ES, AND CE (END OF TRADE)) OF STOCK-ONLY AND STOCK PLUS OPTIONS

Portfolio Composition	Sharpe ratio	ROT	VaR
Stocks (long only)	0.7253	11.2199	-0.0385
Stocks (short selling allowed)	0.1761	1.9404	-0.0449
Stocks and options	1.5504	111.0176	-0.0645
Portfolio Composition	ES	CE	
Stocks (long only)	-0.0423	0.00059	
Stocks (short selling allowed)	-0.0609	0.00016	
Stocks and options	-0.0806	0.0040	

2) Robustness is Necessary: The zero-risk fallacy was mentioned in Sec. III and we want to illustrate the importance of introducing robustness. To simplify the comparison, we only impose stochastic robustness, which is solving (39) with $\varepsilon = 0$. Recall that we need to specify $\mathbf{D}$, which is $\text{Blkdiag}(\{\Sigma_{ii}\mathbf{R}_i\})$. We specify $\mathbf{R}_i$ as $\text{Diag}(\mathbf{v}_i^2)$ (elementwise squared, $\mathbf{v}_i$ is defined after (19)).[5] The parameter pair (λ, ξ) is still obtained via cross validation. We present the simulation result in Fig. 5 and Table IV. As can be seen in the figure and table, the cumulative PnL curve obtained from solving the nonrobust formulation (25) is higher but more volatile; it has a lower Sharpe ratio and ROT, a larger potential loss (lower VaR and ES), though a slightly higher CE. Its drawdown is also larger on most trading days. When stochastic robustness is imposed, we achieve a more stable cumulative PnL curve. Note that we have not realized the full potential of the robust formulation (39). We still can introduce the worst-case robustness as well. We will look into this in the next subsection.

3) Factor Model: Factor-model-based methods are also worth trying to modify the covariance after we introduce stochastic robustness. For a given covariance matrix $\mathbf{M} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^T + \mathbf{D}$, we perform eigenvalue decomposition $\mathbf{M} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ with $\boldsymbol{\Lambda} = \text{Diag}(\lambda)$ and $\lambda_1 \ge \lambda_2 \ge \cdots$. We keep the largest K values in λ. The remaining values naturally form a

[3]For sanity check, we also tried longer periods like 40 or 80 days and we obtained similar results.

[4]As can be found in [30], if short selling is allowed, the general constraint should be $1/\text{Lev}_{\text{long}} \times \mathbf{1}^T \max(\mathbf{w}, \mathbf{0}) + 1/\text{Lev}_{\text{short}} \times \mathbf{1}^T \max(-\mathbf{w}, \mathbf{0}) \le 1$, where Lev_{long} and $\text{Lev}_{\text{short}}$ are the leverage parameters for a long and short position, respectively. (Note that $1/\text{Lev}_{\text{short}}$ is the margin requirement for shorting.) For a long position, one is allowed to borrow $(\text{Lev}_{\text{long}} - 1)/\text{Lev}_{\text{long}}$ of the value of the position from the broker, so one needs to provide $1/\text{Lev}_{\text{long}} \times \mathbf{1}^T \max(\mathbf{w}, \mathbf{0})$ as a long budget; for a short position, one is required to have at least $1/\text{Lev}_{\text{short}}$ of the short value as the initial margin to establish the short position, so one needs to provide $1/\text{Lev}_{\text{short}} \times \mathbf{1}^T \max(-\mathbf{w}, \mathbf{0})$ as a short budget. If we set $\text{Lev}_{\text{long}} = \text{Lev}_{\text{short}} = \text{Lev}$, then the general constraint becomes $\|\mathbf{w}\|_1 \le \text{Lev}$. If we additionally set $\text{Lev} = 1$, then the constraint becomes $\|\mathbf{w}\|_1 \le 1$.

[5]This is a heuristic choice that is simple and we found to perform well.

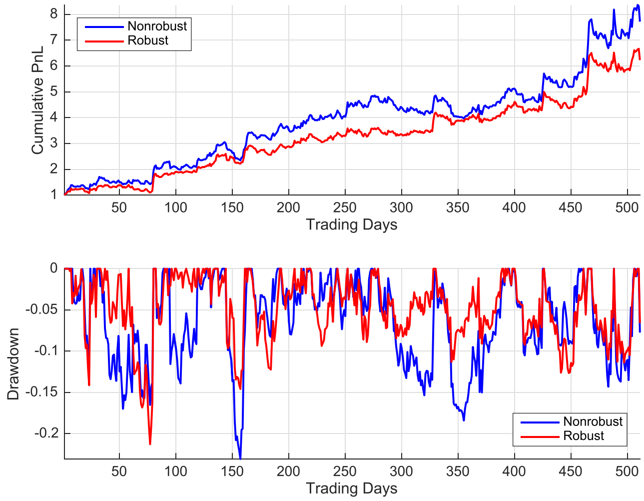


Fig. 5. Performance evaluation (cumulative PnL and drawdown versus trading days) of robust and nonrobust formulation, options included.

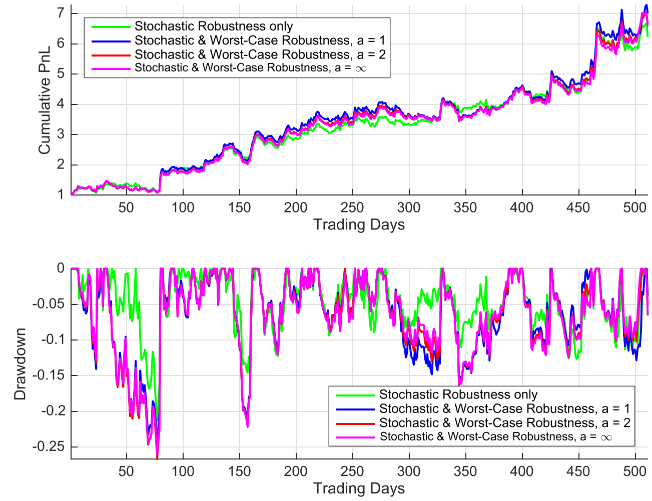


Fig. 7. Performance evaluation (cumulative PnL and drawdown versus trading days) of different types of robustness.

TABLE IV
PERFORMANCE EVALUATION (SHARPE RATIO, ROT, VAR, ES, AND CE (END OF TRADE)) OF ROBUST AND NONROBUST FORMULATION, OPTIONS INCLUDED

Robust or not	Sharpe ratio	ROT	VaR	ES	CE
Nonrobust	1.5504	111.0176	-0.0645	-0.0806	0.0040
Robust	1.6763	117.5772	-0.0642	-0.0716	0.0036

TABLE V
PERFORMANCE EVALUATION (SHARPE RATIO, ROT, VAR, ES, AND CE (END OF TRADE)) OF DIFFERENT TYPES OF ROBUSTNESS

Different types of robustness	Sharpe ratio	ROT	VaR	ES	CE
Stochastic Robustness only	1.6763	117.5772	-0.0642	-0.0716	0.0036
Stochastic & Worst-Case Robustness, $a = 1$	1.8514	103.8731	-0.0561	-0.0650	0.0038
Stochastic & Worst-Case Robustness, $a = 2$	1.7779	100.1020	-0.0572	-0.0661	0.0037
Stochastic & Worst-Case Robustness, $a = +\infty$	1.7541	99.5456	-0.0573	-0.0655	0.0037

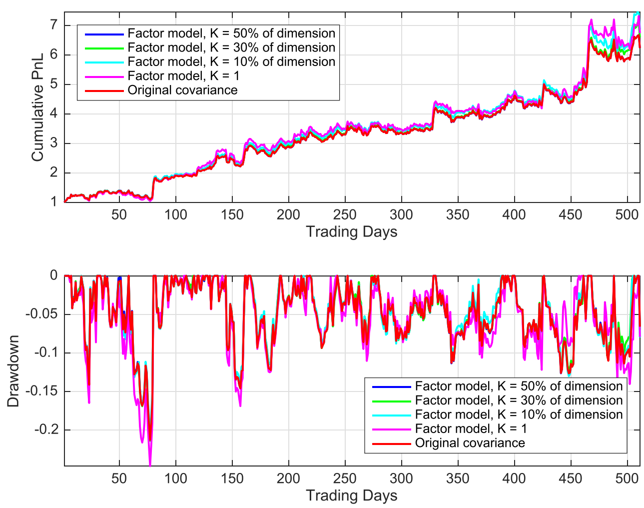


Fig. 6. Performance evaluation (cumulative PnL and drawdown versus trading days) of applying a factor model.

subset $\{\lambda_{K+1}, \lambda_{K+2}, \cdots\}$ and we set each λ_i in the subset with the subset mean. Thus we obtain $\tilde{\lambda}$. The factor-model-based modified covariance is given as $\tilde{\mathbf{M}} = \mathbf{U}\mathrm{Diag}(\tilde{\lambda})\mathbf{U}^T$. We present in the following the simulation results of adopting $\mathbf{M}$ and $\tilde{\mathbf{M}}$ with different choices of K. In the simulation, we set K to be 50%, 30%, and 10% of the matrix dimension, and 1. In Fig. 6, we can see that the curve of cumulative PnL seems to slightly improve as K decreases, indicating that it is worthwhile to adopt a factor-model to modify the covariance.

4) Full Potential of Robustness: To start with, we fix $\varepsilon = 10^{-4}$ as the worst-case robustness parameter, which is the same order as the diagonal elements of $\mathbf{\Sigma}$ (the covariance matrix of the stock returns only). For convenience, the comparison among different types of worst-case robustness is reflected in the choice of parameter a. We present the simulation result in Fig. 7 and Table V. As can be seen in the figure and table, imposing different types of robustness does not change the cumulative PnL and drawdown very much, although imposing stochastic and worst-case robustness simultaneously with $a = 1$ seems to achieve the highest cumulative PnL and Sharpe ratio and the second highest ROT, with slightly larger drawdown. It appears that choosing $a = 1$ is better than choosing $a = 2$ and $+\infty$. Let us verify this claim in the next experiment.

We choose the Sharpe ratio as the performance measure. In Fig. 8, we plot the achieved Sharpe ratio versus the choice of ε. If ε is no larger than 3×10^{-4}, we can safely conclude that $a = 1$ is superior to the other two options. When $\varepsilon \geq 3 \times 10^{-4}$, the Sharpe ratio begins to drop and $a = 1$ is not necessarily the best parameter. Thus, for the sake of performance, we should not set ε too large. If ε is less than 3×10^{-4}, we can safely set $a = 1$.

Finally, let us realize the full potential of robustness. We regard ε as an additional tuning parameter and thus we obtain $(\lambda, \varepsilon, \xi)$ via cross validation. We also confine ε within $[0,\ 3 \times 10^{-4}]$ so that we can safely set $a = 1$ and save a tuning parameter. In Fig. 9 and Table VI, we present the performance after realizing the full potential of robustness. For comparison, we include four additional methods: equally weighted

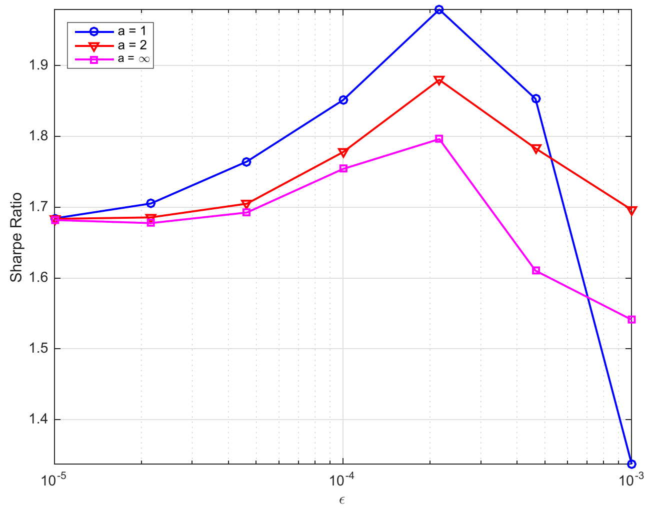


Fig. 8. Performance evaluation (Sharpe ratio versus choice of ε) of different types of robustness.

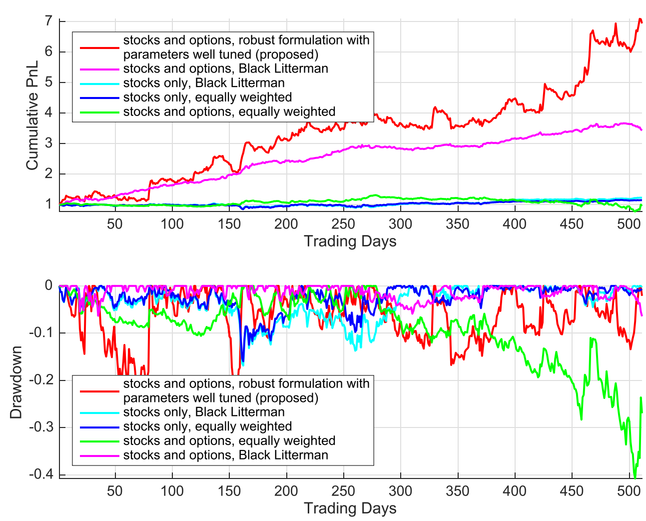


Fig. 9. Performance evaluation (cumulative PnL and drawdown versus trading days) of the proposed portfolio, equally weighted portfolio, and Black-Litterman portfolio.

TABLE VI
PERFORMANCE EVALUATION (SHARPE RATIO, ROT, VAR, ES, AND CE (END OF TRADE)) OF THE PROPOSED PORTFOLIO, EQUALLY WEIGHTED PORTFOLIO, AND BLACK-LITTERMAN PORTFOLIO

Types of portfolio	Sharpe ratio	ROT
Proposed portfolio	1.9388	97.9964
Equally weighted stock-only portfolio	0.4569	1485.7
Equally weighted stock-and-option portfolio	-0.0753	26.9503
Black-Litterman stock-only portfolio	0.6588	26.5550
Black-Litterman stock-and-option portfolio	3.6016	42.6103

Types of portfolio	VaR	ES	CE
Proposed portfolio	-0.0546	-0.0668	0.0038
Equally weighted stock-only portfolio	-0.0301	-0.0366	0.00027
Equally weighted stock-and-option portfolio	-0.0440	-0.0512	-0.00008
Black-Litterman stock-only portfolio	-0.0320	-0.0387	0.00040
Black-Litterman stock-and-option portfolio	-0.0247	-0.0308	0.0025

stock-only portfolio, equally weighted stock-and-option portfolio [31], Black-Litterman stock-only portfolio, and Black-Litterman stock-and-option portfolio with our proposed views as in eq. (43) (the parameter settings follow [20]). The proposed portfolio performs better than the four benchmark portfolios with the highest cumulative PnL and largest CE, although it suffers from a slightly larger potential loss (a slightly lower VaR and ES). The Black-Litterman stock-only portfolio has a slightly larger Sharpe Ratio than the equally weighted stock-only portfolio. The Black-Litterman stock-and-option portfolio achieves the second highest cumulative PnL and the largest Sharpe ratio with the smallest drawdown. This is because the Black-Litterman portfolio takes advantage of financial views to improve performance.

VI. CONCLUSION

In this paper, we have studied the problem of option portfolio design under the Markowitz mean-variance framework. The option returns have been modeled statistically with first- and second-order moments, enriching the conventional delta-gamma approximation. The naive mean-variance formulation allows for the zero-risk fallacy, which has been circumvented with a more realistic robust formulation. Transaction cost has also been considered in the robust formulation. We have proposed an efficient BSUM-M-based algorithm to solve the portfolio design problem. It serves as an efficient alternative to such off-the-shelf solvers as MOSEK. The proposed algorithm can perform as well as the off-the-shelf solvers but with a faster convergence speed, about one half or one order of magnitude faster according to the simulation results. Numerical results based on real data have demonstrated the superior performance of our proposed stock-option portfolio compared to the pure-stock portfolio design.

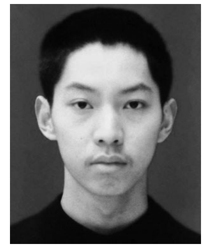
ACKNOWLEDGMENT

The authors would like to thank the Associate Editor, Prof. M. Bengtsson, and the anonymous reviewers for their helpful comments on the paper.

REFERENCES

[1] H. Markowitz, "Portfolio selection," *J. Finance*, vol. 7, no. 1, pp. 77–91, 1952.
[2] J. C. Hull, *Options, Futures, and Other Derivatives*, 9th ed. New York, NY, USA: Pearson, 2015.
[3] N. Sheldon, *Option Volatility & Pricing–Advanced Trading Strategies and Techniques*. New York, NY, USA: McGraw-Hill, 1994.
[4] J. Gatheral, *The Volatility Surface: A Practitioner's Guide*, vol. 357. Hoboken, NJ, USA: Wiley, 2011.
[5] F. Black and M. Scholes, "The pricing of options and corporate liabilities," *J. Political Econ.*, vol. 81, no. 3, pp. 637–654, 1973.
[6] J. C. Cox, S. A. Ross, and M. Rubinstein, "Option pricing: A simplified approach," *J. Financial Econ.*, vol. 7, no. 3, pp. 229–263, 1979.
[7] X. Yu, "Robust portfolio optimization with options under VE constraint using Monte Carlo," *J. Comput.*, vol. 8, no. 6, pp. 1580–1586, 2013.
[8] S. Zymler, B. Rustem, and D. Kuhn, "Robust portfolio optimization with derivative insurance guarantees," *Eur. J. Oper. Res.*, vol. 210, no. 2, pp. 410–424, 2011.
[9] S. Zymler, D. Kuhn, and B. Rustem, "Worst-case value at risk of nonlinear portfolios," *Manage. Sci.*, vol. 59, no. 1, pp. 172–188, 2013.
[10] S. Alexander, T. F. Coleman, and Y. Li, "Minimizing CVaR and VaR for a portfolio of derivatives," *J. Banking Finance*, vol. 30, no. 2, pp. 583–605, 2006.
[11] S. R. Jaschke, "The Cornish–Fisher expansion in the context of delta-gamma-normal approximations," *J. Risk*, vol. 4, pp. 33–52, 2002.

[12] D. Eddelbuettel and K. Nguyen, "Rquantlib: R interface to the quantlib library," *R Package Version 0.3*, vol. 8, 2011.
[13] Y. Feng *et al.*, "A signal processing perspective of financial engineering," *Found. Trends Signal Process.*, vol. 9, no. 1–2, pp. 1–231, 2016.
[14] G. P. Szegö, *Portfolio Theory: With Application to Bank Asset Management*. New York, NY, USA: Academic, 2014.
[15] C. Jiang, "Efficient subset selection in large-scale portfolio with singular covariance matrix," in *Proc. 7th Int. Conf. Manage. Sci. Eng. Manage..*, 2014, pp. 1443–1453.
[16] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Roy. Stat. Soc.. Ser. B (Methodol.)*, vol. 58, pp. 267–288, 1996.
[17] P. Bühlmann and S. Van De Geer, *Statistics for High-Dimensional Data: Methods, Theory and Applications*. New York, NY, USA: Springer, 2011.
[18] X. Zhang, D. P. Palomar, and B. Ottersten, "Statistically robust design of linear MIMO transceivers," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3678–3689, Aug. 2008.
[19] F. Black and R. B. Litterman, "Asset allocation: Combining investor views with market equilibrium," *J. Fixed Income*, vol. 1, no. 2, pp. 7–18, 1991.
[20] A. Meucci, *Risk and Asset Allocation*. New York, NY, USA: Springer, 2009.
[21] "The MOSEK optimization toolbox for Matlab manual, version 7.1 (revision 28)," MOSEK, Copenhagen, Denmark, 2015. [Online]. Available: http://www.mosek.com
[22] J. F. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," *Optim. Methods Softw.*, vol. 11, no. 1–4, pp. 625–653, 1999.
[23] K.-C. Toh, M. J. Todd, and R. H. Tütüncü, "On the implementation and usage of SDPT3–a Matlab software package for semidefinite-quadratic-linear programming, version 4.0," in *Handbook on Semidefinite, Conic and Polynomial Optimization*. New York, NY, USA: Springer, 2012, pp. 715–754.
[24] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," Mar. 2014. [Online]. Available: http://cvxr.com/cvx
[25] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
[26] M. Hong, T.-H. Chang, X. Wang, M. Razaviyayn, S. Ma, and Z.-Q. Luo, "A block successive upper bound minimization method of multipliers for linearly constrained convex optimization," arXiv:1401.7079, to be published.
[27] J. Song, P. Babu, and D. P. Palomar, "Optimization methods for designing sequences with low autocorrelation sidelobes," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3998–4009, Aug. 2015.
[28] Y. Sun, P. Babu, and D. P. Palomar, "Regularized Tyler's scatter estimator: Existence, uniqueness, and algorithms," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5143–5156, Oct. 2014.
[29] J. Bun, J.-P. Bouchaud, and M. Potters, "Cleaning large correlation matrices: Tools from random matrix theory," *Phys. Rep.*, vol. 666, pp. 1–109, 2017.
[30] B. I. Jacobs, K. N. Levy, and H. M. Markowitz, "Portfolio optimization with factors, scenarios, and realistic short positions," *Oper. Res.*, vol. 53, no. 4, pp. 586–599, 2005.
[31] V. DeMiguel, L. Garlappi, and R. Uppal, "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" *Rev. Financial Stud.*, vol. 22, no. 5, pp. 1915–1953, 2007.



Licheng Zhao received the B.S. degree in information engineering from Southeast University, Nanjing, China, in 2014. He is currently working toward the Ph.D. degree with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology. His research interests include optimization theory and fast algorithms, with applications in signal processing, machine learning, and financial engineering.

Daniel P. Palomar (S'99–M'03–SM'08–F'12) received the Electrical Engineering and Ph.D. degrees from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1998 and 2003, respectively. He is currently a Professor with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST), Hong Kong, which he joined in 2006. Since 2013, he has been a Fellow with the Institute for Advance Study, HKUST. He had previously held several research appointments, namely, at King's College London, London, U.K.; Stanford University, Stanford, CA, USA; Telecommunications Technological Center of Catalonia, Barcelona, Spain; Royal Institute of Technology, Stockholm, Sweden; University of Rome "La Sapienza," Rome, Italy; and Princeton University, Princeton, NJ, USA. His current research interests include applications of convex optimization theory, game theory, and variational inequality theory to financial systems, big data systems, and communication systems. He is a recipient of a 2004/06 Fulbright Research Fellowship, the 2004 and 2015 (coauthor) Young Author Best Paper Awards by the IEEE Signal Processing Society, the 2015–2016 HKUST Excellence Research Award, the 2002/2003 best Ph.D. prize in Information Technologies and Communications by UPC, the 2002/2003 Rosina Ribalta first prize for the Best Doctoral Thesis in Information Technologies and Communications by the Epson Foundation, and the 2004 prize for the best Doctoral Thesis in Advanced Mobile Communications by the Vodafone Foundation and COIT. He is a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING 2016 Special Issue on "Financial Signal Processing and Machine Learning for Electronic Trading" and has been an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY and IEEE TRANSACTIONS ON SIGNAL PROCESSING, a Guest Editor for the IEEE SIGNAL PROCESSING MAGAZINE 2010 Special Issue on "Convex Optimization for Signal Processing," the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2008 Special Issue on "Game Theory in Communication Systems," and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2007 Special Issue on "Optimization of MIMO Transceivers for Realistic Communication Networks."