Robust Portfolio Optimization with Respect to Spectral Risk Measures Under Correlation Uncertainty

Man Yiu Tsang[1] · Tony Sit[1] · Hoi Ying Wong[1]

Abstract

This paper proposes a distributionally robust multi-period portfolio model with ambiguity on asset correlations with fixed individual asset return mean and variance. The correlation matrix bounds can be quantified via corresponding confidence intervals based on historical data. We employ a general class of coherent risk measures namely the spectral risk measure, which includes the popular measure conditional value-at-risk (CVaR) as a particular case, as our objective function. Specific choices of spectral risk measure permit flexibility for capturing risk preferences of different investors. A semi-analytical solution is derived for our model. The prominent stochastic dual dynamic programming (SDDP) algorithm adapted with intricate modifications is developed as a numerical method under the discrete distribution setting. In particular, our new formulation accounts for the unknown worst-case distribution in each iteration. We verify the convergence property of this algorithm under the setting of finite scenarios. Our results show that the optimal solution favours a certain degree of anti-diversification due to dependence ambiguity and exhibits its protection ability during the financial crisis period.

1 Introduction

Markowitz's [40] portfolio optimization model which advocates the idea of mean-risk models in modern portfolio theory in 1952 and has received tremendous attention since its proposal. This neat idea is influential upon which many subsequent and

---

✉ Tony Sit
  tony.sit@cuhk.edu.hk

[1]   Department of Statistics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

more sophisticated models are established, including models that incorporate various practical constraints. For instance, constraints on transaction costs are imposed to capture the costs in buying and selling the assets and constraints on maximum holding proportion for each asset to cover the investor perference on risk diversification. A summary of portfolio optimization models can be found in [6], for example. The role of numerical optimization methods to solve these models has also been an active research area. Instead of considering single-period models, a more realistic class of multi-period models are studied as they take the complete investment horizon into consideration with the possibility of adjusting one's investment position at certain times regularly. Under some specific settings, the multi-period problem can be solved analytically (see, for example, [31] and [18]).

Risk measures and related theories have undergone rapid development in recent years as attempts to incorporate the effect due to risk aversion on investment decision. This consideration of risk measures has become more crucial after the 2008 financial crisis, after which investors have developed preference for robust portfolios. In addition to the prevalent risk measure conditional value-at-risk (CVaR) discussed in [52], which has been widely adopted in financial sector as required in Basel Accord III, a more general class of risk measures, namely, the spectral risk measure, is proposed in [1] and further discussed in [11] and [7] amongst others. This class includes some novel risk measures in the literature such as second-order superquantile [51] and Gini-shortfall [17]. Readers may also refer to [46] and [55] for relevant discussion on such risk measures. In this paper, we focus on this general class of spectral risk measures under which many different well-known risk measures are covered.

While the mean-variance model is convenient, variance is a symmetric measure of volatility and it does not provide effective quantification for downside risk. In modern portfolio theory, rational and prudent investors are cautious about downside losses at the cost of sacrificing partially the upward potentials of the associated investments. The use of alternative risk measures for portfolio models has become increasingly attractive. An extensive amount of research effort has been devoted to construct implementable risk-averse portfolio models. Fundamental theories for mean-risk model are developed in [53] and the context under portfolio optimization is discussed in [42]. Consideration of high-order conditional risk measures based upon the scenario generation method is recently investigated in [24] followed by [2] which extends the scope to the use of entropic value-at-risk.

With the growing concerns for reliability of parameter estimation, distributionally robust portfolio models are developed to protect investors from suffering tremendous losses due to misspecification of model parameters (see [15]). Two types of uncertainty set construction prevail in the literature, namely distance-type uncertainty and moment-type uncertainty. For the former, it can be constructed using the empirical distribution as the reference probability measure through specific metrics; a thorough discussion using Wasserstein distance can be found in [5]. For the latter, moment information from the observed data is captured. In this paper, we consider a multi-period problem with the objective function being the worst-case spectral risk measure, characterized by moment-type uncertainty. To be more specific, the moment uncertainty set is quantified by the estimated individual asset's mean and variance while the correlation is quantified by a specific interval, say for instance, a relevant confidence interval so

as to capture the model uncertainty. Unlike most of the papers on moment uncertainty set that focus on an ellipsoidal set on the first two moments (see, for example, [9]), the construction is relatively straight-forward. Ellipsoidal uncertainty set requires the choice of the scale and shape parameters that are hard to interpret compared to the use of confidence intervals. A recent discussion on a deliberate construction appears in [38], which relies on estimates provided by experts. Moreover, the existing ellipsoidal set cannot capture the correlation uncertainty between assets. A similar multi-period portfolio problem with ellipsoidal uncertainty sets is investigated in [35] while the single-period of the problem using CVaR objective is covered in [38]. In this line of research, related works also include [30, 33, 34] and [36], amongst others.

The motivation of correlation ambiguity emerges from the concern about accuracy in estimating correlation coefficients. It is inevitable that correlation affects investment decisions tremendously. It is well known that obtaining accurate estimation of high-dimensional correlation matrix is especially challenging. Also, through experiments, [14] observed a lack of confidence in decision making under correlation uncertainty. It implies that aversion towards correlation uncertainty inheres in the portfolio selection process without which the resulting strategies tend to be mistakenly optimistic. Thus, we are interested in the correlation ambiguity in portfolio models. Fouque et al. [16] studied the situation with two risky assets and one risk-free asset in continuous-time economy while the multiple-asset case in mean-variance framework is studied in [28]. [25] considered a single-period setting with normally distributed payoffs.

Multi-stage optimization is typically challenging. In the discrete setting, it is obvious that the number of scenarios grows exponentially as time proceeds. That is, if we have N scenarios for each period, we would have N^T scenarios over these T periods, which is computationally intractable in general. Under the stagewise independence assumption of the randomness, one can circumvent the problem with the help of stochastic dual dynamic programming (SDDP) method, pioneered in [43]. The method is summarized and analyzed in [54] for the linear case and the risk averse version is described in [56]. Recently, [23] proposed a variant for linear stochastic programs with optimality cut selection steps to enhance computational efficiency. For robust stochastic linear programs, [20] developed a robust version of SDDP, which involves both upper and lower linear approximations of the recourse function. Time-dependent variants could be found in [37] and [4]. The convergence proof of the method for risk neutral and risk averse case in general convex setting appear in [21] and [22], respectively.

In this paper, we consider a multi-period distributionally robust portfolio optimization problem with the spectral risk measure objective under correlation ambiguity. The use of spectral risk measure permits flexibility on preferences from different investors. Our numerical method under discrete setting is considered with historical scenarios. In this area of active research, [26] investigated the risk neutral case with l_∞ uncertainty set while [47] studied the one with l_2 uncertainty set. Duque et al. [12] discussed the use of Wasserstein distance ambiguity set in linear stochastic programs with convergence analysis. To the best of our knowledge, multi-period portfolio optimization problem with moment-type uncertainty set has not been investigated in the literature. Moment information crucial to investors, such as the mean and variance of random

returns, can be captured explicitly while it is difficult to capture correlation ambiguity via distance-type uncertainty set formulations.

The contribution of this paper is three-fold. First, we obtain a semi-analytical solution for the problem, which exploits the explicit formula of worst-case risk measures presented in [32]. Different from [35], our model works with general spectral risk measures and we include the no short-selling constraint in our model. This is a realistic assumption since short selling is not always allowed. Moreover, the worst-case distribution cannot be obtained analytically as in their formulation. Despite all the complexities, our solution can still be computed efficiently via interior-point methods. Second, to incorporate the support of return vectors, we apply the scenario-based approach to our problem and propose a variant of SDDP to solve, which considers the moment-type uncertainty set in the form of correlation ambiguity. This numerical method can incorporate lower bounds for the random return vectors via the generated scenarios. The use of SDDP in distributionally robust optimization using l_2 distance appeared recently in [47] and has been investigated in other contexts. We provide the convergence proof of our proposed algorithm, which extends the previous works [21] and [22] to the distributionally robust context. In the numerical experiment, our formulation performs better than the l_2 formulation in the 2008 crisis during which the market experienced structural changes. Our model with correlation uncertainty set can provide a more secured decision if the correlation structure of the assets changes significantly. Third, we observe that the use approximation of spectral function introduced in [57] can be potentially useful under the distributionally robust situation in the sense that the computational time per iteration would reduce with a reasonable marginal error in optimal value. It is noteworthy that although formulation is developed under a portfolio optimization framework, the algorithm and the theoretical results can be readily extended to a more general setting with linear constraints, which can be readily modified to cover applications such as operation planning problem in [55] and inventory problem in [3].

The remainder of the paper is organized as follows: Sect. 2 introduces the problem, including the notion of spectral risk measures and construction of uncertainty set. A semi-analytic solution is derived in Sect. 3. Section 4 discusses the method to obtain worst-case distribution under uncertainty based on discrete setting, which would be used in the numerical SDDP method. The convergence of the method would be verified in Sect. 5. Some computational results are demonstrated in Sect. 6 and the paper is concluded in Sect. 7. All the proofs are provided in the Supplementary material.

2 Multi-stage Portfolio Optimization Problem Formulation

We consider a multi-period problem with finite scenarios and T stages. Assume that there are n risky assets under investment consideration with individual mean μ_i and variance σ_i^2 for $i = 1, \ldots, n$. The sample space at time t is given by $\Omega_t = \mathbb{R}^n$ for $t \in \mathcal{T} = \{1, \ldots, T\}$. Stagewise independence of return vectors is assumed. Denote $\{\mathcal{F}_t\}$ as a filtration accumulated up to time $t \in \mathcal{T}$ and r_t as the random price ratio vector at time t (i.e., asset price at time t divided by that at time $t - 1$). For simplicity, we also call r_t as the return vector as in the literature. Let W_t be the wealth at time t
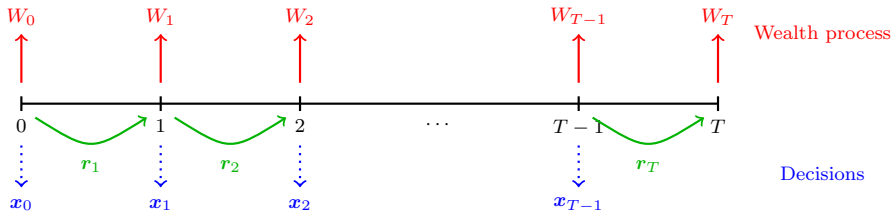
Fig. 1 Illustration of Portfolio Horizon

and the initial investment amount W_0 be, without loss of generality, standardized as 1. The multi-stage portfolio optimization problem considered is formulated as follows:

$$
\begin{aligned}
\operatorname*{minimize}_{x_0, x_1, \ldots, x_{T-1}} \quad & \sup_{\mathbb{P}_1 \in \mathcal{P}_1} \rho_1 \left[\sup_{\mathbb{P}_2 \in \mathcal{P}_2} \rho_2 \cdots \left\{ \sup_{\mathbb{P}_T \in \mathcal{P}_T} \rho_T (-W_T) \right\} \right] \\
\text{subject to} \quad & \mathbf{1}^\mathsf{T} x_t = W_t, \quad t = 0, 1, \ldots, T-1, \\
& r_{t+1}^\mathsf{T} x_t = W_{t+1}, \quad t = 0, 1, \ldots, T-1, \\
& x_t \geq \mathbf{0}, \quad t = 0, 1, \ldots, T-1,
\end{aligned}
\tag{1}
$$

where x_t is the investment decision at time t, $\mathcal{P}_t$ is the uncertainty set, a subset of probability measures on $(\Omega_t, \mathcal{F}_t)$, and ρ_t is a chosen coherent risk mapping conditioned on $\mathcal{F}_{t-1}$. The first and second constraints describe the investment process: the wealth obtained from the previous period is invested in the next period. The third constraint restricts that short selling is not allowed. Figure 1 illustrates the investment process.

Remark 1 In problem (1), we use a recursive (i.e., nested) risk measure as our objective. This facilitates the use of dynamic programming to solve the problem. Note that if a policy satisfies the associated dynamic equations in (6)–(7), then such a policy is also time consistent (see [49] and [55] for relevant discussions).

Different risk measures are proposed in the literature to capture the risks involved. Spectral risk measures is a class of risk measures which covers many of the commonly used risk measures. For a random variable Z, we denote F_Z as its distribution function.

Definition 1 Let $\phi : [0, 1) \to \mathbb{R}$ be a non-negative and non-decreasing function satisfying $\int_0^1 \phi(p) dp = 1$. The spectral risk measure based on the spectral function ϕ is defined as

$$
\rho_\phi(Z) = \int_0^1 \phi(p) F_Z^{-1}(p) dp,
\tag{2}
$$

where $F_Z^{-1}(p) = \inf\{t \in \mathbb{R} | F_Z(t) \geq p\}$ is the left-continuous quantile function.

Remark 2 The choice of the space of random variables such that the spectral risk measure is finite is a complicated issue, which is discussed in [48]. In this paper, we assume that $\|\phi\|_2$ is finite (see Theorem 1). Together with the second moment finiteness of

random return vector under our construction of uncertainty set, the finiteness is always guaranteed. Also, this is guaranteed for the finite number of scenarios considered in the numerical method.

To see that this general class of risk measures covers many existing choices, consider $\phi(p) = (1 - \alpha)^{-1}\mathbf{1}(p \in [\alpha, 1))$. The measure specified in (2) gives the conditional value-at-risk of level α (CVaR$_\alpha$), where $\mathbf{1}(A)$ is the indicator funcion taking value one on A and zero otherwise. Another equivalent formulation for CVaR$_\alpha$ is given as

$$\text{CVaR}_\alpha(Z) = \inf_t \left\{ t + \frac{1}{1 - \alpha} \mathbb{E}\left[(Z - t)_+\right] \right\}. \tag{3}$$

It is well-known that spectral risk measures are coherent risk measures, which satisfy convexity, monotonicity, translation equivariance and positive homogeneity axioms.

The uncertainty sets $\mathcal{P}_t$ are constructed using moment information which are fixed at time 0. We assume that the first two individual moments are known with partial information being used for correlation. To avoid notational ambiguity, we remove the dependence on t and consider the random vector $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$. Then, the uncertainty set is defined as

$$\mathcal{P} = \left\{ \pi \in \mathcal{D} \,\middle|\, \mathbb{E}_\pi(\boldsymbol{\xi}) = \boldsymbol{\mu}, \; \text{Var}_\pi(\xi_i) = \sigma_i^2, \; i = 1, \ldots, n, \right.$$
$$\left. \boldsymbol{C}_\pi := \text{Corr}_\pi(\boldsymbol{\xi}) \succeq 0, \; \boldsymbol{C}_\pi \in [\underline{\boldsymbol{C}}, \overline{\boldsymbol{C}}] \right\}, \tag{4}$$

where $\mathcal{D}$ is the set of all probability distributions and $\boldsymbol{C}_\pi$ is a symmetric (positive semidefinite) correlation matrix. The two-sided bounds for $\boldsymbol{C}_\pi$ are interpreted element-wisely, which capture the correlation ambiguity. The diagonal entries of $\underline{\boldsymbol{C}}$ and $\overline{\boldsymbol{C}}$ are, by default, set to be 1. It is easy to observe that the set $\mathcal{P}$ is convex (see Supplementary material for a proof).

Inspired by [35], due to the recursive nature of the defined risk measure and the monotonicity analyzed in [59], we can rewrite (1) in terms of dynamic programming equations. The equivalence of the multi-stage problem and the dynamic programming formulation is detailed in [49]. First, define the constraint sets

$$\mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{r}_t) = \left\{ \boldsymbol{x}_t \geq \mathbf{0} \mid \mathbf{1}^\mathsf{T}\boldsymbol{x}_t = W_t, \; \boldsymbol{r}_t^\mathsf{T}\boldsymbol{x}_{t-1} = W_t \right\}$$
$$= \left\{ \boldsymbol{x}_t \geq \mathbf{0} \mid \mathbf{1}^\mathsf{T}\boldsymbol{x}_t = \boldsymbol{r}_t^\mathsf{T}\boldsymbol{x}_{t-1} \right\} \tag{5}$$

for $t = 1, \ldots, T$ with $\mathcal{X}_0 = \{\boldsymbol{x}_0 | \mathbf{1}^\mathsf{T}\boldsymbol{x}_0 = W_0\}$. Also, we define the cost functions $f_T(\boldsymbol{x}_T, \boldsymbol{r}_T) = -\mathbf{1}^\mathsf{T}\boldsymbol{x}_T$ and $f_t \equiv 0$ for $t = 1, \ldots, T - 1$. With the stagewise independence assumption, the dynamic equations are established recursively as

$$Q_t(\boldsymbol{x}_{t-1}, \boldsymbol{r}_t) = \inf_{\boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}, \boldsymbol{r}_t)} \{f_t(\boldsymbol{x}_t, \boldsymbol{r}_t) + \mathcal{Q}_{t+1}(\boldsymbol{x}_t)\} \text{ with} \tag{6}$$

$$\mathcal{Q}_t(\boldsymbol{x}_{t-1}) = \sup_{\mathbb{P}_t \in \mathcal{P}_t} \rho_t(Q_t(\boldsymbol{x}_{t-1}, \boldsymbol{r}_t)), \tag{7}$$

for $t = 1, \ldots, T$, where $Q_0 = \inf_{x_0 \in \mathcal{X}_0} Q_1(x_0)$ is the first-stage minimization problem and $Q_{T+1} \equiv 0$. It is noteworthy that $Q_T(x_{T-1}, r_T) = -r_T^{\mathsf{T}} x_{T-1}$ involves no optimization.

3 Semi-analytical Solution

In this section, we derive a semi-analytical solution for problem (1) with uncertainty set (4). An explicit solution is difficult to obtain due to the constraint $x_t \geq 0$. The problem without such a constraint and with a different construction of the uncertainty set has been studied by [35]. Their uncertainty set formulation as some ellipsoidal set is not compatible with the one we consider (see Remark 3). Our formulation explicitly addresses correlation ambiguity, which naturally emerges in decision making processes. Also, our model incorporates the practical no short-selling constraints. These complicate the analysis but the solution obtained can be efficiently computed. To obtain the solution, we first derive the worst-case spectral risk measure formula.

Define $\boldsymbol{\sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$, $\operatorname{Var}(\boldsymbol{\xi}) = \boldsymbol{\Gamma} = \boldsymbol{\sigma}^{\mathsf{T}} \boldsymbol{C} \boldsymbol{\sigma}$ and the set $\mathcal{C} = \{\boldsymbol{C} \in \mathbb{R}^{n \times n} | \boldsymbol{C} \succeq \boldsymbol{0}, \ \boldsymbol{C} \in [\underline{\boldsymbol{C}}, \overline{\boldsymbol{C}}]\}$. It is easy to observe that $\mathcal{C}$ is convex, bounded and closed (with respect to norm topology in $\mathbb{R}^{n^2}$). Adopting the above notation, for any vector $\boldsymbol{a} \in \mathbb{R}^n$, we have

$$\operatorname{Var}(\boldsymbol{a}^{\mathsf{T}} \boldsymbol{\xi}) = \boldsymbol{a}^{\mathsf{T}} \operatorname{Var}(\boldsymbol{\xi}) \boldsymbol{a} = (\boldsymbol{\sigma} \boldsymbol{a})^{\mathsf{T}} \boldsymbol{C} (\boldsymbol{\sigma} \boldsymbol{a}).$$

We can then define $\underline{v}(\boldsymbol{a}) = \min_{\boldsymbol{C} \in \mathcal{C}} \operatorname{Var}(\boldsymbol{a}^{\mathsf{T}} \boldsymbol{\xi})$ and $\overline{v}(\boldsymbol{a}) = \max_{\boldsymbol{C} \in \mathcal{C}} \operatorname{Var}(\boldsymbol{a}^{\mathsf{T}} \boldsymbol{\xi})$, where the existence is guaranteed by the continuity of the variance function in $\boldsymbol{C}$ and the compactness of $\mathcal{C}$.

Theorem 1 *Let $\boldsymbol{a} \in \mathbb{R}^n$ be a fixed vector. Then, the worst-case spectral risk measure defined in (2) using spectral function ϕ with finite $\|\phi\|_2^2$ can be written as*

$$\sup_{\mathbb{P} \in \mathcal{P}} \rho_\phi(-\boldsymbol{a}^{\mathsf{T}} \boldsymbol{\xi}) = -\boldsymbol{\mu}^{\mathsf{T}} \boldsymbol{a} + \kappa \sqrt{\overline{v}(\boldsymbol{a})},$$

where $\kappa = \sqrt{\|\phi\|_2^2 - 1}$ and $\|\phi\|_2^2 = \int_0^1 \phi(p)^2 dp$.

Remark 3 We can compare the use of ellipsoidal sets in the literature and our choice of uncertainty set in terms of worst-case distribution. We assume that the mean is known and $\boldsymbol{\Sigma}_0$ is some reference covariance matrix. Consider

$$\mathcal{P}_1 = \{\pi \in \mathcal{D} \mid \mathbb{E}_\pi(\boldsymbol{\xi}) = \boldsymbol{\mu}_0, \ \operatorname{Var}_\pi(\boldsymbol{\xi}) \preceq k \boldsymbol{\Sigma}_0\}$$

used in [9, 35]. The worst-case distribution over $\rho_\phi(-\boldsymbol{a}^{\mathsf{T}} \boldsymbol{\xi})$ will have a covariance matrix $k \boldsymbol{\Sigma}_0$. From this, we can see that the correlation structure is indeed fixed by the input covariance matrix $\boldsymbol{\Sigma}_0$. In [38], the ellipsoidal uncertainty set is used as

$$\mathcal{P}_2 = \left\{ \pi \in \mathcal{D} \mid \mathbb{E}_\pi(\boldsymbol{\xi}) = \boldsymbol{\mu}_0, \ \frac{S-1}{2} \left\| \boldsymbol{\Sigma}_0^{-\frac{1}{2}} (\operatorname{Var}_\pi(\boldsymbol{\xi}) - \boldsymbol{\Sigma}_0) \boldsymbol{\Sigma}_0^{-\frac{1}{2}} \right\| \leq \delta^2 \right\},$$

where S is the number of scenarios. The covariance matrix of the worst-case distribution is given by $(1 + \delta\sqrt{2/(S-1)})\boldsymbol{\Sigma}_0$. This also shows that the correlation matrix remains the same. Neither $\mathcal{P}_1$ nor $\mathcal{P}_2$ can address our concern over correlation uncertainty. Moreover, the choice of the parameter in the uncertainty set, i.e. k and δ, may not have a direct interpretation while the use of confidence interval can be understood easily, especially for portfolio managers.

With the above results, we can make use of the dynamic programming equations to obtain the expressions for the solution. We first give a lemma, which would be useful throughout the derivation of the semi-analytical solution. It concerns the following minimization problem for any $\kappa > 0$.

$$
\begin{aligned}
& \text{Minimize } -\boldsymbol{\mu}^{\mathsf{T}}\boldsymbol{x} + \kappa\sqrt{\boldsymbol{x}^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{x}} \\
P(\gamma): \quad & \text{subject to } \boldsymbol{x}^{\mathsf{T}}\mathbf{1} = \gamma, \\
& \qquad\qquad \boldsymbol{x} \geq \mathbf{0}.
\end{aligned}
\tag{8}
$$

Lemma 1 *Assume that the problem $P(1)$, namely $P(\gamma)$ with $\gamma = 1$, has an optimal solution $\boldsymbol{x}^*$. Then, $\gamma\boldsymbol{x}^*$ is an optimal solution to the problem $P(\gamma)$ for any $\gamma > 0$. Moreover, the dual solution $\boldsymbol{\lambda}^*$ associated to the constraint $\boldsymbol{x} \geq \mathbf{0}$ is the same for any $\gamma > 0$.*

We then state the main theorem of this section, which gives an expression for the solution under uncertainty set of the form $\mathcal{P}$. For simplicity, all the uncertainty sets $\mathcal{P}_t$ in (1) are assumed to be the same so that we can omit the subscript t. Otherwise, the parameters $\boldsymbol{\mu}$, σ_i and $\boldsymbol{C}$ in the uncertainty set $\mathcal{P}_t$ depend on t, which are specified at time 0, and our formulation is still implementable. For a correlation matrix $\boldsymbol{C}$, define

$$
A_t(\boldsymbol{C}, \boldsymbol{\mu}) = \frac{\boldsymbol{\mu}^{\mathsf{T}}(\sigma\boldsymbol{C}\sigma)^{-1}\mathbf{1} - \sqrt{\left(\boldsymbol{\mu}^{\mathsf{T}}(\sigma\boldsymbol{C}\sigma)^{-1}\mathbf{1}\right)^2 - \mathbf{1}^{\mathsf{T}}(\sigma\boldsymbol{C}\sigma)^{-1}\mathbf{1}\left(\boldsymbol{\mu}^{\mathsf{T}}(\sigma\boldsymbol{C}\sigma)^{-1}\boldsymbol{\mu} - \kappa\right)}}{\mathbf{1}^{\mathsf{T}}(\sigma\boldsymbol{C}\sigma)^{-1}\mathbf{1}}
$$

and the constant $A_t^*(\boldsymbol{\mu}) = \inf_{\boldsymbol{C}\in\mathcal{C}} A_t(\boldsymbol{C}, \boldsymbol{\mu})$. It is possible that the constant $A_t(\boldsymbol{C}, \boldsymbol{\mu})$ is not real as the expression inside the square root can be negative. In this case, the value of $A_t^*(\boldsymbol{\mu})$ is set to be positive infinity. Note that we incorporate the non-negativity constraint on $\boldsymbol{x}_t$ in our derivation since short selling of some assets may be restricted. Instead of having a fully analytical result for each problem in the dynamic programming equations (see, for example, Corollary 2.2 of [10] and Theorem 2 of [35]), our formula includes the Lagrangian dual variables. The solution also applies to the problem without no short-selling constraint by setting all $\boldsymbol{\lambda}_t = 0$ which gives a simpler formula. Technical details for the proof are provided, followed by the theorem.

Theorem 2 *Consider problem (1) with uncertainty set of the form $\mathcal{P}$ defined in (4). Let $\boldsymbol{\lambda}_t^*$ be the optimal dual variable associated to the constraint $\boldsymbol{x}_t \geq \mathbf{0}$ in (6). Assume that for $t = 0, \ldots, T-1$,*

(a) $(\boldsymbol{\mu} + \boldsymbol{\lambda}_t^*)^{\mathsf{T}}(\sigma\boldsymbol{C}\sigma)^{-1}(\boldsymbol{\mu} + \boldsymbol{\lambda}_t^*) - \kappa^2 \geq 0$ *for all $\boldsymbol{C} \in \mathcal{C}$,*

(b) $\left((\boldsymbol{\mu} + \boldsymbol{\lambda}_t^*)^{\mathsf{T}}(\boldsymbol{\sigma C \sigma})^{-1}\mathbf{1}\right)^2 \geq \mathbf{1}^{\mathsf{T}}(\boldsymbol{\sigma C \sigma})^{-1}\mathbf{1}\left((\boldsymbol{\mu} + \boldsymbol{\lambda}_t^*)^{\mathsf{T}}(\boldsymbol{\sigma C \sigma})^{-1}(\boldsymbol{\mu} + \boldsymbol{\lambda}_t^*) - \kappa^2\right)$
   for some $\boldsymbol{C} \in \mathcal{C}$,

(c) *the feasible sets $\mathcal{X}_t$ defined in (5) are non-empty.*

Let $\hat{A}_t^* = \inf_{\boldsymbol{C} \in \mathcal{C}} A_t(\boldsymbol{C}, \boldsymbol{\mu} + \boldsymbol{\lambda}_{t-1}^*)$ with $\boldsymbol{C}_t^*$ being the minimizer and $\boldsymbol{\Sigma}_t^* = \boldsymbol{\sigma C}_t^* \boldsymbol{\sigma}$. Then, the optimal value is given by $-\prod_{t=1}^T \hat{A}_t^*$ with optimal solution $\boldsymbol{x}_0^* = [\mathbf{1}^{\mathsf{T}}(\boldsymbol{\Sigma}_1^*)^{-1}(\boldsymbol{\mu} + \boldsymbol{\lambda}_0^* - \hat{A}_2^*\mathbf{1})]^{-1}(\boldsymbol{\Sigma}_1^*)^{-1}(\boldsymbol{\mu} + \boldsymbol{\lambda}_0^* - \hat{A}_2^*\mathbf{1})$.

Remark 4 Assumption (a) secures the recurrent use of minimax theorem and subsequent arguments. Indeed, if assumption (a) is violated for some $\boldsymbol{C} \in \mathcal{C}$, then $\hat{A}_t^*$ would be negative. For instance, the function in the proof,

$$\hat{A}_T^* \left\{ -\boldsymbol{\mu}^{\mathsf{T}} \boldsymbol{x}_{T-2} + \kappa \sqrt{\boldsymbol{x}_{T-2}^{\mathsf{T}}(\boldsymbol{\sigma C \sigma})\boldsymbol{x}_{T-2}} \right\} ,$$

would not be convex in $\boldsymbol{x}$ and concave in $\boldsymbol{C}$ due to the negative constant. Hence, the minimax theorem fails to apply and the arguments could not follow. For assumption (b), the violation of it would lead to infeasibility of the infimum problem and hence the optimal value could be considered as negative infinity. These two assumptions can be checked via the optimal value of (9) (see Remark 5 and the discussion followed after it).

Remark 5 We can also make the following observations based on Theorem 2. First, Lemma 1 ensures that $\hat{A}_t^*$ involving $\boldsymbol{\lambda}_{t-1}^*$ could be treated as constant so that it can be taken out from supremum and infimum operators. Second, we notice that if the uncertainty sets are the same independent of t, the optimal trading strategy would be a static strategy, meaning that investors only invest in the beginning and do not adjust the portfolio throughout the period.

The solution to the corresponding single-period problem can be reformulated as a semidefinite programming (SDP) problem discussed in [38]. From Theorem 2, we can obtain the optimal solution and value by solving the reduced single-period problem as:

$$
\begin{aligned}
\underset{v \in \mathbb{R}, \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{\Lambda}, \boldsymbol{\Lambda}_+, \boldsymbol{\Lambda}_- \in \mathbb{S}^n}{\text{minimize}} \quad & \kappa^2 v + \operatorname{tr}(\boldsymbol{\Lambda}_+ \overline{\boldsymbol{\Sigma}}) - \operatorname{tr}(\boldsymbol{\Lambda}_- \underline{\boldsymbol{\Sigma}}) - \boldsymbol{\mu}^{\mathsf{T}} \boldsymbol{x} \\
\text{subject to} \quad & \mathbf{1}^{\mathsf{T}} \boldsymbol{x} = 1 , \\
& \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{x}/2 \\ \boldsymbol{x}^{\mathsf{T}}/2 & v \end{bmatrix} \succeq \mathbf{0} , \\
& \boldsymbol{\Lambda} \preceq \boldsymbol{\Lambda}_+ - \boldsymbol{\Lambda}_- , \\
& \boldsymbol{\Lambda}_+, \boldsymbol{\Lambda}_- \succeq \mathbf{0} .
\end{aligned}
\tag{9}
$$

The solution of this optimization problem can be computed efficiently using interior-point methods.

4 The Numerical Algorithm Under Practical Setting

The solution we obtain from the previous section is based on $\Omega_t = \mathbb{R}^n$. Although it is common in the literature (see, for example, [35] and [38]), this may not be realistic for the following reasons. First, the price ratio vector r_t has a natural lower bound of **0**. It is natural to impose the constraint of $r_t \geq \mathbf{0}$ almost surely. In particular, $r_t = \mathbf{0}$ corresponds to a total loss for the investment. Second, it is natural to restrict the support of the return vector r_t on a compact domain. For instance, if r_t represents the daily return, it would be realistic that r_t would not deviate a lot from 1. Thus, with a more confined support, this may avoid over-conservative investment strategy.

However, in the presence of the support information, tractable reformulation of the worst-case risk measure may not be possible. For instance, in [13], they only provide an upper bound for a single-period portfolio optimization problem. In particular, for moment-based ambiguity set, the reformulation of the worst-case expectation typically results in a semi-infinite program and may require specialized numerical algorithm such as cutting-plane methods to solve (see, for example, [9, 19, 39] and [41]).

A common way to handle the situation that we do not have a tractable reformulation, in distributionally robust optimization literature, is through discrete scenarios and this has been successfully employed in various applications (see, for example, [8, 27, 44, 47] and [58]). Our work is the first to address the multi-period portfolio optimization under correlation uncertainty with scenario-based approach and we propose a variant of SDDP to solve the problem. Specifically, we assume that $\Omega_t = \{\omega_1, \ldots, \omega_{N_t}\}$ and consider the uncertainty set of the form

$$
\mathcal{P}^{\mathrm{d}} = \left\{ \boldsymbol{p} \in \mathbb{R}^N \,\middle|\, \sum_{i=1}^N p_i = 1, \ \sum_{i=1}^N p_i \boldsymbol{r}_i = \boldsymbol{\mu}, \ \underline{\boldsymbol{\Sigma}} \leq \sum_{i=1}^N p_i (\boldsymbol{r}_i - \boldsymbol{\mu})(\boldsymbol{r}_i - \boldsymbol{\mu})^\mathsf{T} \leq \overline{\boldsymbol{\Sigma}} \right\},
$$
(10)

where $\overline{\boldsymbol{\Sigma}} = \boldsymbol{\sigma}^\mathsf{T} \overline{\boldsymbol{C}} \boldsymbol{\sigma}$, $\underline{\boldsymbol{\Sigma}} = \boldsymbol{\sigma}^\mathsf{T} \underline{\boldsymbol{C}} \boldsymbol{\sigma}$ and $\boldsymbol{r}_i$ correspond to the return scenarios for $i \in \{1, \ldots, N\}$.

Although it may not be realistic to consider $\Omega_t = \mathbb{R}^n$, the semi-analytical solution derived, which could be obtained efficiently, serves as a valid upper bound when we have support information. It also provides us a way to verify our numerical method proposed. In the following two sections, we discuss and analyze a numerical method that can solve the problem under $\mathcal{P}^{\mathrm{d}}$. In Sect. 4.1, we first discuss our methodology to obtain the worst-case distribution, which will be useful in a part of the proposed numerical method in Sect. 4.2.

4.1 Worst-Case Distribution in a Discrete Setting

The numerical method we introduced is based on a finite number of scenarios and therefore, we are interested in spectral risk measures under discrete distributions; see [57] for instance. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable Z. Suppose that Z takes finite realizations $Z_1, \ldots, Z_N$ with probabilities $p_1, \ldots, p_N$
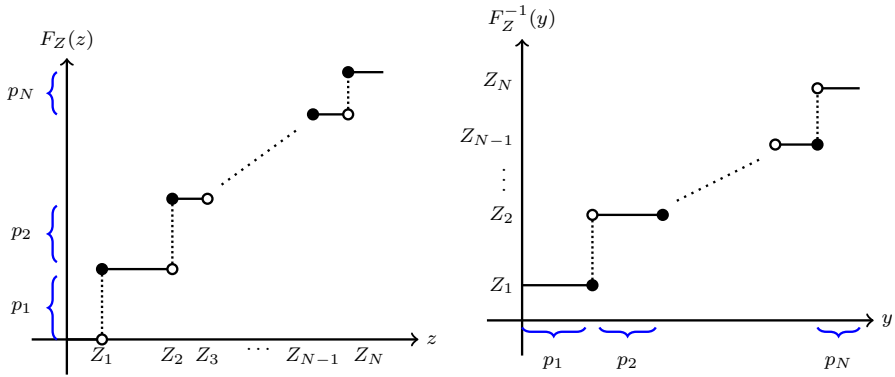
Fig. 2 Cumulative distribution function and quantile function for discrete distributions

respectively. Without loss of generality, assume that $Z_1 < Z_2 < \ldots < Z_N$. Otherwise, if $Z_i = Z_{i+1}$ for some index i, we can combine them and consider $N-1$ losses only. As mentioned in Remark 2, $\rho_\phi(Z)$ is finite. We set $F_Z^{-1}(0) = \operatorname{ess\,inf}(Z) = Z_1$ and $F_Z^{-1}(1) = \operatorname{ess\,sup}(Z) = Z_N$. Figure 2 shows the corresponding cumulative distribution function and quantile function.

Define the points $y_0 = 0$ and $y_i = \sum_{j=1}^{i} p_j$ for $i = 1, \ldots, N$. As the quantile function is piecewise constant, it is possible to write the spectral risk measure defined in (2) as

$$\rho_\phi(Z) = \int_0^1 \phi(p) F_Z^{-1}(p) dp = \sum_{i=1}^{N} Z_i \int_{y_{i-1}}^{y_i} \phi(p) dp \,.$$

Letting $\phi_i = \int_{y_{i-1}}^{y_i} \phi(p) dp$ for $i = 1, \ldots, N$, we have $\rho_\phi(Z) = \sum_{i=1}^{n} \phi_i Z_i$, where we suppress the dependence of ϕ_i on probabilities $\{p_i\}_{i=1}^{N}$. Our next task is to find a distribution $\boldsymbol{p} = (p_1, \ldots, p_N)^\mathsf{T}$ such that the function ρ_ϕ is maximized. The following lemma shows that the objective function is concave and hence, any convex optimization solver can be deployed to obtain the worst-case distribution.

Lemma 2 *Define $\psi : \mathbb{R}^N \to \mathbb{R}$ by the spectral risk measure $\psi(\boldsymbol{p}) = \rho_\phi(Z) = \sum_{i=1}^{N} \phi_i(\boldsymbol{p}) Z_i$. Then, $\psi(\boldsymbol{p})$ is concave.*

Lemma 2 shows that we can employ conventional convex optimization solvers to obtain the global maximizer. However, it may be costly if the number of scenarios N is large. For a special case of spectral risk measure CVaR, the convex problem can be reduced to a linear programming (LP) problem. Notice that

$$\mathrm{CVaR}_\alpha(Z) = \sup_{\boldsymbol{\zeta} \in \mathcal{U}(\mathbb{P})} \mathbb{E}_{\mathbb{P}}(\zeta Z) = \sup_{\boldsymbol{\zeta} \in \mathcal{U}(\mathbb{P})} \sum_{i=1}^{N} p_i \zeta_i Z_i \,,$$

where $\mathcal{U}(\mathbb{P}) = \left\{ \boldsymbol{\zeta} \in \mathbb{R}^N \,|\, \sum_{i=1}^N p_i \zeta_i = 1 \,,\, 0 \leq \zeta_i \leq \frac{1}{1-\alpha} \,,\, i = 1, \ldots, N \right\}$ is a convex compact set. To determine the worst-case distribution, we need to maximize p under the moment constraints as well, which can be cast into following maximization problem.

$$
\begin{aligned}
\underset{\boldsymbol{p}, \boldsymbol{\zeta} \in \mathbb{R}^N}{\text{maximize}} \quad & \sum_{i=1}^N p_i \zeta_i Z_i \\
\text{subject to} \quad & \sum_{i=1}^N p_i \zeta_i = 1 \,, \\
& 0 \leq \zeta_i \leq \tfrac{1}{1-\alpha} \,,\, i = 1, \ldots, N \,, \\
& \boldsymbol{p} \in \mathcal{P}^{\mathrm{d}},
\end{aligned}
$$

where we remark that moment constraints on p are linear. To reformulate it into an LP problem, let $\eta_i = p_i \zeta_i$ for $i = 1, \ldots, N$.

$$
\begin{aligned}
\underset{\boldsymbol{p}, \boldsymbol{\eta} \in \mathbb{R}^N}{\text{maximize}} \quad & \sum_{i=1}^N \eta_i Z_i \\
\text{subject to} \quad & \sum_{i=1}^N \eta_i = 1 \,, \\
& 0 \leq \eta_i \leq \tfrac{1}{1-\alpha} p_i \,,\, i = 1, \ldots, N \,, \\
& \boldsymbol{p} \in \mathcal{P}^{\mathrm{d}}.
\end{aligned}
$$

The computational cost can be significantly reduced by transforming the general convex program into an LP problem. The mixed-CVaR measure defined as $\sum_{i=1}^K \lambda_i \, \mathrm{CVaR}_{\alpha_i}(Z)$ for some $\alpha_i \in [0, 1)$ and λ_i satisfying $\sum_{i=1}^K \lambda_i = 1$ could be argued in a similar way to arrive at an LP problem. This includes the popular risk measure $\lambda \mathbb{E}(Z) + (1 - \lambda) \, \mathrm{CVaR}_\alpha(Z)$ in the literature.

For a more general spectral function ϕ, we may want to approximate the spectral function by step functions so that the resulting risk measure becomes the mixed-CVaR measure. The l^2 approximation is discussed and applied in [57]. We will apply the approximation and examine its performance in this paper on portfolio problem.

Remark 6 It is worthwhile to notice that in Step 3 of the algorithm proposed (Algorithm 4) in [57], the authors suggested using a simple sorting to maintain the order of newly generated iteration points. However, this involves a delicate choice of step sizes. For gradient projection method, it can be modified by replacing the sorting component by a quadratic programming problem. The choice of step sizes would impose less influence to the optimization as long as a sufficiently large number of iterations are permitted. The details of the approximation and one particular example is provided in the Supplementary material (Section B) due to space limitations.

4.2 The Numerical SDDP Method

In this section, we present a variant of SDDP to solve (1) under finite scenarios. It should be noted that the algorithm proposed is not restricted to this particular problem. For instance, the method works for inventory problem or unit commitment problem. For the return random variables $\boldsymbol{r}_t$, $t = 1, \ldots, T$, we assume that they take N_t realizations. Without loss of generality and for notational simplicity, we assume that $N_t = N$ is fixed. Denote $\boldsymbol{r}_t^i$, $i = 1, \ldots, N$ be a particular realization of $\boldsymbol{r}_t$, which

can be represented using a scenario tree with a total number of N^T scenarios. The SDDP algorithm consists of two crucial operations, namely the forward pass and the backward pass. The two operations would repeat alternatively until certain termination criteria are fulfilled.

In the forward pass, a trial solution is obtained, which will be used later in the backward pass. For the first iteration, this refers to the search of any feasible solutions. For instance, one could start with the equally-weighted portfolio or the semi-analytical solution. Assume that we have an approximation $\widehat{\mathcal{Q}}(x_{t-1})$ of $\mathcal{Q}_t(x_{t-1})$, which is obtained in the backward pass. We then sample uniformly from all possible scenarios for a path $(r_1, r_2, \ldots, r_T)$. To obtain a trial point, for $t = 0, \ldots, T-1$, we solve forwardly in time for x_t via the problem of

$$
\begin{aligned}
\underset{x_t}{\text{minimize}} \quad & \widehat{\mathcal{Q}}_{t+1}(x_t) \\
\text{subject to} \quad & \mathbf{1}^\mathsf{T} x_t = r_t^\mathsf{T} \bar{x}_{t-1}, \\
& x_t \geq \mathbf{0},
\end{aligned}
\tag{11}
$$

where $\bar{x}_{t-1}$ is the solution of the problem at $t-1$ and we let $r_0^\mathsf{T} \bar{x}_{-1} = W_0$. Indeed, the approximation functions constructed are piecewise linear taking the form $\widehat{\mathcal{Q}}_{t+1}(x_t) = \max_{l \in \mathcal{L}} \{\alpha_{t+1}^l + (\beta_{t+1}^l)^\mathsf{T} x_t\}$, for some index set $\mathcal{L}$. Then, (11) can be recast as

$$
\begin{aligned}
\underset{x_t, \theta}{\text{minimize}} \quad & \theta \\
\text{subject to} \quad & \theta \geq \alpha_{t+1}^l + (\beta_{t+1}^l)^\mathsf{T} x_t, \ l \in \mathcal{L}, \\
& \mathbf{1}^\mathsf{T} x_t = r_t^\mathsf{T} \bar{x}_{t-1}, \\
& x_t \geq \mathbf{0}, \ \theta \in \mathbb{R}.
\end{aligned}
\tag{12}
$$

A trial point is thus, generated and denoted as $\{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{T-1}\}$.

In the backward pass, we obtain linear approximations of the functions $\mathcal{Q}_t(x_{t-1})$ by proceeding backwardly in time, from $t = T$ to $t = 0$. The idea is the use of subgradient inequality at the trial points. Given $\bar{x}_{t-1}$, we solve, for $i = 1, \ldots, N$,

$$
\begin{aligned}
\underset{x_t}{\text{minimize}} \quad & f_t(x_t, r_t) + \widehat{\mathcal{Q}}_{t+1}(x_t) \\
\text{subject to} \quad & x_t \in \mathcal{X}_t(\bar{x}_{t-1}, r_t^i)
\end{aligned}
\tag{13}
$$

and denote the optimal value as $\widetilde{Q}_t^i$. Notice that $\widetilde{Q}_T^i = -(r_T^i)^\mathsf{T} \bar{x}_{T-1}$. Similar to the case in forward pass, for $t = T-1, \ldots, 1$, we can recast (13) as

$$
\begin{aligned}
\underset{x_t, \theta}{\text{minimize}} \quad & \theta \\
\text{subject to} \quad & \theta \geq \alpha_{t+1}^l + (\beta_{t+1}^l)^\mathsf{T} x_t, \ l \in \mathcal{L}, \\
& \mathbf{1}^\mathsf{T} x_t = (r_t^i)^\mathsf{T} \bar{x}_{t-1}, \\
& x_t \geq \mathbf{0}, \ \theta \in \mathbb{R}.
\end{aligned}
\tag{14}
$$

Denote (π_t^i, μ_t^i) be the corresponding optimal dual solutions to the equality and inequality constraints respectively. More precisely, for $t = T-1, \ldots, 1$,

$$\left(\pi_t^i, \ \boldsymbol{\mu}_t^i\right) \in \underset{\pi, \, \boldsymbol{\mu} \geq 0}{\arg\max} \ \underset{\theta, \, \boldsymbol{x}_t \geq \mathbf{0}}{\min} \left\{\theta + \sum_{l \in \mathcal{L}} \mu_l \left(\alpha_{t+1}^l + (\boldsymbol{\beta}_{t+1}^l)^\mathsf{T} \boldsymbol{x}_t - \theta\right)\right.$$
$$\left. + \pi \left[\mathbf{1}^\mathsf{T} \boldsymbol{x}_t - \left(\boldsymbol{r}_t^i\right)^\mathsf{T} \bar{\boldsymbol{x}}_{t-1}\right]\right\}, \tag{15}$$

and let $\pi_T^i = 1$ for $i = 1, \dots, N$. Next, we need to obtain the worst-case distribution based on the losses $\{\widetilde{Q}_t^i\}$ as discussed in previous section. Denote the optimal weight ϕ_i associated to $\widetilde{Q}_t^i$ by ϕ_i^*. Let $\widetilde{\theta}_t = \sum_{i=1}^N \phi_i^* \widetilde{Q}_t^i$ and $\widetilde{\boldsymbol{g}}_t = -\sum_{i=1}^N \phi_i^* \pi_t^i \boldsymbol{r}_t^i$. The coefficients for the linear approximation function are computed as

$$\alpha_t = \widetilde{\theta}_t - (\widetilde{\boldsymbol{g}}_t)^\mathsf{T} \bar{\boldsymbol{x}}_{t-1} \quad \text{and} \quad \boldsymbol{\beta}_t = \widetilde{\boldsymbol{g}}_t. \tag{16}$$

The optimal value of the first stage problem (14) with $t = 0$ is denoted as $\underline{z}^k$, where k is the iteration number. We would verify later in Lemma 6 in Sect. 5 that $\underline{z}^k$ is an improving lower bound for the problem as k increases. Concerning the termination criterion of the algorithm, it is well known that maintaining an upper bound for risk averse problems in each iteration is difficult (see, for instance, [56]). Hence, we resort to the maximum iteration number and stabilization of the lower bound. To be precise, we fix K as the maximum number of iterations and ϵ as the lower bound tolerance. The complete algorithm is summarized in Algorithm 1.

Algorithm 1: SDDP for distributionally robust problem with moment constraints

    Initialization: $\mathcal{Q}_t^0(\cdot) \equiv -\infty$ for all t, $\mathcal{Q}_{T+1}^k(\cdot) \equiv 0$ for all k, $\epsilon > 0$, $k = 1$, $\underline{z}^0 = -\infty$

1   **while** $\underline{z}^k - \underline{z}^{k-1} \geq \epsilon$ **and** $k < K$ **do**
2   Sample a scenario $\{\boldsymbol{r}_1, \dots, \boldsymbol{r}_T\}$.
3   **for** $t = 0$ **to** $T - 1$ **do**
4   Solve (12) for a feasible policy $\{\bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_2, \dots, \bar{\boldsymbol{x}}_{T-1}\}$.
5   (For $k = 1$, any feasible policy suffices.)
6   **end**
7   **for** $t = T$ **to** 1 **do**
8   Solve (14) for $\widetilde{Q}_t^i$, $i = 1, \dots, N$ and obtain the worst-case distribution ϕ_i^*.
9   Compute α_t^k and $\boldsymbol{\beta}_t^k$ using (16).
10   Update $\widehat{\mathcal{Q}}_t^k(\boldsymbol{x}_{t-1}) = \max\{\widehat{\mathcal{Q}}_t^{k-1}(\boldsymbol{x}_{t-1}), \ \alpha_t^k + (\boldsymbol{\beta}_t^k)^\mathsf{T} \boldsymbol{x}_{t-1}\}$.
11   **end**
12   Update the lower bound $\underline{z}^k$ by solving (14) with $t = 0$. Set $k = k + 1$.
13 **end**

---

5 Convergence Property

In this section, we verify the convergence property of the algorithm. The analysis is established mainly based on the results of [21] and [22]. One should notice that in [21], their contribution is establishing the convergence of multistage risk neutral programs

while [22] extends this to the risk averse setting. However, in [22], the proof is based on the direct use of dual representation of coherent risk measure, which implies the existence of a reference probability while for the current problem, such a reference probability does not exist. Therefore, some modification is necessary in order to justify the convergence of the algorithm.

5.1 Recourse Function and Subgradient

To guarantee the convergence of the algorithm, we utilize some fundamental properties of the recourse function $Q_t(x_{t-1})$ defined in (7). Recall that we have N scenarios in each future time point and they are positive in value. It follows that for any feasible policy $\{x_0, \ldots, x_{t-1}\}$ and realization of $\{r_1, \ldots, r_t\}$, the feasibility set $\mathcal{X}_t(x_{t-1}, r_t)$ is nonempty, convex and compact. We suppress the dependence of ρ on ϕ for simplicity. Next, define, for $t = 1, \ldots, T$, $c_t = \max_{i=1, \ldots, N} \|r_t^i\|_\infty$ and $C_t = \prod_{j=1}^t c_j$ with $c_0 = C_0 = 1$. Under finite scenarios, we have c_t and C_T being finite. The sets $X_t = [0, C_t]^n \subseteq \mathbb{R}^n$ are also nonempty, convex and compact. With initial wealth $W_0 = 1$, we can restrict feasible sets $\mathcal{X}_t$ on X_t.

Lemma 3 *The recourse functions $Q_t(x_{t-1})$ for $t = 1, \ldots, T$ are finite and convex on X_t.*

Remark 7 For any given decision $x_i \in X_i$ for $i = 1, \ldots, t-1$ and random return vector, the feasibility set $\mathcal{X}_t$ is always non-empty. As described in [22], this follows that $Q_t(x_{t-1})$ is continuous on X_t.

Indeed, SDDP method utilizes the convexity of recourse functions $Q_t(x_{t-1})$ and constructs lower approximations for them. The principle is similar to the popular cutting-plane method. The construction of lower approximations is based on subgradient inequality and the following lemma provides a justification for the computation of the coefficients in (16). The lemma extends Proposition 1 in [47] to the use of risk measure instead of expectation.

Lemma 4 *Define $\psi(x) = \sup_{\mathbb{P} \in \mathcal{P}} \rho_\phi(Z(x, \omega))$. Let $Z : \mathbb{R}^n \times \Omega \to \mathcal{Z}$ be a convex function in x for fixed $\omega \in \Omega$ and $z(\bar{x}, \omega) \in \partial Z(\bar{x}, \omega)$ for some $\bar{x} \in \text{dom}(\psi)$. Assume that for $\rho_\phi(Z(\bar{x}, \omega))$, ϕ_i^* is the optimal weight corresponding to the worst-case distribution $\mathbb{P}^* \in \mathcal{P}$ (identified as an $\mathbb{R}^N$ vector). Then, $g := \sum_{i=1}^N \phi_i^* z(\bar{x}, \omega)$ is a subgradient of $\psi(x)$ at $\bar{x}$.*

5.2 Convergence Analysis

In this subsection, we shall verify the theoretical properties of the algorithm, including the convergence. The two assumptions imposed throughout the analysis are stated below.

Assumption 1 The randomnesses $\{r_t\}_{t=1, \ldots, T}$ are stagewise independent and the number of realizations in each stage is finite with support $\{r_t^i\}_{i=1, \ldots, N}$.

Assumption 2 Every node has a positive probability of being selected in forward pass, i.e. $\bar{\mathbb{P}}(\bar{r}_t^k = r_t^i) > 0$ for $t = 1, \ldots, T$, $i = 1, \ldots, N$ and $k \in \mathbb{N}$, where $\bar{r}_t^k$ is the random variable representing the selection process. Also, $\bar{r}_t^k$ is independent of the previous selection process.

We first give a lemma, followed by a corollary that can reduce significantly the computational cost in the backward pass. Without further specification, we shall omit $\bar{\mathbb{P}}$-a.s. in the proof for simplicity. Notice by the fact that the feasibility sets are non-empty for any implementable policies and realizations, the coefficients α_t^l and β_t^l are finite. Recall the dual problem of (14).

$$
\begin{aligned}
\underset{\boldsymbol{\mu}, \pi}{\text{maximize}} \quad & \sum_{l \in \mathcal{L}} \mu_l \alpha_{t+1}^l - \pi \bar{\boldsymbol{x}}_{t-1}^\mathsf{T} \boldsymbol{r}_t^i \\
\text{subject to} \quad & \sum_{l \in \mathcal{L}} \mu_l = 1, \\
& \sum_{l \in \mathcal{L}} \mu_l \boldsymbol{\beta}_{t+1}^l + \pi \mathbf{1} \geq \mathbf{0}, \\
& \boldsymbol{\mu} \geq \mathbf{0}, \; \pi \in \mathbb{R}.
\end{aligned}
$$

Lemma 5 *With $\bar{\mathbb{P}}$-almost surely, for $t = 1, \ldots, T$, $\alpha_t^l = 0$ and $\boldsymbol{\beta}_t^l \leq \mathbf{0}$ for all $l \in \mathbb{N}$.*

Corollary 1 *With $\bar{\mathbb{P}}$-almost surely, the dual optimal solutions are the same within each time step, i.e. $\pi_t^i = \pi_t^j$ for $t = 1, \ldots, T-1$ and $i, j = 1, \ldots, N$.*

Making use of Lemma 5 and Corollary 1, we do not need to compute the coefficient α as it is always equal to zero. Moreover, instead of solving N problems to obtain $\widetilde{Q}_t^i$ and π_t^i, we need to solve only one program. By rescaling, we obtain all other $\widetilde{Q}_t^i$ while the dual solutions are the same. By doing so, the computational cost in backward pass can be substantially reduced.

As a final step, we now justify the convergence of the algorithm: The lower approximation property is going to be verified and the a.s. convergence is established at the end. Throughout the algorithm, we introduce the following two functions at iteration k, namely

$$
\widehat{Q}_t^k(\boldsymbol{x}_{t-1}) = \max_{l=1,\ldots,k} \left\{ \alpha_t^l + \left(\boldsymbol{\beta}_t^l \right)^\mathsf{T} \boldsymbol{x}_{t-1} \right\}, \; t = 1, \ldots, T, \text{ and} \tag{17}
$$

$$
\widetilde{Q}_t^k(\boldsymbol{x}_{t-1}, r_t) = \inf_{\boldsymbol{x}_t \in \mathcal{X}_t(\boldsymbol{x}_{t-1}, r_t)} \left\{ f_t(\boldsymbol{x}_t, \boldsymbol{r}_t) + \widehat{Q}_{t+1}^k(\boldsymbol{x}_t) \right\}, \; t = 0, \ldots, T. \tag{18}
$$

Function (17) is the constructed approximation function of the original recourse functions $\mathcal{Q}_t(\boldsymbol{x}_{t-1})$ while function (18) is the optimization problem in each stage by replacing the unknown recourse function $\mathcal{Q}_t(\boldsymbol{x}_{t-1})$ by its current approximation. We emphasize that the notation $\widetilde{Q}_t^i$ (see description above (14)) should not be confused with the function defined in (18). Precisely, $\widetilde{Q}_t^i = \widetilde{Q}_t^k(\boldsymbol{x}_{t-1}, \boldsymbol{r}_t^i)$ at iteration k. We may denote the value of ρ under distribution $\mathbb{P}$ as $\rho^{\mathbb{P}}$ where necessary. Lemma 6 shows that the algorithm constructs lower approximation to the recourse function and Lemma 7 is a technical lemma which helps develop a crucial inequality in the convergence proof.

Lemma 6 *The algorithm generates lower approximations, i.e. $\bar{\mathbb{P}}$-almost surely, $\mathcal{Q}_t(\cdot) \geq \widehat{Q}_t^k(\cdot)$ for all $k \in \mathbb{N}$ on X_t for $t = 1, \ldots, T$.*

Lemma 7 *For $t = 1, \ldots, T$, $\widehat{\mathcal{Q}}_t^k(\bar{\boldsymbol{x}}_{t-1}^k) = \sup_{\mathbb{P}_t \in \mathcal{P}_t} \rho_t\left(\widetilde{Q}_t^k(\bar{\boldsymbol{x}}_{t-1}^k, \boldsymbol{r}_t)\right) = \widetilde{\theta}_t^k \, \mathbb{P}$-almost surely for all $k \in \mathbb{N}$.*

With the help of previous two lemmas, we can develop the convergence property of the proposed algorithm, which demonstrates the point-wise convergence of the approximation function (17) to the true recourse function and hence, the convergence to the optimal solution.

Theorem 3 *Consider Algorithm 1. The following two statements hold $\bar{\mathbb{P}}$-almost surely:*

(a) *For $t = 1, \ldots, T + 1$, $\lim_{k \to \infty}[\mathcal{Q}_t(\boldsymbol{x}_{t-1}^k) - \widehat{\mathcal{Q}}_t^k(\boldsymbol{x}_{t-1}^k)] = 0$, where $\boldsymbol{x}_{t-1}^k$ is any possible feasible solution at time $t - 1$ and iteration k.*

(b) *Denote z^* as the optimal solution and $\boldsymbol{x}_0^k$ be the optimal solution to (13) at $t = 0$. Then, $\lim_{k \to \infty} \underline{z}^k = z^*$ and any accumulation point of $\{\boldsymbol{x}_0^k\}_{k \in \mathbb{N}}$ is an optimal solution for the first stage problem.*

6 Computational Results

In this section, we first demonstrate the use of approximation and the semi-analytical solutions. Next, we compare problem (1) with the one without correlation ambiguity under simulation. That is, we assume that the first two moments are known completely and the same algorithm can be applied to solve by equating the lower and upper correlation bounds. Finally, we apply the model to a real data set. The distributional robust model, as expected, performs better under shocks. All the computations are performed using MATLAB in a computer using 3.60 GHz Intel Core i7 with 16 GB RAM and the time is presented in second.

6.1 Convergence and Approximation

To visualize the convergence, we consider a case with small numbers of stages. In particular, we consider $T = 1$ and $T = 2$, where the upper bound can be computed by evaluating the objective function directly using the trial solution x_1^k in each iteration. Interior-point methods are used to solve LP problems with the termination criterion being the stabilization of the gap between lower and upper bound. The gap tolerance is set to be 10^{-6}. We use 45 stocks from the Hang Seng Index (HSI) with relatively longer data history in Hong Kong with the daily historical data obtained from either 26/03/2008 or 27/05/2017 to 27/05/2019. The data is obtained from Yahoo! Finance through the MATLAB function `hist_stock_data` developed by [50]. The returns are computed using closing prices. The individual asset mean and variance are estimated using sample quantities while the correlation bounds are quantified by 95% confidence intervals.

The spectral risk measure adopted in this example is the exponential type, which is proposed in [11] and followed by [7]. The spectral function of this type is defined as $\phi(p) = h(1 - e^{-h})^{-1} e^{-h(1-p)}$ for $h > 0$. A larger value of h corresponds to a higher degree of risk aversion. Figure 3 shows the convergence of the case $h = 2$ with $N = 2739$ return scenarios and $T = 1$ while the graphs of other testing cases are
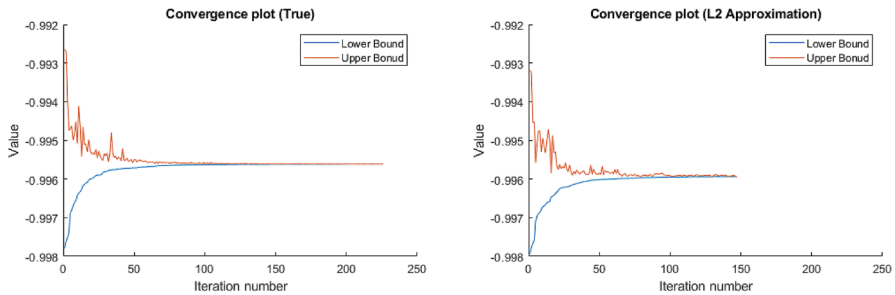
Fig. 3 Convergence plot using $h = 2$ for the exact case (left) and approximation case (right)

similar. To examine the solution quality between the one using spectral risk measure and the one using the approximation, the solution obtained via approximation is used to perform the full evaluation of the objective in spectral risk measure. Relative error is used as a criterion for measuring the performance, which is computed as the absolute difference between the two obtained objective function values divided by the true value. Table 1 summarizes the results under different settings identified as a triple (T, h, N).

We observe that in all the testing cases, convergence is achieved after a relatively small number of iterations. This corroborates the convergence result discussed in Theorem 3. For the computational time, as expected, the use of approximation can remarkably reduce the computational effort by around 6 to 10 times. This can be elucidated by the use of LP solver for CVaR measures while the general spectral risk measure resorts to convex optimization solver. By exploiting the linear structure, LP solver could perform better in time. Though the computational time is much reduced, the solution remains reasonably accurate, with relative errors staying in the order of 10^{-5}. We also remark that the resulting solutions exhibit sparsity. The solutions in all the cases contain less than 20 non-zero weights and in most cases, the number of non-zero weights is around 10. The under-diversification result agrees with recent literature, for example, [45]. Conventionally, robust portfolios are thought to be conservative with a large degree of diversification. Indeed, uncertainty in individual asset performance usually leads to a well diversified portfolio as a result of our confident view towards the estimated correlation structure. However, when we consider the risk associated to the unknown correlation structure, anti-diversification would be counter-intuitively more desirable since the assets, in the worst-case scenario, may fluctuate in a similar fashion. Focusing on a few assets appears to be a more robust choice. In other words, ignoring ambiguity in the correlation matrix concerned may oftentimes result in a misleading conclusion that a more diversified portfolio is robust and risk averse.

As we could expect, by increasing the number of scenarios, it would lead to an increment in the objective value. In the Supplementary material, we conduct a simulation study, based on the same data set, to study the effect of number of scenarios. We also examine the convergence under various choices of the initial guess (with results provided in the Supplementary material). We observe that our algorithm is relatively insensitive to the choice of initial guess. Finally, note that the semi-analytical solution provides an upper bound to the discretized problem. Indeed, in this experiment, we

Table 1 Approximating solution qualities under different settings identified as a triple (T, h, N)

	(1, 2, 489)			(1, 2, 2739)		
	True	Approx	Relative error	True	Approx	Relative error
Value	−0.996936	−0.996908	2.83×10^{-5}	−0.995612	−0.995587	2.47×10^{-5}
Time	1347.40	176.54	(7.63 times)	23,703.35	4165.69	(5.69 times)
	(1, 10, 489)			(2, 2, 489)		
	True	Approx	Relative error	True	Approx	Relative error
Value	−0.988067	−0.988012	5.55×10^{-5}	−0.997656	−0.997628	2.75×10^{-5}
Time	2073.44	333.45	(6.22 times)	30,691.29	2562.99	(11.97 times)

also observe that the optimal value from the numerical scheme is always less than the one given by the semi-analytical solution. We further investigate the difference between the semi-analytical solution and the numerical scheme if we have extra support information on the return vector in Sect. 6.2.

6.2 Use of Support Information

In order to compare the semi-analytical solution and the solution obtained from the numerical method, we conduct a simulation study to examine the effect of support information. We assume that there are 5 assets with mean 1, standard deviations $\{0.1, 0.05, 0.09, 0.1, 0.08\}$ and correlation matrix

$$C = \begin{pmatrix} 1 & 0.3 & -0.3 & 0.4 & 0.2 \\ 0.1 & 1 & 0.1 & -0.1 & 0.2 \\ -0.3 & 0.1 & 1 & 0.15 & 0.3 \\ 0.4 & -0.1 & 0.15 & 1 & 0.05 \\ 0.2 & 0.2 & 0.3 & 0.05 & 1 \end{pmatrix}.$$

The lower (upper) correlation bounds are constructed by subtracting (adding) 0.1 from (to) the correlation matrix C.

In the experiment, we first generate N_1 scenarios from uniform distribution on $(1 - \delta_1, 1 + \delta_1)$ entry-wisely. We use uniform distribution to confine the support of the return vector on a bounded region. Next, we add additional $N_2 - N_1$ scenarios generated from uniform distribution on $(1 - \delta_2, 1 + \delta_2)$ for some $\delta_2 > \delta_1$. That is, we enlarge the scenario set and the support of the scenarios. We repeat this process to generate 4 sets of scenarios with $N_1 = 1000$, $N_2 = 2000$, $N_3 = 4000$, $N_4 = 8000$, $\delta_1 = 0.12, \delta_2 = 0.15, \delta_3 = 0.2$ and $\delta_4 = 0.3$. We solve the resulting problem using the 4 generated sets of scenarios by our proposed numerical method with $T \in \{1, 2, 5\}$ and we use the approximation of exponential type spectral risk measure with $h \in \{2, 10\}$ as the objective. We record the results after 200 iterations, where we observe the lower bound is already stabilized.

Table 2 presents the results of the simulation study. Each column presents one generated scenario set with total number of scenarios N and support $(1 - \delta, 1 + \delta)$. The last column shows the optimal value from the semi-analytical solution via the SDP reformulation. For each pair of (T, h), we present the terminal value from the algorithm and compute the l_1 norm difference between the solution obtained from the algorithm and the one by the semi-analytical solution. We observe that when we enlarge the support (i.e., increase δ), the optimal value, as well as the optimal solution, approach the one obtained from the semi-analytical solution. Also, the optimal value from the semi-analytical solution always serves as an upper bound. We notice that if there exists support information (i.e., δ is small), the optimal solution could behave differently from the one from the semi-analytical solution.

Table 2 Optimal value and the l_1 norm difference in optimal solution between semi-analytical solution and numerical solution

	δ	0.12	0.15	0.2	0.3	SDP
	N	1000	2000	4000	8000	
$T = 1, h = 2$	Opt. val.	-0.9806	-0.9791	-0.9791	-0.9791	-0.9759
	l_1 diff in opt sol.	0.1084	0.0488	0.0753	0.0647	NA
$T = 1, h = 10$	Opt. val.	-0.9440	-0.9359	-0.9236	-0.9233	-0.9138
	l_1 diff in opt sol.	0.8114	1.0981	0.0800	0.0159	NA
$T = 2, h = 2$	Opt. val.	-0.9616	-0.9586	-0.9586	-0.9585	-0.9524
	l_1 diff in opt sol.	0.1116	0.0922	0.0492	0.0462	NA
$T = 2, h = 10$	Opt. val.	-0.8912	-0.8760	-0.8529	-0.8525	-0.8351
	l_1 diff in opt sol.	0.8114	1.0979	0.0559	0.0345	NA
$T = 5, h = 2$	Opt. val.	-0.9066	-0.8997	-0.8996	-0.8996	-0.8852
	l_1 diff in opt sol.	0.1111	0.0474	0.1246	0.0551	NA
$T = 5, h = 10$	Opt. val.	-0.7497	-0.7182	-0.6718	-0.6710	-0.6373
	l_1 diff in opt sol.	0.8110	1.0978	0.0755	0.0161	NA

6.3 Use of Correlation Ambiguity

In order to examine the importance of correlation ambiguity, we perform a simulation study on three different models: (a) model (1) which excludes correlation ambiguity (complete knowledge in second moments), (b) model (1) which includes correlation ambiguity and (c) a simple single-period global minimum variance model using sample estimates. In this study, we use multivariate normal distribution for scenario generation under Setting I, which is provided in the Supplementary material. First batch of 2000 scenarios based on the settings are generated, which are considered as historical data and they are used to obtain the investment decision. Afterwards, second batch of 10,000 future scenarios are generated and the terminal wealth under three different decisions are compared to investigate the performance. The second batch of scenarios is generated using the worst-case correlation matrix obtained when solving problem (1).

For the model setting, we consider $T = 21$ as a monthly trading exercise such that the period is long enough to observe the alteration in correlation structure. The obtained portfolio weight is adopted throughout the whole period. We employ the approximation of exponential type spectral risk measure with $h = 10$, which is given by $0.3546\mathbb{E}(Z) + 0.6454\,\mathrm{CVaR}_{0.884}(Z)$. The histogram of the simulated terminal wealth is provided in Fig. 4. For other values of h, similar patterns are observed.

We observe that the terminal wealth using the model with correlation ambiguity result in a right shift of the empirical terminal wealth distribution when compared with the other two models. This demonstrates the superiority of the model using correlation ambiguity in terms of protection under change in correlation structure. Similar observations can also be found under other settings and the corresponding results are summarized in the Supplementary material. If the period T is not long
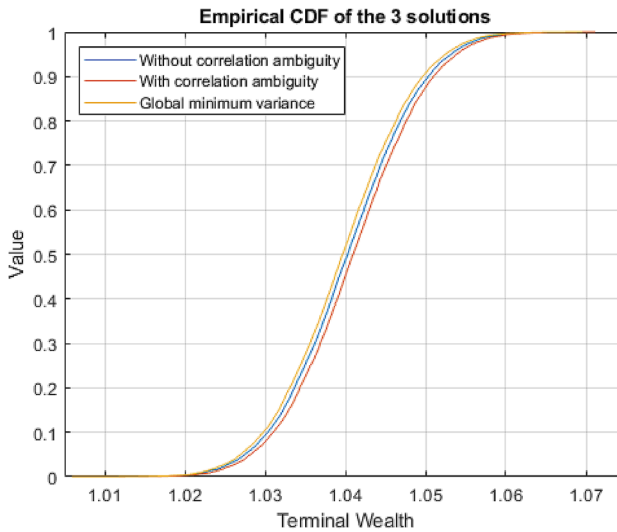
Fig. 4 Simulation study under setting I

enough, the advantage of our proposal tends to become less noticeable, which is reasonable as the model aims at protecting investors from changes in the underlying correlation structure amongst individual assets. To summarize, this study indicates the importance of correlation ambiguity in portfolio optimization model.

6.4 Real Data Application

This section presents a real data study with reference to that presented in [29]. We use historical data of six major indices representing different parts of the world, namely Nikkei225, FTSE100, Nasdaq, DAX30, Sensex and Bovespa, from 01/11/1998 to 16/06/2019. The indices are normalized to 1 on 01/07/2003. Figure 5 displays the historical paths of the selected six indices. We compare the following three models: the proposed one with correlation ambiguity, single-period global minimum variance model and distributionally robust model with the expectation objective under l_2 norm in [47] with the parameter $r = 0.05$ indicating the maximum l_2 distance from nomial probability. The parameter is chosen with reference to the three choices (0.033, 0.066, 0.133) in [47]. We also test for some other choices of r and the results are presented in the Supplementary material.

First, we compare the overall performance of the proposed model under different values of parameter h. To examine the protective effect under a shock in financial market, prices prior to 04/01/2007 are used for training purpose and the trading ends on 31/12/2013. We also perform another analysis using a longer period, which is from 01/07/2003 to 16/06/2019 and the results are demonstrated in the Supplementary material. The approximation of exponential type spectral risk measure with three different parameters $h = 2, 10, 15$ are employed, which are given in Table 3. The numerical result is based on the monthly rebalancing portfolio management strategy.
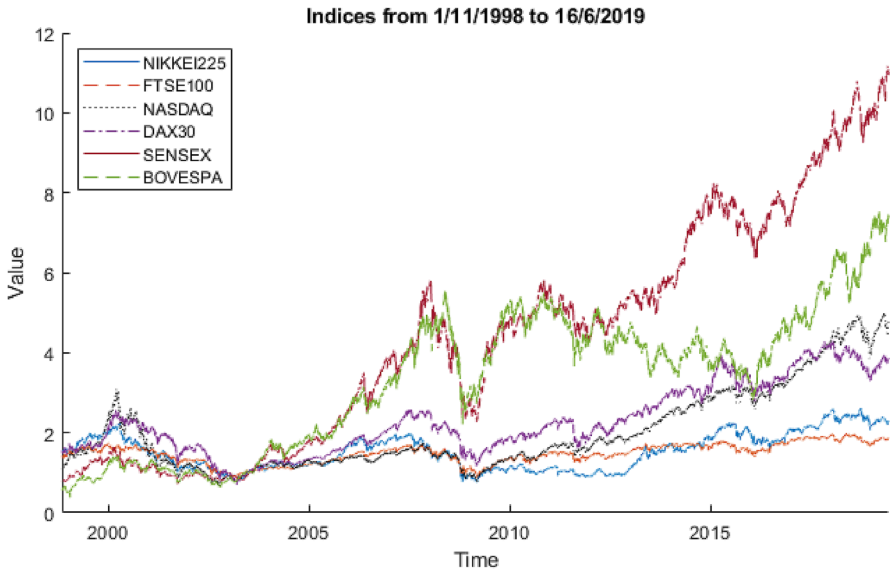
Fig. 5 Historical data for the six indices

At the first day of each month, a T period problem (1) is solved with correlation ambiguity using all previously observed data with T being the number of calendar days. Number of calendar days is used instead of trading days since the non-trading days vary with each index.

The result is shown in Fig. 6. The upper panel indicates the wealth process with unit initial wealth while the lower panel gives the difference between the reference model $h = 15$ and the other two choices. We observe that the three models perform similarly with profits at the end of the testing period. However, it could be noticed that $h = 15$ eventually outperforms the other two models, which is reasonable since this is the most risk-averse one when compared with the other two. The difference between $h = 10$ and $h = 15$ is not significantly distinguishable mainly due to the similarity of the approximated risk measures. For the model with $h = 2$, it is natural that, under a huge market downturn, as a result of a smaller degree of risk aversion, the model ends up with the poorest performance among the three, while it is surprising that the difference is more distinct at the end of testing period, where an upward trend starts. In particular, we examine the portfolios starting from 2013 for the three models. We notice that from Fig. 7, the weight on BOVESPA for that model is larger than the other two models most of the time. Indeed, the arithmetic return over the whole period of the last index is the largest among all indices while it performs the poorest in terms of index value normalized at the beginning of 2013. This could be served as one possible explanation for such a deviation.

The out-of-sample performance statistics are summarized in Table 3. We observe that the overall mean using model $h = 15$ is the largest the variance is the lowest among the three. Also, the mean-to-standard deviation (SD) ratio outperforms the other two models. Another possible measure for out-of-sample performance is the
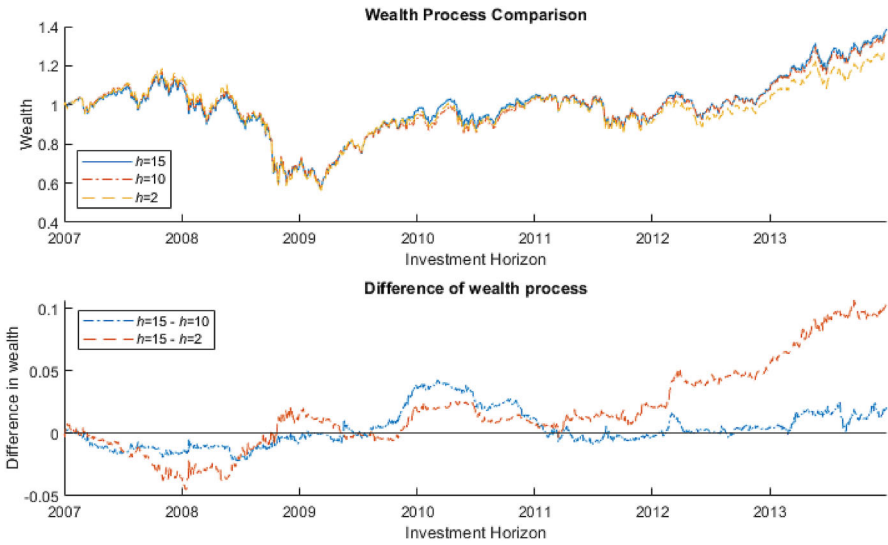
Fig. 6 Wealth process and difference for different values of h: solid blue line for $h = 15$, dash-dot orange line for $h = 10$ and dashed yellow line for $h = 2$ in the upper panel
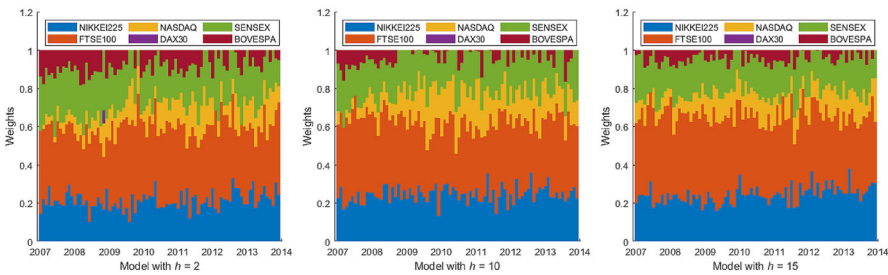

Fig. 7 The weights distribution in each period for different choices of h

drawdown over the period, which is defined as the current value subtracting the historical high and normalized by the historical high. The normalization is used to ensure fair comparison in relative sense. This can measure how shocks may adversely affect the wealth performance. The colored regions indicate that model using $h = 15$ has the smallest drawdown in Fig. 8. It is observable that the shaded region covers the financial crisis and the concomitant recovery period, which indicates that the model has a strong protective power against shocks.

Finally, we compare our proposed models with the aforementioned two models. The single-period global minimum variance portfolio is computed at the beginning of each month while the multi-period model with the same value of T in [47] is employed. Figure 9 shows the computational result. The upper panel demonstrates the wealth process while the lower panel gives the differences. Each column represents a choice of h described previously: $h = 15$ for the left-most column, followed by $h = 10$ and $h = 2$. As expected, both risk-averse model $h = 10$ and $h = 15$

Table 3 Out-of-sample daily return statistics

h	Risk measure	Mean ($\times 10^{-4}$)	Variance ($\times 10^{-4}$)	Mean-to-SD
2	$0.6549 \mathbb{E}(Z) + 0.3541 \, \mathrm{CVaR}_{0.654}(Z)$	1.7467	1.1270	0.01645
10	$0.3546 \mathbb{E}(Z) + 0.6454 \, \mathrm{CVaR}_{0.884}(Z)$	2.0333	1.1041	0.01935
15	$0.3274 \mathbb{E}(Z) + 0.6726 \, \mathrm{CVaR}_{0.920}(Z)$	2.0545	1.0852	0.01972
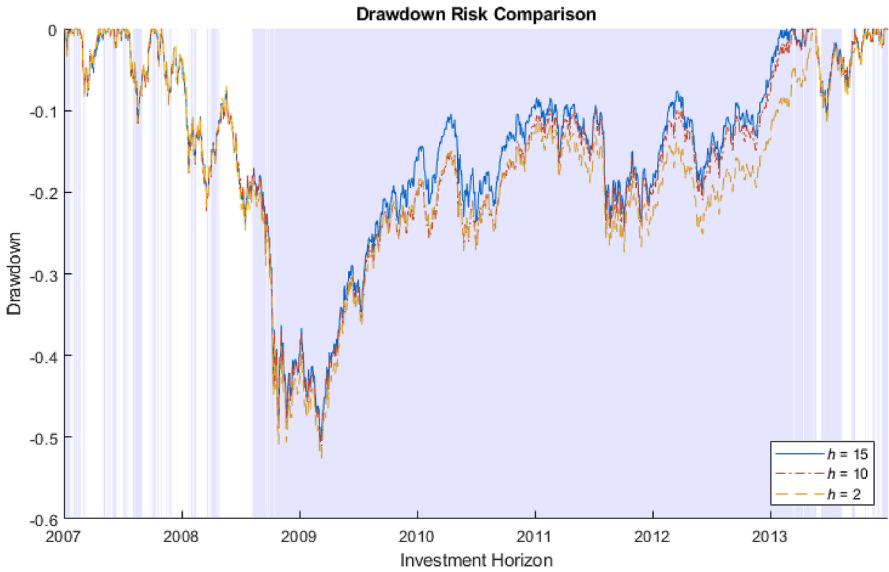
Fig. 8 Drawdown of the model with three choices of h (Colored regions indicate that the model using $h = 15$ has the smallest drawdown)

outperform the other two models. In particular, a larger deviation can be observed in the comparison with single-period global minimum variance model, implying the importance of robustness. The advantage of our proposed model over the one in [47] using l_2 norm could be explained by the change in dependence structure during the period, which coincides with our intention for proposing the model. More importantly, their model uses expectation in the objective function while we use a risk measure. The greater level of risk averseness could also account for the edge. On the right-most column, we observe that the model with $h = 2$ underperforms all the other two models. The global minimum variance model appears to be a less aggressive model which contributes to this result. For the second one, it performs better since they do not have a moment constraint. In our construction of uncertainty set, the means of the returns are fixed while in their construction, moment constraints are not required. Therefore, the flexibility in moment offered from their model yields a better result.
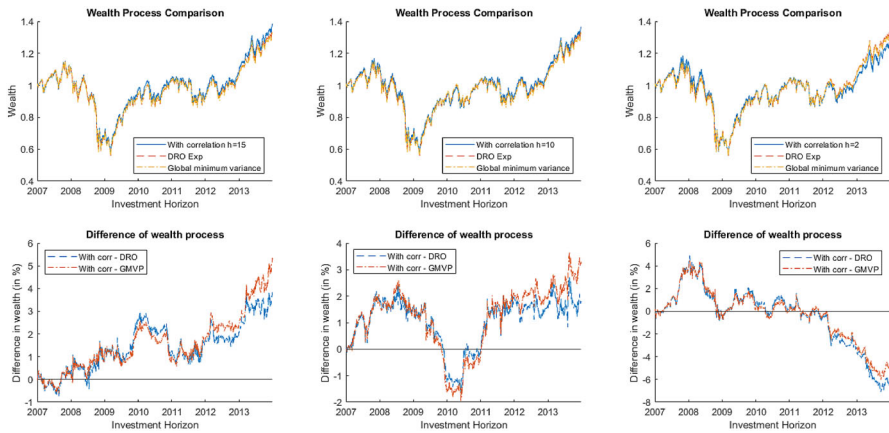
Fig. 9 Starting from the left panel, our model with $h = 15$, $h = 10$ and $h = 2$ are compared with the other two models. Upper panel shows the wealth process in the following order: our prposed model, l_2 uncertainty model (DRO Exp) and the global minimum variance portfolio (GMVP). Lower panel shows the difference between our model and the remaining two models in percentage with respect to initial unit investment in the following order: out-performance of our model compared with DRO Exp and that compared with GMVP

7 Conclusion

Distributionally robust portfolio optimization originates from the uncertainty in the correlation structure amongst different assets, hence the return distributions. Such kind of models is of surging interest due to their protection against unfavorable outcomes. In this paper, we study the multi-period portfolio problem under correlation uncertainty. While previous works focus mainly on mean-variance or utility objective, we incorporate the use of spectral risk measure in our model so that investors can incorporate their risk preference in the portfolio optimization exercise.

To achieve the goal, we first extend the work [35] by considering the moment-type uncertainty set instead of ellipsoidal set. The use of the existing ellipsoidal sets cannot capture explicitly the correlation ambiguity, which is the key concern in this paper. Also, our semi-analytical solution is derived for spectral risk measure objective, which includes their mean-CVaR objective as a special case. Second, we discuss a numerical SDDP method to solve the problem under the discrete distribution setting. The method relies on convex optimization subproblems and can be applied to a more general distributionally robust problem with moment constraints with the almost sure convergence of the algorithm being guaranteed. Our work extends previous literature in the convergence of SDDP method to a distributionally robust setting.

In numerical exercises, we demonstrate that the approximation of spectral risk measure is feasible, which can significantly reduce the computational burden of the algorithm. We also illustrate the importance and usefulness of correlation ambiguity in our model. Specifically, consideration of such ambiguity leads to a higher terminal wealth as demonstrated in both the simulation study and the real data analysis. We also observe the robustness of the model during the period of financial shocks through the drawdown plot, which highlights the purpose of our proposed model.

While the almost sure convergence of the proposed algorithm is guaranteed, it is noteworthy that historical data is employed in the numerical setting and this may not be able to fully evaluate the worst-case spectral risk measures. Possible future research directions include the convergence of the solution under the discrete grid to the true solution and the possibility of using generated scenarios instead of historical data. The quantification of such a convergence may improve the efficiency of the numerical procedure by reducing the computational burden.

Declarations

Competing interest The authors have not disclosed any competing interests.

References

1. Acerbi, C.: Spectral measures of risk: a coherent representation of subjective risk aversion. J. Bank. Financ. **26**(7), 1505–1518 (2002)
2. Ahmadi-Javid, A., Fallah-Tafti, M.: Portfolio optimization with entropic value-at-risk. Eur. J. Oper. Res. **279**(1), 225–241 (2019)
3. Ahmed, S., Çakmak, U., Shapiro, A.: Coherent risk measures in inventory problems. Eur. J. Oper. Res. **182**(1), 226–238 (2007)
4. Asamov, T., Powell, W.B.: Regularized decomposition of high-dimensional multistage stochastic programs with markov uncertainty. SIAM J. Optim. **28**(1), 575–595 (2018)
5. Blanchet, J., Chen, L., Zhou, X.Y.: Distributionally robust mean-variance portfolio selection with wasserstein distances. arXiv:1802.04885 (2018)
6. Boyd, S., Busseti, E., Diamond, S., Kahn, R.N., Koh, K., Nystrup, P., Speth, J., et al.: Multi-period trading via convex optimization. Found. Trends® Optim. **3**(1), 1–76 (2017)
7. Brandtner, M.: "spectral risk measures: properties and limitations": comment on dowd, cotter, and sorwar. J. Financ. Serv. Res. **49**(1), 121–131 (2016)
8. Chen, Z., Sim, M., Xiong, P.: Robust stochastic optimization made easy with RSOME. Manag. Sci. **66**(8), 3329–3339 (2020)
9. Delage, E., Ye, Y.: Distributionally robust optimization under moment uncertainty with application to data-driven problems. Oper. Res. **58**(3), 595–612 (2010)
10. Dentcheva, D., Stock, G.J.: On the price of risk in a mean-risk optimization model. Quantitative Financ. **18**(10), 1699–1713 (2018)
11. Dowd, K., Cotter, J., Sorwar, G.: Spectral risk measures: properties and limitations. J. Financ. Serv. Res. **34**(1), 61–75 (2008)
12. Duque, D., Morton, D.P.: Distributionally robust stochastic dual dynamic programming. SIAM J. Optim. **30**(4), 2841–2865 (2020)

13. El Ghaoui, L., Oks, M., Oustry, F.: Worst-case value-at-risk and robust portfolio optimization: a conic programming approach. Oper. Res. **51**(4), 543–556 (2003)
14. Epstein, L.G., Halevy, Y.: Ambiguous correlation. Rev. Econ. Stud. **86**(2), 668–693 (2018)
15. Fabozzi, F.J., Kolm, P.N., Pachamanova, D.A., Focardi, S.M.: Robust Portfolio Optimization and Management. Wiley, New York (2007)
16. Fouque, J.P., Pun, C.S., Wong, H.Y.: Portfolio optimization with ambiguous correlation and stochastic volatilities. SIAM J. Control Optim. **54**(5), 2309–2338 (2016)
17. Furman, E., Wang, R., Zitikis, R.: Gini-type measures of risk and variability: gini shortfall, capital allocations, and heavy-tailed risks. J. Bank. Financ. **83**, 70–84 (2017)
18. Gao, J., Li, D.: Multiperiod mean-variance portfolio optimization with general correlated returns. IFAC Proc. **47**(3), 9007–9012 (2014)
19. Gao, S.Y., Sun, J., Wu, S.Y.: A semi-infinite programming approach to two-stage stochastic linear programs with high-order moment constraints. Optim. Lett. **12**(6), 1237–1247 (2018)
20. Georghiou, A., Tsoukalas, A., Wiesemann, W.: Robust dual dynamic programming. Oper. Res. **67**(3), 813–830 (2019)
21. Girardeau, P., Leclere, V., Philpott, A.B.: On the convergence of decomposition methods for multistage stochastic convex programs. Math. Oper. Res. **40**(1), 130–145 (2014)
22. Guigues, V.: Convergence analysis of sampling-based decomposition methods for risk-averse multi-stage stochastic convex programs. SIAM J. Optim. **26**(4), 2468–2494 (2016)
23. Guigues, V., Bandarra, M.: Single cut and multicut sddp with cut selection for multistage stochastic linear programs: convergence proof and numerical experiments. arXiv:1902.06757 (2019)
24. Gülten, S., Ruszczyński, A.: Two-stage portfolio optimization with higher-order conditional measures of risk. Ann. Oper. Res. **229**(1), 409–427 (2015)
25. Huang, H.H., Zhang, S., Zhu, W.: Limited participation under ambiguity of correlation. J. Financ. Mark. **32**, 97–143 (2017)
26. Huang, J., Zhou, K., Guan, Y.: A study of distributionally robust multistage stochastic optimization. arXiv:1708.07930 (2017)
27. Huang, R., Qu, S., Yang, X., Liu, Z.: Multi-stage distributionally robust optimization with risk aversion. J. Ind. Manag. Optim. **17**(1), 233 (2021)
28. Ismail, A., Pham, H.: Robust markowitz mean-variance portfolio selection under ambiguous covariance matrix. Math. Financ. **29**(1), 174–207 (2019)
29. Kakouris, I., Rustem, B.: Robust portfolio optimization with copulas. Eur. J. Oper. Res. **235**(1), 28–37 (2014)
30. Kang, Z., Li, X., Li, Z., Zhu, S.: Data-driven robust mean-CVaR portfolio selection under distribution ambiguity. Quantitative Financ. **19**(1), 105–121 (2019)
31. Li, D., Ng, W.L.: Optimal dynamic portfolio selection: multiperiod mean-variance formulation. Math. Financ. **10**(3), 387–406 (2000)
32. Li, J.Y.M.: Closed-form solutions for worst-case law invariant risk measures with application to robust portfolio optimization. Oper. Res. **66**(6), 1533–1541 (2018)
33. Ling, A., Sun, J., Wang, M.: Robust multi-period portfolio selection based on downside risk with asymmetrically distributed uncertainty set. Eur. J. Oper. Res. **285**(1), 81–95 (2020)
34. Liu, J., Chen, Z.: Time consistent multi-period robust risk measures and portfolio selection models with regime-switching. Eur. J. Oper. Res. **268**(1), 373–385 (2018)
35. Liu, J., Chen, Z., Lisser, A., Xu, Z.: Closed-form optimal portfolios of distributionally robust mean-CVaR problems with unknown mean and variance. Appl. Math. Optim. **79**(3), 671–693 (2019)
36. Liu, J., Chen, Z.P., Hui, Y.C.: Time consistent multi-period worst-case risk measure in robust portfolio selection. J. Oper. Res. Soc. China **6**(1), 139–158 (2018)
37. Löhndorf, N., Shapiro, A.: Modeling time-dependent randomness in stochastic dual dynamic programming. Eur. J. Oper. Res. **273**(2), 650–661 (2019)
38. Lotfi, S., Zenios, S.A.: Robust Var and CVaR optimization under joint ambiguity in distributions, means, and covariances. Eur. J. Oper. Res. **269**(2), 556–576 (2018)
39. Luo, F., Mehrotra, S.: Distributionally robust optimization with decision dependent ambiguity sets. Optim. Lett. **14**(8), 2565–2594 (2020)
40. Markowitz, H.: Portfolio selection. J. Financ. **7**(1), 77–91 (1952)
41. Mehrotra, S., Papp, D.: A cutting surface algorithm for semi-infinite convex programming with an application to moment robust optimization. SIAM J. Optim. **24**(4), 1670–1697 (2014)

42. Miller, N., Ruszczyński, A.: Risk-adjusted probability measures in portfolio optimization with coherent measures of risk. Eur. J. Oper. Res. **191**(1), 193–206 (2008)

43. Pereira, M.V., Pinto, L.M.: Multi-stage stochastic optimization applied to energy planning. Math. Program. **52**(1–3), 359–375 (1991)

44. Pflug, G., Wozabal, D.: Ambiguity in portfolio selection. Quantitative Financ. **7**(4), 435–442 (2007)

45. Pflug, G.C., Pohl, M.: A review on ambiguity in stochastic portfolio optimization. Set-Valued Var. Anal. **26**(4), 733–757 (2018)

46. Pflug, G.C., Werner, R.: Modeling, Measuring and Managing Risk. World Scientific, Singapore (2007)

47. Philpott, A.B., de Matos, V.L., Kapelevich, L.: Distributionally robust SDDP. Comput. Manag. Sci. **15**(3–4), 431–454 (2018)

48. Pichler, A.: The natural banach space for version independent risk measures. Insurance **53**(2), 405–415 (2013)

49. Pichler, A., Liu, R.P., Shapiro, A.: Risk-averse stochastic programming: Time consistency and optimal stopping. Oper. Res. (2021)

50. Renfree, J.: hist_stock_data(start_date, end_date, varargin). MATLAB Central File Exchange, https://www.mathworks.com/matlabcentral/fileexchange/18458-hist_stock_data-start_date-end_date-varargin (2020). Accessed 28 May 2019

51. Rockafellar, R.T., Royset, J.O.: Superquantile/cvar risk measures: second-order theory. Ann. Oper. Res. **262**(1), 3–28 (2018)

52. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. J. Risk **2**, 21–42 (2000)

53. Ruszczyński, A., Shapiro, A.: Optimization of convex risk functions. Math. Oper. Res. **31**(3), 433–452 (2006)

54. Shapiro, A.: Analysis of stochastic dual dynamic programming method. Eur. J. Oper. Res. **209**(1), 63–72 (2011)

55. Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming: Modeling and Theory, 2nd edn. SIAM, Philadelphia (2014)

56. Shapiro, A., Tekaya, W., da Costa, J.P., Soares, M.P.: Risk neutral and risk averse stochastic dual dynamic programming method. Eur. J. Oper. Res. **224**(2), 375–391 (2013)

57. Su, L., Sun, L., Karwan, M., Kwon, C.: Spectral risk measure minimization in hazardous materials transportation. IISE Trans. **51**(6), 638–652 (2019)

58. Yu, X., Shen, S.: Multistage distributionally robust mixed-integer programming with decision-dependent moment-based ambiguity sets. Math. Program. 1–40 (2020)

59. Zhu, S., Fukushima, M.: Worst-case conditional value-at-risk with application to robust portfolio management. Oper. Res. **57**(5), 1155–1168 (2009)