Conditional Value-at-Risk Robust Optimization

Daniel Mitchell
P.A. Nguyen

Abstract

We propose using the well-known conditional value at risk (CVaR) risk measure as a new methodology for incorporating robustness into portfolio optimization. Robustness in portfolio optimization can address the poor out-of-sample performance of the classical mean-variance optimization problems. Using many estimates of the covariance matrix and mean return vector, we incorporate robustness by finding a portfolio that performs well, on average, for a worst-case subset of these estimates, rather than for a single estimate. This becomes a bilevel integer program that we reformulate into a tractable form under appropriate conditions. We present numerical results that compares this CVaR robust method to a distributionally robust optimization approach that uses the Wasserstein metric to measure robustness. Theoretically, we extend the existing work of stochastic programming by linking the CVaR robustness to the sample average approximation. Specifically, we show that the CVaR robustness problem provides an upper bound, in expectation, to stochastic programming problems. We derive various asymptotic convergence results.

Keywords: Robust optimization, conditional Value-at-risk, portfolio optimization, factor model

1. Introduction

The mean-variance portfolio optimization problem (MVO) has been an active area of research in the finance literature since the work of Markowitz (1952). In his paper, the author considers the portfolio construction problem of an investor who has access to a set of assets and would like to minimize the variance of their portfolio returns while maintaining a threshold on the expectation of portfolio returns. Although MVO has been around for over half a century, its use is limited in the investment community due to the "error-maximization" effect and "unstable optimal solutions," as documented in Michaud (1989). That is, MVO works well only if the true mean and variance of returns are known. In reality, we can only obtain the empirical estimators for mean and variance from a given historical data set, which results in the optimal portfolio allocation x^*_{emp}. While x^*_{emp} is optimal to one particular data set, x^*_{emp} often has a disappointing out-of-sample performance on a different data set. This phenomenon is usually

Preprint submitted to journal

referred to as the optimizer's curse and is well documented in Michaud (1989), Best & Grauer (1991), Chopra & Ziemba (1993), Tütüncü & Koenig (2004), Santos (2010), Ban et al. (2018), and references therein. The optimizer's curse is not only relevant to portfolio optimization but is also seen in many applications where the parameters of the optimization problem are estimated using historical data (see Smith & Winkler (2006)).

In this paper, we pose the MVO problem differently in that, instead of optimizing over the mean and variance estimators of returns of a single data set, we want to optimize over various mean and variance estimators across multiple data sets. Those data sets could come from (i) different forecast models or different data sources, (ii) parametric Monte Carlo (MC) simulation if some assumption on the distribution of the returns random variable is made, or (iii) nonparametric bootstrapping if no assumption on the distribution of the returns random variable is required. To impose extra conservatism on the newly proposed problem, we define a new type of robustness based on the well-known concept of conditional value at risk (CVaR). The intuition is that instead of finding an optimal decision that performs well on average for all empirical estimators for mean and variance, we seek an optimal solution that fares well in a subset of some worst-case estimators. Nevertheless, we avoid being too conservative here by omitting a few good estimators to avoid overfitting while still retaining sufficient data to make an informed optimal portfolio allocation. We also incorporate factor model in our CVaR robust framework and present numerical experiments to compare the performance of the CVaR robust strategies with other robust strategies.

We also study a simpler (with respect to the feasible set) yet more general CVaR robust form of some general function and connect the CVaR robust problem to stochastic programming (SP), and the sample average approximation (SAA) approach. While the SAA problem in expectation is a lower bound to the SP problem (see Chapter 5 of Shapiro et al. (2021) for further details), the proposed CVaR robust problem in expectation is an upper bound for SP. The smallest CVaR robust upper bound of SP is achieved when we include every data set/data point except for the best one. We prove this upper bound property of the CVaR robust method as well as various other asymptotic and consistency results in Section 5.

1.1. Literature

Various papers have studied methods to combat the optimizer's curse phenomenon. One stream of literature on this problem is to use robust optimization (RO) where the optimization problem is performed over some uncertainty sets of the mean and variance of returns. While there exist different interpretations of robustness (Jen (2003), Ben-Tal et al. (2009), Bertsimas & Copenhaver (2018), and references therein), we take the pessimistic view and find an optimal solution that perform the best for some worst-case scenarios. In this spirit, Lobo & Boyd (2000) and Tütüncü & Koenig (2004) solve the robust MVO by constructing the interval and ellipsoid uncertainty sets on the mean and covariance parameters. Ceria & Stubbs (2006) derive the lowest possible value of the actual expected return over some confidence region around the true expected return and use this term in place of the empirical mean return parameter.

A more recent stream of research focuses on distributionally robust optimization (DRO). In the DRO setting, the investor chooses a portfolio allocation that performs well over the worst case of all plausible sources of data (distributions of returns), instead of just the observed historical data set. In this way, if the optimal portfolio can perform well under the most adversarial of those plausible distributions, then it will probably perform well when tested out-of-sample also. Bertsimas et al. (2018) constructs the distributional ambiguity set in terms of confidence regions of goodness-of-fit hypothesis tests. Calafiore & Ghaoui (2006), Popescu (2007), and Delage & Ye (2010) consider distributions with specified properties such as symmetry and moment information. Alternatively, the ambiguity set can be defined as a ball in the space of probability distributions with some probability distance function such as the Prohorov metric (Erdoğan & Iyengar (2006)), ϕ-divergences (Calafiore (2007), Bayraksan & Love (2015)), or the Wasserstein metric (Wozabal (2014), Gao & Kleywegt (2016), Esfahani & Kuhn (2018)).

Besides RO and DRO, there are also other statistical techniques to reduce the sensitivity of MVO to the input uncertainties. Fama & French (1993) and Zhao et al. (2019) seek to separate the signal from the noise in the estimation of the mean and variance of returns. Klein & Bawa (1976) and Black & Litterman (1991) propose using the Bayesian method to explicitly incorporate estimation errors of the MVO parameters. The James-Stein estimator and the Bayesian-Stein estimator have been suggested by Chopra et al. (1993) and Jorion (1986) respectively. Those approaches fall under the shrinkage class where the estimators are obtained by "shrinking" towards some proposed values and hence decreasing the estimation errors. Taking a nonparametric approach, Michaud (1998) proposes bootstrapping to solve MVO multiple times after which the various optimal weight solutions are averaged. A part of our numerical experiment will take advantage of bootstrapping, but instead of resolving MVO multiple times and averaging the optimal solutions as in Michaud (1998) we use bootstrapping to generate various data sets and find one optimal solution that performs well across all data sets.

Our proposed approach of CVaR robustness is different from all of the aforementioned methods. The idea for the CVaR robust framework starts with the SAA version of the empirical MVO since we do not know the true mean and variance of the stock portfolio, and have to estimate the mean and variance from various empirical data sets. These data sets could come from (i) different forecast models or different data sources, (ii) parametric Monte Carlo (MC) simulation if some assumption on the distribution of the returns random variable is made, or (iii) nonparametric bootstrapping if no assumption on the distribution of the returns random variable is required. To impose robustness and to further hedge against the uncertain nature of the empirical data, we consider optimizing over some set of the worst-case data sets. Nevertheless, it should also be noted that the CVaR here should not be thought of as the extreme tail conditional expectation where quantile β greater or equal to 90%. As a matter of fact, Lim et al. (2011) provide numerical evidence to illustrate that the CVaR at a quantile level β at 95% or above tend to produce results with higher estimation errors. Intuitively, this makes sense since a higher level of the CVaR (i.e., a more extreme tail) should require a larger number of data points in order to attain good accuracy. Therefore, the definition of CVaR robustness in our proposed framework does not focus on the extreme β levels. Instead, we only want to omit a few good data sets such that we can avoid overfitting while still retaining sufficient data, albeit being the worst-case data points, to make an informed optimal decision. The quantile level β to be used in the CVaR robust problem should be at most 50%.

Our approach also incorporates factor models. In this line of research, Goldfarb & Iyengar (2003) take the classical RO approach to the Markowitz problem. They assume a linear regression model of the stock returns and construct uncertainty sets for the mean returns, the factor loadings, and the covariance matrix of the regression residuals. The authors then show that their proposed problems can be reformulated into a second-order cone program. Although we also incorporate a factor model like Goldfarb & Iyengar (2003), we apply the concept of CVaR in place of RO. Another closely related work is the paper by Garlappi et al. (2007). They take a similar approach to that of Goldfarb & Iyengar (2003) by constructing the ambiguity sets of the mean and covariance estimators where the size of the ambiguity set is interpreted as a confidence level under the normal distribution assumption of returns. Nevertheless, the framework in Garlappi et al. (2007) is more general than that proposed by Goldfarb & Iyengar (2003) because it covers both factor and no factor models. While the analytical solutions in Garlappi et al. (2007) are derived for various types of ambiguity sets, the normal assumption is required. Thus, unlike the model proposed by Garlappi et al. (2007), our CVaR robust framework can be agnostic of the distribution assumption and can cater to both factor and non-factor models.

1.2. Summary of main contributions

The main contributions of our work are as follows:

- We introduce a CVaR robustness framework to address the effect of the estimation error of the MVO problem. This framework can be applied in more general problem settings than MVO.

- Mathematical formulation and computational algorithm are provided to solve the proposed problem. A procedure is outlined to select the quantile hyperparameter.
- We study asymptotic properties of the CVaR robust problem that extends the work of Rockafellar & Uryasev (2000) and Shapiro et al. (2021).
- Our numerical experiments investigate the performance of the CVaR robust MVO problem and validate some of our theoretical results.

We shall refer to CVaR as discrete conditional value at risk (DCVaR) in the discrete, data-driven setting. Section 2 formulates the main problem and proposes an optimization algorithm. Sections 3 is devoted to the numerical analysis while the numerical results are presented in Section 4 for different portfolio optimization strategies. Section 5 discusses various theoretical properties of the CVaR/DCVaR robustness beyond the portfolio optimization problem. We conclude in Section 6.

1.3. Notations

Let ξ be the random variable with support Ξ and $\mathbb{P}(\Xi)$ the set of (Borel) probability distributions over Ξ. For any $P \in \mathbb{P}(\Xi)$, $P(A)$ denotes the probability of the event $\xi \in A$. Let $\mathcal{X}$ be the feasible region for x and f be a mapping to the extended real line $f : \mathcal{X} \times \Xi \to \overline{\mathbb{R}}$. Assume that $\mathcal{X} \in \mathbb{R}^S$ is closed and for any $x \in \mathcal{X}$, $f(x, \xi)$ is finite almost surely. Define the SP and SAA problems to be

$$v^* \triangleq \inf_{x \in \mathcal{X}} \left\{ F(x) \triangleq \mathbb{E}_\Xi[f(x, \xi)] \right\} \text{ and } v_{SAA}^M \triangleq \inf_{x \in \mathcal{X}} \left\{ F_{\mathrm{SAA}}^M(x) \triangleq \frac{1}{M} \sum_{m=1}^{M} f(x, \xi_m) \right\}.$$

For some $\beta \in [0, 1)$, we want to find an optimal solution in

$$v_{\mathrm{CVaR}}^\beta \triangleq \inf_{x \in \mathcal{X}} \left\{ F_{\mathrm{CVaR}}^\beta(x) \triangleq \frac{1}{1-\beta} \int_{\mathrm{VaR}_\beta}^{\infty} f(x, \xi) p_f d\xi \right\}, \tag{1}$$

where $\mathrm{VaR}_\beta = \inf \{t \in \mathbb{R} : \mathbb{P}(f(x, \xi) \leq t) \geq \beta\}$ and p_f is the density of the random variable $f(x, \xi)$ induced by ξ. The special case of $\beta = 0$ corresponds to the classic SP problem. Since the true distribution of ξ is not known and the best knowledge that we have of $p(\xi)$ is obtained through some observed data, we consider the DCVaR approximation of v_{CVaR}

$$v_{\mathrm{DCVaR}}^k \triangleq \inf_{x \in \mathcal{X}} \left\{ F_{\mathrm{DCVaR}}^k(x) \triangleq \sup_{y \in \mathcal{Y}_k} \frac{1}{k} \sum_{m=1}^{M} y_m f(x, \xi_m) \right\} \tag{2}$$

where $k = \lfloor M(1-\beta) \rfloor$ and $\mathcal{Y}_k = \left\{ y \in \mathbb{R}^M \mid y_m \in \{0,1\},\ 1^\top y = k \right\}$. The term $F^k_{\text{DCVaR}}(x)$ signifies the average of k largest elements of some set. We use the subscript M for k to emphasize the dependence of k on the data sample size M for a given β. This specification will be important in the subsequent asymptotic proofs of the DCVaR robust problem. It will also be useful to note that

$$\lim_{M\to\infty} \frac{k}{M} = \lim_{M\to\infty} \frac{\lfloor M(1-\beta) \rfloor}{M} = 1 - \beta.$$

In the formulation of (2), k can be defined as $\lfloor M(1-\beta) \rfloor$ if we consider the case of a "lower" CVaR while $k \triangleq \lceil M(1-\beta) \rceil$ corresponds to the "upper" CVaR case. A more detailed discussion of the upper and lower CVaR can be found in Rockafellar & Uryasev (2002). Without loss of generality in our asymptotic analysis, we assume the "lower" CVaR definition in this paper. β can be thought of as the percentage of data thrown out and we are only interested in throwing away a small amount of good data to make the solutions more robust. When $\beta = 0$, the DCVaR problem is the SAA problem. We denote S and S^k_{DCVaR} to be the sets of optimal solutions for v^* and v^k_{DCVaR}.

2. DCVaR robust portfolio optimization

2.1. Problem setup

Denote Σ and r to be the true covariance and mean of the stock returns. We are interested in the following problem

$$\min_x \ \max_y \ \frac{1}{k} \sum_{m=1}^{M} y_m \ x^\top \Sigma_m x \tag{3a}$$

$$\text{s.t.} \quad 1^\top y = k,\ y_m \in \{0,1\}, \qquad m = 1, \cdots, M \tag{3b}$$

$$1^\top x = 1,\ x \geq 0, \tag{3c}$$

$$\frac{1}{k} \sum_{m=1}^{M} z^*_m r^\top_m x \geq \alpha \tag{3d}$$

$$z^* \in \arg\min_z \left\{ \frac{1}{k} \sum_{m=1}^{M} z_m \ r^\top_m x :\ 1^\top z = k,\ z_m \in \{0,1\}, m = 1, \cdots, M \right\}.$$

where M is the total number of the data sets, $k \leq M$ refers to a subset of some worst case scenarios, Σ_m is the estimator of Σ given by data set m, and r_m is the estimator of r given by data set m. The binary variables y choose which Σ's are included in the objective of (3a), where maximizing over y guarantees that only the k worst Σ's are included (as presented in the constraint (3b)). This implies that for a fixed x the portfolio variance is worse for the covariance Σ's whose y values are equal to 1 than for the covariance Σ's whose y values are equal to 0.

Here maximizing over y can be interpreted as the discrete CVaR of the portfolio variance given by the selected Σ's. In addition to the CVaR robust variance, we also have a CVaR robust mean return threshold constraint (3d) where we require the CVaR of expected returns to be greater than a threshold α. It should be noted that we use different integeter variables, y and z respectively, for the CVaR robust variance objective function (3a) and the CVaR robust mean returns threshold constraint (3d). Calafiore (2007) considers a DRO portfolio optimization with the Kullback–Leibler divergence distance function where the worst-case discrete distribution for the variance is the same as that for the mean returns. This corresponds to $y = z$ in our formulation of (3). On the other hand, the DRO portfolio optimization problem with the Wasserstein metric proposed by Blanchet et al. (2021a) allows the worst case continuous distributions in the variance and the mean returns to be different. This corresponds to the case where the solution z can be different from y in (3). Thus, the problem formulation proposed in (3) is more general such that if the subset of k worst case scenarios for the variance coincides with that for the mean returns, then we must obtain $y = z$ as a result of solving (3). Furthermore, in restricting y and z to be the same instead of the general formulation as in (3), we are limiting the set of the worst case estimators for Σ. That is, the k worst case data sets that ensure the minimum required CVaR mean returns (3d) is at least α may not correspond to the k worst case data sets that yield the worst possible CVaR variance of the objective function (3a). For all of these reasons, we will consider different integer variables for the CVaR robust variance objective function and the CVaR robust mean returns threshold constraint as proposed in (3). In the common vocabulary of CVaR, we can think of $\beta = 1 - \frac{k}{M}$ as the percentage level of the CVaR, as in Rockafellar & Uryasev (2002). Taken all together, we want to find portfolio that makes the CVaR of the variance as small as possible while meeting the constraint threshold of the CVaR of the mean returns. Finally, the constraint (3c) ensures that the total asset weights sum to 0 and that no short selling is allowed.

In DRO, the size of the set of plausible distributions is often chosen as a hyper-parameter that determines how robust the investor wants to be. In the same way in our CVaR robust framework, k (or β) is chosen as a hyper-parameter that defines the desired level of robustness. Larger β (i.e., smaller k) corresponds to more robust portfolios because the optimal portfolio must perform well over the more extreme scenarios. In most scenarios where CVaR is computed, one is often interested in relatively high β (i.e., small k) as a measure of how risky a portfolio is. Nevertheless, the best out-of-sample performance is often observed when the decision maker is not very robust (see Lim et al. (2011) and references therein). We therefore will focus on scenarios where k is relatively close to M (i.e., small β), so that the optimal portfolio is evaluated over all but a few of the best-case Σ's and r's. Indeed, in our numerical studies, we find the most consistently good out-of-sample performance by arbitrarily picking $\beta = 10\%$.

2.2. Algorithm

While problem (3) is a bilevel integer program, it can be relaxed and reformulated into a tractable form under some appropriate assumptions. We develop an algorithm to solve the reformulated problem by solving a sequence of linear program (LP) problems.

Proposition 1. *The optimization problem in* (3) *can be reformulated as*

$$\min_{x,\lambda,\mu} \; k\lambda + \sum_{m=1}^{M} \mu_m \tag{4a}$$

$$s.t. \; \lambda + \mu_m \geq x^\top \Sigma_m x, \qquad m = 1, \cdots, M \tag{4b}$$

$$1^\top x = 1, \; x \geq 0, \; \mu \geq 0 \tag{4c}$$

$$\frac{1}{k} \sum_{m=1}^{M} z_m^* r_m^\top x \geq \alpha \tag{4d}$$

$$z^* \in \arg\min_{z} \left\{ \frac{1}{k} \sum_{m=1}^{M} z_m \; r_m^\top x : \; 1^\top z = k, \; 0 \leq z_m \leq 1, m = 1, \cdots, M \right\}. \tag{4e}$$

The proof of this proposition is provided in the Appendix. While (4) is still a bilevel optimization problem, solving (4) is easier than solving (3) as we do not have to optimize over the space of integer solutions. In what follows, we propose an algorithm to solve (4). We will refer to (4a)-(4d) as the upper problem and (4e) as the lower problem. Since the lower problem is an LP, the simplex algorithm (available in any commercial solver) will visit corners of the polyhedron of the feasible region to descend to the optimal solution. In that sense, we denote $z_\ell^{(i)}$ to be the ℓ best optimal solution and $\alpha_\ell^{(i)}$ be its optimal objective value at the i-th iteration. The optimal (i.e., the best) solution and objective value are then $z_1^{(i)}$ and $\alpha_1^{(i)}$ respectively. The upper problem is a quadratically constrained program (QCP). In Algorithm 1, we define $\overline{r}$ to be the term $\dfrac{r_m}{k}$ in (4e).

The idea of Algorithm 1 is that we start with a larger feasible region without being concerned with the minimum required returns constraint. As we solve the lower level optimization problem, we continually add a cut at each iteration until the minimum required returns threshold is met. That is, if the ℓ-th optimal objective value of the lower problem at the iteration i is less than α, then adding the half space cut $\mathcal{H} : \left\{ x \in \mathcal{X} : \left(z_\ell^{(i)\top} \overline{r} \right) x \geq \alpha \right\}$ will shrink the feasible region to consider only x that yields a higher average returns value than α for a given $z_\ell^{(i)}$. Ideally, we want ℓ to be 1, i.e., the optimal solution of the lower problem. Nevertheless, it is possible that such value $z_1^{(i)}$ creates a cut that already exists, in which case we continue to evaluate to add the second best, third best, and so on cut in place of the best cut. If no new cut can be found and the minimum required threshold α is not met, then the problem must be infeasible for the given α level. This is a limitation of the CVaR robust problem compared to the Wassertein DRO problem formulation in Blanchet et al. (2021a).

Algorithm 1 Bilevel DCVaR robust algorithm

let $x^{(1)}$ be the minimizer of the upper problem
repeat
    given $x^{(i)}$, solve the lower problem $\rightarrow z_\ell^{(i)}$
    set $\alpha_\ell^{(i)} = z_\ell^{(i)\top} \overline{r} x^{(i)}$
    if $\alpha_1^{(i)} < \alpha$ **then**
        set $\ell' = \min\{\ell : z_\ell^{(i)\top} \overline{r} \leq \alpha,\ z_\ell^{(i)\top} \overline{r} \neq z^{(j)\top} \overline{r} \text{ for } j = 1, \cdots, i-1\}$
        if $\ell \in \emptyset$ **then**
            terminate (infeasible problem)
        else
            add the cutting plane constraint $z_{\ell'}^{(i)\top} \overline{r} x \geq \alpha$ to the upper problem
            set $z^{(i)} = z_{\ell'}^{(i)}$
            solve the upper problem $\rightarrow x^{(i+1)}$
        end if
    end if
until $\alpha_1^{(i)} \geq \alpha$

It is possible that Algorithm 1 does not terminate with a minimum returns meeting the α threshold since the problem (3) and its equivalent reformulation (4) are not feasible for all values of α. The infeasibility is more likely for a more aggressive (higher) α. This phenomenon is due to the discrete nature of the support of the empirical random variable. In contrast, the Wassertein DRO problem in Blanchet et al. (2021a) allows for more flexibility with respect to the probabilities as well as the support of the random variable. Such flexibility allows their problem formulation to be feasible when α is high. Nevertheless, while meeting a higher minimum required returns threshold translates to a higher out-of-sample mean returns performance, it also comes at the cost of a higher variance. If this happens for a rolling month in the rolling analysis of the numerical experiment, no stock investment will be made and the return for such a rolling period is 0%.

3. Model inputs

3.1. Deriving the mean returns and the covariance matrices

The key to the DCVaR robust MVO problem presented (3) and its equivalent reformulation (4) resides in the derivation of Σ_m's and r_m's. In this section, we describe a few methods for obtaining M samples of Σ_m and r_m, each of which is estimated from its own data set. The availability of the various stock returns data sets could be due to different forecast models, MC simulation, and/or bootstrapping. In the first scenario, one uses a historical returns data set and apply various factor models to fit and generate numerous estimates for the historical covariance matrix and the average stock returns. Then DCVaR robustness can be applied to find an optimal weight allocation that performs well on some worst case subset of these estimates. Our paper does not focus on this application though many useful references of the abundant factor models in finance can be found in Hou et al. (2014), Ahmed et al. (2019), and the cited papers therein. We instead choose a factor model, and then use the MC simulation or the bootstrapping method. Specifically, we first consider a no-factor model setting with MC and bootstrapping. Then we extend the approach to the Fama and French 5-factor model also with MC and bootstrapping. The methodologies will be described first followed by the numerical results.

In the non-factor MC approach, we are given a historical data set of S stocks and N data points. Assuming that the stock returns follow the multivariate normal distribution, the covariance matrix is a random variable from the Wishart distribution with $S - 1$ degree of freedom and the scale matrix parameter being the historical covariance matrix (Kent (1984)). Then, M i.i.d. samples can be trivially generated from the Wishart distribution. We shall refer to this method as the non-factor i.i.d. strategy. In the non-factor bootstrap strategy, instead of making the normal distribution assumption, we bootstrap the historical returns data set to obtain M bootstrapped samples. Given those M samples, we obtain M different estimates for the covariance matrix and mean returns which allow us to solve the DCVaR robust problems as in (3).

In the factor model extension to the aforementioned strategies, we pick the Fama and French 5-factor model among the top performing factor models in the literature (Hou et al. (2014) and Ahmed et al. (2019)). With respect to the factor MC strategy, we first perform the ordinary least squares (OLS) fitting on the historical data to get the factor loadings. We then use those loadings to obtain an estimator for Σ. Assuming that the residuals are normally distributed and independent among stocks, we simulate M new sets of residuals and use the previously calculated factor loadings to get M new returns data sets. Those new data sets yield M new mean stock returns estimators for r. Furthermore, the new factor loadings are also estimated and used to calculate M new estimators for Σ. In the factor boostrap model setting, the returns and factors data are bootstrapped M times before being fitted which gives M distinct estimates of the historical covariance matrices and average returns. Then, DCVaR robustness can be applied as in (3).

3.2. Choice of k

The choice of the parameter k is also important in the DCVaR robust problem. While this parameter can be chosen qualitatively based on the subject matter experts, there also exists quantitative justification for choosing an optimal k based on the best out-of-sample performance of the training data. The bigger k is, the closer the DCVaR robust problem is to SAA and we will start the lose the benefit of DCVaR robustness. On the other hand, a small k may result in a higher estimation error of the problem parameters. In RO and DRO, a lot of theory has been developed to derive the conditions under which the best hyperparameters can be chosen to ensure some out-of-sample performance guarantee in their respective robust problems (Esfahani & Kuhn (2018), Blanchet et al. (2021a), and the references therein). However, as also noted in those papers, the numerical experiments have shown that the theoretical "best" choices for those hyperparameters tend to be conservative in practice. Instead, cross validation tends to perform better. As a result, we will use cross validation to select the hyperparameter k. Generally speaking, the q-fold cross validation procedure involves partitioning the full data sets into q subsets. The first $q-1$ folds are used to train the DCVaR robust portfolio optimization problem and obtain some optimal portfolio weights. These optimal weights are then evaluated using the holdout q-th fold. This process is repeated until every fold has been given an opportunity to be used as the holdout validation set. As a result of this procedure, a total of q DCVaR robust problems are solved on the training set before the optimal solutions are evaluated on the test set. The performance of the DCVaR robust model for a given hyperparameter k is assessed by the variance of the validation set results. The chosen k is the hyperparameter value that corresponds to the lowest out-of-sample variance. For a more detailed introduction to cross validation, one can refer to Chapter 5 of James et al. (2013).

Figure 1 shows the difference in the variance of the out-of-sample performance when 10-fold cross validation, leave one out cross validation, and iterative cross validation are used. To generate left panel of Figure 1, we randomly choose a month between January 2000 and December 2019 and. The 10-year monthly historical data prior to the randomly selected month becomes the data sets on which we perform 10-fold cross validation several times. We consistently see a higher variance for the out-of-sample performance when k is smaller. The phenomenon is due to the nature of the objective function of the DCVaR robustness measure. The variability problem is further reduced in leave-one-out cross validation and iterative cross validation (see the middle and right panels in Figure 1 respectively). In leave-one-out cross validation, q is set to be the size of the training data set while for iterative cross validation, we perform 10-fold cross validation numerous times before taking the average of the out-of-sample performance.

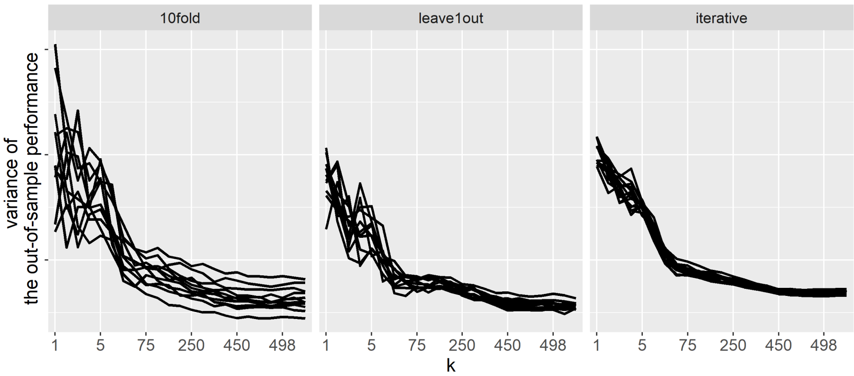


Figure 1: The variances of the out-of-sample performance via different cross validation methods.
Model parameters: the annual mean returns threshold $\alpha \approx 6\%$, the number M of the data sets $= 500$, and the training sample size N of the weekly stock returns $= 120$.

While leave-one-out cross validation and iterative cross validation result in a reliable and unbiased estimate of the model performance, the procedure can be computationally expensive to perform. Nevertheless, the commonly observed trend among the three cross validation strategies is that lower k values tend to exhibit more variance. The observed reduced variability in higher k values in our numerical experiments justifies narrowing the search space for k by only considering a set of high enough values for k. This is also consistent with the result of Lim et al. (2011) and the intention of using the CVaR robustness to find the optimal decisions by excluding only a few good data points. As a result, we will use the range of k values that corresponds to $\beta \in [0, 0.5]$ for the cross validation procedure in the numerical results of Section 4. It should be noted that when $\beta = 0$, the DCVaR robust problem becomes the SAA problem where no robustness is introduced.

3.3. Data, benchmarks, and the comparison methods

We perform our numerical study on the Dow Jones index. The historical data of the Dow Jones index as well as its constituent stocks between January 1990 and December 2019 are obtained from the Center for Research in Security Prices (CRSP). The idea is to compare the performance of actively investing in the constituent stocks in the Dow Jones index against the passive strategy of investing in the index itself. Throughout the analysis time horizon, the Dow replaced some existing stocks with new ones; our analysis makes sure to adjust for those updates accordingly. With respect to the factor model, we use the Fama-French five factor model. The data for those factors, which dates back as far as July 1963, are publicly available at `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

Since the motivation of the DCVaR robust approach stems from several weaknesses of the classical MVO problem, it is natural to include the MVO results in the out-of-sample performance benchmark comparison. Furthermore, Van Parys et al. (2021) prove that DRO delivers the best predictor-prescriptor solution pair that guarantees a certain level of out-of-sample performance and Blanchet et al. (2021a) study the specific portfolio optimization problem in the Wasserstein DRO setting similar to ours with respect to imposing the minimum required mean returns in the constraints. Therefore, we will compare our method to the Wasserstein DRO approach, specifically the results of Blanchet et al. (2021a).

The comparison will be based on a rolling horizon analysis. In the rolling horizon method, we start the rolling horizon at January 2000 and use the prior 10-year monthly stock data between January 1990 and December 1999 as the training set. The 10-fold cross validation procedure is performed on this training set to attain the best hyperparameter k. Also with the same training set, we compute the estimators for Σ and μ using the estimation methods described in Section 3.1. Given those Σ's, μ's, and k, we solve for the optimal stock weights via the approach in Section 2. Then based on the actual stock returns of January 2000, we compute the out-of-sample portfolio return. We then move forward to February 2000, shift the training data set to be between February 1990 and January 2000, and repeat the analysis process described above. This procedure is continued for every month until the end of the rolling horizon. The rolling horizon analysis results in a series of out-of-sample portfolio returns over the time frame between January 2000 and December 2019. The goal is to compare the out-of-sample portfolio performance among the various robust strategies and with the empirical index itself based on the 5 metrics that we will describe shortly.

We compare 6 main strategies: the traditional MVO problem, a DCVaR robust non-factor model with the i.i.d. assumption, a DCVaR robust non-factor model using bootstrap, a DCVaR robust factor model with the i.i.d. assumption, a DCVaR robust factor model using bootstrap, and the DRO approach with the Wasserstein metric as proposed in Blanchet et al. (2021a). We shall refer to those 6 strategies as the active investment strategies to differentiate from the passive investment strategy by investing directly in the Dow Jones index. We consider 6 strategies with and without the minimum required returns thresholds, as well as with and without using cross validation to pick the hyperparameter k. We study the out-of-sample portfolio performance in terms of the mean and variance of returns but also examine the concentration risk and turnover. To do this, we consider 5 metrics: the annualized average returns, the annualized standard deviation of returns, the average largest holding, the average deviation from the equally weighted portfolio, and the average portfolio turnover. The first 2 of the 5 metrics (columns 3 and 4 of Table 1 and Table 2) compute the annualized average returns and the annualized standard deviation of those out-of-sample returns data points. The third and fourth metrics measure portfolio diversification. In particular, the average largest holding (columns 5 of Table 1 and Table 2) computes the average weight of the largest holdings over the 240 rolling months. Also for each rolling month, we calculate the total deviation of a strategy's portfolio allocation from the an equally weighted portfolio. That is, let n denote the n-th rolling month, M^n be the total number of active stocks in the rolling analysis for that month, and x^n be the optimal allocation solution for some strategy in month n, then the corresponding deviation from the equally weighted portfolio in month n is computed as $\frac{\sum_{m=1}^{M^n} |x_m^n - \frac{1}{M^n}|}{M^n}$. Averaging this quantity across the rolling horizon yields the metric for the average deviation from the equally weighted portfolio in column 6 of Table 1 and Table 2. Finally, the fifth metric measures the portfolio turnover via the weight changes of the constituent stocks from one rolling month to another. If one stock is active in month n but inactive in month $n+1$ and vice versa, then we exclude that stock from the calculation of the metric. The turnover for month n is calculated as $\frac{\sum_{m=1}^{\overline{M}^n} |x_m^n - x_m^{n-1}|}{\overline{M}^n}$ where $\overline{M}^n \triangleq \min(M^n, M^{n-1})$. Averaging this monthly quantity across the rolling horizon yields the metric for the average portfolio turnover in column 7 of Table 1 and Table 2.

4. Numerical results

4.1. With cross validation

First, let us discuss the results that compare the performance of the 6 active investment strategies where cross validation is used to select the hyperparameters. For the DCVaR robust models, since we only want to omit a few estimators for the mean and covariance of the returns, we limit the search space for k to be between 250 and 500 with an increment of 50. For further discussion on the search space for k in the DCVaR robust approach, one can refer to Section 3.2. The search space for the hyperparameter ϵ of the ambiguity set size for the Wasserstein DRO strategies is between 10^{-9} and 10 with a multiplication increment by 10^2. Figure 2 shows the cumulative returns over the rolling horizon when cross validation is used and the minimum required returns threshold α is set to be 0.5% per month, which equates to approximately 6% annually. This threshold α remains the same for the rest of our numerical experiments. Based on Figure 2, not only do all strategies now outperform the actual cumulative returns of the passive strategy but they also has lower standard deviations. All active strategies have higher out-of-sample mean returns than when the minimum required returns threshold is inactive, as shown in the slight right shift of the distribution of returns in Figure 3. Within the active strategies, Table 1 shows that the robust methods with the exception of the DCVaR robust no factor i.i.d. model produces better out-of-sample mean returns compared to those of MVO although the out-of-sample standard deviations of the robust strategies are also higher. With the minimum required returns threshold constraint, the DRO Wasserstein method yields higher out-of-sample means returns than most of the other DCVaR strategies albeit at a slightly higher standard deviation level.

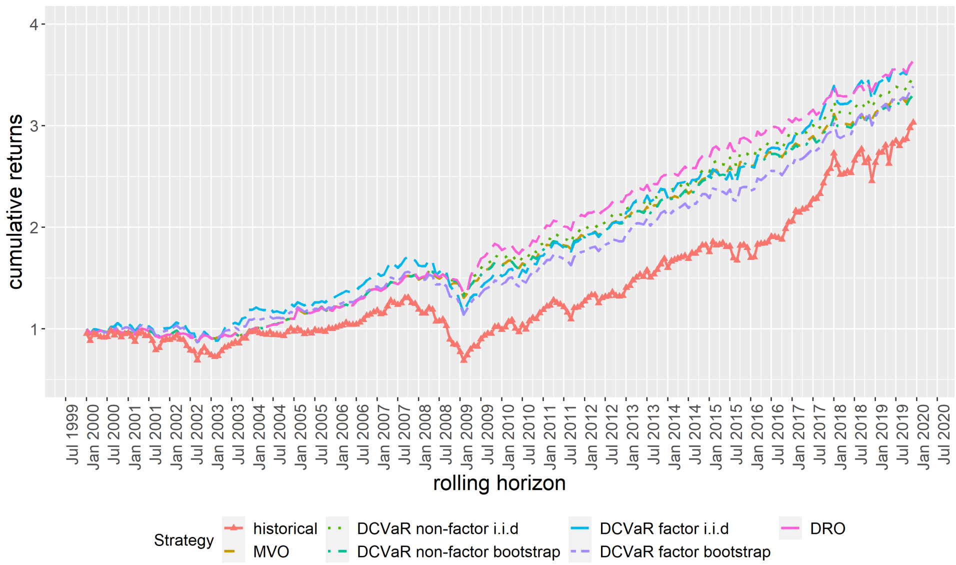


Figure 2: Cumulative returns throughout the rolling horizon when cross validation is used.

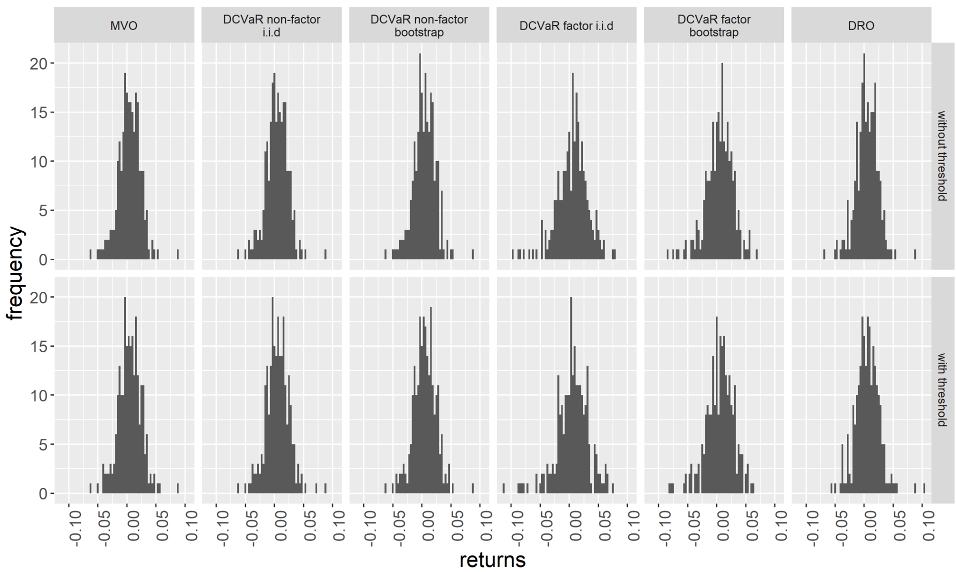


Figure 3: Histogram of monthly returns throughout the rolling horizon when cross validation is used.

strategy	annualized mean returns	annualized standard deviation of returns	average largest holding	average deviation from the equally weighted portfolio	average portfolio turnover
historical data	5.69	14.02			
MVO	6.20	6.85	59.67	5.31	0.30
no factor i.i.d.	6.40	6.90	59.78	5.32	0.36
no factor bootstrap	6.15	6.80	59.44	5.31	0.35
factor i.i.d.	6.67	9.71	26.43	3.37	0.63
factor bootstrap	6.27	8.25	29.71	3.19	0.20
DRO	6.65	7.16	58.52	5.34	0.53

* Units of the metrics are in %.

Table 1: Summarized results with cross validation

In terms of diversification, the DCVaR robust factor models again produce a more diversified portfolio since the average largest holding as well as the average deviation from the equally weighted portfolio metrics are lowest for these 2 strategies. When the minimum required returns threshold constraint is imposed, the average portfolio turnover are higher across all active investment strategies compared to when the constraint is inactive. The DCVaR robust factor i.i.d. model, despite its good diversification, has the highest average portfolio turnover. The average portfolio turnover of the DRO Wasserstein is the second highest. So, the DRO Wasserstein model, while delivering high average returns at a comparable level of standard deviation, suffers from high turnover and a lack of diversification. Overall, whether the minimum returns threshold is imposed or not, the DCVaR robust factor models can produce portfolios that have higher out-of-sample mean returns, comparable out-of-sample standard deviation of returns, better diversification, and lower turnover than those from the MVO and the Wasserstein metric based DRO strategies.

4.2. Without cross validation

In the next set of numerical results, we want to reduce the variance caused by cross validation, and thus set k to be 450 and the radius of the Wasserstein ambiguity to be 10^{-7}. The minimum required returns threshold is still 0.5% monthly (or 6% annually). The intention of choosing $k = 450$ is to introduce some conservatism in the optimization problem by excluding the best 10% of the stock returns data. While a common criticism of robust optimization is that it can be too conservative, the literature has shown that a little bit of robustness can deliver better performance than a lot of robustness. In DRO statistical learning using the Wasserstein metric, Blanchet et al. (2019), Blanchet et al. (2021b), and Gao (2020) rigorously prove that the optimal size choice of the Wasserstein ambiguity set is of the order $O(1/\sqrt{N})$ instead of the convervative estimate of $O(N^{-1/d})$ in Esfahani & Kuhn (2018) where N is the i.i.d. sample size and d is the covariate dimension. For the DRO problem using the Kullback-Leibler divergence, Supandi et al. (2017) and Gotoh et al. (2021) numerically show that a small amount of robustness significantly reduces the variance of the out-of-sample performance whereas the impact on the mean of the out-of-sample performance is negligible. The choice for $k = 450$ is justified by those results and is further confirmed by our numerical results. Furthermore, with $k = 450$ it is intuitively clear that we are aiming for 10% CVaR robustness by excluding the best 10% of good data. In contrast, gaining an intuition for the choice for the radius of the Wasserstein ambiguity set is more difficult and must rely on cross validation. The value 10^{-7} that we pick for the DRO Wasserstein strategy in this numerical experiment is based on the cross validation exercise from the previous subsection.

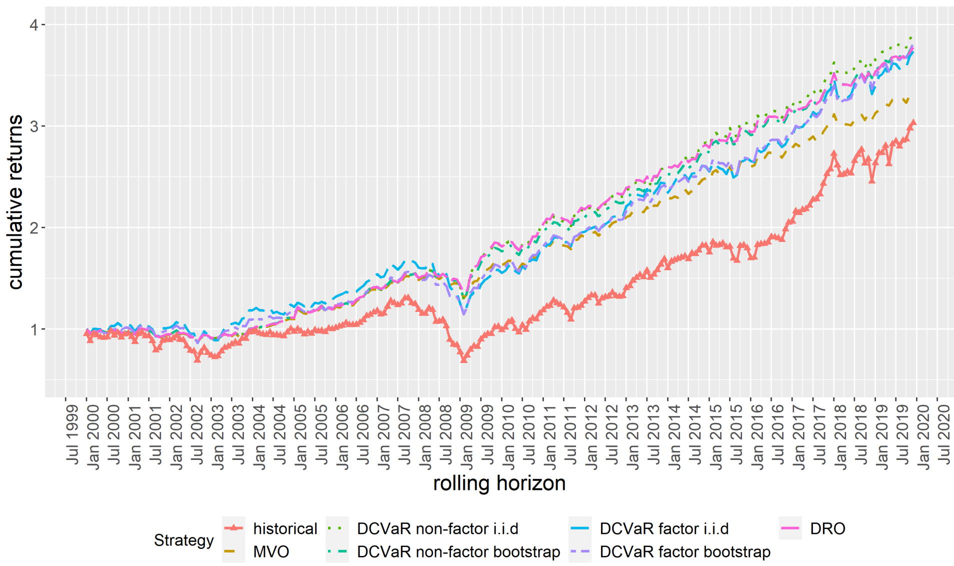


Figure 4: Cumulative returns throughout the rolling horizon when cross validation is not used.

The active investment strategies in Figure 4 have higher cumulative returns than that of the actual Dow Jones index. The DRO Wasserstein problem has better cumulative returns than most other DCVaR robust models. Again, the active strategies have higher out-of-sample mean returns when the minimum required returns threshold is active, as shown in the slight right shift of the distribution of returns in Figure 5. Table 2 shows that the active strategies have better out-of-sample mean returns and standard deviation than the passive strategy. All of the robust models also perform better than MVO with respect to the out-of-sample mean returns and standard deviation metrics. Within the robust models, the DRO Wasserstein strategy is comparable to the other DCVaR robust strategies. Nevertheless, when it comes to the diversification metrics, the DCVaR robust factor models produce a more diversified portfolio since the average largest holding as well as the average deviation from the equally weighted portfolio metrics are lowest for these 2 strategies. Again, the DCVaR robust factor i.i.d. model despite its good diversification has the highest average portfolio turnover. The DRO Wasserstein model also suffers from the same drawbacks.

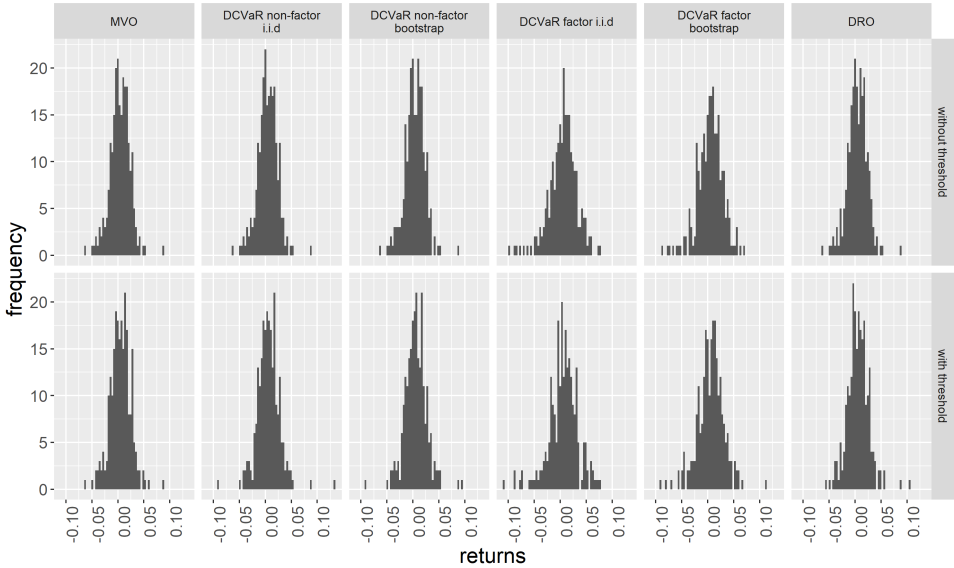


Figure 5: Histogram of monthly returns throughout the rolling horizon when cross validation is not used.

strategy	annualized mean returns	annualized standard deviation of returns	average largest holding	average deviation from the equally weighted portfolio	average portfolio turnover
historical data	5.69	14.02			
MVO	6.20	6.85	59.67	5.31	0.30
no factor i.i.d.	7.02	7.59	57.84	5.28	0.38
no factor bootstrap	6.86	7.34	57.88	5.28	0.37
factor i.i.d.	6.79	9.79	26.49	3.39	0.63
factor bootstrap	6.90	8.78	29.08	3.15	0.22
DRO	6.85	7.17	58.86	5.35	0.55

* Units of the metrics are in %.

Table 2: Summarized results without cross validation

Overall, when the minimum returns threshold constraint is not active, all active investment strategies yield slightly lower mean returns than the passive strategy does but the standard deviations of returns of the active strategies are significantly lower. The CVaR robust models have comparable out-of-sample performance compared to the DRO Wasserstain model. Specifically, the CVaR robust factor methods have better out-of-sample mean returns, albeit at the expense of a slightly higher standard deviation, than the CVaR robust non-factor models and the the DRO Wasserstein approach do. Furthermore, the CVaR robust factor methods also demonstrate less concentration risk and lower turnover. When the minimum returns threshold constraint is active, all active investment strategies deliver both higher mean returns and lower standard deviation of returns than the passive strategy does. The CVaR robust factor models also perform comparably to the DRO Wassertein model, and better than the CVaR robust non-factor models in terms of the mean and standard deviation of the out-of-sample returns. For diversification and turnover, the CVaR robust factor models are superior to the other robust active investment strategies.

5. Asymptotic analysis

In this section, we show some asymptotic results of the CVaR robust problem such as consistency of the optimal objective function value and the optimal solution, the decreasing upper bound for SP given by the CVaR robust problem, and the optimality gap analysis. The CVaR robust problem of interest is in a more simplified yet general setting beyond the scope of the portfolio optimization problem presented thus far. In particular, we consider the general DCVaR robust form of some general function $f(\cdot)$ under the data-driven setting

$$
\begin{aligned}
\min_{x\in\mathcal{X}} \max_{y} \ & \frac{1}{k}\sum_{m=1}^{M} y_m f(x,\xi_m) \\
\text{s.t.} \ & 1^\top y = k \\
& y_m \in \{0,1\}, \quad m = 1,\cdots,M,
\end{aligned}
\tag{5}
$$

where $\mathcal{X}$ is a convex set. We study various asymptotic and consistency properties of problem (5), and draw relevant connections of problem (5) with SP and SAA. The proofs for all of the results in this section can be found in the appendix.

5.1. Some properties of the DCVaR robust problem

The following proposition shows the asymptotic convergence of DCVaR to its continuous version CVaR. The result has been proved in Trindade et al. (2007) and shall be restated here for the sake of completeness.

Theorem 2. *Assume that the random variables ξ_m $(m = 1,\cdots,M)$ are independent and identically distributed (i.i.d.) with some continuous cumulative distribution. Let $\mathcal{X}$ be the convex feasible set. For any $x \in \mathcal{X}$, DCVaR converges to CVaR almost surely pointwise*

$$\lim_{M\to\infty} F^{k}_{DCVaR}(x) = F^{\beta}_{CVaR}(x).$$

Next, we state the dual of the general CVaR robust problem (5). The significance of this result is that it presents an alternative proof to Equation (5) in Theorem 1 of Rockafellar & Uryasev (2000) for the discrete case.

Proposition 3. *The reformulation of* (5) *is*

$$\begin{aligned} \min_{x,\lambda,\mu} \; & k\lambda + \sum_{n=1}^{N} \mu_n \\ s.t. \; & \lambda + \mu_n \geq f(x, \xi_n), \quad n = 1, \cdots, N \\ & \mu \geq 0, \; x \in \mathcal{X}. \end{aligned} \tag{6}$$

The proof of Proposition 3 is very similar to the dual derivation of the inner maximization of the upper problem in (4), which can be found in the appendix. In deriving (6), no convexity assumption of $f(x, \xi)$ is required since the objective function in (5) is linear in y. The function $f(x, \xi_n)$ in Rockafellar & Uryasev (2000) is linear in x and hence their reformulation version of (6) is just an LP. In the portfolio optimization problem where the minimum required returns threshold constraint is ignored, $f(x, \xi_n)$ is a quadratic function and hence (6) will be a quadratically constrained linear program (QCLP). Generally speaking, if $f(x, \xi_n)$ is convex in x, (6) is a convex optimization problem that can be solved using any commercially available solvers.

5.2. Consistency of the DCVaR robust estimators

In this section, we study the convergence of the DCVaR robust estimators v^k_{DCVaR} and S^k_{DCVaR}. The following proposition states the consistency of the DCVaR robust estimator of the objective value where k is chosen such that the DCVaR robust problem is the lowest upper bound for v^* (i.e., $k = M - 1$).

Proposition 4. *Fix $k = M-1$. Let $\mathbb{P}$ denote the cumulative distribution of the random variable $f(x, \xi)$ satisfying*

$$\mathbb{P}\left(\left|f\left(x, \xi\right)\right| \geq t\right) \leq \exp\left(-\alpha t\right)$$

for all $t \geq 0$ and some $\alpha > 0$. Then v^{M-1}_{DCVaR} converges to v^* with probability (w.p.) 1 as M goes to infinity.

The next proposition presents the convergence result for the CVaR robust optimal solutions. We include the proof here since it is short.

Proposition 5. *Fix $k = M - 1$. Assume that (i) the set S of the optimal solutions to the SP problem is nonempty, (ii) the function $F(x)$ is finite valued and continuous on $\mathcal{X}$, (iii)*

$F^k_{DCVaR}(x)$ *converges to* $F(x)$ *w.p. 1 uniformly on* $\mathcal{X}$*., and (iv) the st* S^k_{DCVaR} *is nonempty w.p. 1 for large enough* M*. Then* $v^k_{DCVaR} \to v^*$ *and the deviation between* S^k_{DCVaR} *and* S *goes to 0 w.p. 1 as* $M \to \infty$*.*

Proof. The result is due to Proposition 4 above and Theorem 5.3 in Shapiro et al. (2021). □

5.3. Asymptotic properties of the DCVaR robust optimal value

In the previous section, we study the consistency of the CVaR robust estimators which assures that the error of the estimation asymptotically goes to 0 as the sample size grows to infinity. We now study the magnitude of the error for a given sample size. First, we observe that v^k_{DCVaR} is an upward biased estimator of v^*.

Proposition 6. *Assume that we have i.i.d. data. Then, for all* $k \in \{1, 2, \cdots, M-1\}$ *we have*

$$\mathbb{E}_\Xi\left[v^M_{SAA}\right] \leq v^* \leq \mathbb{E}_\Xi\left[v^k_{DCVaR}\right].$$

The numerical illustrations of this proposition can be found in Section Appendix F. Proposition 6 connects DCVaR robustness directly to SP and SAA. While SAA provides a downward biased estimator for v^*, the DCVaR robust problem gives an upward biased estimator. Furthermore, this upward bias decreases monotonically as k increases for a fix sample size M.

Proposition 7. *For all* $k \in \{1, 2, \cdots, M-1\}$*, it holds that*

$$v^* \leq \mathbb{E}_\Xi[v^{M-1}_{DCVaR}] \leq \cdots \leq \mathbb{E}_\Xi[v^k_{DCVaR}] \leq \cdots \leq [v^1_{DCVaR}].$$

This proposition suggests that $\mathbb{E}_\Xi[v^{M-1}_{\text{DCVaR}}]$ is the smallest upper bound for the SP problem given a fixed M. In a similar manner, if we fix $k = M-1$, the upward bias decreases monotonically as M increases. We investigate the order of the bias $\mathbb{E}_\Xi[v^{M-1}_{\text{DCVaR}}] - v^*$ in the next proposition.

Proposition 8. *Assume that, for a fixed* x*, the random variable* $-f(x,\xi)$ *induced by* ξ *has finite mean and variance. Then* $\lim_{M\to\infty} \mathbb{E}_\Xi\left[v^{M-1}_{CVaR}\right] = v^*$ *and the rate of convergence is of order* $O(M^{-1/2})$*.*

The convergence rate of the DCVaR robust problem to SP is of the same order as the convergence rate of SAA to SP in Shapiro et al. (2021) when the set S is not a singleton. While the analysis in Shapiro et al. (2021) focuses on the lower bound SAA problem of SP, our DCVaR robust formulation provides a nice and convenient upper bound to SP without requiring additional data. This naturally leads to the optimality gap analysis in the next section.

5.4. Otimality gap

For a fixed M, Proposition 6 suggests that v^k_{DCVaR} and v^M_{SAA} provide valid statistical upper and lower bound respectively for v^*. Furthermore, by Proposition 7, we specifically should consider $\mathbb{E}[v^{M-1}_{\text{DCVaR}}]$ since it is the smallest lower bound given by the DCVaR robust problem. The expectations $\mathbb{E}[v^{M-1}_{\text{DCVaR}}]$ and $\mathbb{E}[v^M_{\text{SAA}}]$ can be estimated by averaging over the same set of data sets. That is, we can solve Q times the DCVaR robust and SAA problems on independently generated data sets each of size M. Let $v^{M-1,q}_{\text{DCVaR}}$ and $v^{M,q}_{\text{SAA}}$ denote the optimal objective values of the DCVaR robust and SAA problems respectively for $q = 1, \cdots, Q$. Then, the unbiased estimators of $\mathbb{E}[v^{M-1}_{\text{DCVaR}}]$ and $\mathbb{E}[v^M_{\text{SAA}}]$ respectively are $U^M_Q = \dfrac{1}{Q}\displaystyle\sum_{q=1}^{Q} v^{M-1,q}_{\text{DCVaR}}$ and $L^M_Q = \frac{1}{Q}\sum_{q=1}^{Q} v^{M,q}_{\text{SAA}}$. If the Q samples are independent then $v^{M-1,q}_{\text{DCVaR}}$ and $v^{M,q}_{\text{SAA}}$ are also independent. As a result, the estimated variances of U^M_Q and L^M_Q respectively become

$$\sigma^2_U = \frac{1}{Q}\Big[\frac{1}{Q-1}\sum_{q=1}^{Q}(v^{M-1,q}_{\text{DCVaR}} - U^M_Q)^2\Big] \text{ and } \sigma^2_L = \frac{1}{Q}\Big[\frac{1}{Q-1}\sum_{q=1}^{Q}(v^{M,q}_{\text{SAA}} - L^M_Q)^2\Big].$$

By Law of Large Numbers (LLN), we have

$$\sqrt{Q}\;(U^M_Q - \mathbb{E}[v^{M-1}_{\text{CVaR}}]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_U) \text{ and } \sqrt{Q}\;(L^M_Q - \mathbb{E}[v^M_{\text{SAA}}]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_L).$$

Thus, for a given γ, we can define the $(1-\gamma)\%$ confidence intervals for L^N_M and U^N_M as

$$\Big[U^N_M - z_\gamma\sigma_U,\; U^N_M + z_\gamma\sigma_U\Big] \text{ and } \Big[L^N_M - z_\gamma\sigma_L,\; L^N_M + z_\gamma\sigma_L\Big].$$

With respect to the objective function gap, the true gap between the SAA and the CVaR problem for data sets of size M is $\text{GAP}^M_{\text{true}} = \mathbb{E}[v^{M-1}_{\text{CVaR}}] - \mathbb{E}[v^M_{\text{SAA}}]$. The gap estimator from solving Q instances of the CVaR robust and SAA problems is $\text{GAP}^M_Q = U^M_Q - L^M_Q$. Again, by LLN and assuming independence, we have $\sqrt{Q}\;(\text{GAP}^M_{\text{true}} - \text{GAP}^M_Q) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_U + \sigma^2_L)$ and the confidence interval is $\Big[\text{GAP}^M_Q - z_\gamma\sqrt{\sigma^2_L + \sigma^2_U},\; \text{GAP}^M_Q + z_\gamma\sqrt{\sigma^2_L + \sigma^2_U}\Big]$. Increasing Q will reduce the variability in estimating the true gap while increasing M will tighten the true gap in which the true v^* reside. There exists an inherent trade-off between the computational complexity of the DCVaR robust and SAA problems as the same size M increases and the number of instances of the DCVaR robust and SAA problems that need to be resolved.

6. Conclusions

This paper considers a new approach to robustness where we want to find some optimal asset allocation that minimizes MVO across multiple data sets, instead of a single data set. To impose extra conservatism, we define a new type of robustness based on the well-known concept of CVaR. The intuition is that instead of finding an optimal decision that performs well on average for all empirical estimators for the mean and variance of the portfolio, we seek an optimal solution that fares well in a subset of some worst-case estimators. Nevertheless, we can avoid being too conservative by omitting a few good estimators. While the problem of interest is a bilevel integer program, we reformulate it into a tractable form under appropriate conditions and provide algorithms to solve the reformulated problem. Our numerical results show that the performance of the CVaR robust method is comparable to the Wasserstein DRO approach proposed by Blanchet et al. (2021a).

In addition, our reformulation of the CVaR robustness provides an alternative proof to the reformulation of Rockafellar & Uryasev (2000) under the discrete case. Extending further beyond the work of Rockafellar & Uryasev (2000), we link the CVaR robustness to SP and SAA. Specifically, we show that the CVaR robustness problem in expectation provides an upper bound to SP and is non-decreasing in the number of worst case scenarios and the size of the data set. The smallest upper bound provided by the CVaR robustness is achieved by excluding one best data point.

References

Ahmed, S., Bu, Z., & Tsvetanov, D. (2019). Best of the best: A comparison of factor models. *Journal of Financial and Quantitative Analysis*, *54*, 1713 – 1758. doi:10.1017/S0022109018000947.

Arnold, B. C., & Groeneveld, R. A. (1979). Bounds on expectations of linear systematic statistics based on dependent samples. *The Annals of Statistics*, *7*, 220 – 223. doi:10.1214/aos/1176344567.

Ban, G. Y., El Karoui, N., & Lim, A. E. (2018). Machine learning and portfolio optimization. *Management Science*, *64*, 983–1476. doi:10.1287/mnsc.2016.2644.

Bayraksan, G., & Love, D. K. (2015). Data-driven stochastic programming using Phi-divergences. In *The Operations Research Revolution* chapter 1. (pp. 1–19). INFORMS TutORials in Operations Research. doi:10.1287/educ.2015.0134.

Ben-Tal, A., Ghaoui, L. E., & Nemirovski, A. (2009). *Robust Optimization*. Princeton University Press.

Bertsimas, D., & Copenhaver, M. S. (2018). Characterization of the equivalence of robustification and regularization in linear and matrix regression. *European Journal of Operational Research*, *270*, 931–942. doi:10.1016/j.ejor.2017.03.051.

Bertsimas, D., Gupta, V., & Kallus, N. (2018). Robust sample average approximation. *Mathematical Programming*, *171*, 217–282. doi:10.1007/s10107-017-1174-z.

Best, M. J., & Grauer, R. R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *The Review of Financial Studies*, *4*, 315–342. doi:10.1093/rfs/4.2.315.

Black, F., & Litterman, R. (1991). Asset allocation: Combining investor views with market equilibrium. *The Journal of Fixed Income*, *1*, 7–18. doi:10.3905/jfi.1991.408013.

Blanchet, J., Chen, L., & Zhou, X. Y. (2021a). Distributionally robust mean-variance portfolio selection with Wasserstein distances. *Management Science*, (pp. 1–29). doi:10.1287/mnsc.2021.4155.

Blanchet, J., Kang, Y., & Murthy, K. (2019). Robust Wasserstein profile inference and applications to machine learning. *Journal of Applied Probability*, *56*, 830–857. doi:10.1017/jpr.2019.49.

Blanchet, J., Murthy, K., & Si, N. (2021b). Confidence regions in Wasserstein distributionally robust estimation. *Biometrika*, . doi:10.1093/biomet/asab026.

Calafiore, G. C. (2007). Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, *18*, 853–877. doi:10.1137/060654803.

Calafiore, G. C., & Ghaoui, L. E. (2006). On distributionally robust chance-constrained linear programs. *Journal of Optimization Theory and Applications*, *130*, 1–22. doi:10.1007/s10957-006-9084-x.

Ceria, S., & Stubbs, R. A. (2006). Incorporating estimation errors into portfolio selection: Robust portfolio construction. *Journal of Asset Management*, *7*, 109–127. doi:10.1057/palgrave.jam.2240207.

Chopra, V. K., Hensel, C. R., & Turner, A. L. (1993). Massaging mean-variance inputs: Returns from alternative global investment strategies in the 1980s. *Management Science*, *39*, 845–855. doi:10.1287/mnsc.39.7.845.

Chopra, V. K., & Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Journal of Portfolio Management*, *19*, 6–11. doi:10.3905/jpm.1993.409440.

Delage, E., & Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, *58*, 595–612. doi:10.1287/opre.1090.0741.

Erdoğan, E., & Iyengar, G. (2006). Ambiguous chance constrained problems and robust optimization. *Mathematical Programming*, *107*, 37–61. doi:10.1007/s10107-005-0678-0.

Esfahani, M., & Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*, *171*, 115–166. doi:10.1007/s10107-017-1172-1.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*, 3–56. doi:10.1016/0304-405X(93)90023-5.

Gao, R. (2020). Finite-sample guarantees for Wasserstein distributionally robust optimization: breaking the curse of dimensionality. *arXiv e-prints*, . doi:10.48550/arXiv.2009.04382.

Gao, R., & Kleywegt, A. J. (2016). Distributionally robust stochastic optimization with Wasserstein distance. *arXiv: Optimization and Control*, (pp. 1–43). doi:10.48550/arXiv.1604.02199.

Garlappi, L., Uppal, R., & Wang, T. (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies*, *20*, 41–81. doi:10.1093/rfs/hhl003.

Goldfarb, D., & Iyengar, G. (2003). Robust portfolio selection problems. *Mathematics of Operations Research*, *28*, 1–38. doi:10.1287/moor.28.1.1.14260.

Gotoh, J., Kim, M. J., & Lim, A. E. B. (2021). Calibration of distributionally robust empirical optimization models. *Operations Research*, *69*, 1630–1650. doi:10.1287/opre.2020.2041.

Hou, K., Xue, C., & Zhang, L. (2014). A comparison of new factor models. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2520929. doi:10.2139/ssrn.2520929.

James, G., Witten, D., Hastie, T., & Tibishirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. Springer.

Jen, E. (2003). Stable or robust? What's the difference? *Complexity*, *8*, 12–18. doi:10.1002/cplx.10077.

Jorion, P. (1986). Bayes-Stein estimation for portfolio analysis. *The Journal of Financial and Quantitative Analysis*, *21*, 279–292. doi:10.2307/2331042.

Kent, J. T. (1984). Aspects of multivariate statistical theory. *Journal of the Royal Statistical Society*, *33*, 251–251. doi:10.2307/2987858.

Klein, R. W., & Bawa, V. S. (1976). The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics*, *3*, 215–231. doi:10.1016/0304-405X(76)90004-0.

Lim, A. E., Shanthikumar, J. G., & Vahn, G. Y. (2011). Conditional Value-at-Risk in portfolio optimization: Coherent but fragile. *Operations Research Letters*, *39*, 163–171. doi:10.1016/j.orl.2011.03.004.

Lobo, M., & Boyd, S. (2000). The worst-case risk of a portfolio. https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.136.5286.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7*, 77–91. doi:10.2307/2975974.

Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, *45*, 31–42. doi:10.2469/faj.v45.n1.31.

Michaud, R. O. (1998). *Efficient asset management: A practical guide to stock portfolio optimization and asset allocation.*. Financial Management Association, Survey and Synthesis Series. HBS Press, Boston, MA.

Popescu, I. (2007). Robust mean-covariance solutions for stochastic optimization. *Operations Research*, *55*, 98–112. doi:10.1287/opre.1060.0353.

Rockafellar, R. T., & Uryasev, S. (2000). Optimization of Conditional Value-at-Risk. *The Journal of Risk*, *2*, 21–41. doi:10.21314/jor.2000.038.

Rockafellar, R. T., & Uryasev, S. (2002). Conditional Value-at-Risk for general loss distributions. *Journal of Banking and Finance*, *26*, 1443–1471. doi:10.1016/S0378-4266(02)00271-6.

Santos, A. (2010). The out-of-sample performance of robust portfolio optimization. *Brazilian Review of Finance*, *8*, 141–166.

Shapiro, A., Dentcheva, D., & Ruszczynski, A. (2021). *Lectures on Stochastic Programming: Modeling and Theory*. (3rd ed.). SIAM - Society for Industrial and Applied Mathematics.

Smith, J. E., & Winkler, R. L. (2006). The optimizer's curse: skepticism and postdecision surprise in decision dnalysis. *Management Science*, *52*, 311–322. doi:10.1287/mnsc.1050.0451.

Supandi, E. D., Rosadi, D., & Abdurakhman (2017). An empirical comparison between robust estimation and robust optimization to mean-variance portfolio. *Journal of Modern Applied Statistical Methods*, *16*, 589–611. doi:10.22237/jmasm/1493598720.

Trindade, A. A., Uryasev, S., Shapiro, A., & Zrazhevsky, G. (2007). Financial prediction with constrained tail risk. *Journal of Banking and Finance*, *31*, 3524–3538. doi:10.1016/j.jbankfin.2007.04.014.

Turner, J. (2013). Linear programming and network optimization. URL: https://www.arl.wustl.edu/~jst/cse/542/text/sec14.pdf.

Tütüncü, R. H., & Koenig, M. (2004). Robust asset allocation. *Annals of Operations Research*, *132*, 157–187. doi:10.1023/B:ANOR.0000045281.41041.ed.

Van Parys, B. P. G., Esfahani, P. M., & Kuhn, D. (2021). From data to decisions: Distributionally robust optimization is optimal. *Management Science*, *67*, 3387–3402. doi:10.1287/mnsc.2020.3678.

Wozabal, D. (2014). Robustifying convex risk measures for linear portfolios: A nonparametric approach. *Operations Research*, *62*, 1302–1315. doi:10.1287/opre.2014.1323.

Zhao, L., Chakrabarti, D., & Muthuraman, K. (2019). Portfolio construction by mitigating error amplification: The bounded-noise portfolio. *Operations Research*, *67*, 965–983. doi:10.1287/opre.2019.1858.

Appendix A. Proof of Proposition 1.

Proof. Since the decision variable y in the upper problem of (3) is independent of the decision variable z in the lower problem, we obtain the dual of the inner maximization of the upper problem. For a fixed k, consider a relaxation of the inner maximization of the upper problem.

$$\begin{aligned}
\max_{y} \ & \frac{1}{k}\sum_{m=1}^{M} y_m \ x^\top \Sigma_m x \\
\text{s.t.} \ & 1^\top y = k \\
& 0 \le y_m \le 1 \qquad\qquad m = 1, \cdots, M.
\end{aligned}$$

Since the constraint matrix of this relaxed version is totally unimodular, its optimal solutions are integers and coincide with the optimal solutions of the original integer linear program. As a result, even if the original problem is an integer linear program, we can solve its relaxed problem in polynomial time and obtain the same optimal solutions (Turner (2013)). Furthermore, the strong duality holds and the dual of the inner maximization problem with respect to y is

$$\begin{aligned}
\min_{\lambda,\mu} \ & k\lambda + \sum_{m=1}^{M} \mu_m \\
\text{s.t.} \ & \lambda + \mu_m \ge x^\top \Sigma_m x \qquad\qquad m = 1, \cdots, M \\
& \mu \ge 0.
\end{aligned}$$

We further observe that the lower integer problem of (3) can be relaxed to be a linear program. Therefore, we have the desired result. □

Appendix B. Proof of Proposition 4.

Proof. We want to show that for all $\epsilon > 0$ there exists $\widetilde{M}$ depending on ϵ and ω such that

$$\sup_{x\in\mathcal{X}} \left| F^{M-1}_{\text{DCVaR}}(x) - F(x) \right|$$

for all $M \ge \widetilde{M}$. The triangle inequality gives

$$\sup_{x\in\mathcal{X}} \left| F^{M-1}_{\text{DCVaR}}(x) - F(x) \right| \le \sup_{x\in\mathcal{X}} \left| F^{M-1}_{\text{DCVaR}}(x) - F^{\beta}_{\text{CVaR}}(x) \right| + \sup_{x\in\mathcal{X}} \left| F^{\beta}_{\text{CVaR}}(x) - F(x) \right|. \tag{B.1}$$

Let us analyze the first term of the right hand side of (B.1). Since $\mathcal{X}$ is compact, there exists a finite number of points $x_1, \cdots, x_n \in \mathcal{X}$ such that the corresponding neighborhoods $\mathcal{X}_1, \cdots, \mathcal{X}_n$ cover $\mathcal{X}$. Applying the triangle inequality again yields

$$\sup_{x\in\mathcal{X}} \left| F^{M-1}_{\text{DCVaR}}(x) - F^{\beta}_{\text{CVaR}}(x) \right|$$

$$\leq \sup_{x\in\mathcal{X}_j\cap\mathcal{X}} \left|F^{M-1}_{\text{DCVaR}}(x) - F^{M-1}_{\text{DCVaR}}(x_j)\right| + \left|F^{M-1}_{\text{DCVaR}}(x_j) - F^{\beta}_{\text{CVaR}}(x_j)\right| + \sup_{x\in\mathcal{X}_j\cap\mathcal{X}} \left|F^{\beta}_{\text{CVaR}}(x_j) - F^{\beta}_{\text{CVaR}}(x)\right|$$

for some $j = 1, \cdots, n$.

Due to Proposition 2, we have

$$\left|F^{M-1}_{\text{DCVaR}}(x) - F^{M-1}_{\text{DCVaR}}(x_j)\right| \leq \epsilon, \qquad j = 1, \cdots, n \tag{B.2}$$

with probability 1 for M large enough. Next, let x_ℓ be a sequence of points in $\mathcal{X}$ that converge to x. Using the dominated convergence theorem and then the fact that for all $x \in \mathcal{X}$ the function $f(\cdot, \xi)$ is continuous at x almost everywhere for $\xi \in \Xi$, we get

$$\lim_{\ell\to\infty} \int_{\text{VaR}_\beta}^{\infty} f(x_\ell, \xi)\, p(\xi)\, d\xi = \int_{\text{VaR}_\beta}^{\infty} \lim_{\ell\to\infty} f(x_\ell, \xi)\, p(\xi)\, d\xi = \int_{\text{VaR}_\beta}^{\infty} f(x, \xi)\, p(\xi)\, d\xi,$$

implying that $F^{\beta}_{\text{CVaR}}(x)$ is continuous in x. Hence, some neighborhood around x_j can be chosen such that if $d(x, x_j) \leq \delta$ then for $j \in \{1, \cdots, n\}$

$$\sup_{x\in\mathcal{X}_j\cap\mathcal{X}} \left|F^{\beta}_{\text{CVaR}}(x) - F^{\beta}_{\text{CVaR}}(x_j)\right| \leq \epsilon. \tag{B.3}$$

Now, we consider x_j and its δ-neighborhood $\mathcal{X}_j$. Let $1 \leq m \leq M$. Recall the following holds for all $x \in \mathcal{X}$

$$F^{k}_{\text{DCVaR}}(x) - F^{M}_{\text{SAA}}(x) = \frac{-\min_m f(x, \xi_m)}{M-1} + \frac{F^{M}_{\text{SAA}}(x)}{M-1},$$

implying that $F^{k}_{\text{DCVaR}}(x) = \dfrac{-\min_m f(x, \xi_m)}{M-1} + \dfrac{M}{M-1} F^{M}_{\text{SAA}}(x)$. As a result,

$$\begin{aligned}
&\sup_{x\in\mathcal{X}_j\cap\mathcal{X}} \left|F^{k}_{\text{DCVaR}}(x) - F^{k}_{\text{DCVaR}}(x_j)\right| \\
&= \sup_{x\in\mathcal{X}_j\cap\mathcal{X}} \left|\frac{-\min_m f(x, \xi_m)}{M-1} + \frac{M}{M-1} F^{M}_{\text{SAA}}(x) + \frac{\min_m f(x_j, \xi_m)}{M-1} - \frac{M}{M-1} F^{M}_{\text{SAA}}(x_j)\right| \\
&\leq \sup_{x\in\mathcal{X}_j\cap\mathcal{X}} \frac{M}{M-1} \left|F^{M}_{\text{SAA}}(x) - F^{M}_{\text{SAA}}(x_j)\right| + \sup_{x\in\mathcal{X}_j\cap\mathcal{X}} \frac{1}{M-1} \left|\min_m f(x, \xi_m) - \min_m f(x_j, \xi_m)\right|
\end{aligned}$$

Since $f(x, \xi)$ is continuous in x, taking the minimum of continuous functions is still continuous. Therefore, if we choose a δ_1-neighborhood small enough around x_j, we can obtain

$$\sup_{x\in\mathcal{X}_j\cap\mathcal{X}} \frac{1}{M-1} \left|\min_m f(x, \xi_m) - \min_m f(x_j, \xi_m)\right| \leq \frac{1}{M-1}\epsilon.$$

Similarly, a finite average over continuous function is still a continuous function. We choose a δ_2-neighborhood to be small enough around x_j such that there exists an M for

$$\sup_{x\in\mathcal{X}_j\cap\mathcal{X}} \frac{M}{M-1} \left|F^{M}_{\text{SAA}}(x) - F^{M}_{\text{SAA}}(x_j)\right| \leq \frac{M}{M-1} \cdot \frac{\epsilon}{2}.$$

So we choose the points x_j with its corresponding neighborhoods $\mathcal{X}_j$ of $\delta = \min(\delta_1, \delta_2)$ such that they cover the whole $\mathcal{X}$. In that case, we have

$$\sup_{x \in \mathcal{X}_j \cap \mathcal{X}} \left| F^k_{\text{DCVaR}}(x) - F^k_{\text{DCVaR}}(x_j) \right| \leq \frac{M}{M-1} \cdot \frac{\epsilon}{2} + \frac{1}{M-1}\epsilon$$

and there will exist an M such that the right hand side quantity is less than ϵ and

$$\sup_{x \in \mathcal{X}_j \cap \mathcal{X}} \left| F^k_{\text{DCVaR}}(x) - F^k_{\text{DCVaR}}(x_j) \right| \leq \epsilon \tag{B.4}$$

for all $1 \leq j \leq n$. It follows from (B.2), (B.3), and (B.4) that w.p. 1 for M large enough

$$\sup_{x \in \mathcal{X}} \left| F^{M-1}_{\text{DCVaR}}(x) - F^{\beta}_{\text{CVaR}}(x) \right| \leq 3\epsilon. \tag{B.5}$$

Next, we analyze the second term of the right hand side of (B.1). Let Z be the rv $f(x, \xi)$ with the probability density function $p_Z(z)$ and cumulative distribution function $P_Z(z)$. Now, for all $x \in \mathcal{X}$, we have

$$\left| F^{\beta}_{\text{CVaR}}(x) - F(x) \right| = \left| \frac{1}{1-\beta} \int_{\text{VaR}_\beta}^{\infty} z\, p_Z(z)\, dz - \int_{-\infty}^{\infty} z\, p_Z(z)\, dz \right| = \left| \frac{\beta}{1-\beta} \mathbb{E}[Z] - \frac{1}{1-\beta} \int_{-\infty}^{\text{VaR}_\beta} z\, p_Z(z)\, dz \right|.$$

Applying integration by parts on the second term yields

$$\begin{aligned}
\left| F^{\beta}_{\text{CVaR}}(x) - F(x) \right| &= \left| \frac{\beta}{1-\beta} \mathbb{E}[Z] - \frac{1}{1-\beta} z\, P_Z(z) \Big|_{-\infty}^{\text{VaR}_\beta} + \frac{1}{1-\beta} \int_{-\infty}^{\text{VaR}_\beta} P_Z(z)\, dz \right| \\
&= \left| \frac{\beta}{1-\beta} \mathbb{E}[Z] - \frac{1}{1-\beta} \text{VaR}_\beta\, P_Z(\text{VaR}_\beta) + \frac{1}{1-\beta} \int_{-\infty}^{\text{VaR}_\beta} P_Z(z)\, dz \right| \\
&\leq \left| \frac{\beta}{1-\beta} \mathbb{E}[Z] \right| + \left| \frac{1}{1-\beta} \text{VaR}_\beta\, P_Z(\text{VaR}_\beta) \right| + \left| \frac{1}{1-\beta} \int_{-\infty}^{\text{VaR}_\beta} P_Z(z)\, dz \right|.
\end{aligned}$$

Note that since $\beta \approx \frac{1}{M}$, we can find a M big enough such that $VaR_\beta < 0$ and

$$P_Z\left(VaR_\beta\right) \leq P\left(|Z| \geq -VaR_\beta\right) \leq \exp\left(\alpha\, VaR_\beta\right),$$

where the last inequality is due to the assumption on the tail property of the distribution of Z. As a result, we obtain

$$\begin{aligned}
\left| F^{\beta}_{\text{CVaR}}(x) - F(x) \right| &\leq \left| \frac{\beta}{1-\beta} \mathbb{E}[Z] \right| + \frac{1}{1-\beta} \left| \text{VaR}_\beta \right| \exp\left(\alpha\, \text{VaR}_\beta\right) + \frac{1}{1-\beta} \left| \int_{-\infty}^{\text{VaR}_\beta} \exp\left(\alpha z\right) dz \right| \\
&= \left| \frac{\beta}{1-\beta} \mathbb{E}[Z] \right| + \frac{1}{1-\beta} \left| \text{VaR}_\beta \right| \exp\left(\alpha \text{VaR}_\beta\right) + \frac{1}{\alpha(1-\beta)} \exp\left(\alpha \left| \text{VaR}_\beta \right|\right).
\end{aligned}$$

We can make M big enough so that β gets very close to 0, which diminishes the first term. On the other hand, with M big enough, VaR_β tends to $-\infty$ linearly but $e^{\alpha\ \text{VaR}_\beta}$ tends to 0 exponentially fast so the second term and the third term also get very close to 0. Therefore, we have for all $x \in \mathcal{X}$

$$\left| F^{\beta}_{\text{CVaR}}(x) - F(x) \right| \leq \epsilon. \tag{B.6}$$

From (B.5) and (B.6), we have $F^k_{\text{DCVaR}}(x)$ converges to $F(x)$ w.p. 1 uniformly on $\mathcal{X}$. The convergence of v_{CVaR} to v^* follows from Theorem 5.2 of Shapiro et al. (2021). $\square$

Appendix C. Proof of Proposition 6.

Proof. The first inequality is well-known from the SAA result. For the second inequality, let x^*_{DCVaR} be the optimal solution of $\inf_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} \frac{1}{k}\sum_{n=1}^{N} y_n f(x,\xi_n)$. Then $\forall k \in \{1, 2, \cdots, M-1\}$

$$\begin{aligned}
&\inf_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} \frac{1}{k}\sum_{m=1}^{M} y_m f(x,\xi_m) \geq \frac{1}{M}\sum_{m=1}^{M} f(x^*_{\text{DCVaR}},\xi_m) \\
\Rightarrow\ & \mathbb{E}_\Xi\Big[\inf_x \sup_y \frac{1}{k}\sum_{m=1}^{M} y_m f(x,\xi_m)\Big] \geq \mathbb{E}_\Xi\Big[\frac{1}{M}\sum_{m=1}^{M} f(x^*_{\text{DCVaR}},\xi_m)\Big] \\
&\qquad = \mathbb{E}_\Xi[f(x^*_{\text{DCVaR}},\xi)] \qquad\qquad \text{(i.i.d assumption)} \\
&\qquad \geq \inf_x \mathbb{E}_\Xi[f(x,\xi)].
\end{aligned}$$

This completes the result of Proposition 6. □

Appendix D. Proof of Proposition 7.

Proof. The first inequality in the proposition is due to Proposition 6. For the second inequality, let $\widetilde{x}^*_{\text{DCVaR}}$ be the optimal solution to $v^{k_M-1}_{\text{DCVaR}}$. Then we have

$$\begin{aligned}
v^{k_M-1}_{\text{DCVaR}} = \sup_{y\in\mathcal{Y}_{k_M-1}} \frac{1}{k_M-1}\sum_{n=1}^{N} y_n f(\widetilde{x}^*_{\text{DCVaR}},\xi_n) &\geq \sup_{y\in\mathcal{Y}_{k_M}} \frac{1}{k_M}\sum_{n=1}^{N} y_n f(\widetilde{x}^*_{\text{DCVaR}},\xi_n) \\
&\geq v^{k_M}_{\text{DCVaR}}
\end{aligned}$$

Taking expectation over Ξ completes the proof. □

Appendix E. Proof of Proposition 8.

Proof. Let $(x^*_{\text{DCVaR}}, y^*)$ and x^*_{SAA} be the optimal solutions to v^{M-1}_{DCVaR} and v^M_{SAA} respectively. We evaluate

$$\begin{aligned}
&\mathbb{E}_\Xi\left[v^{M-1}_{\text{CVaR}}\right] - v^* \\
\leq &\mathbb{E}_\Xi\left[v^{M-1}_{\text{CVaR}}\right] - \mathbb{E}_\Xi\left[v^M_{\text{SAA}}\right] \\
= &\mathbb{E}_\Xi\left[\frac{1}{M-1}\sum_{m=1}^{M} y^*_m f\left(x^*_{\text{DCVaR}},\xi_m\right) - \frac{1}{M}\sum_{m=1}^{M} f\left(x^*_{\text{SAA}},\xi_m\right)\right] \\
\leq &\mathbb{E}_\Xi\left[\frac{1}{M-1}\sum_{m=1}^{M} y^*_m f\left(x^*_{\text{SAA}},\xi_m\right) - \frac{1}{M}\sum_{m=1}^{M} f\left(x^*_{\text{SAA}},\xi_m\right)\right] \\
= &\mathbb{E}_\Xi\left[\frac{\sum_{m=1}^{M} y^*_m f\left(x^*_{\text{SAA}},\xi_m\right) - \sum_{m=1}^{M} f\left(x^*_{\text{SAA}},\xi_m\right)}{M-1} + \frac{1}{M-1}\frac{\sum_{m=1}^{M} f\left(x^*_{\text{SAA}},\xi_m\right)}{M}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\mathbb{E}_{\Xi}\left[v^{M}_{\mathrm{SAA}}\right]}{M-1} - \frac{\mathbb{E}_{\Xi}\left[\min_{m} f(x^{*}_{\mathrm{SAA}}, \xi_m)\right]}{M-1} \\
&= \frac{\mathbb{E}_{\Xi}[v^{M}_{\mathrm{SAA}}]}{M-1} + \frac{\mathbb{E}_{\Xi}[\max_{m} -f(x^{*}_{\mathrm{SAA}}, \xi_m)]}{M-1} \\
&\leq \frac{\mathbb{E}_{\Xi}[v^{M}_{\mathrm{SAA}}]}{M-1} + \frac{\mu + \sigma\sqrt{M-1}}{M-1},
\end{aligned}
$$

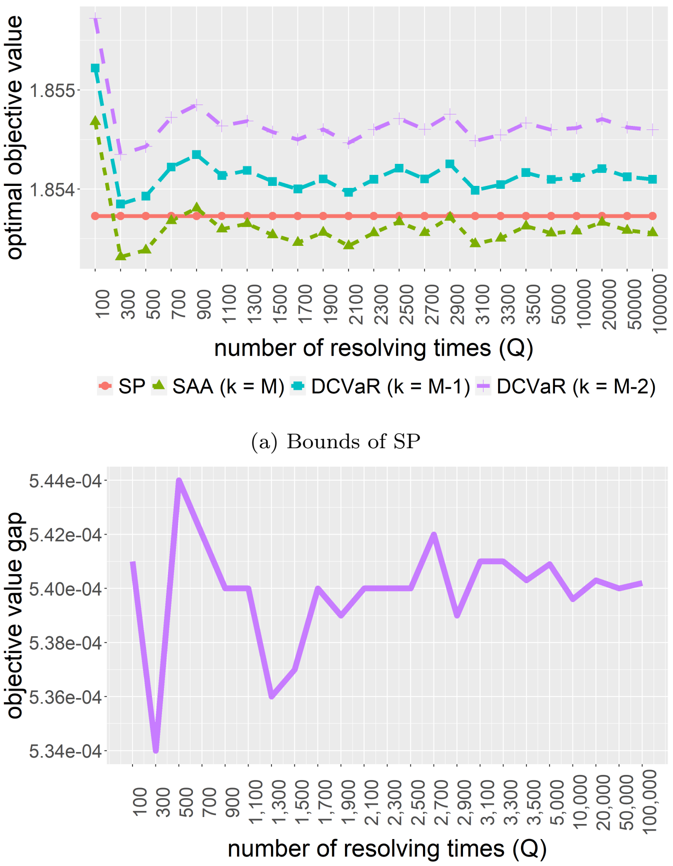
which converges to 0 sublinearly. In the last inequality, we use the result for a tight bound on the expectation of the order statistics from Arnold & Groeneveld (1979). □

Appendix F. Numerical illustration of the asymptotic behaviors of the DCVaR robust problem

We present numerical illustrations of the theoretical properties of the CVaR/DCVaR robust problems. For a fixed M, suppose we have M distinct data sets that give M different estimations for the covariance parameter. We solve the SAA and the DCVAR robust problems to obtain one realization for each of v^{M}_{SAA}, v^{M-1}_{DCVaR}, and v^{M-2}_{DCVaR}. Repeating this process Q times yields the average estimates for v^{M}_{SAA}, v^{M-1}_{DCVaR}, and v^{M-2}_{DCVaR}. One can refer to Section 3.1 for more details on the process of obtaining the new covariance estimates. Figure F.6a shows that the mean of v^{M}_{SAA} (teal dashed line) is a lower bound for SP (solid purple line) and the means of v^{M-1}_{DCVaR} (red dashed line), and v^{M-2}_{DCVaR} (green dotted line) provide the upper bounds when we use the MC procedure. Since M is fixed, the optimality gap is positive and hence as Q gets bigger, the estimated gap will converge to the positive true optimalility gap as shown in Figure F.6b.

Next, we fix Q, the number of resolving times for the SAA and the DCVAR robust problems, and vary M, the number of different estimates for the covariance parameter. Using MC simulation, Figure F.7a shows that asymptotically the mean of v^{M}_{SAA} (teal dashed line) is still a lower bound for SP (solid purple line) and the means of v^{M-1}_{DCVaR} (red dashed line), and v^{M-2}_{DCVaR} (green dotted line) are still the upper bounds. Nevertheless, v^{M}_{SAA}, v^{M-1}_{DCVaR}, and v^{M-2}_{DCVaR} converges to 0 as M gets bigger. This phenomenon is also reflected in the optimality gap in Firgure F.7b.

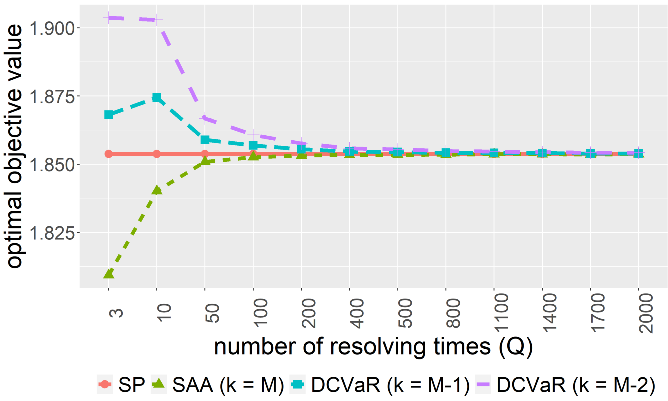
Those observations of the asymptotic bounds, gaps, and convergence of the optimal solutions are consistent with Proposition 6 and Proposition 7. The next section will provide the technical proofs for those asymptotic behaviors of the DCVaR robust problem.



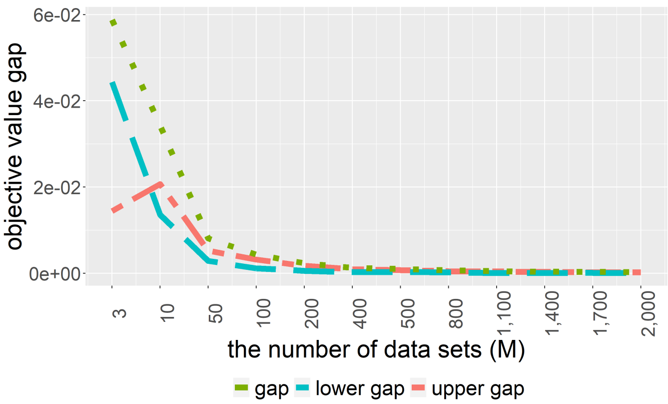
(a) Bounds of SP

(b) The optimality gap with respect to SP

Figure F.6: The bounds and optimality gap when the number M of data sets is fixed and the number of resolving instances varies.



(a) Bounds of SP



(b) The optimality gap with respect to SP

Figure F.7: The bounds and optimality gap when the number M of data sets varies and the number of resolving instances is fixed.