Dynamic Distributionally Robust Portfolio Optimization

David Islip
Mechanical and Industrial Engineering
University of Toronto
Toronto, Canada
ryan.islip@mail.utoronto.ca

Gurjot Dhaliwal
Mechanical and Industrial Engineering
University of Toronto
Toronto, Canada
gurjot.dhaliwal@utoronto.ca

***Abstract*—Portfolio optimization suffers from two main challenges; the amplification of estimation error through optimization, and decision difficulties related to the dynamic behaviour of financial markets. To combat these difficulties, in this report, we propose the use of a min-max changepoint detection algorithm combined with a distributionally robust portfolio optimization routine. The proposed methodology is formulated as the solution of two convex programs and is tested on financial data. It is shown that our proposed approach addresses the two challenges and exhibits desired risk-return performance relative to other methods.**

***Index Terms*—Portfolio Optimization; Distributionally robust optimization; Data-driven optimization; Hypothesis testing**

I. Introduction

A. Motivations and Challenges

Portfolio optimization is concerned with investor's decisions to buy or sell specific assets in an attempt to achieve an objective for a portfolio. The seminal work on portfolio optimization by [1] proposes selecting the portfolio that minimizes the portfolio's variance subject to an expected return constraint (α) over some time horizon. That is:

$$\min_{w \in \mathbb{R}^d} \{ w^\intercal \Sigma_{\mathbb{P}}(R) w : w^\intercal = 1,\ w^\intercal \mathbb{E}_{\mathbb{P}}(R) = \alpha \} \tag{1}$$

Where $R \in \Omega \subseteq \mathbb{R}^d$ is the random vector of stock returns; $\mathbb{P}$ is the probability measure of R; $\mathbb{E}$ and Σ denote expectation and variance respectively. There have been many extensions to the framework proposed by [1], which include the use of different risk measures, constraints and models of transaction costs. Practitioners of portfolio optimization commonly identify two challenges: i) The financial risk stemming from errors in parameter estimates, and ii) difficulties related to the dynamic and non-stationary nature of return distributions. Estimation error is amplified by the optimization procedure because any decision-maker who consistently chooses alternatives based on the maximal (minimal) estimated objective value is likely to select an option that has the highest estimation error, even when the estimates are unbiased. This phenomenon is known as the Optimizer's curse [2]. Further, with regards to (ii), it is well documented that economic time-series exhibit dramatic and sudden changes due to world events, and other external drivers [3].

B. Contributions

In this work, we provide background on existing techniques that attempt to solve the challenges above and discuss some of their potential shortcomings. We propose a novel approach to dynamic portfolio optimization that overcomes the disadvantages of previous literature via the application of robust changepoint detection in sequence with distributionally robust portfolio optimization. Our proposed approach: is entirely data-driven, allows for user-specified levels of robustness and is tractable via convex programming. Lastly, we compare our proposed methodology against other methods for the S&P 500 universe of stocks and find improved risk-return performance on a comparative basis.

II. Background

A. Optimization Approaches to Combating Estimation Error

Several approaches attempt to reduce the sensitivity of portfolio optimization outputs to parameter estimates along with providing performance guarantees. The approaches to robust portfolio optimization typically fall into the categories of; Bayesian parameter estimation, robust deterministic optimization, stochastic programming and sampling, and distributionally robust optimization (DRO). The first three approaches above have drawbacks. The Bayesian estimation approach, while known to improve performance, does not provide guarantees on the risk-return performance of the portfolio. In the case of deterministic-robust optimization, [4] points out, that although tractable, deterministic-robust optimization does not exploit available distributional knowledge of the uncertain parameters. Stochastic programming, on the other hand, accounts for distributional information through the use of expectations with respect to a distribution for the model parameters but has the disadvantage that it assumes a distribution for the parameters is known. [5].

Lastly, DRO loosens the distributional assumption present in stochastic programming while still leveraging distributional information and providing performance guarantees. DRO proceeds by specifying an ambiguity set of probability distributions and replacing the portfolio optimization risk-return with the worst-case risk-return evaluated over the ambiguity set. The question of designing an ambiguity set is one that has gathered significant attention. The next subsection introduces the Wasserstein ambiguity set in detail and outlines its strengths as a choice to capture distributional uncertainty. From this point on, the general risk minimization problem for a decision w is defined as: $\inf_{w \in \mathcal{W}} \mathcal{R}(\mathbb{P}, w)$ where $\mathcal{R}$ is the risk that depends on the probability measure $\mathbb{P}$ and decision $w \in \mathcal{W}$

B. Wasserstein Ambiguity Sets Mitigating the Optimizer's Curse

A notable example of an ambiguity set that has desirable properties is the statistical distance based ambiguity which considers all distributions within a specified Wasserstein distance metric δ of an empirical distribution $\mathbb{P}_n$ defined by n data points:

$$\mathcal{P}^o_\delta(\mathbb{P}_n) := \{\mathbb{P} : D^o_c(\mathbb{P}, \mathbb{P}_n) \leq \delta\} \tag{2}$$

where $D^o_c(.)$ is the q norm - Wasserstein distance of order $o = \{1, 2\}$, defined for any two probability distributions $\mathbb{P}$, $\mathbb{Q}$ supported on $\mathbb{R}^d$

$$\begin{aligned} D^o_c(\mathbb{P}, \mathbb{Q}) := \inf\{&\mathbb{E}_\pi[\|U - W\|^o_q] \\ &: \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d), \pi_U = \mathbb{P}, \pi_W = \mathbb{Q}\} \end{aligned} \tag{3}$$

where $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ denotes the space of Borel probability measures supported on $\mathbb{R}^d \times \mathbb{R}^d$ and for a given element $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$; a vector from the support of π takes the form (U, V), $U \in \mathbb{R}^d$, $V \in \mathbb{R}^d$. Thereby the marginals distributions are denoted as π_U and π_V.

The suitable distributionally robust optimization (DRO) problem is then formed by selecting w to minimize the worst case risk over all distributions in $\mathcal{P}^o_\delta(\mathbb{P}_n)$, that is to solve:

$$w^* = \arg\inf_{w \in \mathcal{W}} \sup_{\mathbb{P} \in \mathcal{P}^o_\delta(\mathbb{P}_n)} \mathcal{R}(\mathbb{P}, w) \tag{4}$$

[6] shows that for a wide class of objective functions (including proper, convex and lower-semicontinuous functions), that the optimal value of the DRO problem defined in (4) (i.e., the worst-case optimal risk) provides upper confidence bounds on the out-of-sample performance. The performance bounds therefore, reduce the disappointment caused by the Optimizer's curse.

C. Changepoint Detection and Portfolio Optimization

Changepoint detection is a necessary component of dynamic asset allocation and is most frequently achieved via portfolio rules based on heuristics or Markov models of regime changes. However, [7] points out that the assumption of a fixed number of regimes, each with assumed stationarity (conditional on the regime), is unlikely to lead to optimal risk-reward performance. [7] proposes a novel approach to the dynamic asset allocation problem by using univariate non-parametric changepoint estimators on the S&P 500 and VIX indices to determine changepoints to rebalance a portfolio that consists of the S&P index or cash. Inspired by this approach, our work builds on [7] to detect changepoints of multivariate high dimensional financial datasets using a more flexible changepoint detection scheme.

D. Changepoint Detection as Hypothesis Testing

The problem of detecting a changepoint can be formulated as a hypothesis testing problem. Assuming that some sequence of observations $\{R_i\}_{i=1}^\infty$ (for example returns) is observed and that there are known two distributions $\mathbb{P}_1$ and $\mathbb{P}_2$, the detection of a changepoint at index κ is formulated as the following hypothesis testing problem:

$$\begin{aligned} H_0 &: R_i \sim \mathbb{P}_1, i = 1, 2, \ldots \\ H_1 &: R_i \sim \mathbb{P}_1, i = 1, 2, \ldots, \kappa, \\ &\quad R_i \sim \mathbb{P}_2, i = \kappa + 1, \kappa + 2, \ldots \end{aligned} \tag{5}$$

which can be constructed while controlling for multiple comparisons error from the single observation hypothesis test for a given i:

$$H_0 : \; R_i \sim \mathbb{P}_1 \quad \text{or} \quad H_1 : \; R_i \sim \mathbb{P}_2 \tag{6}$$

There are three methodologies used to conduct hypothesis tests: non-parametric, parametric, and min-max hypothesis testing, respectively. Non-parametric hypothesis tests make no assumptions regarding parameter distribution, which leads to robust statistical tests, however similar to robust deterministic optimization, it fails to account for distributional information that may be available [8]. Similarly, for parametric hypothesis tests, a probability distribution for the hypothesis is assumed to be known, which is invalid for the case of asset returns. Lastly, in the same spirit as DRO, min-max hypothesis testing allows the distributions for each hypothesis to lie within an ambiguity set, therefore, providing control over the level of robustness against distributional uncertainty.

III. Distributionally Robust Mean-Variance

Although (4) is seemingly complex, there are tractable formulations for specific problems. [9] introduces the order - two Wasserstein distributionally robust mean variance problem (DRMVO) as:

$$\inf_{w \in \mathcal{W}_{\delta,\alpha}} \sup_{\mathbb{P} \in \mathcal{P}^2_\delta(\mathbb{P}_n)} w^\intercal \Sigma_{\mathbb{P}}(R) w \tag{7}$$

Where the robust feasibility set is defined as:

$$\mathcal{W}_{\delta,\alpha} = \{w^\intercal 1 = 1, \; \inf_{\mathbb{P} \in \mathcal{P}^2_\delta(\mathbb{P}_n)} w^\intercal \mathbb{E}_{\mathbb{P}}(R) \geq \alpha\} \tag{8}$$

Fortunately, [9] shows that (7) can be solved as a second order cone program:

$$\inf_{w \in \mathcal{W}_{\delta,\alpha}} (\sqrt{w^\intercal \Sigma_{\mathbb{P}}(R) w} + \sqrt{\delta}\, \|w\|_p)^2 \tag{9}$$

With the robust feasibility set, being equivalent to:

$$\begin{gathered} \mathcal{W}_{\delta,\alpha} = \{w^\intercal 1 = 1, \; w^\intercal \mathbb{E}_{\mathbb{P}_n}(R) \geq \alpha + \sqrt{\delta}\, \|w\|_p\} \\ \frac{1}{q} + \frac{1}{p} = 1 \end{gathered} \tag{10}$$

The minimum expected return threshold α and radii δ for the Wasserstein set are pre-specified hyper-parameters. DRMVO enjoys out of sample guarantees, is based on the empirical distribution of the data allowing it to use distributional information and has a similar computational complexity compared to standard portfolio optimization methods due to its conic reformulation (9).

IV. Robust Changpoint Detection

Recently, [10] propose the use of Wasserstein ambiguity sets for robust hypothesis testing because they allow for a suitable level of robustness via controlling the radii of the set and amount to computing the solution of convex programs that do not depend on the dimension of the data.

[10] formulates the simple test (6) as the problem of finding a detector $\phi : \Omega \rightarrow \mathbb{R}$, which for a given $R \in \Omega$ accepts H_1 and rejects H_2 whenever $\phi(R) \geq 0$ and vice versa. The approach is to select a detector ϕ that minimizes the maximum of the worst case probabilities of hypothesis classification error, where the probability distributions used to evaluate the error probabilities are allowed to vary within their own order-one Wasserstein ambiguity sets centered at the assumed $\mathbb{P}_1$ and $\mathbb{P}_2$.

$$\inf_{\phi:\Omega\rightarrow\mathbb{R}} \max(\sup_{P_1\in\mathcal{P}^1_{\theta_1}(\mathbb{P}_1)} P_1[R:\phi(R)<0], \sup_{P_2\in\mathcal{P}^1_{\theta_2}(\mathbb{P}_2)} P_2[R:\phi(R)\geq 0]) \tag{11}$$

(11) is a non-convex infinite dimensional program over all detectors and distributions within their feasible sets - making it quite challenging to solve. Fortunately, [10] introduces a safe convex approximation to (11) that is within a specified gap of the true solution by: i) invoking the use of a non-negative, nondecreasing convex function called the generating function l which satisfies $l(0) = 1$ and $\lim_{t\to\infty} l(t) = 0$ and ii) Bounding the worst case risk (the inner supremum of (11)) by the functional $\epsilon(.|\mathcal{P}^1_{\theta_1}(\mathbb{P}_1), \mathcal{P}^2_{\theta_2}(\mathbb{P}_2))$ which maps ϕ to $\mathbb{R}$

$$\epsilon(\phi|\mathcal{P}^1_{\delta_1}(\mathbb{P}_1),\mathcal{P}^2_{\delta_2}(\mathbb{P}_2)) = \sup_{P_1\in\mathcal{P}^1_{\theta_1}(\mathbb{P}_1),P_2\in\mathcal{P}^1_{\theta_2}(\mathbb{P}_2)} \mathbb{E}_{P_1}[l\circ(-\phi)(R)] + \mathbb{E}_{P_2}[l\circ(\phi)(R)] \tag{12}$$

By replacing (11) with $\inf_{\phi:\Omega\rightarrow\mathbb{R}} \epsilon(\phi|\mathcal{P}^1_{\delta_1}(\mathbb{P}_1), \mathcal{P}^2_{\delta_2}(\mathbb{P}_2))$, [10] show that the structure of the optimal detector for a given convex approximation is not dependant on the worst case distribution and it only depends on the choice of l. That is; the $\inf$ and the $\sup$ can be exchanged, which allows for the optimal detector to be computed for any distribution pair: P_1, P_2 including the worst case distribution pair P^*_1, P^*_2. Further, [10] also shows that the optimality gap between (12) and (11) is bounded by $\psi(\epsilon) - \epsilon$ where $\psi : \mathbb{R} \rightarrow \mathbb{R}$ and ϵ denotes the optimal value attained by (11). The tightest convex approximation they derive is for the case that $l(t) = (t+1)_+$ which corresponds to $\psi(p) = 2\min(p, 1-p)$, and the following optimal detector: $\phi(R) = \text{sign}(P_1(R) - P_2(R))$. Next, we consider the case where the centers of the Wasserstein balls are given by two empirical distributions $\mathbb{P}_{n_1}$, $\mathbb{P}_{n_2}$ with $\{R^{(1)}_k\}^{n_1}_{k=1}$ and $\{R^{(2)}_k\}^{n_1}_{k=1}$ samples respectively. To ease notation the two sets of samples are combined by setting $R_l = R^{(1)}_l$ for $l = 1, 2, \ldots n_1$ and $R^{(2)}_{l-n_1}$ for $l = n_1+1, n_1+2, \ldots n_1+n_2$. The evaluation of the worst case distribution amounts to solving the following convex program:

$$\begin{aligned} \max \quad & \sum_{l=1}^{n_1+n_2} (p^l_1 + p^l_2)\psi\Big(\frac{p^l_1}{p^l_1+p^l_2}\Big) \qquad (13) \\ \text{s.t.} \quad & \sum_{l=1}^{n_1+n_2}\sum_{m=1}^{n_1+n_2} \gamma^{lm}_k ||R_l - R_m|| \leq \theta_k, \quad k = 1,2 \\ & \sum_{m=1}^{n_1+n_2} \gamma^{lm}_1 = \frac{1}{n_1},\ 1 \leq l \leq n_1, \\ & \sum_{m=1}^{n_1+n_2} \gamma^{lm}_1 = 0,\ n_1+1 \leq l \leq n_1+n_2 \\ & \sum_{m=1}^{n_1+n_2} \gamma^{lm}_2 = 0,\ 1 \leq l \leq n_1, \\ & \sum_{m=1}^{n_1+n_2} \gamma^{lm}_2 = \frac{1}{n_2},\ n_1+1 \leq l \leq n_1+n_2 \\ & \sum_{l=1}^{n_1+n_2} \gamma^{lm}_k = p^m_k,\ 1 \leq m \leq n_1+n_2,\ k = 1,2 \end{aligned}$$

where $p_1, p_2 \in \mathbb{R}^{n_1+n_2}_+$ and $\gamma_1, \gamma_2 \in \mathbb{R}^{n_1+n_2}_+$ are the decision variables in the optimization problem.

After solving (13) for p_1 and p_2 the signal can be calculated:

$$\phi(R_l) = \text{sign}(p_1(R_l) - p_2(R_l)) \quad l = 1, 2, \ldots n_1+n_2 \tag{14}$$

Given the optimal detector ϕ for the single hypothesis test, in the case that the n_1+n_2 observations are sequential, two pieces are required to extend to the sequential change-point hypothesis testing problem (5). The first is a method to detect a changepoint stopping time T given the observations and ϕ, and the second is some measure of the error rate. Average run length is a standard error measure that is used and is given by the mean time between false detections: $\mathbb{E}^{\mathbb{P}_1}[T]$. [11] points out that the cumulative sum (CuSum) method is the optimal method to minimize the "worst-case" expected detection delay for a given average run length (ARL) [1]. The cumulative sum method specifies a parameter b whereby a change is detected at the following stopping time T_1

$$T_1 = \inf\{t > 0 | \max_{1\leq k\leq t} \sum_{i=k}^{t} (-\phi(\omega^i)) \geq b\} \tag{15}$$

Furthermore, building on [11] we show in an external appendix that the ARL can be bounded below via the following inequality:

$$\frac{1}{ARL} \leq \sum_{k=1}^{\infty} \Big(\mathbb{E}^{\mathbb{P}_1}[l(-\phi^*_k(\omega))]\Big)^k \tag{16}$$

Where $\phi^*_k = \phi + \frac{b}{k}$. Also, (16) can be computed efficiently to arbitrary precision by bounding the tail of the sum using a geometric series.

[1]Please see [11] for more details regarding the CuSum detector and it's optimality.

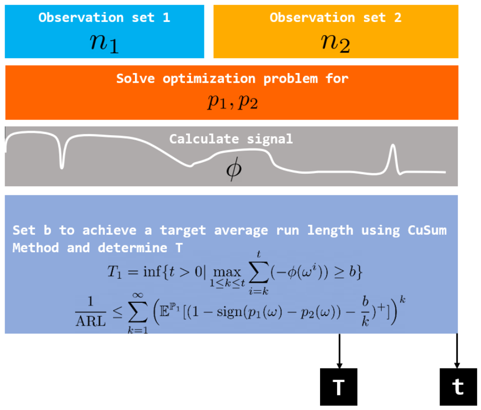

Fig. 1. For a given sequence of $n_1 + n_2$ observations the steps in the figure describe the changepoint detection method.

Therefore, to detect a change point among given set of $n_1 + n_2$ observations for a specified minimum ARL, i) split the dataset into n_1 and n_2 samples, ii) solve the optimization problem (13), iii) form the signal ϕ, and make use of the CuSum detection method while setting b to achieve a target ARL. This procedure is displayed in Fig. 1. This entire procedure is referred to as the Wasserstein Robust Change Point Detection method (WRCP). Going forward, the Wasserstein radii and splits between the observations are set to be equal: $\theta_1 = \theta_2 = \theta$, and $n_1 = n_2 = n$ allowing a single instance of the WRCP procedure to be specified by the tuple (n, θ, ARL). In the case of WRCP, the l_2 norm is used to define the order-one Wasserstein distance.

V. Combined Methodology

Given the theoretical benefits of WRCP and DRMVO, We propose: i) the iterated use of WRCP to detect changepoints to indicate when rebalancing should occur and ii) the use of DRMVO to rebalance the portfolio. To test our proposed approach, first, we proceed by generating rebalancing times and a-posteriori changepoint detection times via the iterated application of the WRCP method. The algorithm is as follows:

Step 1: Given a time series of n observations, specify the parameters n_1 and n_2, and initialize a variable t_{mid} that represents the time value of the first observation after n_1. Initialize a variable t_{now} that represents the current time, which is given by n_2 time steps after t_{mid}.

Step 2: Iterate through the dataset by applying the WRCP algorithm repeatedly through sequences of the dataset while updating the separation point t_{mid} based on the detection time and storing t_{now} as a rebalance time if there was a change in the subsequence.

VI. Experiments

Various experiments are performed using a dataset of S&P 500 stocks. The data consists of 468 stocks over from 2012-01-01 to 2020-03-31, obtained using Yahoo Finance [2]. We solved the underlying convex optimization problems using CVXPY [12, 13].

A. Experiment Description

We compare portfolio strategies by considering different optimization and rebalancing approaches. First, for simplicity, three strategies are compared: the equal weight portfolio strategy, the baseline mean-variance strategy, and the DRMVO strategy. All portfolio optimizations considered included a long-only constraint. The annualized return (AR), annualized volatility (SD), unadjusted Sharpe Ratio (SR) and the maximum drawdown (MDD) are measured for each optimization strategy. Second, we compare the risk-return performance for each of the optimization strategies with and without dynamic rebalancing rules. Since we are not modelling transaction costs, we compare a particular instance of the WRCP method against the fixed rebalancing strategy by setting the fixed rebalancing period such that the number of rebalancing periods in the fixed strategy is equal to the number of balancing periods in the dynamic strategy.

For the mean-variance and DRMVO problems, the target return $\alpha = 14\%$ was selected to be slightly higher than the annual market return since 2012. Furthermore, for the DRMVO problem, the Wasserstein metric was defined based on the infinity norm ($q = \infty, p = 1$) because that selection provides a natural scale for δ; since for feasible w, combining the long-only constraint and unit balance constraint implies $\|w\|_1 = \mathbf{1}^\intercal w = 1$, implying that $\mathbb{E}_{\mathbb{P}_n}(R) \ge \alpha + \sqrt{\delta}$. Based on this, δ is set to be squared multiples of α. To compare the dynamic rebalancing (via WRCP) against fixed rebalancing, three sets of parameters for the change-point detection algorithm are selected to represent different frequencies of rebalancing. It was found that several different parameter settings would detect identical change-points and that the parameters presented represent the most common change patterns.

B. Results and Discussion

Fig. 2 shows a series of rebalancing times outputted by WRCP with $(n, \theta, \text{ARL}) = (60, 0.1, 10000)$. The stopping times are overlaid on the stock return time series plotted with transparency. It is an interesting exercise to select a rebalancing time and search the corresponding headlines. For example, the second detected distributional change in Fig. 2 occurs on the same day as the Sandy Hook shooting.

Table I, displays the risk-return performance results for the tested strategies. It is apparent that the equally weighted portfolio exhibits similar performance across any rebalancing strategy. First, Table I shows that the use of WRCP leads to higher annualized returns along with lower portfolio volatility in eight of the nine mean-variance type cases. Furthermore, the use of WRCP leads to smaller maximum drawdowns for six of

[2] In this study, we drop any tickers with missing values or incomplete series. This introduces survivorship bias, but for comparative purposes, it is expected to impact all methods equally

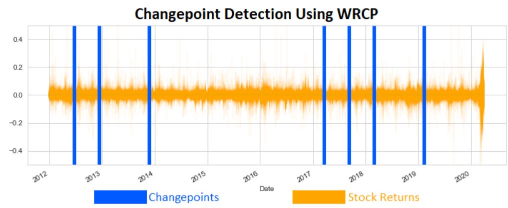

Fig. 2. A series of detected changepoints using the Wasserstein Robust Changepoint method. The average time between false detection is greater than 10,000, the Wasserstein radius is set to be 0.1, and the rolling window size is 60

TABLE I
PERFORMANCE TESTS

Target Return $\alpha = 0.14/250$					
Rebalance Schedule	**Optimization Type**	**AR**	**SD**	**SR**	**MDD**
Rebalanced every 44 days	Equal Weight	0.135	0.163	0.829	-0.390
47 Rebalancing Points	MVO	0.113	0.136	0.834	-0.340
	DRMVO: $\delta = (0.5\alpha)^2$	0.108	0.136	0.788	-0.340
	DRMVO: $\delta = (1.0\alpha)^2$	0.106	0.137	0.774	-0.340
WRCP	Equal Weight	0.135	0.163	0.828	-0.392
47 Rebalancing Points	MVO	0.160	0.129	1.240	-0.210
$n = 30$, $\theta = 0.1$,	DRMVO: $\delta = (0.5\alpha)^2$	0.160	0.129	1.239	-0.216
ARL $= 10000$	DRMVO: $\delta = (1.0\alpha)^2$	0.157	0.131	1.203	-0.228
Rebalanced every 172 days	Equal Weight	0.140	0.164	0.857	-0.372
12 Rebalancing Points	MVO	0.094	0.131	0.714	-0.311
	DRMVO: $\delta = (0.5\alpha)^2$	0.091	0.132	0.688	-0.311
	DRMVO: $\delta = (1.0\alpha)^2$	0.095	0.134	0.708	-0.318
WRCP	Equal Weight	0.135	0.163	0.828	-0.390
12 Rebalancing Points	MVO	0.122	0.128	0.950	-0.273
$n = 90$, $\theta = 0.075$,	DRMVO: $\delta = (0.5\alpha)^2$	0.123	0.128	0.957	-0.258
ARL $= 1000$	DRMVO: $\delta = (1.0\alpha)^2$	0.123	0.129	0.951	-0.249
Rebalanced every 296 days	Equal Weight	0.137	0.163	0.841	-0.384
7 Rebalancing Points	MVO	0.138	0.142	0.970	-0.316
	DRMVO: $\delta = (0.5\alpha)^2$	0.148	0.144	1.024	-0.318
	DRMVO: $\delta = (1.0\alpha)^2$	0.156	0.147	1.062	-0.322
WRCP	Equal Weight	0.136	0.162	0.844	-0.384
7 Rebalancing Points	MVO	0.149	0.152	0.978	-0.337
$n = 60$, $\theta = 0.1$,	DRMVO: $\delta = (0.5\alpha)^2$	0.150	0.152	0.987	-0.337
ARL $= 10000$	DRMVO: $\delta = (1.0\alpha)^2$	0.148	0.152	0.972	-0.337

the nine mean-variance type strategies. Furthermore, the effect of δ is less clear. The only cases where δ has an impact on the annual return of more than 3% are the fixed 44 and 296 day rebalancing strategies. In the case of fixed rebalancing every 296 days, DRMVO significantly outperforms the base mean-variance case in terms of annual return and Sharpe ratio. Conversely, for the strategy of rebalancing every 44 days, increasing δ has a negative impact on all performance metrics, however re-running the experiments at a finer grid of points shows that smaller δ are required to improve performance and that the improvement is smaller for shorter time horizons.

VII. CONCLUSIONS

Robust changepoint detection used in conjunction with distributionally robust optimization forms an approach to portfolio optimization that has several advantages over traditional methods. Particularly, the use of WRCP and DRMVO forms a tractable methodology - due to its formulation as convex programs, and is flexible - since it allows for a controllable level of robustness via the specification of Wasserstein radii. This study is the first to apply the proposed style of asset allocation to high dimensional financial data and shows evidence that distributionally robust changepoint and optimization may have performance benefits over traditional approaches in certain settings. There are several next steps of analysis, which include: i) Enhancement of the testing procedures to capture transaction costs, ii) comparisons of different risk measures and distributionally robust extensions, iii) comparing different methods of dynamic asset allocation, and iv) and extensions to multi-period problems.

REFERENCES

[1] Harry Markowitz. "Portfolio Selection". In: *The Journal of Finance* 7.1 (1952), pp. 77–91.

[2] James E. Smith and Robert L. Winkler. "The Optimizer's Curse: Skepticism and Postdecision Surprise in Decision Analysis". In: *Management Science* 52.3 (2006), pp. 311–322.

[3] James D Hamilton. *Macroeconomic Regimes and Regime Shifts*. Working Paper 21863. National Bureau of Economic Research, Jan. 2016.

[4] Wolfram Wiesemann, Daniel Kuhn, and Melvyn Sim. "Distributionally robust convex optimization". In: *Operations Research* 62.6 (2014), pp. 1358–1376.

[5] Donald Goldfarb and Garud Iyengar. "Robust portfolio selection problems". In: *Mathematics of operations research* 28.1 (2003), pp. 1–38.

[6] Daniel Kuhn et al. "Wasserstein distributionally robust optimization: Theory and applications in machine learning". In: *Operations Research & Management Science in the Age of Analytics*. INFORMS, 2019, pp. 130–166.

[7] Peter Nystrup et al. "Detecting change points in VIX and S&P 500: A new approach to dynamic asset allocation". In: *Journal of Asset Management* 17.5 (2016), pp. 361–374.

[8] G. Gül and A. M. Zoubir. "Minimax Robust Hypothesis Testing". In: *IEEE Transactions on Information Theory* 63.9 (2017), pp. 5572–5587.

[9] Jose Blanchet, Lin Chen, and Xun Yu Zhou. *Distributionally Robust Mean-Variance Portfolio Selection with Wasserstein Distances*. 2018. arXiv: 1802.04885.

[10] Rui Gao et al. *Robust Hypothesis Testing Using Wasserstein Uncertainty Sets*. 2018.

[11] Yang Cao and Yao Xie. "Robust sequential change-point detection by convex optimization". In: *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2017, pp. 1287–1291.

[12] Steven Diamond and Stephen Boyd. "CVXPY: A Python-Embedded Modeling Language for Convex Optimization". In: *Journal of Machine Learning Research* 17.83 (2016), pp. 1–5.

[13] Akshay Agrawal et al. "A Rewriting System for Convex Optimization Problems". In: *Journal of Control and Decision* 5.1 (2018), pp. 42–60.

APPENDIX

An external appendix for the proof of the average run length inequality (16) is available @ https://github.com/davidislip/ECE1505-Project/Appendix.pdf