Regularized Hypervolume Selection for Robust Portfolio Optimization in Dynamic Environments

Carlos R. B. Azevedo  
School of Electrical and Computer Engineering  
University of Campinas  
Campinas, São Paulo, Brazil  
Email: azevedo@dca.fee.unicamp.br

Fernando J. Von Zuben  
School of Electrical and Computer Engineering  
University of Campinas  
Campinas, São Paulo, Brazil  
Email: vonzuben@dca.fee.unicamp.br

Abstract—This paper proposes a regularized hypervolume (S-Metric) selection algorithm. The proposal is used for incorporating stability and diversification in financial portfolios obtained by solving a temporal sequence of multi-objective Mean Variance Problems (MVP) on real-world stock data, for short to long-term rebalancing periods. We also propose the usage of robust statistics for estimating the parameters of the assets returns distribution so that we are able to test two variants (with and without regularization) on dynamic environments under different levels of instability. The results suggest that the maximum attaining Sharpe Ratio portfolios obtained for the original MVP without regularization are unstable, yielding high turnover rates, whereas solving the robust MVP with regularization mitigated turnover, providing more stable solutions for unseen, dynamic environments. Finally, we report an apparent conflict between stability in the objective space and in the decision space.

Keywords—Multi-objective optimization, indicator-based search, portfolio optimization, mean-variance problem, robust statistics, regularization, dynamic environments.

I. INTRODUCTION

Portfolio optimization [11] addresses how to distribute wealth among risky investment assets in a potentially profitable market. In computational finance, the formulation of the Markowitz's Mean-Variance Problem (MVP) [19] and its diversification principle have been the essence of theoretical investment risk management for nearly half a century. The solution to the MVP leads to an optimal wealth allocation strategy which maximizes the expected return rate under a maximum risk constraint.

However, despite its intuitive appeal, the practical application of the Markowitz portfolio approach is not reliable and has been reported to behave poorly in real-world out-of-sample data, e.g. [10], [14]. This is mainly due to the prevalent presence of outliers in real-world data which makes the tails of the estimated returns distribution fatter than those of the Gaussians assumed in the MVP [12]. Hence, approaches to robustly estimate the returns distribution parameters have been investigated for mitigating such model risk, e.g. [11], [14].

The problem becomes NP-hard when practical constraints are considered in the MVP, such as bounded cardinality or transaction lots thresholds, for which classical convex quadratic programming techniques to analytically compute the efficient frontier (the geometric location where lie all efficient portfolios in the objective space) no longer work [8]. In those scenarios, multi-objective metaheuristics, such as Evolutionary Multi-Objective Algorithms (EMOAs), have been reported to yield good approximations of the efficient frontier for in-sample data (e.g. [17], [13], [20]).

When matching a portfolio at the efficient frontier to a certain preference profile, however, investment managers will want the recommended portfolio to remain stable upon the arrival of new data, in two senses: (a) the rebalanced portfolio which matches the preference profile should not differ too much from the previously recommended one; and (b) the expected return and risk of the old portfolio should change as little as possible from one environment to another.

Stable portfolios of the first kind usually possess lower turnover rates than unstable ones. It turns out that higher turnover rates imply higher transaction costs (e.g. stockbrokers commissions). Hence, buy-and-hold strategies will have minimum transaction costs, but will be more susceptible to the market volatility. On the other hand, very active management strategies can undermine the potential gains of the portfolio by incurring higher costs.

The second form of stability is also important because it implicitly means that the solution is adaptable over time, even if the individual stocks returns processes are non-stationary. In other words, as new data arrive, the estimated expected return and risk will likely change much less than what it would for unadaptable portfolios. In this sense, being adaptable means being resilient to new and possibly unpredictable scenarios.

This paper yields the following original contributions: (i) it proposes a robustness integrating hypervolume framework for obtaining temporal stable efficient frontiers by using a regularized version of the S-Metric Selection-EMOA (SMS-EMOA) [5]; (ii) it investigates the usage of a robust estimator [1] for the covariance matrix of returns; and (iii) it assess the stability of the resulting portfolios on real-world ex-post stock data, reporting the obtained turnover rates.

The main goal of this paper is thus to assess whether the Regularized SMS-EMOA (RSMS-EMOA) proposal will outperform the original SMS-EMOA on producing stable solutions in dynamic environments. For evaluating the merit of
each algorithm, we selected two MVP scenarios: one for which unstable maximum likelihood estimators are used, and other for which robust estimators provide a more smooth change between investment rounds. We also compare the RSMS-EMOA with the NSGA-II algorithm, which has been reported to perform satisfactorily for MVP frameworks (e.g. [17]).

It is expected that RSMS-EMOA proposal will be able to incorporate structural preferences over the candidate solutions into the optimization process. In the case of portfolio selection, a few of such desirable structural features may include: lower cardinality, balanced weight distribution, and stability in training and validation data sets.

This paper is organized as follows: Section II discusses robust statistics for real-world financial data; Section III presents the original and the proposed robust MVP formulations, as well as the rolling horizon version; Section IV outlines the proposed metaheuristics for MVP; Section V describes the experimental setup; Section VI presents the simulation results and discussion. Finally, Section VII provides conclusions and highlights ongoing and future research.

II. ROBUST STATISTICS

Robustness was defined by Huber and Ronchetti [16] as “insensitivity to small deviations from (distributional) assumptions”, when estimating population parameters from finite samples. This insensitivity is usually traded off by the efficiency of the estimator. Statistically, maximum likelihood estimators, such as the sample mean and variance for the Gaussian, carry optimal asymptotic efficiency when the data follow the assumed distribution. However, the presence of outliers in the data may lead to large estimation errors.

Robust statistics may not have optimal efficiency for the assumed model, but they can usually tolerate larger proportions of outliers in the sample before diverging from the true population parameters. For instance, the sample median can tolerate levels up to 50% of contamination (outliers) before breaking down, whereas the sample mean may tolerate none. Venables and Ripley [21] pointed out that the goal of robust statistics is to maintain high (although suboptimal) efficiency in the neighborhood of the assumed distributions.

A. Why Robustness in Financial Data?

The motivation for using robust statistics to estimate the MVP parameters is to improve estimation efficiency when the returns data deviates from the Gaussian distribution. In fact, real-world financial data hardly fit the normality assumption. Figure 1 shows the normal Q-Q plot for the 25 lagged FTSE 100 index fund daily data which is utilized in our experiments, from November 2006 to November 2012. We observe that the positive returns (profits) portion of the data presents a good fit to normality, while the probability of extreme negative returns (losses) is underestimated by the normality assumption. Remarkably, these are the same observations reported for NASDAQ and S&P500 by Frahm and Jaekel [12], even considering a different time frame. Such violation of the elliptical assumption causes the sample standard deviation to underestimate downside risk and, thus, one can expect the estimation error to impair the performance of the MVP portfolios in practice.

B. Pairwise Covariance Matrix Estimation

In this section, we describe the proposal of Alqallaf et al. [1] for a pairwise robust estimation of the covariance matrix. The main advantages of their approach are: (i) its quadratic complexity on the dimensionality of the sample vectors ($N^2$) when compared to the exponential complexity ($2^N$) of other methods; (ii) the possibility of using any robust univariate statistic for the first and second moments of the distribution; and (iii) the resulting matrix is guaranteed to be positive definite, what is not true for other pairwise approaches in the literature.

Given a $T \times N$ dataset matrix $X$, the $N \times N$ robust covariance matrix estimator is given as follows [1]:

- Compute the univariate robust statistics for the central tendency and dispersion for each column:
  \[
  \hat{m}_j = Q^{(2)}_j(x_j) \\
  \hat{s}_j = b \cdot \left( Q^{(3)}_j(x_j) - Q^{(1)}_j(x_j) \right),
  \]

  with $b = 0.7413$ chosen to provide an unbiased estimate of the standard deviation when the columns follow a Gaussian distribution. Note that Eq. (1) corresponds to the computation of the median (breakdown point of 50%), while Eq. (2) corresponds to the interquartile range, defined as the difference between the upper ($Q^{(3)}$) and lower ($Q^{(1)}$) quartiles (breakdown point of 25%).

- Compute the quadrant correlation estimate:
  \[
  \hat{c}_{lk} = \frac{1}{T \cdot \hat{m}_l} \sum_{t=1}^{T} \psi(x_{lt} - \hat{m}_l) \psi(x_{tk} - \hat{m}_k),
  \]

  in which
  \[
  \psi(y) = \begin{cases} 
  1, & y > 0 \\
  -1, & y < 0 
  \end{cases}
  \]
and $T_{i,k,0}$ is the number of rows with non-zero entries for both $(x_{il} - m_l)$ and $(x_{ik} - m_k)$. Note that $\tilde{c}_{lk}$ is effectively a robust estimative of the correlation coefficient using rank statistics, which has an intrinsic bias. For $x_{il}$ and $x_{ik}$ jointly Gaussian, the bias is removed by computing (see [16]):

$$\hat{c}_{lk} = \sin\left(\frac{\pi}{2} \tilde{c}_{lk}\right), l \neq k; \hat{\rho}_{lk} = 1, l = k. \quad (5)$$

The initial robust covariance matrix estimate is $\hat{\Sigma}_0 = \{s_{lk}\hat{\rho}_{lk}\}$, although not necessarily positive definite.

- Compute the spectral decomposition, $\hat{\Sigma}_0 = U\Lambda U^\top$, in which the columns of $U_{N \times N}$ form the orthonormal basis of eigenvectors and $\Lambda$ is the diagonal matrix whose entries are the corresponding eigenvalues, which are not necessarily positive. By (a) row-wisely projecting the original dataset $X$ onto the eigenvectors basis, i.e., $\hat{x}_t = U \cdot x_t$ ($x_t$ is the $t$-th row of $X_t$); (b) computing the univariate robust dispersion statistics $\hat{s}_j = 0.7414 \cdot \left(\hat{Q}_j^2(\hat{x}_j) - \hat{Q}_j^1(\hat{x}_j)\right)$; and (c) noting that, for $\hat{\Sigma}_0$, the entries of $\Lambda$ are the variances of the projected data on the eigenvectors basis, one can simply replace all $\lambda_j$ by $\hat{s}_{j}^2$ to form $\Lambda = \{\hat{s}_{j}^2\}$.

- The final step consists on applying a series of column permutations over $\Lambda$ so that the new eigenvalues are ordered from the largest to the smallest one in the main diagonal. The final unbiased robust estimative for the covariance matrix, with the positive definiteness property restored is $\hat{\Sigma} = U\Lambda U^\top$.

### III. The Mean-Variance Problem

Let $r = (r_1, r_2, \ldots, r_N)^\top \in \mathbb{R}^N$ be a random vector composed by random returns of $N$ risky assets, in which $\mu_r$ and $\Sigma_r$ are its mean and covariance matrix, and let $w = (w_1, w_2, \ldots, w_N)^\top \in S^N = \left\{w \in \mathbb{R}^N : \sum_{j=1}^N w_j = 1\right\}$ be the weight vector of the portfolio, denoting the proportion of wealth to be invested in each available asset, in which $S^N$ denotes the $N$-simplex. Then, the classical Mean-Variance Problem (MVP) was formulated by Markowitz [19] as the following constrained quadratic program:

$$w^* = \arg \min_w \left\{w^\top \Sigma w : \mu^\top w \geq \mu_0, \sum_{j=1}^N w_j = 1\right\}, \quad (6)$$

in which $\mu_0$ is the minimum expected return for $w^*$. If the first and second moments are known in advance, the MVP can be analytically solved. However, $\{\mu, \Sigma\}$ must be estimated from the available data in real-world applications.

#### A. The Dynamic Multi-Objective Robust MVP

In this paper, we transform (6) into a time-varying mixed-binary bi-objective optimization problem:

$$\begin{align*}
\min_{A_t, w_t} & \quad w_t^\top A_t \Sigma_t A_t w_t \\
\max_{A_t, w_t} & \quad \hat{\mu}_t^\top A_t w_t \\
\text{s.t.} & \quad 1_t^\top A_t w_t = 1, w_{t,j} \geq 0, \quad \text{tr}(A_t) \leq a_u \\
& \quad \hat{a}_{t,j} = 1, w_{t,j} > 0, \quad i, j \in \{1, 2, \ldots, N\}.
\end{align*} \quad (7)$$

in which $A_t$ is a $N \times N$ binary diagonal matrix so that $A_t(j,j) = 1 \iff w_{t,j} > 0$, i.e., the $j$-th asset belongs to the portfolio, and $A_t(j,j) = 0 \iff w_{t,j} = 0$. The motivation for including the binary variables into problem (7) is to allow for an increased search power over the feasible solution space. Hence, the usage of search operators over the binary diagonal matrix $A_t$ complies with the bounded cardinality constraint of Eq. (10) (in which $\text{tr}(A_t)$ denotes the trace of $A_t$), allowing for the optimization procedure to quickly reduce (or increase) the cardinality of the portfolios (see Section IV for details). In addition, $\{\hat{\mu}_t, \Sigma_t\}$ are estimated by using the sample median and the robust covariance estimator discussed in Section II-B.

1) The Rolling Horizon Plugin-Rule: We now set the current time index as $t \leftarrow t_0(t_0 \geq K)$ and adopt the following three-step procedure:

1. Given the time series of observed returns from the latest $K$ time-steps for each asset, i.e., $R_{t-K}^{-1} = \{r_k\}$, estimate $\{\hat{\mu}_t, \hat{\Sigma}_t\}$ from $R_{t-K}^{-1}$.
2. Solve (7) after replacing the unknown parameters with those estimated from the data in the previous step.
3. Implement the efficient portfolio $\hat{w}_t^\top$ at time $t$. This procedure is called the plugin-rule (see e.g. [11]).

The efficient portfolio can be hold unchanged for the next $H$ time-steps, in which $H$ is the investment horizon. During period $\Delta_t(H) = [t, t+H]$, the Return Over Investment (ROI) associated with $\hat{w}_t^\top$ at the end of $\Delta_t(H)$ is given by

$$\text{ROI}(\hat{w}_t^\top, \Delta_t(H)) = \frac{V_{t+H}(\hat{w}_t^\top) - V_t(\hat{w}_t^\top)}{V_t(\hat{w}_t^\top)}, \quad (11)$$

in which $V_t(w)$ is the net asset value of $w_t$ at time $t$. In the rolling horizon approach [9], the time index is then set as $t \leftarrow t + H$, the returns matrix $R_{t-K}$ is updated to include the new observations arrived during $\Delta_t(H)$, discarding the oldest $H$ data points, and the procedure goes on as the portfolios are rebalanced for the new environments.

### IV. Metaheuristics for the MVP

When practical constraints such as bounded cardinality (Eq. (10)) are integrated into the MVP, the problem becomes a mixed-binary quadratic programming, which is NP-hard [18]. Therefore, in realistic settings, stochastic approximation methods such as metaheuristics are often considered.

Metaheuristics such as simulated annealing, memetic, genetic, and estimation of distribution algorithms have been applied for the MVP [2], [6], [20]. Fewer studies assessed multi-objective versions (e.g. [17], [13], [8], [15]) but none of them considered dynamic environments settings, or the integration of robust statistics and regularization terms. Some even failed to report performance in ex-post data.

#### A. NSGA-II

Due to its good reported performance for the standard MVP [17], we use NSGA-II as a baseline for solving (7).

---

2 We follow Deb and Agrawal’s [7] definition of search power, which corresponds to the probability of moving to an arbitrary location of the solution space when applying one search operator on any given solution.
to (10). The idea behind NSGA-II is to take advantage of dominance relations among vectors in the objective space, $F_M$ ($M$ is the number of objectives), by partitioning the population into $v$ classes $F_1, \ldots, F_v$, so that the non-dominated solutions ($x \in F_i$) are never replaced by dominated ones ($y \in \bigcup_i F_i \setminus \{F_i\}$). Within classes, solutions with high crowding distance (CD, a neighborhood density estimate) values are preferred, and the extreme solutions of each class are preserved, providing good coverage of the Pareto Frontier (PF). The role of CD in NSGA-II is to eliminate redundant solutions so that vectors are well spread over the PF.

### B. SMS-EMOA

In the S-Metric Selection EMOA (SMS-EMOA) [5], the same framework of NSGA-II is utilized (e.g., partitioning into classes, etc.), except for: (a) NSGA-II is a generational algorithm (i.e., updates all population at once), whereas SMS-EMOA is a steady-state one (i.e., updates population as new solutions are sampled); and (b) the selective pressure within classes is guided by the solutions hypervolume ($S$) contribution, instead of the CD values. The idea is that the least contributing solutions should be iteratively discarded.

The $S$ contribution of a solution $z'$ in class $F_i$ is given by

$$\Delta_S(z', F_i) = S(F_i) - S(F_i \setminus \{z'\}).$$  

Beume et al. [5] argued that $S(P_{t+1}) \geq S(P_t)$ holds after one iteration of SMS-EMOA (see Algorithms 1 and 2).

1) Robustness Integrating Hypervolume ($S^\phi$): The hypervolume ($S$, S-Metric) [3] is a theoretical performance indicator that captures both the proximity of $F_i$ to the true PF and its distribution over $F_M$. Given a reference point $z_{ref} \in F_M$, $S(F_i, z_{ref})$ is the hypervolume of the polytope formed by the union of all dominance regions of each $z' \in F_i$, bounded by the hyperbox that has $z_{ref}$ and the ideal vector in $F_M$ as opposite vertices. Mathematically,

$$S(F_i, z_{ref}) = \int_{z \in F_M} 1_{H(F_i, z_{ref})}(dz),$$  

where $H(F_i, z_{ref}) = \{z \mid \exists z' \in F_i : z' \prec z \prec z_{ref}\}$,  

in which $1_{H(F_i, z_{ref})}$ is the characteristic function of $H(F_i, z_{ref})$, that is 1 if $z \in H(F_i, z_{ref})$ and 0 otherwise. Its robustness integrating version was introduced by Bader and Zitzler as [4]:

$$S^\phi(F_i, z_{ref}) = \int_{z \in F_M} \alpha^\phi_{F_i}(z)dz,$$  

in which the characteristic function in $S$ (Eq. (13)) is replaced by the attainment function $\alpha^\phi_{F_i}(z)$:

$$\alpha^\phi_{F_i}(z) = \begin{cases} \phi \left( \min_{z' \in F_i} r(z') \right), & \text{if } z \in H(F_i, z_{ref}) \\ 0, & \text{otherwise.} \end{cases}$$  

The function $r : F_M \to [0, 1]$ is designed such that the maximally robust solutions are assigned 0 and the least robust ones are assigned 1. The desirability function $\phi : [0, 1] \to [0, 1]$ can then assume various forms to account for preferences toward robustness. In Eq. (15), $z \in H(F_i, z_{ref})$ contributes 100% to $S^\phi$ only if it is fully desired in terms of robustness. For $r(z) > 0$, though, the contribution of $z$ to $S^\phi$ is discounted by the desirability assigned to the most robust solution $z' \in F_i$ that dominates $z$.

### C. Regularized SMS-EMOA

We propose integrating $r$ into Algorithm 2 as a multiplicative regularization term by means of the desirability function, i.e., $\Delta S(z, F_i, \phi(r(z))$. This leads to the Regularized SMS-EMOA (RSMS-EMOA), which is equivalent to maximizing $S^\phi$ within classes. The choice of $r$ allows for integrating selective pressure toward the following three desirable aspects of portfolio optimization: diversification, cardinality reduction, and stability. For the original SMS-EMOA, we define $r_0(z) = 0, \forall z \in F_M$. For all versions of the algorithm, the desirability function is a linear decreasing mapping: $\phi(r(z)) = 1 - r(z)$.

1) Entropy Regularization: In order to improve the diversification of the candidate portfolios, we define the regularization function $r_E$ as the normalized (negative) entropy $^5$:

$$r_E(A, w) = 1 + \frac{1}{\text{tr}(A)} \sum_{i=0}^{N} a_{ii} w_i \log a_{ii} w_i,$$  

in which portfolio $(A, w)$ attains maximum diversification when its weights are equally distributed among the assets that take part in it, i.e., $w_i = \frac{1}{\text{tr}(A)} \forall i$ such that $a_{ii} = 1 \implies r_E(A, w) = 0$. Also, $\text{tr}(A) = 1 \implies r_E(A, w) = 1$, i.e., imbalanced portfolios are less desired. It is worth noting that $r_E$ is insensitive to the portfolio cardinality and, therefore, if an equally-weighted two-assets portfolio attains maximum diversification, so do the $N$-assets index portfolio.

2) Cardinality Regularization: Besides diversification, another desirable feature is low cardinality portfolios, inducing sparse weight vectors. Such portfolios may be easier to manage in practice, but may also conflict with risk minimization. The regularization function $r_C$ is given as

$$r_C(A, w) = \frac{1}{N} \text{tr}(A),$$  

in which $N$-assets portfolios are the least desired ones.

---

$^3$In fact, solutions from superior classes $(F_i)$ are preferred over those from inferior ones $(F_j)$, with $i < j$.

$^4$We omit the reference point, $z_{ref}$, for simplicity.

$^5$We abuse notation to simplify the exposition. Hence, the objective vector $z = f(A, w) \in \mathbb{R}^2$ is $(w^T A \Sigma A w \mu^T A w)^T$ and $r(z) = r(A, w)$.
3) **Stability Regularization**: Finally, we focus on the second form of portfolio stability discussed in Sec. I. This is important because it means that the solution is adaptable over time, even if the market enters a more volatile phase (see e.g. [15]). As new data arrive, the estimated expected return and risk of such stable portfolios will likely change less than what it would for unadaptable portfolios.

For such, we evaluate the change in performance of the candidate portfolio in a validation dataset which is temporally correlated with the training set. If $R_{t-K}^S$ is the returns data matrix in period $\Delta_{t-K}(K)$, we partition the matrix into $T = R_{t-K}^SH^{-1}$ and $V = R_{t-K}^I$ ($H \ll K$), in which $T, V$ are the training and validation returns datasets for period $\Delta_{t-K}(K)$. Let $z_T, z_V$ be the objective vectors of the portfolio evaluated in $T$ and in $V$, respectively. Then,

$$r_S(z) = 1 - \frac{1}{1 + ||z_T - z_V||^2}, \quad (19)$$

in which the objective vectors that remain stable under the same time varying conditions are most desired.

V. **Experimental Setup**

We assessed two RSMS-EMOA($r$) versions$^6$ in our experiments. The RSMS-EMOA($rs$) utilizes the regularization function of Eq. (19) and tries to regularize the portfolios so that their estimables are stable in subsequent environments. For the second version, we mixed $r_E$ and $r_C$ to form $r_{EC} = (1 - \alpha) r_E + \alpha r_C$, yielding the RSMS-EMOA($r_{EC}$). whose goal is to achieve lower cardinality portfolios with a diversified wealth allocation, as a plausible ad-hoc preemptive strategy for mitigating risk and for making the portfolios easier to manage. We set $\alpha = 1/2$ for the experiments.

Because NSGA-II, SMS-EMOA, and RSMS-EMOA($r$) all share the same evolutionary algorithmic framework, we were able to render all parameters values the same for each version. The setup utilized are as follow: population size was set as $\mu = 100$; mutation rate of 0.2; crossover probability of 1.0; and binary tournament selection. Because our focus is on analysing the effects of robustness and regularization integration, we utilized a very simple uniform crossover in which two offspring inherit each weight of the portfolios from either parents with 1/2 probability.

For mutation, we randomly choose between two operators: (1) modify the non-zero weights; or (2) modify the binary diagonal matrix (A). If operator (1) is selected, then, with probability 1/2, we either increase or decrease the investment on a randomly chosen asset by a uniformly drawn factor from 10 to 50%. If (2) is selected, then, with probability 1/2, we either add or remove a randomly chosen asset from the portfolio. If it is removed, we simply set its weight (and its entry in A) to zero. If it is added, we set its entry to one in A and randomly set its weight within a ±10% range from an equally-balanced weight, $1/\text{tr}(A)$. After any operation, the portfolio weights are renormalized.

Other parameters for the simulations were: 100 generations for each algorithm in each environment; 30 runs; 89 FTSE 100 assets, using the adjusted close prices from 20/11/2006 to 20/11/2012 (including the financial crisis period); three investment horizons ($H \in \{25, 50, 100\}$); the training periods length $K$ were chosen such that $H/K \approx 0.13$, meaning that the rate of change between environments is constant for all horizons. We also used Mann-Whitney paired significance tests at a 5% level where appropriate. In addition, we narrowed the scope of the experimental study, relaxing the cardinality constraint and allowing all 89 assets to take part in the portfolios, but added a minimum investment constraint of 0.5%, which is easily handled by pruning the asset which violates that condition and renormalizing the portfolio.

Finally, we adopted a seeding approach in which, each time the environment changes, the previous efficient portfolios serve as starting points for rebalancing as part of the new populations of candidate solutions.

A. **Evaluation Metrics**

We describe in this section additional metrics used for discussing and comparing the performance of each algorithm.

1) **Ranks Correlation Between Periods**: We compute the Spearman Correlation coefficient between the portfolios in the efficient frontier, between the expected return estimated before and after each rebalancing operation. Stable frontiers are expected to possess high rank correlation, keeping the relative order of the portfolios so that e.g. high ranked portfolios will remain so in a new investment environment.

2) **Turnover Rates**: For measuring the Turnover Rate (TR) of a given investment strategy $s$ over time, we follow DeMiguel and Nogales [9] and define TR as the average rate of wealth traded between subsequent investment periods:

$$\text{TR}(w^s_1, w^s_{t+1}) = \frac{1}{T-K-1} \sum_{t=K}^{T-1} \sum_{j=1}^{N} |w^s_{t+1,j} - w^s_{t,j}|. \quad (20)$$

High TR has direct consequences for trading, since it generates high transaction costs.

3) **Sharpe Ratio**: The Sharpe Ratio (SR), is defined as the expected return of $w$ in excess of the risk-free return divided by the volatility (standard deviation):

$$\text{SR}(w) = \frac{1}{\sigma_w} (R_w - R_0), \quad (21)$$

in which $R_w = \mu^T w$ and $R_0$ is the return of a risk-free asset with a fixed, low return (we use a zero return asset).

B. **Decision-making Rule**

In real-world portfolio management, it is common for investors to look after the portfolio which is the most rewarding for each risk unit taken. Therefore, at the beginning of each investment period, we identify the efficient portfolio which possess the highest SR value (Eq. (21)), $w^*_SR$, and implement it in a rolling horizon, ex-post data simulation, rebalancing the portfolio for 6, 22, and 56 periods of $H = 100, 50$, and 25 days, respectively. Besides the maximum SR value achieved by $w^*_SR$, we also report the turnover rates, entropy, and the cardinality of the portfolio. The results are contrasted to those of the naïve equally-weighted portfolio, $w^*_\text{Index}$, implementing a buy-and-hold strategy for the whole test period.

---

$^6$Source codes at http://www.researchgate.net/profile/Carlos_Azevedo2
We separate the results reporting in two parts: first, in Sect. VI-A, we point out the observed benefits of solving the robust MVP over the original non-robust version in terms of portfolio diversification and stability; then, in Sect. VI-B, we discuss the effects of integrating regularization terms under the RSMS-EMOA(\(r\)) framework.

### VI. Results and Discussion

We separate the results reporting in two parts: first, in Sect. VI-A, we point out the observed benefits of solving the robust MVP over the original non-robust version in terms of portfolio diversification and stability; then, in Sect. VI-B, we discuss the effects of integrating regularization terms under the RSMS-EMOA(\(r\)) framework.

#### A. The Effects of Robust Estimation

The first interesting observation we can draw regards what happens to the robustly estimated return and risk\(^7\) when considering the portfolios obtained by the algorithms solving the non-robust MVP, and vice-versa. For instance, we observe that, despite solving MVP by utilizing the robust covariance matrix for estimating risk, both NSGA-II(\(R\)) and SMS-EMOA(\(R\)) produced portfolios with much lower risk when evaluated using the non-robust sample covariance estimator (Figure 2(c)). The same did not happen with regard to return (Figure 2(d)), as those algorithms evolved efficient portfolios with lower estimated (sample average) return, although the estimated median return was statistically significantly higher than for the non-robust portfolios. This result suggests that (i) solving the original MVP may lead to risk overestimation; and (ii) the robust MVP may lead to less risk-inclined portfolios.

Another remarkable effect of integrating the robust statistics into the original MVP was observed for the portfolios diversification. While the algorithms solving the non-robust MVP, NSGA-II(\(NR\)) and SMS-EMOA(\(NR\)), obtained low cardinality (see Table I) and low entropy portfolios, the opposite happened with NSGA-II(\(R\)) and SMS-EMOA(\(R\)), which favored medium cardinality portfolios with more balanced wealth allocations (Fig. 3(a)). This observation strongly complies with the observed lower estimated risk for the robust versions previously noted.

Regarding the effects of solving the robust and original MVP formulations on stability in the objective space, we note from Fig 4(d)–(f) that the robust versions did not always lead to the most stable Pareto frontiers, as measured by the offline hypervolume, which is also a remarkable outcome.

Finally, from the ex-post simulations, in which the maximum SR (Max\(SR\)) portfolio is implemented in each period, we observe from Table I that the non-robust versions were outperformed by all other robust versions in terms of stability

---

\(^{7}\)All boxplots represent offline measures, meaning that we collect the statistics at the end of each training period, before rebalancing the portfolios.
in the decision space, as measured by the turnover rates. Here, we find intuitive the fact that, due to the lower cardinality portfolios evolved by NSGA-II(NR) and SMS-EMOA(NR), more effort is needed to rebalance the portfolio in order to handle changes in the non-robust return/risk estimates as new data arrive, what was somewhat expected.

B. The Effects of Regularization

Among the algorithms solving the robust MVP formulation, the RSMS-EMOA($r_{EC}$) was successful on obtaining the highest diversified portfolios (with statistical significance), as expected (Fig. 3(b)), followed by the other regularized version, RSMS-EMOA($r_S$). The plain SMS-EMOA(R), on the other hand, appeared to have greedily traded off diversification by return maximization, what may explain the lowest entropy portfolios for the robust MVP while attaining higher hypervolume values than those achieved by NSGA-II(R) (Fig. 3(d)). In terms of offline hypervolume, RSMS-EMOA($r_S$) outperformed its contenders, followed by the other SMS-EMOA versions, which outperformed NSGA-II. The striking performance of RSMS-EMOA($r_S$) in terms of hypervolume is coherent with the fact that it also obtained the portfolios with the lowest median class values among all versions. After the non-dominated sorting procedure is applied for all population, each portfolio is assigned a class value, ranging from zero to the number of partitions found. It can be seen that RSMS-EMOA($r_S$) is the version that keeps the highest number of mutually non-dominated portfolios in the population, what may contribute for a more effective exploitation of the efficient frontier, thus greatly improving its population hypervolume.

The version that seeks to incorporate stability in the objective space, RSMS-EMOA($r_S$), has achieved its goal (see Figure 4(a)–(c)), although tied with NSGA-II(R) for $H = 25$, and, therefore, for investors seeking portfolios which are less likely to be affected by market volatility over time in terms of the estimated return and risk, it is the recommended choice among the algorithms we tested. What is interesting about this result, however, is that such objective space stability lead to higher instability in the decision space (as measured by turnover rates, Table I), when compared to the other algorithms solving the robust MVP.

This apparent conflict between the two forms of stability makes sense when recalling that RSMS-EMOA($r_S$) also achieved the highest hypervolume values: when attempting to stabilize the return and risk estimatives between rebalancing periods, the algorithm may have become more effective at improving over local Pareto sets in the decision space, given that it showed the highest levels of improvement in hypervolume between training periods. However, this goal was attained at the cost of more dramatically modifying the weights of the seeded population of efficient portfolios from one investment environment to another.

VII. Final Remarks

The regularized, robust algorithms proposed in this paper stands under the class of preemptive approaches for model (the robust part) and structural (the regularization part) risk mitigation. The results obtained with the proposed rolling horizon multi-objective portfolio framework for the robust Mean-Variance Problem (MVP) support the following remarks:

---

8The reference point was taken as (-50,50) for return/risk.
The plain SMS-EMOA have outperformed NSGA-II on approximating the MVP efficient frontier for real-world stock data, both under robust and non-robust estimators for the mean and covariance matrix of returns, but the Pareto frontiers obtained with NSGA-II were the most stable ones;

The consideration of robust statistics in the MVP formulation, specifically the pairwise robust covariance matrix estimation, allowed for more stable portfolios in the decision space, as significantly lower turnover rates have been observed;

The proposed RSMS-EMOA framework was effective on incorporating desired structural aspects in the final portfolios, such as reduced cardinality and improved diversification and stability; and

The two concepts of stability discussed for both spaces (objective and decision) are desirable, but, as indicated in our results, they may be conflicting for different investment profiles.

Finally, despite the focus on the MVP application, we claim that the RSMS-EMOA can be used in virtually any problem for which regularization plays an essential role, such as on avoiding overfitting in machine learning multi-objective model building (e.g. in neural networks architecture optimization).

A. Future and Ongoing Work

Motivated by such positive results, we suggest assessing a RSMS-EMOA version explicitly incorporating the turnover rate (measured in reduced validation sets) as the regularization term. The higher computational effort of such task, however, would not pose a problem, since this step can be done offline within reasonable time, even for databases containing thousands of assets.

Ongoing investigation on the evaluation of such preemptive algorithms is being pursued on controlled scenarios with generative models. We are also researching anticipatory stochastic multi-objective approaches for this application, in which the effects of the estimation error (model risk) on an hypervolume-based loss function can be determined for assessing the suitability of the assumed predictive models.

ACKNOWLEDGMENT

The authors would like to thank Capes for the financial support. The first author has been supported by two PhD scholarships thanks to CNPq, and FAPESP, process numbers 141058/2011-0 and 2012/16504-5. The second author is partially supported by CNPq, process number 304622/2010-9.

REFERENCES


