Analysis of risk measures in multiobjective optimization portfolios with cardinality constraint
(Análise de risco em otimização multiobjetivo de carteiras com restrição de cardinalidade)

Rodrigo T. N. Cardoso†
Bruno Cândido Barroso‡
Mariana dos Santos de Oliveira*
Felipe Dias Paiva**

Abstract
Portfolio selection has been the subject of extensive study to obtain increased returns, minimizing the investment risk. However, the most appropriate risk measure to consider is still an open problem. The aim of this work is to study different risk measures in the multiobjective portfolios optimization with cardinality constraint and rebalancing. The in-sample analysis compares the fronts of each algorithm, metric and range of cardinality, and out-of-sample analysis compares the results of each measure of risk with each other and with two benchmarks. The returns of portfolios are compared in terms of assets choice and assignment of weights. Statistical tests are performed to verify if any measure of risk shows superiority. Results indicate that downside risk measures can reduce the cardinality and the risk of financial drawdown without reducing drawup, once they are able to reduce just the negative historical returns scenarios.

Keywords: Risk measures; Cardinality constraint; Portfolio optimization; Multiobjective optimization
JEL Code: G11, G32.

1. Introduction

The financial market generally allows investors to obtain higher profits, with the counterpoint of being exposed to greater risks. Therefore, investors are interested in simultaneously maximizing profits and minimizing risks. Using mathematical and computational models is proven to be decisively helpful in achieving optimal investments in stock markets. Markowitz (1952) was the first to establish a model based on diversification, with the premise that stock prices are normally distributed. In his work, the return of a portfolio is considered as the weighting of the expected values of the individual assets' returns, and the risk is considered as the variance or dispersion of the assets returns series in relation to the expected value (Nawrocki, 1999).

In portfolio selection, the higher the return, the greater the risk incurred, and there are two conflicting objectives. Thus, the investor must make a tradeoff between risk and return. Therefore, a multiobjective model is considered here, which presents solutions of compromise between risk and return as the final answer. The model also considers a cardinality constraint, with a given minimum and maximum quantity of assets to be included in the portfolio. The inclusion of the cardinality constraint greatly increases the algorithmic complexity of the solution (Cheng & Gao, 2015), so that the use of computational techniques as evolutionary algorithms is advisable to ensure good solutions in a reasonable time.

A risk measure can be defined as a function that associates to each distribution a number which describes its riskiness (Roman & Mitra, 2009). The study of new measures of risk is important since several financial catastrophes have shown the need to better evaluate the risk of investments. In fact, several risk models have been proposed taking into account that the return distributions may have heavy tails and fluctuations in the second moment of prices (Rosario, Mantegna & Stanley, 2000).
This work considers the comparison of the following risk measures in the Brazilian market case: variance, Exponentially Weighted Moving Average (EWMA), Generalized Autoregressive Conditional Heteroskedasticity (GARCH), semivariance, Value at Risk (VaR), and Conditional Value at Risk (CVaR).

In fact, the variance as a measure of risk, as proposed by Markowitz, is subject to many criticisms. For example, it considers the price variation as the risk and, thus, returns below and above expectations are equally treated. Given this scenario, downside risk measures, such as VaR and CVaR, which consider just the risk from the perspective of losses, have gained ground over the years (Guo, Chan, Wong & Zhu, 2019). In view of that, risk aversion is one of the premises of this work, assuming that when the return is constant, investors always choose the lowest risk portfolio. Furthermore, some articles have also dealt with traditional volatility measures, such as EWMA and GARCH, as the risk measure in portfolio selection problems (Sahamkhadam, Stephan & Östermark, 2018; Tang & Do, 2019).

Portfolio optimization problems with cardinality constraints cannot be solved by global techniques from discrete optimization (Burdakov, Kanzow & Schwartz, 2016). Furthermore, genetic algorithms have been used in the multiobjective portfolio optimization literature, both because they represent a good alternative in the face of hard optimization problems and also because they simultaneously generate several nondominated solutions in each execution run (Ferreira, Barroso, Hanaoka, Paiva & Cardoso, 2017).

To solve each biobjective risk-return optimization problem with cardinality constraint, two standard genetic algorithms are proposed here: based on Non-dominated Sorting Genetic Algorithm (NSGA-II) (Deb, Pratap, Agarwal & Meyarivan, 2002) and based on Strength Pareto Evolutionary Algorithm (SPEA2) (Zitzler, Laumanns, Thiele et al., 2001). For both algorithms, specific operators are considered, according to the nature of the problem. Statistical tests are performed to evaluate the consistency of the return, risk, drawdown, and drawup of each risk measure.


Comparing the performance of multiobjective algorithms, considering the variance as the risk measure, Anagnostopoulos & Mamanis (2011) show that algorithms SPEA2, NSGA-II, and e-MOEA performed better. Deb, Steuer, Tewari & Tewari (2011) propose a new version of NSGA-II with local search in a portfolio optimization problem with cardinality constraint. Considering just CVaR as the risk measure, Branda, Bucher, Červinka & Schwartz (2018) show that NSGA-II performs better in a problem with cardinality constraint. Other algorithms have also been used in real-world portfolio optimization, such as those based on differential evolution (Krink & Paterlini, 2011).

Synthetically, this article proposes:

- a comparison of optimizing with cardinality constraint and several risk measures: variance, semivariance, EWMA, GARCH, VaR with two approaches, under normal distribution of returns and their robust counterparts under moment conditions, and CVaR, considering two multiobjective genetic algorithms, one based on NSGA-II and another based on SPEA2, in an in-sample analysis; and

- a comparison of optimizing with these risk measures in an out-of-sample analysis, performing a realistic case study, with rebalancing, considering fifty-three shares of B3, the Brazilian stock market, between the years 2012 to 2015.

The in-sample study shows that the algorithm SPEA2 performed better. With this algorithm, the relationship between the cardinality of the portfolio and each risk measured is evaluated, considering three cardinality intervals. Finally, the Pareto fronts obtained by each risk measure are compared, and the optimal portfolios of
the greater ratio of risk and return as well. In the out-of-sample analysis, the stock marketing trades show the superiority of the risk measures based on quantiles, CVaR, and VaR, because they provide better monthly and accumulated returns and they are able to reduce negative returns scenarios, promoting a reduction of drawdown risk without limiting the financial drawup.

The article is organized as follows. Section 2 shows the theoretical basis of the considered risk measures. Section 3 shows the model and the proposed algorithms to solve the problem, and explains the metrics chosen to compare the algorithms. Section 4 presents the methodological details of the comparison between the two proposed algorithms (the in-sample analysis) and the behavior of the best algorithm in a stock market simulation with the comparison of the risk measures (the out-of-sample analysis). Section 5 shows and discusses the results. Finally, Section 6 summarizes and analyzes the implications of the results.

2. Risk measures

Risk models can be divided into two categories: dispersion measures and measures based on quantiles (Roman & Mitra, 2009). Dispersion measures are based on a targeted return and generate only positive values. They can be divided into symmetrical and asymmetrical measures. The symmetrical measures evaluate the returns dispersion treating positive and negative values in the same way, while asymmetrical ones capture the risk considering the downside risk, the undesired returns. The dispersion measures considered here are variance, EWMA, semivariance, and GARCH. Measures based on quantiles quantify the risk by targeting the severity of losses, focusing on the left tail of the returns probability distribution. The measures based on quantiles considered here are VaR and CVaR.

2.1 Variance

The formula of variance is expressed by:

$$\sigma^2 = \frac{\sum_{t=1}^{T} (r_{ti} - \mu)^2}{T},$$

(1)

where $r_{ti}$ is the returns series of asset $i$ in time $t$, $\mu$ the mean of returns, and $T$ the number of observations in the series of returns. The variance is used in the popular Markowitz model expressed in terms of covariance between assets.

2.2 Semivariance

Semivariance was considered by Markowitz since 1959 as a more adequate risk measure, but the mean-variance model has gained more space for its ease of implementation. The semivariance can be treated as a particular case of the Lower Partial Moment (LPM) family, a model that considers returns smaller than a target return or minimum acceptable return, which characterizes the model as based on downside risk.

The semivariance with respect to a benchmark $B$, chosen by the investor, can be expressed by

$$SV = E[\min(r_i - B, 0)]^2 = \frac{1}{T} \sum_{t=1}^{T} [\min(r_{ij} - B, 0)]^2,$$

(2)

where $r_{ij}$ is the return of assets $i$ in time $t$. The semivariance based in semi-covariance expressed in Eq. (2) is than defined as follows:

$$SV = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j \Sigma_{ij} B,$$

(3)

where $x_i$ and $x_j$ are the weights of assets $i$ and $j$ in the portfolio.
2.3 EWMA

The Exponentially Weighted Moving Average model (EWMA), proposed by Riskmetrics™, calculates the conditional volatility at the time $t$, that is, it assumes that the returns series is heteroscedastic. The EWMA measures the variance of the current instant as a function of the variance in the previous instant according to the parameter $\lambda$. It usually has the following formulation:

$$\sigma_t^2 = (1 - \lambda) r_{t-1}^2 + \lambda \sigma_{t-1}^2,$$

where $r_{t-1}$ and $\sigma_{t-1}^2$ are the return and variance values at time $t - 1$, respectively. The variable $\lambda$ belongs to the range $[0, 1]$ and in this work, it assumes the value 0.94 (Morgan & Reuters, 1996).

2.4 GARCH

The Generalized Autoregressive Conditional Heteroscedasticity (GARCH) represents a generalization of the ARCH model (Bollerslev, 1986). Like EWMA, the GARCH model assumes heteroscedasticity of the returns series, that is, the variability of the series is conditioned at time $t$. A GARCH model $(m, n)$ is expressed as follows:

$$r_t = \sqrt{\sigma_t^2} \epsilon_t; \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{m} \alpha_i r_{t-i}^2 + \sum_{j=1}^{n} \beta_j \sigma_{t-j}^2,$$

where $\epsilon_t$ are independent identically distributed random variables, $m \geq 0$, $n > 0$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$.

2.5 VaR

Value at Risk (VaR) is a measure based on quantiles, interpreted as the maximum potential loss of a portfolio over a defined time horizon, with a certain confidence degree (Kimura, Suen, Perera & Basso, 2009). The $\beta$-VaR of the portfolio is the lowest amount $\alpha$ such that, with probability $\beta$, the loss will not exceed $\alpha$ (Rockafellar & Uryasev, 2000). The VaR is a downside risk measure since it focuses on the left tail of the assets returns distribution. The formulation is given by:

$$\Psi(x, \alpha) = \int_{f(x,y) \leq \alpha} p(y) dy$$

where $x$ is a vector of asset weights in a portfolio, $y$ the series of losses, the loss $f(x,y)$ is a random variable with distribution in $R$. The function $f'(x,y)$ can be described as

$$f'(x,y) = -[x_1 y_1 + \cdots + x_n y_n] = -x^T y.$$ 

There are several ways to calculate VaR, such as considering the historical simulation and the parametric version. In the first case, the historical simulation generates a series of losses and the $\beta$-VaR is obtained as the $\beta$-series percentile (Baixauli-Soler, Alfaro-Cid & Fernandez-Blanco, 2012), denoted here $VaR(h)$. The parametric version, based in variance, in turn, assumes a normal distribution for returns. Its formulation is given by

$$VaR_\beta = Z_\beta \cdot \sigma_p,$$

where $Z_\beta$ is the critical value of the normal for the adopted confidence level and $\sigma_p$ is the portfolio standard deviation calculated through the variance. In this approach, the confidence level of 95% is applied in the parametric VaR, commonly adopted in literature, making the $Z_\beta$ value equal to 1.64. In this paper, this approach is called $VaR(v)$. 

Revista Brasileira de Finanças (Online) 17(3), 2019
2.6 CVaR

The Conditional Value at Risk (CVaR) model is a risk measure defined in the terms of VaR. The $\beta$-CVaR can be defined as the conditional expectation of losses on the amount $\alpha$, where $\alpha$ represents the VaR (Rockafellar & Uryasev, 2000). Therefore, CVaR can be understood as an expected value of losses when the losses exceed VaR. The formulation is given by

$$CVaR_\beta(X) = E[-X; -X \geq VaR_\beta(X)]$$

(8)

CVaR can then be represented as

$$\phi_\beta(x) = \frac{1}{1-\beta} \int_{f(x,y)\alpha(x)}^{} f(x,y)p(y)dy,$$

(9)

where $f(x,y)$ is a function of loss defined by Eq. (6) and $\alpha(x)$ is the VaR defined by Eq. (5). The formulation adopted in this work is a discrete version of CVaR (Rockafellar & Uryasev, 2000):

$$F_\beta(x,\alpha) = \alpha + \frac{1}{1-\beta} \sum_{t=1}^{j} \pi_t [f(w,y) - \alpha]^+,$$

(10)

where $\alpha$ is the VaR, $\beta$ is the confidence level, $x$ is a vector of asset weights in a portfolio, $f(x,y)$ the function defined in Equation (6) and $\pi$ is the probability of occurrence of the scenarios $y_1,y_2,...,y_j$ (considered here as having uniform distribution) and $j$ is the size of the historical price series.

3. Portfolio Optimization

This section explains the model, the operators of the algorithms to solve the problem, and the metrics chosen to compare the algorithms.

3.1 Portfolio Optimization Model

The general portfolio optimization model considered here is

$$\begin{align*}
\max_{w_1,\ldots,w_N} & \sum_{i=1}^{N} w_i \mu_i \\
\text{min}_{w_1,\ldots,w_N} & \text{risk measure}
\end{align*}$$

(11)

subject to:

$$\begin{align*}
\sum_{i=1}^{N} w_i &= 1 \\
w_i &= 0 \text{ or } \alpha \leq w_i \leq \beta \\
d^{\text{min}} &\leq \sum_{i=1}^{N} d_i(w) \leq d^{\text{max}}
\end{align*}$$

(12)

where $w_i$ is the weight of the $i$-th asset in the portfolio $w$, $\mu_i$ is the expected value of the return of asset $i$, considered as the mean of the past values, $[\alpha, \beta]$ represents the range for the weight $w_i$, $d^{\text{min}}$ and $d^{\text{max}}$ are the minimum and maximum portfolio cardinality, and $d_i(w)$ is a binomial variable which takes value 1 when the asset $i$ belongs to the portfolio $w$ and 0 otherwise. The first objective corresponds to the return maximization, while the second corresponds to the minimization of risk: variance, semivariance, EWMA, GARCH, VaR with two approaches, and CVaR.

3.2 Multiobjective Genetic Algorithms


In this problem, the process of optimizing portfolios using GAs starts with a certain set of randomly-generated portfolios, that obeys the constraints of the problem. The whole population is divided according to the number of different cardinalities, so that the initial population covers them equally. The weights of each
individual are allocated by randomly choosing the positions of the vector and assigning them a random value evenly distributed in the interval \([\alpha, \beta]\) until cardinality is complete.

After that, some portfolios are selected according to their risks and returns. The vectors representing such portfolios can be altered by mutation and crossover. The whole process is repeated several times in order to obtain, in the end, the best possible non-dominated set of portfolios according to their risks and returns. The operators considered in this work are described below.

3.2.1 Crossover operator

The crossover operator considered here is proposed in Deb et al. (2011), where two parents generate two children. The cardinality of the child is randomly selected in the range \([C_{p1}, C_{p2}]\), where \(C_{p1}\) and \(C_{p2}\) are the cardinality of parents \(p1\) and \(p2\). Let \(n_0\) be the number of positions of the vectors \(w_{p1}\) and \(w_{p2}\) that have the common zero weight, and \(n_c\) be the number of positions that have greater than zero common weights. The child vectors generated by these parents will have zero weight in all positions \(n_0\) and weights other than zero in all positions \(n_c\). In the other positions, where one parent has zero weight and the other parent’s value is greater than zero, the weight is determined according to the cardinality chosen for the child. As long as the cardinality of the child is not reached, the weight of the parent whose corresponding position is different from zero is assigned.

3.2.2 Mutation operators

Three mutation operators are used here, considering an equal probability of selection.

The first operator changes a weight greater than zero of a randomly chosen asset with uniform distribution in the range \([0, 1]\) (Deb et al., 2011). The second operator exchanges an asset that belongs to the portfolio by another by another asset chosen uniformly at random. Both assets are randomly chosen, choosing positions in the vector of solutions whose weight is greater than zero and another position with zero weight. The third operator changes the portfolio cardinality defined by \(C\). If the \(C\) value is equal to \(d_{\text{min}}\), a position whose weight is equal to zero is randomly selected and receives weight greater than zero, raising the \(C\) value by one unit. If the \(C\) value is equal to \(d_{\text{max}}\), a position whose weight is greater than zero is randomly selected and receives weight equal to zero. If \(C\) is in the open interval \((d_{\text{min}}, d_{\text{max}})\), \(C\) increases or decreases by a random selection.

3.2.3 Selection operators

This article considers the selection operators of two standard multiobjective algorithms: the Elitist Non-Dominated Sorting Genetic Algorithm (NSGA-II) (Deb et al., 2002), and the Strength Pareto Evolutionary Algorithm (SPEA2) (Zitzler et al., 2001).

NSGA-II is a method for multiobjective optimization where individuals are divided into dominance fronts. The selection is made using binary tournament. If individuals belong to different fronts, the individual who is dominated by fewer individuals wins the tournament. In case the selected solutions belong to the same front, the winner is the one that presents the greatest crowded distance, evaluated according to the density of solutions, in order to provide greater diversity at the efficient front. Here, the algorithm with this selection operator is called NSGA-II.

SPEA2 was proposed as the first multiobjective genetic algorithm to incorporate the structure of archive, with previously stipulated size, which stores the non-dominated solutions. The aptitude of the individual is given by the concept of the force of all the elements that dominate it, added to a measure of density, considering the nearest neighbors. Here, the algorithm with this selection operator is called SPEA2.

3.3 Algorithm Evaluation Metrics

To compare the behavior of the optimization algorithms, two standard measures are used in this paper: the \(S\)-metric and the Coverage metric (Deb, 2001).

The \(S\)-metric, also known as Hypervolume (HV) metric, evaluates the convergence and diversity of the non-dominated front through the volume in the space of objectives, generated by the solutions of the efficient
front. It has the following formulation:

$$HV = \sum_{i=1}^{Q} v_i$$  \hspace{1cm} (13)$$

where $Q$ is the set of front points, $W$ is a reference point, obtained considering each coordinate as the worst value of the objective functions, and $v_i$ is the hypervolume between the points $i$ and $W$.

The Coverage metric calculates, given sets $A$ and $B$, the proportion of solutions in $B$ that are weakly dominated by the solutions of $A$.

$$C(A,B) = \frac{|\{b \in B; \exists a \in A : a \succ b\}|}{|B|}$$  \hspace{1cm} (14)$$

This dominance relation, denoted by $a \succ b$, means that given two different variable vectors, $a$ and $b$ (which lead respectively to the cost vectors $F(a)$ and $F(b)$), that it has $F_i(a) \leq F_i(b)$, $\forall \; i \in \{1,2\}$, and $F_i(a) < F_i(b)$, for some $i \in \{1,2\}$. In this case, $C(A,B) = 0$ indicates that no element of $B$ is weakly dominated by elements of $A$ and $C(A,B) = 1$ indicates the dominance of all elements.

### 4. Methodology

The assets used in this work are those belonging to the portfolio of the B3 stock exchange index (Ibovespa) in December 2014. The portfolio was constituted by 69 assets in the period from March/2014 to December/2014. In addition, the returns corresponding to the period up to December/2015 were used to validate the out-of-sample analysis. Prices were adjusted for dividends, interest on shareholders’ equity, split and insplit. The missing data in the price series were filled by the quotation available on the immediately preceding date, generating a null return related to the date on which there was no quotation.

The in-sample analysis aims to compare the behavior of optimization algorithms NSGA-II and SPEA-2 for each risk measure: variance, semivariance, EWMA, GARCH, CVaR and VaR with two approaches: VaR(h) and VaR(v), based on historical simulation and on variance, respectively. The proposal is to verify if they generate statistically the same non-dominated front, for robustness, or if one algorithm performs, on average, better than the other.

Meanwhile, the agreement between the daily risk attributed to the different measures is verified, considering all the assets in all time periods, and the fronts are compared according to cardinality ranges.

The parameters employed in the multiobjective algorithms are chosen empirically by means of several executions. Table 1 shows these parameters. The range for the weight $w_i$ is considered as $[\alpha, \beta] = [0.01, 0.99]$. First, portfolios could choose between $d_{\text{min}} = 2$ and $d_{\text{max}} = 69$ assets (the maximum number of assets).

The obtained fronts are compared in terms of diversity and convergence by the $S$-metric and the Coverage metric, calculated using the combined efficient front, obtained at the end of each execution, considering both algorithms, for each risk measure. Descriptive statistics are generated from $S$-metrics and statistical comparison tests are performed. For the $2 \times 2$ comparison, the Mann-Whitney and $t$-Student hypothesis tests are used. The assumptions verifications are performed using the Anderson Darling tests for normality, Levene or Bartlett for

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Parameters used in multiobjective algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>parameter</td>
<td>value</td>
</tr>
<tr>
<td>population size</td>
<td>100</td>
</tr>
<tr>
<td>archive size</td>
<td>100</td>
</tr>
<tr>
<td>generations</td>
<td>500</td>
</tr>
<tr>
<td>executions</td>
<td>30</td>
</tr>
<tr>
<td>crossover</td>
<td>0.8</td>
</tr>
<tr>
<td>mutation</td>
<td>0.2</td>
</tr>
</tbody>
</table>
equality of variance and test of runs for the randomness of the samples achieved by the two algorithms for each risk measure.

Using the best algorithm according to the previous experiment, the relationship between the cardinality of the portfolios and each risk measured is evaluated, considering the combined Pareto fronts. The experiment is performed considering three cardinality intervals: 2-23, 24-46 and 47-69 assets. Finally, the Pareto fronts obtained by each risk measure and the optimal portfolios of the greater ratio of risk and return are compared, with greater Sharp ratio as in Sharpe (1994).

The objective of the out-of-sample analysis is to evaluate the performance of the nondominated portfolios created by the different risk measures. For each measure, the portfolios with greater Sharp ratio are chosen for cardinalities 2-23. The results of these portfolios are compared with the Ibovespa (Ibov) and with the IBX100, another important index of the Brazilian market, but with more assets. Those 9 optimal portfolios (Ibov, compared over 48 months, between January 2012 and December 2015. Portfolios from optimization are rebalanced on the first day of each month, with a new optimization process. The indexes Ibov and IBX100 are rebalanced every four months when the portfolios are updated.

The optimal portfolios are compared according to three indicators: return, drawdown risk, and drawup. Drawdown corresponds to the largest percentage drop in the value of the yield curve and drawup corresponds to the highest percentage increase in the period in the value of the yield curve, i.e., the opposite of drawdown. Each metric is calculated monthly for each portfolio, totaling 48 samples. Thus, the return corresponds to the variation between the opening price of the first day of the month and the closing price of the last day of the month, considering a transaction cost of 0.5% on the purchase and sale of assets. After obtaining the samples of those indicators, the Analysis of Variance is performed, chosen according to assumptions of normality, independence and homoscedasticity of the samples, and later the multiple comparison test is performed to find out which portfolio was statistically superior.

5. Results

The computational results of the in-sample and the out-of-sample analysis are presented in this section.

5.1 In-samples analysis

5.1.1 Correlation of the risk measures

First, to verify the agreement between the daily risk attributed by different measures, Table 2 shows the results of the Kendall correlation, considering all the assets in all time periods. All correlations are significantly different from zero at the 5% level of significance.

It can be seen that all the risk measures have high concordances. The highest concordance is obtained between VaR(v) and variance, as expected, and with GARCH. On the other hand, measures that are less concordant are VaR(h) and EWMA, as well as VaR(h) with the other ones.

<table>
<thead>
<tr>
<th>measure</th>
<th>CVaR</th>
<th>EWMA</th>
<th>GARCH</th>
<th>SV</th>
<th>Var(h)</th>
<th>Var(v)</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVaR</td>
<td>1</td>
<td>0.79</td>
<td>0.79</td>
<td>0.88</td>
<td>0.78</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>EWMA</td>
<td>0.79</td>
<td>1</td>
<td>0.86</td>
<td>0.84</td>
<td>0.74</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>GARCH</td>
<td>0.79</td>
<td>0.86</td>
<td>1</td>
<td>0.89</td>
<td>0.75</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>SV</td>
<td>0.88</td>
<td>0.84</td>
<td>0.89</td>
<td>1</td>
<td>0.78</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td>Var(h)</td>
<td>0.78</td>
<td>0.74</td>
<td>0.75</td>
<td>0.78</td>
<td>1</td>
<td>0.76</td>
<td>0.76</td>
</tr>
<tr>
<td>VaR(v)</td>
<td>0.80</td>
<td>0.88</td>
<td>0.94</td>
<td>0.89</td>
<td>0.76</td>
<td>1</td>
<td>0.99</td>
</tr>
<tr>
<td>Variance</td>
<td>0.80</td>
<td>0.88</td>
<td>0.94</td>
<td>0.89</td>
<td>0.76</td>
<td>0.99</td>
<td>1</td>
</tr>
</tbody>
</table>

Revista Brasileira de Finanças (Online) 17(3), 2019
Table 3
Mean and coefficient of variation CV of S-metrics for NSGA-II and SPEA2 considering each risk measure

<table>
<thead>
<tr>
<th>measure</th>
<th>NSGA-II mean</th>
<th>NSGA-II CV</th>
<th>SPEA2 mean</th>
<th>SPEA2 CV</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVaR</td>
<td>1.62E-09</td>
<td>2.50E-01</td>
<td>3.05E-05</td>
<td>1.15E-01</td>
</tr>
<tr>
<td>EWMA</td>
<td>1.99E-07</td>
<td>1.94E-01</td>
<td>4.94E-07</td>
<td>6.65E-02</td>
</tr>
<tr>
<td>GARCH</td>
<td>1.66E-07</td>
<td>2.88E-01</td>
<td>3.73E-07</td>
<td>7.40E-02</td>
</tr>
<tr>
<td>SV</td>
<td>9.85E-08</td>
<td>1.97E-01</td>
<td>1.93E-07</td>
<td>6.42E-02</td>
</tr>
<tr>
<td>VaR(v)</td>
<td>1.12E-09</td>
<td>1.77E-01</td>
<td>1.71E-05</td>
<td>8.41E-02</td>
</tr>
<tr>
<td>VaR(h)</td>
<td>1.16E-09</td>
<td>3.19E-01</td>
<td>2.02E-05</td>
<td>1.19E-01</td>
</tr>
<tr>
<td>Variance</td>
<td>1.99E-07</td>
<td>1.40E-01</td>
<td>3.28E-07</td>
<td>7.96E-02</td>
</tr>
</tbody>
</table>

Table 4
Normality test of the S-metrics for each risk measure

<table>
<thead>
<tr>
<th>measure</th>
<th>NSGA-II p-value</th>
<th>NSGA-II normality</th>
<th>SPEA2 p-value</th>
<th>SPEA2 normality</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVaR</td>
<td>0.473</td>
<td>Yes</td>
<td>0.319</td>
<td>Yes</td>
</tr>
<tr>
<td>EWMA</td>
<td>0.015</td>
<td>No</td>
<td>0.622</td>
<td>Yes</td>
</tr>
<tr>
<td>GARCH</td>
<td>0.001</td>
<td>No</td>
<td>0.553</td>
<td>Yes</td>
</tr>
<tr>
<td>SV</td>
<td>0.177</td>
<td>Yes</td>
<td>0.637</td>
<td>Yes</td>
</tr>
<tr>
<td>VaR(v)</td>
<td>0.666</td>
<td>Yes</td>
<td>0.804</td>
<td>Yes</td>
</tr>
<tr>
<td>VaR(h)</td>
<td>0.647</td>
<td>Yes</td>
<td>0.637</td>
<td>Yes</td>
</tr>
<tr>
<td>Variance</td>
<td>0.746</td>
<td>Yes</td>
<td>0.498</td>
<td>Yes</td>
</tr>
</tbody>
</table>

5.1.2 Comparison of algorithms

Table 3 shows the descriptive statistics of S-metrics for NSGA-II and SPEA2 over the executions: mean and coefficient of variation CV, for each risk measure: CVaR, VaR(h), VaR(v), EWMA, SV, GARCH, and variance.

The fronts obtained by SPEA2 present, in all cases, both a higher mean value of S-metric and a lower coefficient of variation than the NSGA-II results. This indicates that, in addition to generating a higher mean, the runs of the SPEA2 are more homogeneous, generating greater reliability for the results.

Table 4 shows the results of the normality of the S-metrics using the Anderson Darling test. The normality hypothesis of S-metrics is verified only for the series generated by CVaR, SV, VaR(v), VaR(h) and variance.

Table 5 shows the results of the equality evaluation between the variances of the series. The results show the equality of variances just for the S-metric series generated by the EWMA, GARCH and variance measures. Table 6 shows the statistic comparison of S-Metrics. Because of the assumptions, the series of the variance measure of S-metrics are evaluated by t-Student while CVaR, Semivariance and VaRs were evaluated using the Welch test. The other comparisons were made using the Mann-Whitney test.

The results show statistically significant differences detected between the fronts generated by the two algorithms using all the risk measures. The negative and zero-values within the confidence intervals indicate that the distribution of S-metrics generated by the two algorithms were different in all cases, with SPEA2 generating higher values in this experiment.

Table 7 shows the results of Coverage Metrics. In this case, considering all the risk measures evaluated, the fronts generated by SPEA2 dominate the greater proportion of the NSGA-II points. These results motivate the choice of SPEA2 as an algorithm to be used from now on in the experiments.
5.1.3 Analysis of cardinalities

Here, the relationship between the cardinality of the portfolio and each risk measured is evaluated, considering each combined Pareto front from SPEA2. The experiment is performed considering three cardinality intervals. In interval 1, the cardinality of the portfolios is restricted to the range 2-23, interval 2 has cardinality restricted to the range 24-46, and interval 3 to the range 47-69.

Figures 1 to 7 show the fronts of these three intervals of cardinality for each risk measure. The results of this experiment indicate that portfolios of lower cardinality are preferable to portfolios with large numbers of assets in all risk measures evaluated, since in all the risk measures, the fronts of interval 1 dominate the other fronts, and that the fronts of interval 2 dominate those of interval 3.

The fronts of the first two intervals are shown to be closer in the region of minimization of risk, in relation to the region of return maximization. This is because, generally, fronts of interval 1 maximize the expected return by concentrating 100% of the weights in the two assets with the highest past average return, reaching its minimum cardinality of 2 assets. So it is able to allocate up to 99% of the weights in the asset with the...
Figure 1
Combined Pareto front for variance

Figure 2
Combined Pareto front for SV
Figure 3
Combined Pareto front for VaR(h)

Figure 4
Combined Pareto front for VaR(v)
Figure 5
Combined Pareto front for CVaR

Figure 6
Combined Pareto front for GARCH
highest average expected return, generating a portfolio of high expected return and risk. Fronts of interval 2, because they have to choose at least 24 assets, can just allocate up to 77% to the asset with the highest past average return, generating portfolios that maximize the return lower than that of the greatest return in the front of interval 1. Meanwhile, interval 3 generates considerably lower return portfolios in relation to the other intervals because it needs to reach the minimum cardinality of 47 shares. As 32 of the 69 assets used in the analysis show a negative average return in this time period, the fronts of interval 3 necessarily choose assets with such unwanted expected returns.

For interval 1, at the front optimized by the variance, the maximum cardinality obtained was 14 assets, with a homogeneous distribution of portfolios in the range 6-13 and a portfolio concentration in cardinalities 3 and 4. The front optimized by VaR(v) shows cardinality maximum of 12 assets, a homogeneous distribution of portfolios in the range of cardinality 6-11 and a concentration of portfolios in the cardinality 3. The front corresponding to GARCH presents a maximum cardinality of 14, a homogeneous distribution of assets between the interval 5-12, and a concentration in portfolios with 3 and 4 assets. The front obtained through EWMA reaches a maximum cardinality of 15 assets, with a large part of the portfolio distributed homogeneously in the set 2-5. Semivariance generates portfolios of a maximum of 11 assets, with most of the portfolios distributed among cardinalities 4 and 6. The fronts obtained through the measures CVaR and VaR(h) obtain a cardinality maximum of 10 assets. Most of the portfolios obtained by CVaR are concentrated in cardinalities 4 and 5, while in the VaR(h) the portfolios are distributed homogeneously between cardinalities 2 and 5. The downside risk measures tend to reach lower cardinalities than the other measures. Despite this, CVaR, semivariance and VaR(h) present a higher proportion of portfolios with 5 or 6 assets, while most portfolios of the other measures are concentrated in cardinalities 3 and 4. Therefore, the fronts of all measures concentrated portfolios in the cardinalities between 3 and 6, representing 13 to 26% of the maximum cardinality allowed.

Fronts of intervals 2 and 3, for all risk measures, just generated fronts with the minimum cardinality (24 or 47 assets per portfolio). For all risk measures, the fronts of interval 2 choose the same assets as those of
interval 1 and distribute the weights equal to or close to the minimum weight of 1% in some assets, until the minimum cardinality is reached. The same occurs in the choice of assets in interval 3 in relation to interval 2. In other words, the fronts of interval 3 contain all the assets chosen in intervals 1 and 2, and some others for which weights are placed close to the minimum to complete the minimum cardinality of 24 or 47 assets.

### 5.1.4 Comparison of Pareto fronts and optimal portfolios

Finally, this subsubsection compares the combined Pareto fronts obtained by each risk measure and the optimal portfolios of the greater ratio of risk and return (the Sharp ratio).

Figure 8 shows the combined Pareto front for all risk measures optimization considering SPEA2, for interval 1. In this figure, the variance of each optimal portfolio is considered as the risk measure, for comparison purposes. Figure 8a shows the front and Figure 8b shows a zoom of the front at the tail in return, corresponding to values $1.24$ to $1.42$. It is remarkable that converting to another measure will get at basically the same behavior. In fact, the fronts have very similar behavior, with small variations, that can, and will, impact the quality of the final out-of-sample response.

Table 8 shows the portfolios with the greater Sharp ratio, for each risk measure, showing the asset codes and the respective percentage of the portfolio, in descending order.

The optimal portfolios are very similar to each other, for all risk measures. The lowest difference is observed
Analysis of risk measures in multiobjective optimization portfolios with cardinality constraint

Table 9

<table>
<thead>
<tr>
<th>Risk Measure</th>
<th>mean</th>
<th>sd</th>
<th>skew</th>
<th>kurt</th>
<th>med</th>
<th>max</th>
<th>min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ibov</td>
<td>-0.45</td>
<td>5.50</td>
<td>-0.08</td>
<td>-0.34</td>
<td>-0.77</td>
<td>9.97</td>
<td>-11.86</td>
</tr>
<tr>
<td>IBX100</td>
<td>-0.07</td>
<td>4.92</td>
<td>-0.18</td>
<td>-0.23</td>
<td>0.27</td>
<td>9.59</td>
<td>-11.25</td>
</tr>
<tr>
<td>CVaR</td>
<td>1.16</td>
<td>3.90</td>
<td>-0.83</td>
<td>0.44</td>
<td>1.82</td>
<td>7.18</td>
<td>-8.40</td>
</tr>
<tr>
<td>VaR(h)</td>
<td>0.99</td>
<td>3.81</td>
<td>-0.15</td>
<td>0.20</td>
<td>0.77</td>
<td>8.79</td>
<td>-8.32</td>
</tr>
<tr>
<td>VaR(v)</td>
<td>0.99</td>
<td>3.41</td>
<td>-0.82</td>
<td>0.54</td>
<td>1.43</td>
<td>6.61</td>
<td>-8.19</td>
</tr>
<tr>
<td>EWMA</td>
<td>0.87</td>
<td>3.38</td>
<td>-0.83</td>
<td>0.50</td>
<td>1.38</td>
<td>6.50</td>
<td>-8.71</td>
</tr>
<tr>
<td>SV</td>
<td>0.48</td>
<td>3.64</td>
<td>-0.58</td>
<td>-0.18</td>
<td>1.22</td>
<td>6.31</td>
<td>-8.21</td>
</tr>
<tr>
<td>GARCH</td>
<td>0.78</td>
<td>0.37</td>
<td>-0.75</td>
<td>0.40</td>
<td>1.20</td>
<td>6.26</td>
<td>-8.31</td>
</tr>
<tr>
<td>Variance</td>
<td>0.89</td>
<td>3.38</td>
<td>-0.87</td>
<td>0.62</td>
<td>1.14</td>
<td>6.50</td>
<td>-8.26</td>
</tr>
</tbody>
</table>

Figure 9

Boxplot and Bonferroni’s test for out-of-sample monthly returns

(a) Boxplot for the monthly returns

(b) Bonferroni’s test for the monthly returns

5.2 Out-of-sample analysis

5.2.1 Monthly and Cumulative Returns

Table 9 shows the descriptive statistics of the percentage average monthly return of 9 considered portfolios: CVaR, VaR(h), VaR(v), EWMA, SV, GARCH, and variance, optimized by SPEA-2, chosen as the ones with greater Sharp ratio come from the combined Pareto front, and for Ibov and IBX100. The mean returns of the optimized portfolios are all positive, even when considering transaction costs. However, in the same period, Ibov and IBX100 have a negative average return. In both the mean and the median, the best out-of-sample return result is obtained considering the CVaR risk model.

To check if the superiority of the optimized portfolios is statistically verified, an Analysis of Variance test was performed. Bonferroni’s test is shown in Figure 9, with the boxplot of returns. The multiple comparison test shows an intersection between the confidence interval of all methods. Thus, even if the optimized portfolios have superior return performance in the 48 months tested, it cannot be concluded that they will generate higher returns compared to the Ibov and IBX100 portfolios.

It is interesting to look closer to identify which risk measures present the best results in the period studied.
The monthly returns results are compared to each other, as shown in Figure 10. The 48 months of simulation provide a total of 48 monthly returns for each method. The figure shows the percentage of times that each method presents a higher monthly return in relation to the other methods. Each row and column presents a method, where the row represents the reference method and the column, the methods to be compared. The figure shows a heat-map scale, in which colder colors (green tones) show better returns and warmer colors (red tones) show worse returns. The figure shows that in most months the optimized portfolios have a higher return than the benchmarks. The risk measures CVaR, VaR(h) and VaR(v) present better returns than the other metrics. The worst result among the optimized portfolios is SV, which surpasses only the benchmarks.

Finally, it is important to analyze the cumulative return of those portfolios over the years, in relation to a risk-free asset (CDI), as shown in Figure 11. As can be seen, all the optimized portfolios exceed the results of benchmarks. In addition, all risk measures exceed the CDI, with the exception of the semivariance. The best return is obtained for CVaR, but it is important to notice that the best performance alternates between some risk measures over the period, mainly those based on quantiles: CVaR, VaR(h) and VaR(v).

5.2.2 Drawdown

Maximum drawdown corresponds to the largest percentage drop in the value of the yield curve. Table 10 shows the descriptive statistics of the out-of-sample drawdown.
Table 10
Descriptive statistics of out-of-sample monthly drawdown for each risk measure

<table>
<thead>
<tr>
<th>Method</th>
<th>mean</th>
<th>sd</th>
<th>skew</th>
<th>kurt</th>
<th>med</th>
<th>max</th>
<th>min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ibov.</td>
<td>6.13</td>
<td>3.21</td>
<td>0.92</td>
<td>0.56</td>
<td>5.85</td>
<td>14.91</td>
<td>1.38</td>
</tr>
<tr>
<td>IBX100</td>
<td>5.33</td>
<td>3.01</td>
<td>1.03</td>
<td>0.32</td>
<td>4.53</td>
<td>12.73</td>
<td>1.10</td>
</tr>
<tr>
<td>CVaR</td>
<td>3.46</td>
<td>2.56</td>
<td>1.64</td>
<td>2.04</td>
<td>2.57</td>
<td>11.13</td>
<td>0.62</td>
</tr>
<tr>
<td>VaR(h)</td>
<td>3.63</td>
<td>2.49</td>
<td>1.26</td>
<td>0.86</td>
<td>2.64</td>
<td>11.00</td>
<td>0.53</td>
</tr>
<tr>
<td>VaR(v)</td>
<td>3.28</td>
<td>2.46</td>
<td>1.72</td>
<td>2.52</td>
<td>2.50</td>
<td>11.44</td>
<td>0.78</td>
</tr>
<tr>
<td>EWMA</td>
<td>3.35</td>
<td>2.53</td>
<td>1.87</td>
<td>3.07</td>
<td>2.50</td>
<td>11.74</td>
<td>1.05</td>
</tr>
<tr>
<td>SV</td>
<td>3.73</td>
<td>2.71</td>
<td>1.74</td>
<td>2.47</td>
<td>2.90</td>
<td>12.46</td>
<td>0.85</td>
</tr>
<tr>
<td>GARCH</td>
<td>3.33</td>
<td>2.48</td>
<td>1.67</td>
<td>2.40</td>
<td>2.47</td>
<td>11.50</td>
<td>0.85</td>
</tr>
<tr>
<td>Variance</td>
<td>3.30</td>
<td>2.45</td>
<td>1.76</td>
<td>2.59</td>
<td>2.41</td>
<td>11.53</td>
<td>1.10</td>
</tr>
</tbody>
</table>

Figure 12
Boxplot and Bonferroni’s test for the out-of-sample monthly drawdown

(a) Boxplot for drawdown
(b) Bonferroni’s test for drawdown

The average monthly values of drawdown of all optimized portfolios are lower than Ibov and IBX100. Therefore, optimized portfolios also reduce the median drawdown of benchmarks by more than 50%. To check the consistency of these results, Bonferroni’s test was performed, as shown in Figure 12, with the boxplot of drawdown. The Kruskal-Wallis test shows statistical difference between the portfolios, with a $p$-value less than 0.001.

This result indicates that the optimized portfolios can statistically reduce the risk of drawdown in a portfolio of assets, in relation to these benchmarks.

5.2.3 Drawup

Optimized portfolios are seen to be able to reduce the risk of drawdown, but it is not desirable that these portfolios also reduce the market upside moments, the financial drawup. Table 11 shows the descriptive statistics of the out-of-sample monthly drawup.

The average monthly values of drawup of all optimized portfolios are lower than the Ibov and IBX100. This result indicates that the optimized portfolios are responsible for reducing some moments of the market upturn. However, the reduction of drawdown occurs in a much higher proportion than the reduction of drawup. Analyzing the median, it can be seen that the optimized portfolios reduce benchmark drawup by 20%, versus a 50% reduction in drawdown. To check the consistency of these results, Bonferroni’s test is performed, as shown in Figure 13. The Kruskal-Wallis test shows a statistical difference between the scenarios, with a $p$-value less than 0.001. As can be seen, the multiple comparison test shows that only the optimized portfolios VaR (v), EWMA, Semivariance, GARCH, and variance have lower drawup than the benchmarks. The downside risk measures CVaR and VaR (h) do not statistically reduce financial drawup.
de Oliveira et al., 2019

Table 11
Descriptive statistics of out-of-sample monthly drawup for each risk measure

<table>
<thead>
<tr>
<th>Risk Measure</th>
<th>mean</th>
<th>sd</th>
<th>skew</th>
<th>kurt</th>
<th>med</th>
<th>max</th>
<th>min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ibov.</td>
<td>6.10</td>
<td>2.84</td>
<td>0.14</td>
<td>-1.18</td>
<td>5.45</td>
<td>12.12</td>
<td>1.55</td>
</tr>
<tr>
<td>IBX100</td>
<td>5.43</td>
<td>2.50</td>
<td>0.29</td>
<td>-0.80</td>
<td>5.00</td>
<td>11.54</td>
<td>1.41</td>
</tr>
<tr>
<td>CVaR</td>
<td>4.53</td>
<td>2.09</td>
<td>0.49</td>
<td>-0.23</td>
<td>4.35</td>
<td>10.38</td>
<td>0.92</td>
</tr>
<tr>
<td>VaR(h)</td>
<td>4.77</td>
<td>2.17</td>
<td>0.90</td>
<td>0.23</td>
<td>4.34</td>
<td>10.40</td>
<td>1.61</td>
</tr>
<tr>
<td>VaR(v)</td>
<td>4.14</td>
<td>1.76</td>
<td>0.20</td>
<td>-0.73</td>
<td>4.00</td>
<td>7.78</td>
<td>1.06</td>
</tr>
<tr>
<td>EWMA</td>
<td>4.08</td>
<td>1.63</td>
<td>0.35</td>
<td>-0.13</td>
<td>4.04</td>
<td>8.16</td>
<td>1.28</td>
</tr>
<tr>
<td>SV</td>
<td>4.20</td>
<td>1.84</td>
<td>0.51</td>
<td>-0.29</td>
<td>4.22</td>
<td>8.50</td>
<td>1.33</td>
</tr>
<tr>
<td>GARCH</td>
<td>3.97</td>
<td>1.59</td>
<td>0.41</td>
<td>-0.32</td>
<td>4.00</td>
<td>7.69</td>
<td>1.20</td>
</tr>
<tr>
<td>Variance</td>
<td>4.05</td>
<td>1.68</td>
<td>0.27</td>
<td>-0.55</td>
<td>4.03</td>
<td>7.69</td>
<td>1.05</td>
</tr>
</tbody>
</table>

Figure 13
Boxplot and Bonferroni’s test for the out-of-sample monthly drawup

(a) Boxplot for drawup

(b) Bonferroni’s test for drawup

This result indicates that risk measures based on quantiles that do not require data normality (CVaR and VaR(h)) were able to reduce just negative returns scenarios.

6. Conclusions

This article proposes a comparison of multiobjective optimization of portfolios models with cardinality constraint for different risk measures: CVaR, VaR(h), VaR(v), EWMA, SV, GARCH, variance. For this, NSGA-II and SPEA2 optimization algorithms are considered.

First, the in-sample analysis leads to the conclusion that the SPEA2-based algorithm performs better, considering the S-metric and the Coverage metric. This is because the non-dominated fronts generated by SPEA2 have greater convergence and diversity in this experiment. The results indicate that portfolios of lower cardinality are preferable to portfolios with large numbers of assets in all risk measures evaluated. They also indicate that the downside risk measures tend to reach a cardinality lower than the others. In any case, the optimal portfolios obtained for each measure are similar to each other, with small variations that impact the quality of the final response.

The out-of-sample analysis compares the behavior of the portfolios of each risk measure, looking for those with the greatest Sharp ratio. Results indicate that the risk measures based on quantiles CVaR and VaR have better-accumulated returns and present better monthly returns in most periods analyzed. The results and statistical tests show that the optimization considering all risk measures can reduce drawdown compared to benchmarks. However, only the risk measures based on quantiles that do not require data normality (CVaR and VaR with historical simulation) are able to reduce only the negative returns scenarios, i.e., reducing drawdown risk.
without limiting financial drawup.

Acknowledgements

The authors would like to thank to CAPES, CNPq and CEFET-MG.

References


**URL:** http://doi.acm.org/10.1145/1219092.1219093
