sciendo

Portfolio Optimization Efficiency Test Considering Data Snooping Bias

Aleš Kresta
VSB – Technical University of Ostrava, Faculty of Economics, Czech Republic
Anlan Wang
VSB – Technical University of Ostrava, Faculty of Economics, Czech Republic

Abstract

Background: In the portfolio optimization area, most of the research is focused on in-sample portfolio optimization. One may ask a rational question of what the efficiency of the portfolio optimization strategy is and how to measure it. **Objectives:** The objective of the paper is to propose the approach to measuring the efficiency of the portfolio strategy based on the hypothesis inference methodology and considering a possible data snooping bias. The proposed approach is demonstrated on the Markowitz minimum variance model and the fuzzy probabilities minimum variance model. **Methods/Approach:** The proposed approach is based on a statistical test. The null hypothesis is that the analysed portfolio optimization strategy creates a portfolio randomly, while the alternative hypothesis is that an optimized portfolio is created in such a way that the risk of the portfolio is lowered. **Results:** It is found out that the analysed strategies indeed lower the risk of the portfolio during the market's decline in the global financial crisis and in 94% of the time in the 2009-2019 period. **Conclusions:** The analysed strategies lower the risk of the portfolio in the out-of-sample period.

Keywords: data snooping bias, financial crisis, hypothesis test, minimum-risk portfolio, portfolio optimization

JEL main category: G11
JEL classification: G11, G17
Paper type: Research article

Received: Jan 31, 2020
Accepted: Jul 06, 2020

Acknowledgments: The authors were supported through the Czech Science Foundation (GACR) under the project no. 18-13951S and the SGS research project of VSB-TU Ostrava under the project no. SP2020/11. The support is greatly appreciated.

Citation: Kresta, A., Wang, A. (2020), "Portfolio Optimization Efficiency Test Considering Data Snooping Bias", Business Systems Research, Vol. 11 No. 2, pp. 73-85.
DOI: 10.2478/bsrj-2020-0016

Introduction

Since the pioneering work of Markowitz (1952) that put out the foundations of modern portfolio theory, it is still a lively and open area with a lot of attention from both academics and practitioners. While academics generally focus more on the question of different models of portfolio selection, paying attention mostly to the theories and assumptions, practitioners focus also on the verification of particular strategies. The question they ask is what the efficiency of the portfolio optimization strategy is.

The standard procedure is that the parameters of the model are estimated from the time series of returns in a historical period, which is usually called the in-sample period. The portfolio performance is then observed and measured in the period following the in-sample period. This observed period is usually called the out-of-sample period and the portfolio performance is generally lower in the out-of-sample period than in the in-sample period. In order to confirm the suitability of the proposed strategy, the out-of-sample performance is usually compared to the chosen benchmark. If the proposed strategy delivers better out-of-sample results than the benchmark, it is considered suitable.

Although the application of the benchmark is commonly accepted, its informative value is rather poor – it answers the question of whether and how much the tested strategy overperformed the benchmark in one particular period. However, it does not say whether the overperformance is high enough to be considered as significant or whether it is just due to the randomness in data and luck. Also, another drawback of this simple approach is that it does not address the data snooping bias – if more strategies are analyzed, then it is more probable that the best strategy overperforming the benchmark is found.

In this paper, we address this gap by proposing a rather different approach to verify the efficiency of the portfolio strategies. The approach is based on the hypothesis inference methodology. Put it simply, the principle is to generate many random portfolios and compare the out-of-sample performance of the analyzed strategy with performances of these random portfolios. Data snooping bias is also considered in the proposed approach. Moreover, in the empirical part of the paper, we investigate whether the classical minimum variance model (Markowitz, 1952) and fuzzy probabilities minimum variance model (Tanaka et al., 2000) decrease the risk in the out-of-sample period.

The paper is divided as follows. The following section provides a brief literature review of applied benchmarks. Then, in the following section, the analyzed portfolio strategies are introduced and the method of testing the efficiency of portfolio strategies is presented. Empirical results are provided in the following section. First, the focus is given to the global financial crisis period. Then, the results on the rolling window basis in the period 2006-2019 are presented. The last sections provide a brief discussion of the results and conclusion to the paper.

Literature review

Various benchmarks can be found in scientific literature; however, there are few groups of benchmarks, which are generally applied. The first benchmark is 1/N strategy, which is also called a naive diversification strategy. The 1/N strategy is easy to implement because it is not necessary to estimate the future returns of the assets, the assets of the naive portfolio are invested with equal weights. DeMiguel et al. (2009) analyzed the differences in the performance of several optimization methods with that of the 1/N strategy, and they found that the effect of estimation error on return probability distribution is large in those optimization models and this type of error can be avoided by using the 1/N weights.

The second benchmark is the classical Mean-Variance model or Mean-VaR model. Since more real-life conditions have been considered, for example, the investors' subjective preference, the transaction costs, the liquidity of the portfolio and so on, the classical Mean-Variance model is improved by incorporating various additional constraints, so the classical model becomes a widely used benchmark to compare with those enhanced approaches, see e.g. Fulga (2016), Ranković et al. (2016), Lwin et al. (2017), or Babazadeh and Esfahanipour (2019).

It is also common to apply market indices as a benchmark to compare the performance of proposed methods. Solares et al. (2019) pointed out the main problem of applying market indices as the benchmark: the performance of portfolios is often compared to that of popular indices, but it is hard to reach the performance of the indices because there is a difference between characteristics of the stocks in the portfolios and the stocks contained in the index. So, to avoid this trap, it is recommended to construct the portfolio with only the stocks, which are selected from the components of the benchmark index.

Besides considering one particular portfolio optimization approach as a benchmark, it is also necessary to apply a benchmark dataset when developing a new approach in the portfolio optimization problem. Most of the studies are based on either case studies or publicly available benchmarking datasets, see Kalayci et al. (2019). In the empirical analysis of this paper, the dataset of Dow Jones Industrial Average (DJIA) index is applied; it is a small dataset of 30 stocks while it is persuasive enough to explain the investment environment in the optimization problem. What's more, the advantage of the small dataset is also its relatively smaller computational complexity. In the future, the small dataset of DJIA could be used as a benchmark dataset to compare with large-scale datasets, which can verify the applicability of the proposed models in complex environments and global markets.

Methodology

Applied Portfolio Models

Portfolio models applied in the paper follow the classical mean-variance framework, i.e. only expected return (mean), its variance, and their inter-relationship are considered. Let us denote x_i as the weight of ith asset in the portfolio. Short sales are excluded from the models, so the values of x_i satisfies $0 \leq x_i \leq 1$ for all assets. If the expected return of ith asset is denoted as $E(R_i)$, then the expected return of a portfolio $E(R_p)$ is the weighted average of $E(R_i)$:

$$E\left(R_p\right) = \sum_{i=1}^{N} x_i \cdot E(R_i) = x^T \cdot E(R), \tag{1}$$

where N is the total number of assets in the portfolio, $x = [x_1, x_2, \dots, x_N]^T$ is portfolio composition and $E(R) = [\,E(R_1), E(R_2), \dots, E(R_N)]^T$ is the vector of expected returns. The variance and standard deviation of the portfolio return are calculated by means of covariances $\sigma_{i,j}$ of the asset returns for all asset pairs (i, j),

$${\sigma_p}^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} x_i \cdot \sigma_{i,j} \cdot x_j = x^T \cdot \mathbf{Q} \cdot x, \tag{2}$$

$$\sigma_p = \sqrt{{\sigma_p}^2}, \tag{3}$$

where $\mathbf{Q}$ denotes covariance matrix, $\mathbf{Q} = [\sigma_{i,j}, i = 1, 2, \dots, N, j = 1, 2, \dots, N]$. The minimum-variance portfolio can be found by solving the following quadratic optimization problem,

$$\text{minimize } {\sigma_p}^2 \tag{4}$$

subject to

$$\sum_{i=1}^{N} x_i = 1 \tag{5}$$

$$x_i \geq 0, i = 1, \dots, N \tag{6}$$

The question is how to estimate the parameters of the return probability distributions. In the simplest approach, let us call it historical estimation, the characteristics of the observed sample distribution are calculated. Specifically, the expected returns are estimated as means of m observed historical returns,

$$E(r_i) = \hat{r}_i = \frac{1}{m}\sum_{i=1}^{m} r_{i,m}. \tag{7}$$

The covariance matrix is estimated in the same way,

$$\boldsymbol{Q} = \widehat{\boldsymbol{Q}} = [\hat{\sigma}_{i,j}] \tag{8}$$

$$\hat{\sigma}_{i,j} = \frac{1}{m}\sum_{k=1}^{m}(r_{i,k} - \hat{r}_i)(r_{j,k} - \hat{r}_j) \tag{9}$$

The problem of historical estimation is the excessive sensitivity of the portfolio composition to errors in parameter estimates (due to the length of time series, the number of assets, etc.). The illustrative example can be found in DeMiguel et al. (2009).

In the Bayesian approach, the estimation of the parameters of return probability distributions, i.e. the vector of expected returns $E(R_i)$ and the covariance matrix **Q**, considers the subjective (a priori) assumption of the shape of this distribution. The resulting (a posteriori) probability distribution is then a combination of the a priori assumption and the probability distribution of the observed sample. Although there are different possibilities for applying this approach, the methodology commonly referred to as the Bayes-Stein portfolio (BS) is applied in the paper. The foundations of this approach have already been laid by Stein (1956) and James and Stein (1961). The estimation suggested in Jordon (1986) is applied:

$$E(r_i^{BS}) = (1-\xi)\cdot\hat{r}_i + \xi\cdot\bar{r}, \tag{10}$$

$$\xi = \frac{N+2}{(N+2) + m\cdot(\hat{r}_i - \bar{r})^T\cdot\widehat{\boldsymbol{Q}}^{-1}\cdot(\hat{r}_i - \bar{r})^T}, \tag{11}$$

$$\boldsymbol{Q}^{BS} = \widehat{\boldsymbol{Q}}\left(1 + \frac{1}{m+\zeta}\right) + \frac{\zeta}{m\cdot(m+1+\zeta)}\frac{\mathbf{1}_N\cdot\mathbf{1}_N{}^T}{\mathbf{1}_N{}^T\times\widehat{\boldsymbol{Q}}^{-1}\times\mathbf{1}_N}, \tag{12}$$

$$\zeta = \frac{m\cdot\xi}{1-\xi}, \tag{13}$$

where $\hat{r}_i$ is the historical estimate of the expected return of the ith asset, see (7), $\bar{r}$ is the a priori expected return on the assets, N is the number of assets and m is the number of historical observations of returns (sample size), $\widehat{\boldsymbol{Q}}$ is the historically estimated covariance matrix, see (8). Jorion (1986) proposes to determine $\bar{r}$ as the return of the minimum-variance portfolio. However, we rather consider the average expected return, which saves one optimization in the procedure. As can be seen, the shrinkage factor ξ depends on the number of assets N (with an increasing number of assets the estimation error increases), the number of historical observations m (the shorter the history, the higher the estimation error) and on the dispersion of estimated expected returns from a priori assumption (the greater the dispersion, the greater the estimation error).

Another approach of how to handle uncertainty in probability distribution parameters is to apply the fuzzy theory. Tanaka et al. (2000) proposed two types of portfolio optimization models. The first model is based on fuzzy probabilities and aims to minimize the variance of the portfolio return while the latter utilizes possibility distributions and minimize the spread of the portfolio return. The first model is applied in the paper. The model considers not only historical returns $\{r_i, i = 1, \dots, N\}$, but also possibility grades $\{h_i, i = 1, \dots, N\}$, which represent a similarity between the future state of the stock market and the state of ith sample offered by experts.

Given the historical returns and possibility grades, the fuzzy weighted expected return can be calculated as follows,

$$E\left(r_i^F\right)=\frac{\sum_{j=1}^{m} h_j \cdot r_{i,j}}{\sum_{j=1}^{m} h_j}, \tag{14}$$

and fuzzy weighted covariance matrix $\mathbf{Q}^{F}=[\sigma_{i,j}^{F}, i=1,\dots,N, j=1,\dots,N]$ can be defined as:

$$\sigma_{i,j}^{F}=\frac{\sum_{k=1}^{m}\left(r_{i,k}-E\left(r_i^F\right)\right)\cdot\left(r_{j,k}-E\left(r_j^F\right)\right)\cdot h_k}{\sum_{k=1}^{m} h_k}. \tag{15}$$

These estimates can be directly applied in the portfolio optimization model (4)-(6).

Portfolio efficiency test

The rational question one can ask is what the efficiency of the portfolio optimization strategy is. Let's consider the portfolio strategy with observed out-of-sample performance p. The observed out-of-sample performance is, in fact, a realization of random variable P. Let us consider the following null hypothesis: the performance of the portfolio strategy is the same as of random strategy. The alternative hypothesis is that the performance of the portfolio strategy is better than of random strategy. The distribution of random variable P under the null hypothesis can be obtained by generating random portfolio compositions and calculating their performances in the out-of-sample period. Then, the classical statistical inference approach is to calculate one-sided p-value:

$$p^s=\Pr(P>p), \tag{16}$$

in case that the performance measure should be maximized (Sharpe ratio, etc.) and

$$p^s=\Pr(P<p), \tag{17}$$

in case that the performance measure should be minimized (maximum drawdown, variance, etc.). If p^s is smaller than the chosen significance level (0.05 or 0.01), the conclusion can be made that analyzed strategy performs better than random in the out-of-sample, i.e. the null hypothesis should be rejected and the alternative hypothesis should be accepted.

The one-sided version of the test is applied because the aim is to prove that the performance of the analyzed strategy is better than random. Alternatively, a two-sided version of the test can be applied, i.e. alternative hypothesis would take the form that the performance of the portfolio strategy is not the same as of random portfolio. However, in this case, rejecting the null hypothesis would only mean that the performance of the portfolio strategy is non-random, not knowing whether it is better or worse than random.

However, there is one more issue that must be taken into account – data snooping bias also named as data mining bias or backtest overfitting, see e.g. White (2000) or Bailey et al. (2014, 2017). This bias occurs when more than one strategy is analyzed, which is typically the case. Let us consider k portfolio strategies, e.g. three above mentioned approaches to parameters estimation. The out-of-sample performances $\{p_i, i=1,\dots k\}$ are known for these strategies. Of course, only the strategy with the best out-of-sample performance measure $p=\max\{p_i, i=1,\dots k\}$ is usually considered further. However, applying the simple test (16)-(17) would be a mistake as the distribution of a random variable (now the random variable is the performance of the best out of k strategies) differs from P. We denote this new random variable as P_k, where k is the number of strategies originally analyzed. Note that simple random variable P is a special case of P_k for $k=1$. Then, the bias-free statistical test should be,

$$p^m=\Pr\left[P_k>\max\{p_i, i=1,\dots k\}\right] \tag{18}$$

for maximization performance measures and

$$p^m=\Pr\left[P_k<\max\{p_i, i=1,\dots k\}\right] \tag{19}$$

for minimization performance measures.

The only problem left is the construction of the statistics P_k. In order to do so, its meaning must be kept in mind – it is the best performance out of k-tuples of random portfolios. Thus, a large number of k-tuples of random weights are generated.

The procedure of test statistics calculation is as follows. First, k-tuples of random numbers $y \in [0,1]^{N-1}$ from standard uniform distribution are generated and sorted so that $0 \leq y_1 \leq \cdots \leq y_{N-1}$. Then, the simulated weights x of the random portfolios are calculated as

$$x = (y_1, y_2 - y_1, y_3 - y_2, \dots, y_{N-1} - y_{N-2}, 1 - y_{N-1}). \quad (20)$$

This approach guarantees that the sum of weights is equal to one. For each portfolio composition from given k-tuples the out-of-sample performance is calculated and only the best performance is recorded, i.e. the maximum for maximizing performance measures and minimum for minimizing performance measures. The probability distribution of the statistics is obtained by repeating this procedure many times (in empirical analysis 50,000 times).

The last issue, which should be addressed, is the choice of performance measure. The strategies can be evaluated in many ways. The following examples are mentioned:

- The annual return corresponds to the out-of-sample return recalculated to annual basis. Investors want to maximize this measure.
- The volatility of (daily) returns – the more volatile the returns are, the riskier the investment is. Different measures for the volatility can be applied; the examples are standard deviation (SD), mean absolute deviation (MAD), Value at Risk (VaR), Conditional Value at Risk (CVaR), etc.
- Investors usually analyze also maximum drawdown – the maximum relative decline in the portfolio value over the analyzed period, see e.g. Chekhlov et al. (2005) or Magdon-Ismail et al. (2004).
- There is also the variety of performance ratios, which are simply the ratios of the reward and risk. The well-known performance ratios are Sharpe ratio (Sharpe, 1966, 1994), Gini ratio (Shalit and Yitzhaki, 1984), mean absolute deviation ratio (Konno and Yamazaki, 1991), mini-max ratio (Young, 1998), Rachev ratio (Biglova et al., 2004) and others. For the summary, see Farinelli et al. (2008).

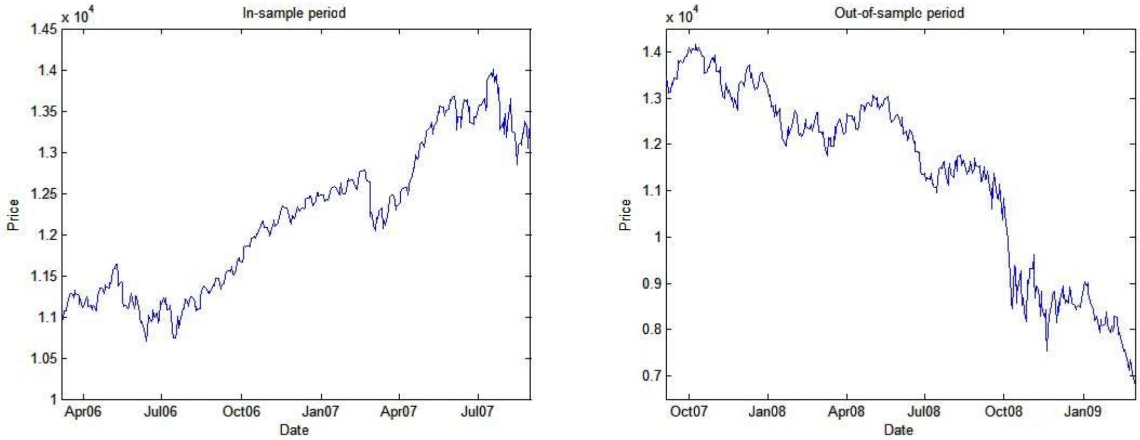
In our paper, the focus is on the portfolio strategies minimizing the risk. Thus, only the performance measures quantifying the riskiness of the portfolio are applied.

Results

Global Financial Crisis

In this section, we study one specific period, which covers the market's decline phase during the global financial crisis (GFC) to see how efficiently the portfolio strategies worked during the decline of the markets. In order to do so, the period prior to GFC, concretely March 7, 2006 – August 31, 2007, is reserved as the in-sample period. The period with the market's decline phase is the out-of-sample period (September 1, 2007 – March 2, 2009). The dataset obtained from finance.yahoo.com consists of daily closing prices of the components of Dow Jones Industrial Average index (DJIA). These prices are adjusted for paid dividends and splits. There are 29 stocks included in our analysis and the missing one is the stock of Visa Inc. due to the incomplete data in the chosen period. The dataset covers 3 years, and is evenly split into the in-sample period and the out-of-sample period. In Figure 1, it can be seen that in the in-sample period the DJIA shows an increasing trend, however, in the out-of-sample period, it keeps decreasing due to the 2007-2008 financial crisis. The minimum risk portfolios are calculated from the in-sample data. Then, the verification of the obtained portfolios is performed with the out-of-sample data.

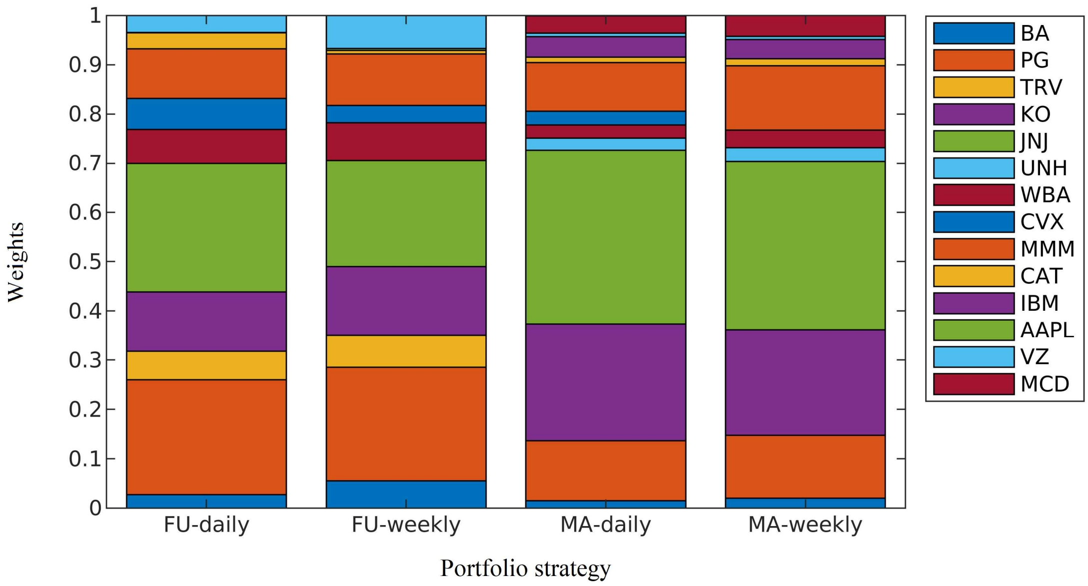
Figure 1
Historical Values of DJIA Index in In-sample and Out-of-sample Periods



Source: http://finace.yahoo.com

Four methods of portfolio strategy creation are analyzed: two models each estimated from daily and weekly returns. The models are Markowitz minimum variance model with classical historical estimation (MA) and fuzzy probabilities minimum variance model (FU). The Bayesian approach is not utilized, because for minimum variance portfolio it suggests the same weights as in case of classical historical estimation. The weights of the portfolios are depicted in Figure 2.

Figure 2
Compositions of minimum variance model (MA) and fuzzy probabilities minimum variance (FU) model estimated from daily and weekly returns



Source: own calculation

As can be seen from Figure 2, which depicts only non-zero weights, the classical historical estimation suggests a more diversified portfolio in terms of the number of assets; however, more than 50% of the portfolio is made up by two stocks (The Coca-

Cola Company and Johnson & Johnson). The portfolio of fuzzy probabilities minimum variance model is more evenly diversified, although with fewer assets.

For these four portfolios, the performance measures are calculated in the out-of-sample period. Only risk measures are considered. Specifically, the chosen risk measures are maximum drawdown (MDD), which is a commonly applied measure among practitioners, standard deviation of daily returns (SD), which is minimized in-sample, and mean absolute deviation of daily returns (MAD) as an alternative measure to the standard deviation. The values of the measures for analyzed portfolios together with two commonly applied benchmarks are depicted in Table 1.

Table 1
Out-of-Sample Performance Measures with Corresponding P-values

Model	MDD	SD	MAD
MA – daily data	35.67% (.001)	1.65% (<.0001)	1.05% (<.0001)
MA – weekly data	36.57% (.019)	1.64% (<.0001)	1.05% (<.0001)
FU – daily data	37.58% (.040)	1.75% (<.0001)	1.12% (<.0001)
FU – weekly data	37.95% (.051)	1.75% (<.0001)	1.13% (<.0001)
DJIA index	52.25%	2.14%	1.46%
1/N strategy	45.24%	2.23%	1.50%

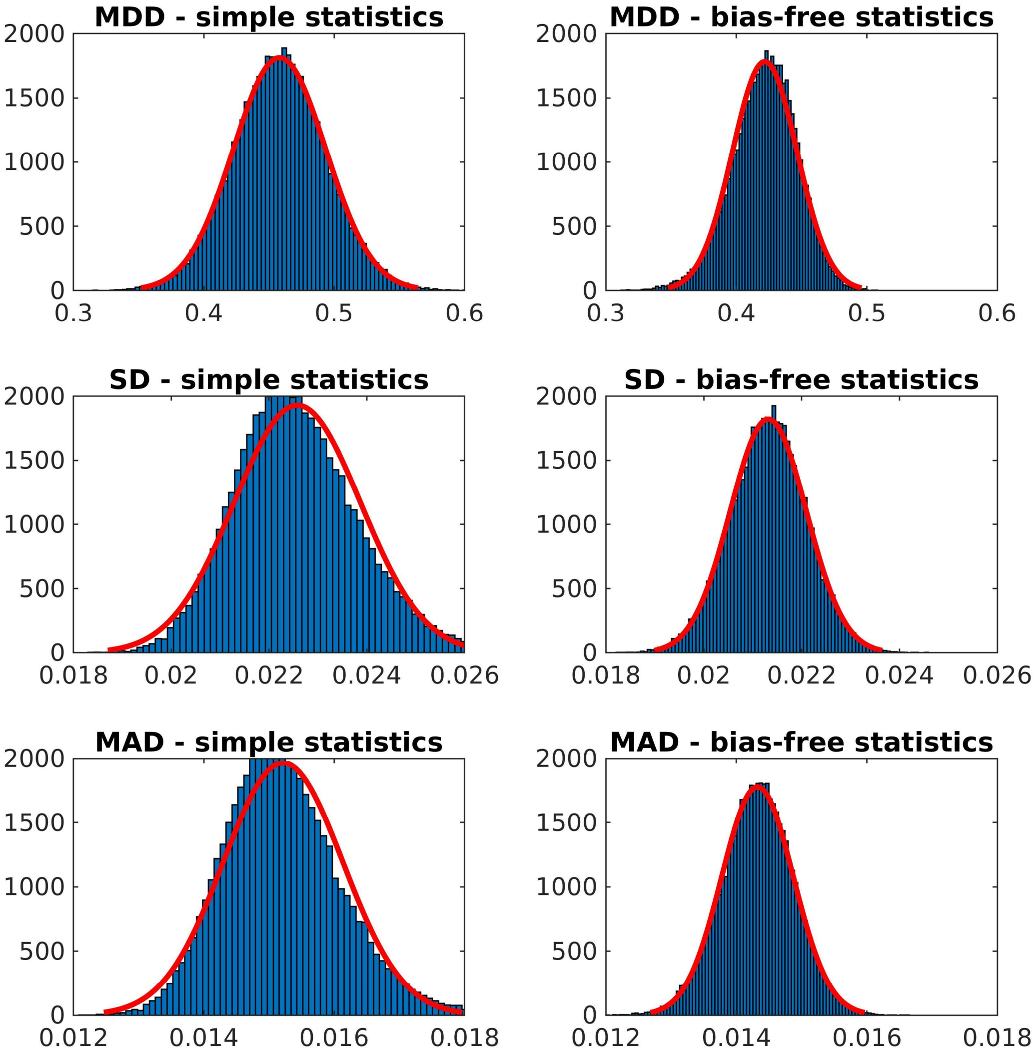
Source: own calculation

As can be seen, for all three considered measures, the portfolios perform better than the benchmarks, thus, according to the classical rule of thumb, the conclusion is that these strategies work. However, is the claim that these strategies lower the risk in the out-of-sample period correct from the statistical point of view? To find this out, the distribution of the test statistics must be calculated. In order to do so, 50,000 times 4 (1 respectively) random portfolio compositions are simulated according to (20), for each portfolio the out-of-sample MDD, SD and MAD are calculated, and finally, for every foursome (respectively for each) portfolios the minimums of these risk measures are recorded. In this way, the distributions of the statistics are numerically obtained, see Figure 3. In the figure, the left column corresponds to the simple statistics (17) and the right column depicts the bias-free statistics (19) for four portfolio strategies. The range on the x-axes is the same to make the comparisons easier. Moreover, the fitted Gaussian distributions are added into the graphs.

It can be noticed that the bias-free statistics are shifted to the left, which means that for the same observed out-of-sample performance the higher p-value is obtained – there are more randomly generated portfolios with better performances. Moreover, it can be noticed that MDD statistics can be approximated by Gaussian distribution in the case of the simple test. In the case of the bias-free test, the statistics is negatively skewed with a heavier left tail. On the other hand, for SD and MAD the bias-free statics are almost normally distributed, but simple statistics are positively skewed with a heavier right tail.

When the mean and standard deviation are known, the classical t-test can be applied. Nevertheless, the p-values for all three bias-free statistics are calculated as the number of random-weights portfolios, which are better than analyzed strategy, divided by 50,000. The calculated p-values are depicted in Table 1. From the results, the following conclusions can be made. Firstly, the portfolio strategies do not create the portfolios randomly (i.e. the null hypothesis is rejected). These strategies lower SD and MAD in the out-of-sample period. Secondly, it can be concluded that the strategy minimizing the variance with standard historical estimation lower the maximum drawdown in the out-of-sample period (p-value for daily and weekly returns is lower than 0.01 and 0.05 respectively). Thirdly, the null hypothesis cannot be rejected for fuzzy probabilities minimum variance model at 1% (5%) significance level for the MDD test and we conclude that this portfolio strategy generates the portfolio composition more or less randomly. To sum it up, the fuzzy probabilities minimum variance model performs slightly worse than Markowitz minimum variance model.

Figure 3
Histograms of Risk Measures from 50,000 Randomly Generated Portfolios with Fitted Gaussian Distributions



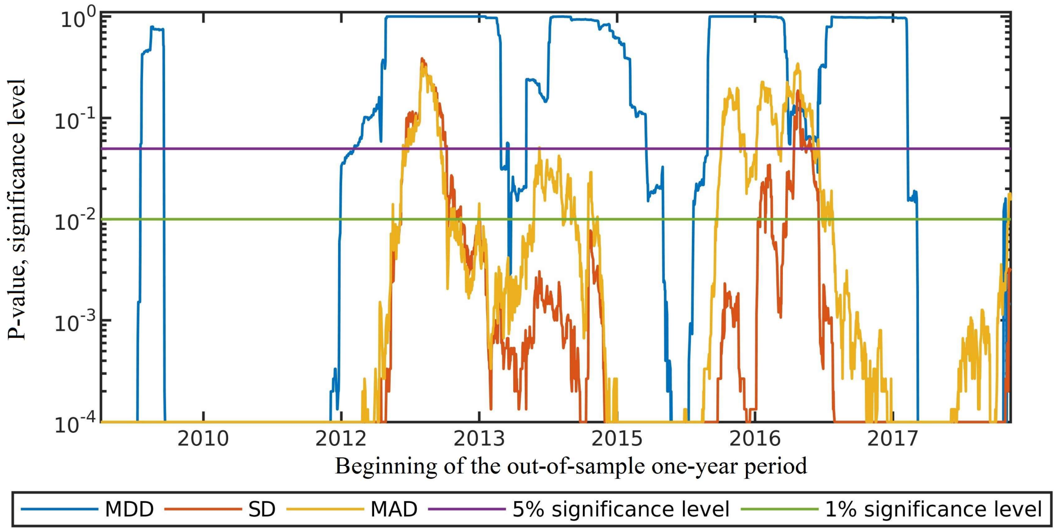
Source: own calculation

Rolling Window Tests in Period 2006-2019

In order to prove that the results are robust to the change of the period considered, the tests are performed on a five-year rolling window basis – four years as the in-sample period and one year as the out-of-sample period. The dataset is the daily-adjusted closing prices of the DJIA components in the period from January 3, 2006, to December 31, 2019. There are only 27 stocks included in our analysis as components Dow Inc., NIKE Inc, and Visa Inc. are excluded due to the incomplete data. The source of the data is finance.yahoo.com. The in-sample period consists always of 1,000 daily returns (approximately four years) and the out-of-sample period consists always of 250 daily returns (approximately one year).

We proceed as follows. First, the portfolio is calculated based on the in-sample period from January 3, 2006, to December 22, 2009. Only Markowitz's minimum variance portfolio is considered, as it is the best strategy according to the results in Table 1. Second, risk measures, specifically maximum drawdown (MDD), standard deviation (SD) and mean absolute deviation (MAD) are calculated in the out-of-sample period from December 22, 2009, to December 21, 2010. Then, the simple statistical tests for these risk measures are performed and p-values are recorded. The simple statistical test (17) is applied because only one portfolio strategy is analyzed. These three steps are repeated 3,273 times moving the beginning of the periods day-by-day from January 3, 2006, to January 2, 2018 (four-year in-sample period) and from December 22, 2009, to January 2, 2019 (one-year out-of-sample period). The recorded p-values of the simple statistical test are shown in Figure 4. For better clarity, the y-axis is exponential and two significance levels (1% and 5%) are added into the graph.

Figure 4
P-values of Bias-free Test for Selected Risk Measures Considering Different Out-of-Sample Periods



Source: own calculation

Mixed results can be seen from the figure 5. First, the strictest statistic is based on MDD. For this measure, there are long periods in which the strategy does not lower the risk in the out-of-sample period efficiently (it is about 50% of the analyzed period).

Secondly, SD and MAD behave similarly, although better results are obtained applying SD. This can be explained by the fact that SD (and not MAD) is minimized in-sample. According to this criterion, the strategy does not lower the risk in the one-year out-of-sample periods from December 11, 2012, to July 18, 2013 (January 2, 2013 – May 28, 2013) and from July 1, 2016, to February 2, 2017 (November 7, 2016 – January 12, 2017) at 1% (5%) significance level. It can be seen that for a 1% (5%) significance level the period of inefficiency is relatively short – around 11% (6%) of the total analyzed period.

Discussion

In the paper, an alternative approach is proposed to evaluate the efficiency of the portfolio optimization strategies. The approach is based on statistical inference methodology, i.e. it statistically tests whether the out-of-sample performance is non-random.

The proposed test statistics are close to the normal distribution, thus also the classical t-test can be applied, however, the mean and standard deviation of the test statistics must still be estimated numerically. Comparing the test statistics in the case of single and multiple analyzed strategies, it was found out that the bias-free statistics are shifted to the left. This finding is in line with the explanation of the bias due to the data snooping provided by Aronson (2011). Both in the paper and in the book the test statistics are shifted in the favorable direction of the considered performance measure – in the paper risk measures are shifted to the left as we consider the minimizing performance measures while according to Aronson (2011) the annualized return is shifted to the right as it is maximizing performance measure. It confirms the need to consider the data snooping bias when analyzing the historical performance of portfolio strategies by investors. This phenomenon was already described by White (2000) and well explained by Taleb (2007) under the term survivorship bias.

Based on the presented results, it was confirmed that minimizing the variance of portfolio return in-sample also lower the out-of-sample risk measures. Thus, for investors seeking the minimum risk portfolio, it is worth analyzing the distribution of historical returns and constructs a minimum variance portfolio. This finding is in contradiction with the findings of DeMiguel et al. (2009), who found that none of the strategies, which they analyzed, is persistently better than the 1/N strategy. Thus, it seems that when focusing solely on the risk, compared to the performance ratios in the mentioned study, the strategies are efficient.

The advantage of the proposed approach is that it is stricter than the simple comparison with the benchmark. It is illustrated in the empirical part – maximum drawdown of fuzzy probabilities minimum variance model is lower than that of the index and 1/N strategy, but its efficiency cannot be statistically accepted at a 1% significance level. Also, the quantity of analyzed strategies is considered, thus, avoiding data snooping bias.

The disadvantage of the approach is its computational complexity as a huge quantity of random portfolios must be simulated and their historical performance must be evaluated.

Conclusion

In the evaluation of strategy portfolio performance, the simple and straightforward way is to test whether the strategy outperform the benchmark. This approach answers the question of whether and how much the analyzed strategy overperformed the benchmark in one particular period, but it does not say whether the overperformance is high enough to be considered as significant or whether it is just due to the randomness in data and luck. In the paper, an alternative approach based on the statistical test is proposed in order to evaluate the efficiency of the portfolio optimization strategies. Based on this approach the statistical significance can be confirmed.

The presented methodology considers only one performance measure – in the paper, the risk minimization is considered. This can be limiting and further research should address the question of how to statistically test strategies aiming at two goals, e.g. minimizing the risk while assuring some minimum expected return.

The goal of the paper is to propose and illustrate the test statistics calculation considering possible data snooping bias. The empirical paper studying more strategies as well as more performance measures should follow so that the findings can be easily generalized. The choice of DJIA index as the dataset was due to the easy availability of the data considering dividends and splits. More markets and periods can be considered.

References

1. Aronson, D. (2011). Evidence-based technical analysis: applying the scientific method and statistical inference to trading signals, Wiley, New Jersey.
2. Babazadeh, H., Esfahanipour, A. (2019), "A novel multi period mean-VaR portfolio optimization model considering practical constraints and transaction cost", Journal of Computational and Applied Mathematics, Vol. 361, No. 1, pp. 313-342.
3. Bailey, D. H., Borwein, J. M., de Prado, M. L., Zhu, Q. J. (2017), "The probability of backtest overfitting ", Journal of Computational Finance, Vol. 20, No. 4, pp. 39-69.
4. Bailey, D. H., Borwein, J. M., de Prado, M. L., Zhu, Q. J. (2014), „Pseudomathematics and financial charlatanism: the effects of backtest over fitting on out-of-sample performance", Notices of the AMS, Vol. 61, No. 5, pp. 458-471.
5. Biglova, A., Ortobelli, S., Rachev, S., Stoyanov, S. (2004), "Different approaches to risk estimation in portfolio theory", Journal of Portfolio Management, Vol. 31, No. 1, pp. 103-112.
6. Chekhlov, A., Uryasev S., Zabarankin M. (2005), "Drawdown measure in portfolio optimization", International Journal of Theoretical and Applied Finance, Vol. 8, No. 1, pp. 13-58.
7. DeMiguel, V., Garlappi, L., Uppal, R. (2009), "Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy", Review of Financial Studies, Vol. 22, No. 5, pp. 1915-1953.
8. Farinelli, S., Ferreira, M., Rossello, D., Thoeny, M., Tibiletti, L. (2008), "Beyond Sharpe ratio: optimal asset allocation using different performance ratios", Journal of Banking & Finance, Vol. 32, No. 10, pp. 2057-2063.
9. Fulga, C. (2016), "Portfolio optimization under loss aversion", European Journal of Operational Research, Vol. 251, No. 1, pp. 310-322.
10. James, W., Stein, C. (1961), "Estimation with quadratic loss", in Neyman, J. (Ed.), Fourth Berkeley Symposium on Mathematical Statistics and Probability, 20-30 June, University of California Press, Berkeley, pp. 361-379.
11. Jorion, P. (1986), "Bayes-Stein estimation for portfolio analysis", The Journal of Financial and Quantitative Analysis, Vol. 21, No. 3, pp. 279-292.
12. Kalayci, C., Ertenlice, O., Akbay, M. (2019), "A comprehensive review of deterministic models and applications for mean-variance portfolio optimization", Expert Systems with Applications, Vol. 125, pp. 345-368.
13. Konno, H., Yamazaki, H. (1991), "Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock-market", Management Science, Vol. 37, No. 5, pp. 591-531.
14. Lwin, K. T., Qu, R., MacCarthy, B. L. (2017), "Mean-VaR portfolio optimization: a nonparametric approach", European Journal of Operational Research, Vol. 260, No. 2, pp. 751-766.
15. Magdon-Ismail, M., Atiya, A. F., Pratap, A., Abu-Mostafa, Y. S. (2004), "On the maximum drawdown of a Brownian motion", Journal of Applied Probability, Vol. 41, No. 1, pp. 147-161.
16. Markowitz, H. (1952), "Portfolio selection", The Journal of Finance, Vol. 7, No. 1, pp. 77-91.
17. Ranković, V., Drenovak, M., Urosevic, B., Jelic, R. (2016), "Mean-univariate GARCH VaR portfolio optimization: actual portfolio approach", Computers & Operations Research, Vol. 72, No. 1, pp. 83-92.
18. Shalit, H., Yitzhaki, S. (1984), "Mean Gini, portfolio theory and pricing of risky assets", Journal of Finance, Vol. 39, No. 5, pp. 1449-1468.
19. Sharpe, W. F. (1966), "Mutual fund performance", Journal of Business, Vol. 39, No. 1, pp. 119-138.
20. Sharpe, W. F. (1994), "The Sharpe ratio", Journal of Portfolio Managemet, Vol. 21, No. 1, pp. 49-58.

21. Solares, E., Coello, C., Fernandez, E., Navarro, J. (2019), "Handling uncertainty through confidence intervals in portfolio optimization", Swarm and Evolutionary Computation, Vol. 44, No. 1, pp. 774-787.
22. Stein, C. (1956), "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution", in Neyman, J. (Ed.), Third Berkeley Symposium on Mathematical Statistics and Probability, December 1954 and July-August 1955, University of California Press, Berkeley, pp. 197-206.
23. Taleb, N. N. (2007). Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets, Penguin Books, London.
24. Tanaka, H., Guo, P., Türksen, B. (2000), "Portfolio selection based on fuzzy probabilities and possibility distributions", Fuzzy Sets and Systems, Vol. 111, No. 3, pp. 387-397.
25. White, H. (2000), "A reality check for data snooping", Econometrica, Vol. 68, No. 5, pp. 1097-1126.
26. Young, M. R. (1998), "A minimax portfolio selection rule with linear programming solution", Management Science, Vol. 44, No. 5, pp. 673-683.

About the authors

Aleš Kresta is an Associate Professor at the Faculty of Economics, VSB – Technical University of Ostrava, Department of Finance. His main research interests are risk estimation and backtesting, portfolio optimization, financial time series modeling, soft-computing, and other problems related to quantitative finance. He is actively engaged in the number of science projects and can be contacted at **ales.kresta@vsb.cz**

Anlan Wang is a Ph.D. student at the Faculty of Economics, VSB – Technical University of Ostrava, Department of Finance. Her main research area and the topic of her dissertation thesis are problems related to the portfolio optimization. She can be contacted at **anlan.wang@vsb.cz**