Application of Ensemble Learning for Views Generation in Meucci Portfolio Optimization Framework

Alexander Didenko and Svetlana Demicheva

Financial University under the Government of the Russian Federation

September 2013

Application of Ensemble Learning for Views Generation in Meucci Portfolio Optimization Framework*

Alexander DIDENKO, Ph.D.
Deputy Dean, International Finance Faculty, Financial University, Moscow
alexander.didenko@gmail.com

Svetlana DEMICHEVA
International Finance Faculty, Financial University, Moscow
svetlana86d@rambler.ru

Abstract. Modern Portfolio Theory assumes that decisions are made by individual agents. In reality most investors are involved in group decision-making. In this research we propose to realize group decision-making process by application of Ensemble Learning algorithm, in particular Random Forest. Predicting accurate asset returns is very important in the process of asset allocation. Most models are based on weak predictors. Ensemble Learning algorithms could significantly improve prediction of weak learners by combining them into one model, which will have superiority in performance. We combine technical fundamental and sentiment analysis in order to generate views on different asset classes. Purpose of the research is to build the model for Meucci Portfolio Optimization under views generated by Random Forest Ensemble Learning algorithm. The model was backtested by comparing with results obtained from other portfolio optimization frameworks.

Аннотация. Современная портфельная теория предполагает индивидуальность в принятии решений инвесторами. В реальности большинство инвесторов принимают решения в группах. В данном исследовании предлагается реализовать процесс группового принятия решений применением алгоритма ансамбля обучения (Ensemble Learning), в частности метода “Случайный лес” (“Random Forest”). Точность в предсказании доходностей активов играет большую роль в портфельной оптимизации. Большинство методик основывается на слабых гипотезах. Алгоритмы ансамбля обучения помогают значительно улучшить точность предсказания, объединяя слабые гипотезы в одну модель. Для предсказания доходностей активов мы объединили фундаментальный, технический и сентиментальный анализы. Целью данного исследования является создание модели для портфельной оптимизации по Меуччи, основывающейся на алгоритме ансамбля обучения. Оценка данной модели проведена путем сравнения ее с другими методами портфельной оптимизации на исторических данных.

Key words: Random Forest, Ensemble Learning, Meucci portfolio optimization, combination of fundamental technical and sentiment analysis.

1. INTRODUCTION

Portfolio optimization problem always stays in front of investors. The Markowitz mean-variance optimization theory had big impact on Modern Portfolio Theory. However it is rarely implemented by professional investors. There are some drawbacks which cause the investors to refuse using Markowitz optimization. Firstly the model produces highly concentrated portfolio and generates short position, if there is no constraint for it. The second is that the optimization is made in unintuitive way. Investors always have the views on market realization, which are not considered by the Markowitz model.

Modern Portfolio Theory assumes that decisions are made by individual agents, but practically investors are involved in group decision-making. It was shown that group decisions improve the final outcomes in decision-making process and people before making a final decision always look for other opinions. They are weighting individual opinions and combine them in order to reach more reasonable and accurate decisions. Researches in decision-making theory show superiority of group decision making over individual. Hinsz *et*

* Применение алгоритма «ансамбля обучения» для формирования рыночных оценок при портфельной оптимизации по Меуччи

al. (1997) showed that about 56% of investors are involved in team decision-making.

For realization the group decision-making process in generation the views on selected asset classes is proposed to use Ensemble Learning algorithm. Ensemble learning is type of machine learning approach which combines single classifiers in purpose to construct the model which has superiority in performance. Previous researches in Ensemble Learning, such as Hansen and Salamon (1990), Yuehui Chen *et al*. (2007), Myoung-Jong Kim *et al.* (2005), Se-Hak Chuna and Yoon-Joo Park (2005), Tae-Hwy Lee and Yang Yang (2005), Chih-Fong Tsai *et al.* (2010) have proved that such algorithms improve significantly accuracy and stability of prediction.

Most theoretical and practical analysis is set on weak hypothesis. Ensemble Learning is based on weak learnability. It suggests that basic model should provide results which are slightly better than random guess. Attractiveness of Ensemble Learning algorithms is that it could create strong learning algorithm from weak basic learners.

Purpose of the research is creation of the model for portfolio optimization based on the views which are generated by the Random Forest Ensemble Learning algorithm.

There are different types of ensembles algorithms, but no one of them has superiority in performance over different cases. There are such methods as bagging, boosting, staking, random forest, multi stratagem ensembles. To forecast asset returns in this research it is proposed to use Random Forest Ensemble Learning algorithm.

Random Forest is a variation of bagging method. It was first described in the work of Breiman (2001). The algorithm consists of great number of individual decision trees. Each tree is constructed from a random subset of features.

Investors use technical, fundamental, sentiment analysis for forecasting asset returns in the market. In this research we combine fundamental, technical and sentiment analysis by Random Forest Ensemble Learning algorithm in order to predict returns of different asset class.

Technical analysis is based on the idea that all relevant information about a company is reflected in its price and with the passage of time there is no need to analyse company fundamental information. Fundamental analysis is the group of methods for stock valuation to determine its intrinsic value. Fundamental analysis is an alternative technique to technical analysis in investment decision making. It considers macroeconomic factors and fundamental information of a company to forecast stock returns. Sentiment analysis of financial markets expresses the opinion of investors on the situation in market. This analysis allows forecasting the movements in financial market before it is reflected in stock prices.

The views generated by Random Forest model will be the inputs for Meucci portfolio optimization framework. Meucci Copula Opinion Pooling optimization model extends the Black-Litterman model by allowing investors to set the views in various ways. Views could be either normally or not-normally distributed and could be set in market realization, not only in the parameters which determine the realization of the market. Black and Litterman introduced their model (1992) in order to solve the problems of highly concentrated portfolio and unintuitive way of Markowitz optimization framework.

In order to evaluate results of Meucci portfolio optimization framework under Random Forest views we will backtest the model by comparing it with other portfolio investment frameworks, such as Markowitz portfolio optimization, market portfolio, naive diversification, 60–40 Equity-Bond portfolio.

2. METHODOLOGY

2.1. ASSET VIEWS GENERATION BY ENSEMBLE LEARNING

To use Meucci Copula Opinion Pooling framework for portfolio optimization we first need to generate the views on selected asset classes. For purpose of asset allocating we need to pick up the asset class which will provide the optimal portfolio with required rate of return and will give enough diversification to reduce the specific risk of the assets. For this purpose we include in our analysis such asset classes as US equities, US fixed interest, US real estate and commodities. The proxy for big caps is S&P 500 stock index, for small caps is Russell 2000 Index, for fixed interest are 10-years treasury notes and Moody's Seasoned AAA Corporate Bond Yield, and proxy for oil is oil futures. We use monthly data for the period from January 1990 till May 2013.

There are different methods for producing such views for asset classes, such as fundamental, sentiment, and technical analysis. In this research Random Forest Ensemble Learning algorithm will generate the views on selected asset classes by combining fundamental, sentiment, and technical analysis. In order to achieve better accuracy in Ensemble Learning model we need to comply with diversification principle, it means that there should be big diversity between basic predictors. To achieve this purpose we considered 60 fundamental, sentiment and technical factors for constructing basic classifiers. Following factors were included in our anal-

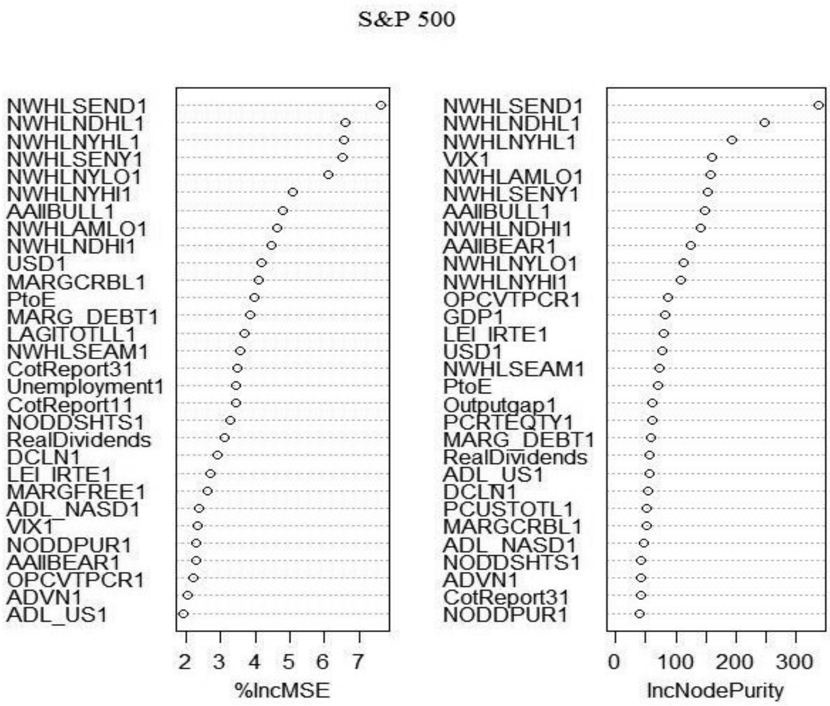


Figure 1. Important variables for S&P 500.

ysis: unemployment, inflation, GDP, output gap, long-term interest rate, U. S. Recession Probabilities, conference board leading and lagging indicators, federal funds rate, volatility index, Michigan Consumer Sentiment Index, commitment of traders, advance –decline indices, sentiment indicators of American Association of Individual Investors, closing arms indices, put-call ratios, new highs- new lows indicators, U. S. Dollar Index, Odd Lot indicators, short interest ratio, NYSE margin, free credits and available cash, S&P 500 EPS, S&P 500 price to earnings ratio, S&P 500 real dividend, S&P 500 real earnings.

The data was processed by using R-programming language.

The dataset which consists of monthly observation of assets returns and monthly values of fundamental, sentiment and technical factors was divided in two samples for training and test purpose. The training sample represents about 70% of dataset and includes the data from January 1990 till December 2005. The test sample represents about 30% of dataset and includes the data from January 2006 till May 2013.

Random Forest constructed the ensemble model by learning from data of training sample. Then the model was applied to the test subset for generating the view on assets returns.

At first Random Forest was built by implying all explanatory variables. Then the variables were evaluated by their ability to explain asset returns. The function "importance" of Random Forest package measures the importance of variables.

The first value (%IncMSE) measures the importance of variable in ability to reduce mean squared error in Random Forest.

The second value (IncNodePurity) shows the importance of variable in ability to decrease of node impurities from splitting on the variable. If the variable is significant in explaining the assets returns then it will have the large value for%IncMSE and IncNodePurity.

Non-significant variables were determined and removed from the dataset for each asset. The significant variables were used for constructing the Ensemble Learning model. Example of important variables for S&P 500 is shown in Figure 1.

The errors in prediction are decreased during increasing the number of trees in Random Forest. It was found out that about 300 trees give minimal error and further rising in number of trees will not improve

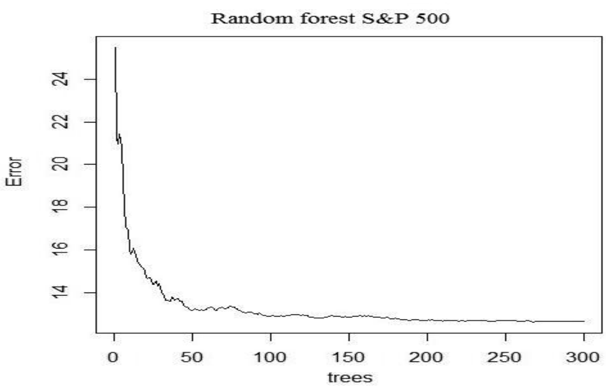


Figure 2. Relation between error and number of trees in Random Forest for S&P500.

a) b)

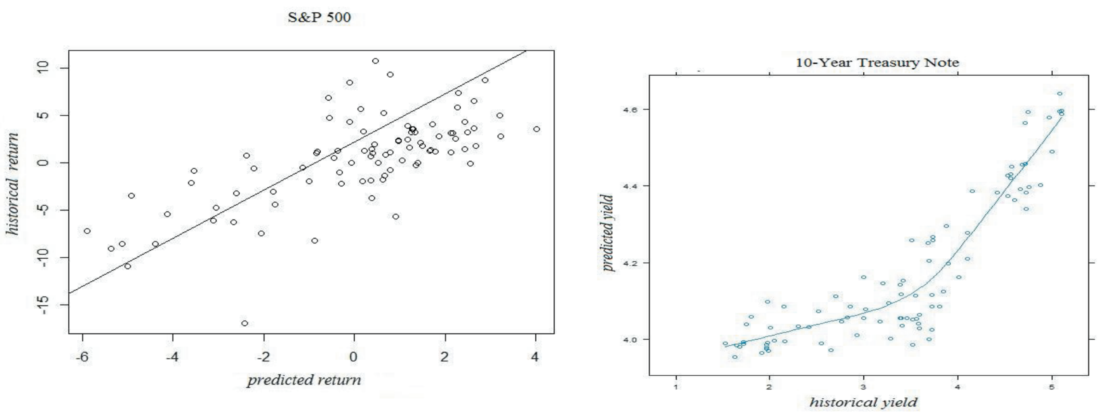


Figure 3. Relationship between historical and predicted values of returns for:
a) S&P500 b) 10 Years Treasure Bonds.

the prediction. Example of chart of error reduction in relation to number of trees for S&P 500 is shown in Figure 2.

Expected returns for each asset class were predicted on test subset of data. Predicted and historical values of returns were plotted on returns scatter diagram. Relations between the predicted and actual values of returns are showed by regression line. Examples for S&P500 and 10-years Treasure Bonds are presented in Figure 3.

The charts above demonstrate that there is a relation between the predicted and actual values of asset returns and we can consider them in portfolio optimization by Meucci as inputs variables.

2.2. PORTFOLIO OPTIMIZATION IN MEUCCI COPULA OPINION POOLING FRAMEWORK

Portfolio optimization by Meucci was made using R-programming language. For the realization of Meucci algorithms for portfolio optimization we firstly generate prior multivariate distribution of returns. Following Meucci recommendations we model prior distribution as multivariate t-Student distribution with five degrees of freedom, Black-Litterman equilibrium returns as means and usual matrices of variance and vectors of standard deviation. Equilibrium returns are calculated by the following formula:

$$\mu = \lambda \bullet \sum W_i$$

Where, μ— equilibrium returns

λ — is risk aversion coefficient

$\sum$ — covariance matrix of asset returns during last 60 months

W_i— current capitalization of asset (%)

We calculate risk aversion λ dynamically for each month as:

$$\lambda = \frac{MR - R_f}{(M\sigma)^2}$$

Where, R_f — risk-free rate.

MR — mean return of market portfolio during last 60 months (cap-weighted return of all 7 assets)

$(M\sigma)^2$ — standard deviation of market portfolio historical returns

Capitalization of asset (%) is calculated as:

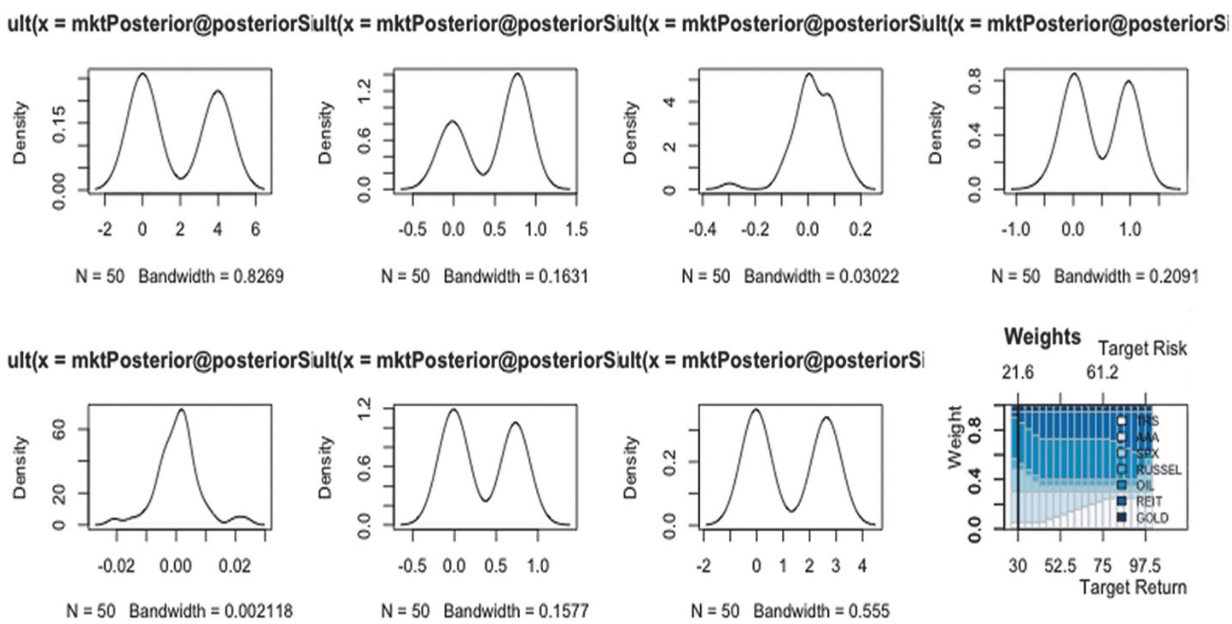
$$W_i = \frac{Cap_i}{\sum Cap_i}$$

Where, Cap_i — Capitalization of the asset class

$\sum Cap_i$ — Sum of the capitalization of all selected asset classes.

Then we introduce views generated by the Random Forest algorithm for each asset class. The views are created as special R-project objects by the COPViews and AddCOPViews functions from BLCOP package. Views on the asset classes are assumed to be normally distributed. Each view is described by mean and standard deviation. Mean equals to return, predicted by Random Forest algorithm, and standard deviation equals to historical standard deviation for assets monthly return.

We then mix views with prior multivariate distribution and generate from this new distribution 500 vectors with 7x1 dimension of possible returns using Monte Carlo Simulation. We calculate means and CVaR risk measures for each of simulated series, and use obtained means and CVaRs as inputs for usual portfolio optimisation. We use portfolioFrontier function from package fPortfolio of R-project statistical software for



10 years Treasury Note; AAA corporate bonds; S&P 500; Russell 2000; Oil; NAREIT; Gold.

Figure 4. Posterior distribution of returns.

constructing the efficiency frontier. Efficiency frontier is thus built basing on CVaR as a coherent risk measure. The example of posterior distribution under applied views of returns is showed in Figure 4. From this figure we can see that presence of bullish views on the asset class, such as Gold or Treasures, increase the weight of the respective asset class in the portfolio. On the contrary, the absence of bullish views for Oil results in relatively small weight of Oil in the portfolio.

We consider six portfolios from efficiency frontier obtained from Meucci optimization for our analysis:

- Tangency Portfolio. This is a portfolio which is located at the tangency point of the efficiency frontier and line drawn from risk-free point;
- Minimum-risk Portfolio;
- Min-mid risk portfolio. It is the portfolio with the average risk between minimum-risk and middle-risk;
- Middle risk portfolio;
- Mid-max risk portfolio. It is the portfolio with the average risk between the middle-risk and maximum-risk of portfolio;
- Maximum risk Portfolio.

For evaluating results of Meucci optimization, we compared the Meucci's portfolios with portfolios obtained from different optimization methods, such as Markowitz, Naive diversification, Market portfolio, 60–40 equity — bonds portfolio.

We consider six portfolios from Markowitz efficiency frontier based on the same principles for risk preference as for Meucci optimization. Market portfolio consists of the asset classes weighted on their market capitalization. 60–40 equity-bond portfolio is a starting point for portfolio optimization for average investor. Equity investments provide growth return opportunities and bonds provide risk-minimization opportunities. Naive diversification suggests to invest in different asset classes with the hope to that diversification will be reached.

Transition maps for optimal portfolios of Meucci optimization under the choosing level of risk are showed in Figure 5. Transition maps for Markowitz optimization are showed in Figure 6.

By comparing the transition maps for Markowitz and Meucci optimization we conclude that Meucci framework provides better diversification across various asset classes. Meucci optimization makes substantial investment in 7 assets for the whole analyzed period. Markowitz portfolio is highly concentrated, and always is allocated between two-three asset classes for considered period.

The box plot of return distribution for portfolios is showed in Figure 7.

Median for each return distribution is showed by vertical line. The boxes show the 50% range of return distribution. Lines limited the 75% range of return distribution. The dots show the outliers of return distribution. We can see that Markowitz maximum risk and max-mid risk portfolios have higher volatility of returns. Portfolios obtained from Meucci optimization have average volatility of returns, which is comparable to the market portfolio, 60–40 Equity-Bonds portfolio and Naive diversification.

Capture Ratio for asset returns is showed in Figure 8. It shows the upside and downside movement

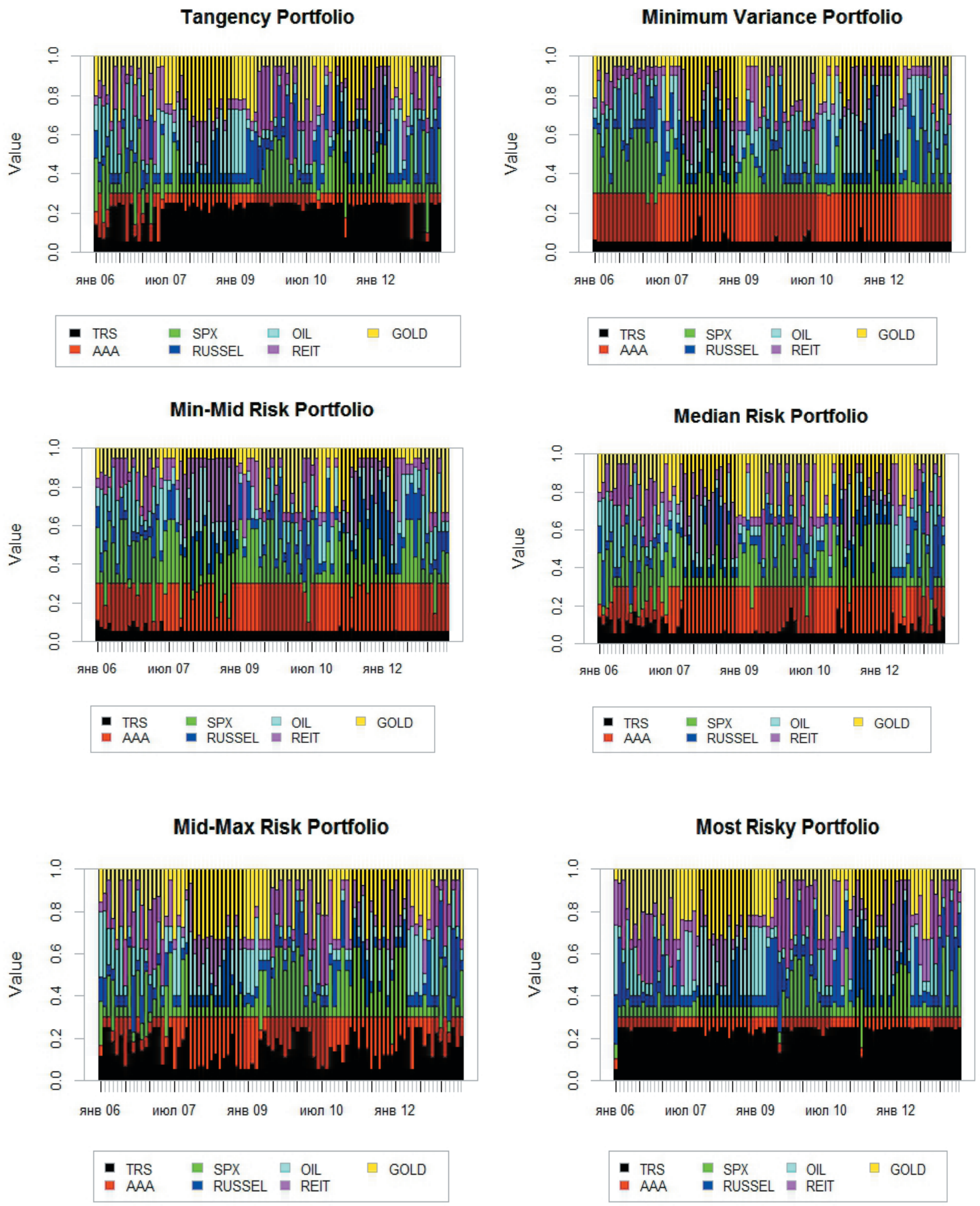


Figure 5. Transition maps of Meucci portfolio optimization.

of portfolio returns in comparison to the market portfolios.

We can see from the chart that when the market moves downside, Gold, Bonds and low-risk Markowitz portfolio go against the market. Equities, REIT, high-risk Markowitz portfolios, all Meucci portfolios move in same direction with market downside movment. When the market moves up Meucci portfolios go against the market same as Equities, REIT and high-risk Markowitz portfolios.

For evaluating the performance of portfolios obtained from different optimization we calculated Sharpe Ratio, Sortino Ratio, and Maximum Drawdown for each portfolio. The results are showed in the Table 1.

The chart for Sharpe Ratio, Sortino Ratio, and Maximum Drawdown measure is shown in Figure 9.

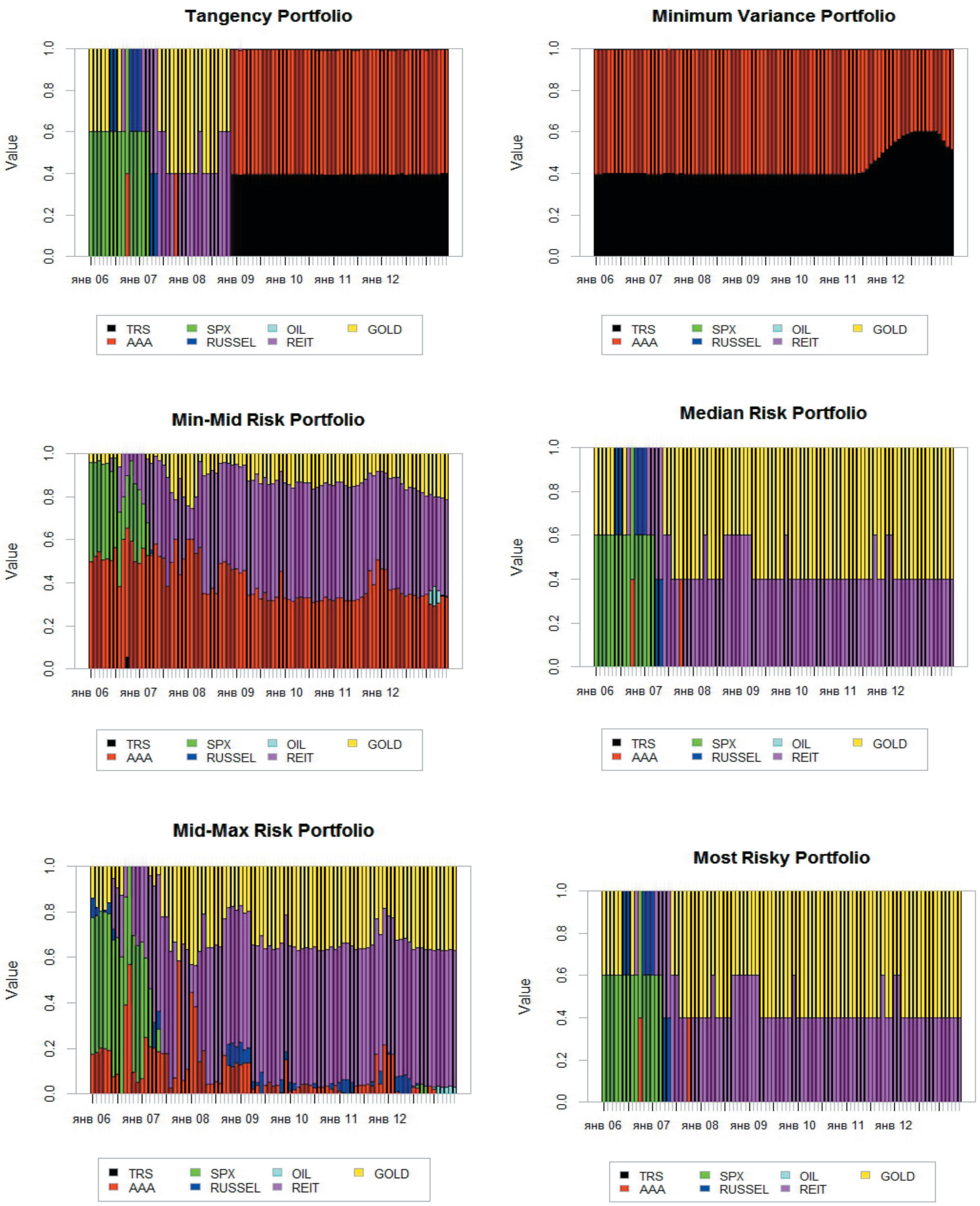


Figure 6. Transition maps of Markowitz portfolio optimization.

Sharpe Ratio is calculated by the following formula:

$$SharpRatio = \frac{R_i - R_f}{\sigma}$$

Where, R_i –return of portfolio
R_f — risk-free rate
σ — Standard deviation

According to Sharpe Ratio Meucci portfolios have good performance. The Sharpe Ratio gives stable results and does not differ significantly across the risk-tolerance. Markowitz's portfolio has good Sharpe Ratio for minimum risk and increasing in risk tolerance leads to decreasing in Sharpe Ratio. For the portfolios with high-risk level Meucci optimization provides better results than Markowitz optimization.

Sortino Ratio based on semi deviation as the risk measure of expected returns. It considers only the

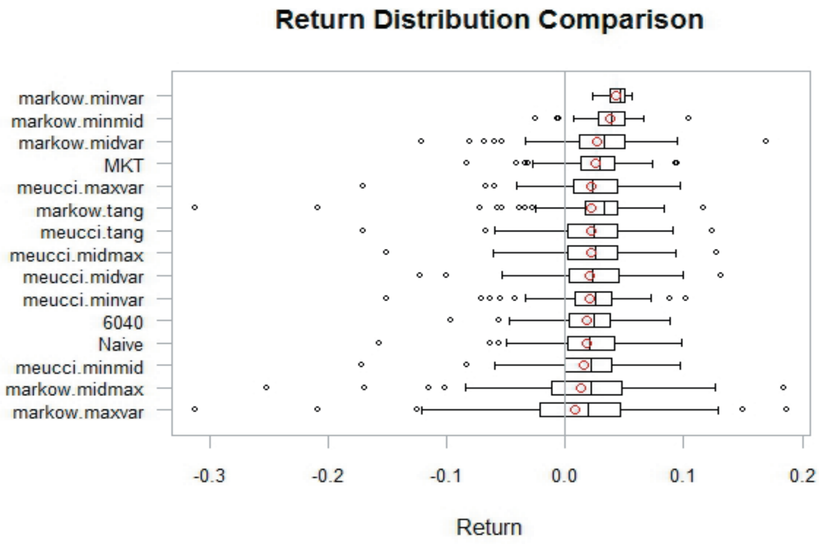


Figure 7. **Box Plot of returns distribution.**

volatility of negative returns. Sortino Ratio is calculated by the following formula:

$$SortinoRatio = \frac{R_i - R_f}{Semideviation}$$

Where, R_i –return of portfolio

R_f — risk-free rate

σ — Standard deviation

Semideviation — Standard deviation of negative returns

Based on analysis of Sortino Ratio, Meucci Portfolios also provides stable results for different risk preferences. There is no big difference in Sortino Ratio for considered Meucci portfolios, while Sortino Ratio for Markowitz portfolio varies significantly under the risk preferences. Meucci optimization provides better results for high risk tolerance, while Markowitz optimization has better results at low-risk tolerance. Sortino Ratio for Markowitz minimum risk portfolio could not be measured because the portfolio consists only of bonds, which provide only positive returns.

Due to the Maximum Drawdown coefficient Meucci portfolios are comparatively better than Markowitz portfolios. All the portfolios of Meucci optimization are stable in Maximum Drawdown and have approximately equal values of drawdown coefficient.

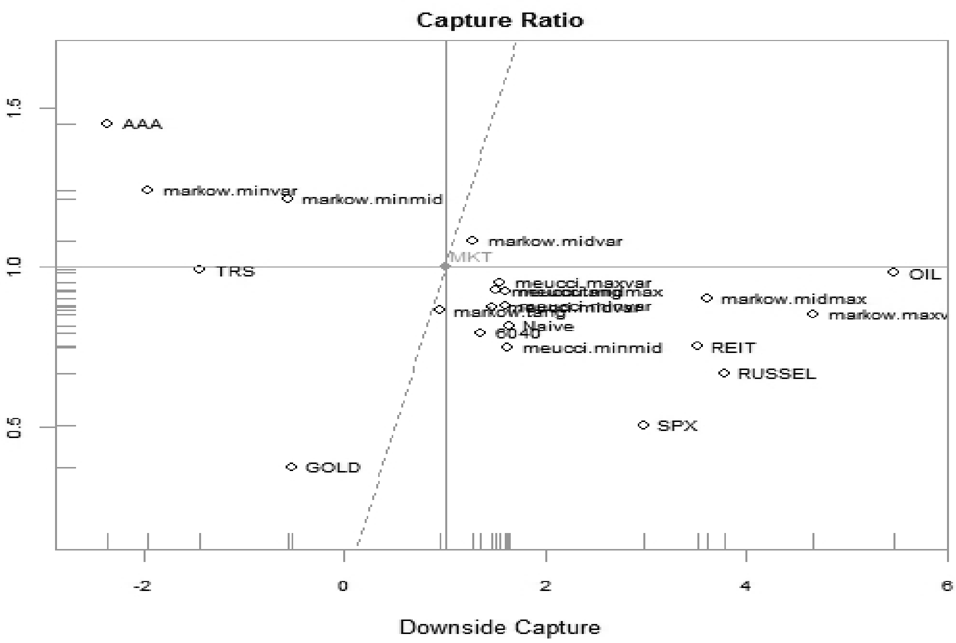


Figure 8. **Capture Ratio of returns**

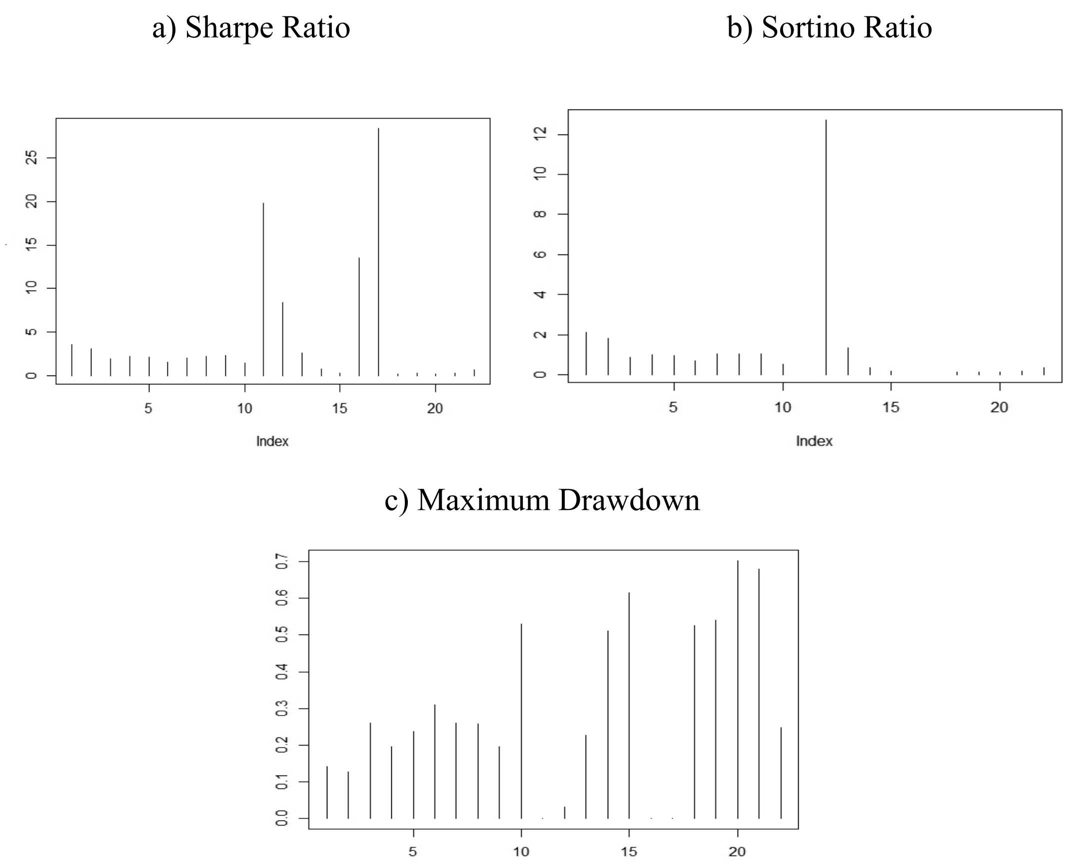
Table 1. Portfolio ratios.

Portfolio	Sharpe Ratio	Sortino Ratio	Maximum Drawdown
Market Portfolio	3.56	2.13	0.14
60–40 Equity — Bond Portfolio	3.04	1.81	0.13
Naive diversification	1.89	0.85	0.26
Meucci tangent portfolio	2.20	1.00	0.20
Meucci minimum risk	2.11	0.94	0.24
Meucci min- mid risk	1.52	0.67	0.31
Meucci medium risk	2.07	1.01	0.26
Meucci mid-max risk	2.17	1.04	0.26
Meucci maximum risk	2.27	1.03	0.20
Markowitz tangent portfolio	1.49	0.52	0.53
Markowitz minimum risk	19.80	Infinity	0.00
Markowitz min- mid risk	8.42	12.74	0.03
Markowitz medium risk	2.64	1.34	0.23
Markowitz mid-max risk	0.72	0.32	0.51
Markowitz maximum risk	0.30	0.16	0.62
10-year Treasury Notes	13.55	Infinity	0.00
Moody's AAA Corporate Bond	28.44	Infinity	0.00
S&P 500	0.23	0.12	0.53
Russell 2000	0.25	0.14	0.54
Oil Futures	0.19	0.14	0.70
REIT	0.30	0.17	0.68
Gold	0.72	0.36	0.25

3. CONCLUSION

The purpose of the research was to test the model of portfolio optimization under the views generated by Ensemble Learning algorithms. For generating such views Random Forest Ensemble Learning algorithm was used.

We made our analysis for the period from 1990 till 2013 for such asset classes as S&P 500, Russell 2000, 10-years Treasury Notes, AAA Moody's Corporate Bonds. Random Forest model was constructed by learning from data for the period from 1990 to 2006. Testing period of the Random Forest is from 2006 till 2013. The Random Forest was based on sixty fundamental, technical and sentiment factors. The analysis of variables for their ability of explanation of expected returns was made. Non-important variables were eliminated and the Ensemble Learning model generated the expected returns for each asset class taking into account only significant variables. Forecast was made at monthly asset return for each asset class. The views obtained from the Random Forest model became the input variables for generating the posterior distribution of returns. Meucci portfolio optimization was made on posterior distribution of the returns and efficiency frontier is the result of this optimization.



1	Market portfolio	12	Markowitz min – mid risk
2	60-40 portfolio	13	Markowitz mid risk
3	Naïve diversification	14	Markowitz mid-max risk
4	Meucci tangency	15	Markowitz max risk
5	Meucci min-risk	16	10 years treasury Notes,
6	Meucci min – mid risk	17	Moody's AAA corporate bond
7	Meucci mid risk	18	S&P 500
8	Meucci mid-max risk	19	RUSSELL 2000
9	Meucci max risk	20	OIL
10	Markowitz tangency	21	REIT
11	Markowitz min-risk	22	GOLD

Figure 9. **Portfolio ratios.**

For evaluating the performance of Meucci optimization under the Random Forest views we made comparative analysis for different optimization frameworks, such as Markowitz optimization, Naive diversification, 60–40 Equity-Bonds investment, Market portfolio. For this purpose we analysed six portfolios obtained from Meucci optimization with different risk level: tangency portfolio, low-risk portfolio, min-mid risk portfolio, middle risk portfolio, mid-max risk portfolio and maximum risk portfolio. Markowitz portfolios considered for analysis have the same risk level as Meucci portfolios.

Meucci portfolio optimization framework under the Random Forest views provides highly-diversified portfolio. Markowitz optimization produces highly concentrated portfolio, for all analyzed period it makes allocation between two asset classes.

We evaluated the performance of optimization by analyzing the Sharpe Ratio, Sortino Ratio and Maximum Drawdown coefficient for portfolios.

Both Meucci and Markowitz optimization beats classic "naive" and 60–40 approaches by almost all measures.

For low-risk tolerance portfolio Markowitz optimization provides better results according to Sharpe and Sortino Ratios and Maximum Drawdown measure.

For high-risk tolerance portfolios, on the contrary, Meucci optimization provides better results according to Sharpe Ratio, Sortino Ratio and Maximum Drawdown coefficient. Moreover, mentioned measures of

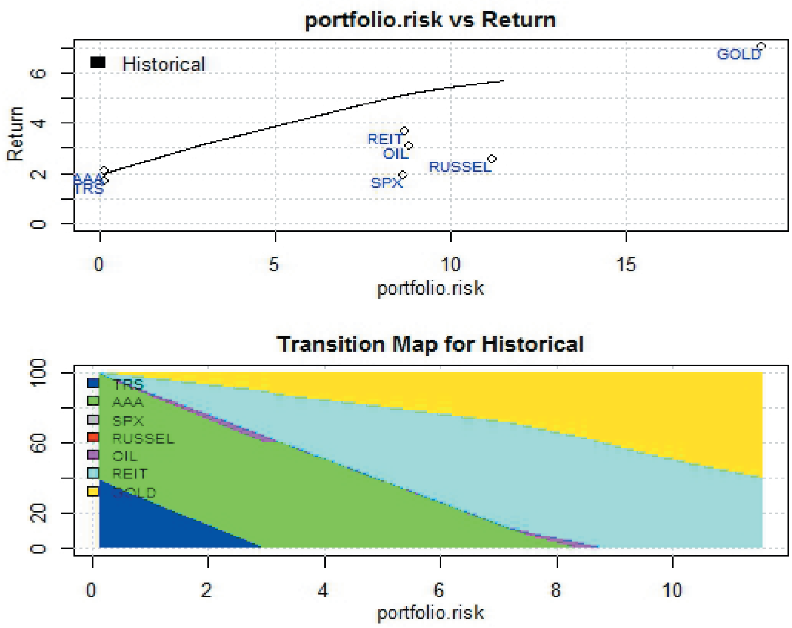


Figure 10. **Efficient frontier of Markowitz optimization.**

Meucci-generated portfolios are not significantly different across risk preferences. That means that while Meucci frontier consists of portfolios with various expected (and realized) risk and return, average payout of each portfolio historical return to historical risk taken (or risk adjusted-return) converges to some market constant, equal for all portfolios. We attribute this to relatively higher level of robustness of Meucci approach as compared to Markowitz approach.

The ratios for Markowitz optimization differ significantly for different levels of risk. Higher absolute performance of Markowitz portfolios could be attributed to the following fact. We make our backtest for the period from 2006 till 2013, and for analyzed period performance of equities was poor. Most Markowitz portfolios avoid investing in equities, which could be explained by usual non-intuitiveness flaws of Markowitz approach (i. e., Markowitz usually invests in two less correlated assets and ignores all others, see Figure 10). Consequently, less exposed to dangerous in 2006–2009 equities, Markowitz portfolios exhibit less drawdowns, less standard deviations and seemingly less risk in general. However this might be just statistical artifact — on longer period well-diversified portfolio would always win.

Meucci portfolio almost always would try to use as wide selection of assets as possible. That makes it more exposed to equity risks of 2007–2009. For better understanding the performance of Meucci optimization future analysis should be applied during economy's healthy period.

The application of Ensemble Learning algorithms for views generation is important topic which needs deeper analysis. Other methods of Ensemble Learning which could be applied for views generation, such as boosting and multi strategy ensembles, stays out of this research. Future research should be done in this sphere for improving the accuracy of predicted returns by Ensemble Learning algorithms.

REFERENCES

Black, F., Litterman, R. (1992), "Global Portfolio Optimization", *Financial Analysts Journal,* 48 (5), 28–43.

Breiman, L. (2001), "Random forests", *Machine Learning*, 45 (1), 5–32.

Chih-Fong Tsaia, Yuah-Chiao Linc, David C. Yenb, Yan-Min Chen (2011), "Predicting stock returns by classifier ensembles", *Applied Soft Computing*, 11, 2452–2459.

Hansen, L.K., Salamon, P. (1990), "Neural network *ensembles", IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12 (10), 993–1001.

Hinsz, V., Tindale, R., Vollrath, D. (1997), "The emerging conceptualization of groups as information processors", *Psychological Bulletin,* 121 (1), 43–64.

Myoung-Jong Kim, Sung-Hwan Min, Ingoo Han (2005), "An evolutionary approach to the combination of multiple classifiers to predict a stock price index", *Expert Systems With Applications — ESWA,* 31 (2), 241–247.

Se-Hak Chuna, Yoon-Joo Park (2005), "Dynamic adaptive ensemble case-based reasoning: application to stock market prediction", *Expert Systems with Applications: An International Journal*, 28 (3), 435–443.

Tae-Hwy Lee, Yang Yang (2005), "Bagging binary and quantile predictors for time series", *Journal of Econometrics,* 135 (1–2), 465–497.

Yuehui Chen, Bo Yang, Ajith Abraham (2007), "Flexible Neural Trees Ensemble for Stock Index Modeling", *Neurocomputing* 70 (4–6), 697–703.