Schedae Informaticae Vol. 25 (2016): 177–188
doi: 10.4467/20838476SI.16.014.6195

tfml 2017
theoretical foundations
of machine learning, Kraków

Portfolio Inputs Selection from Imprecise Training Data

Sarunas RAUDYS[1], Aistis RAUDYS[1], Zidrina PABARSKAITE[1],
Gene BIZIULEVICIENE[1,2]
[1]Department of Mathematics and Informatics, Vilnius University
Didlaukio 47, 08303, Vilnius, Lithuania
e-mail: *sarunas.raudys@mif.vu.lt*
[2]State Research Institute Centre for Innovative Medicine
Zygimantu 9, 01102, Vilnius, Lithuania

Abstract. This paper explores very acute problem of portfolio secondary overfitting. We examined the financial portfolio inputs random selection optimization model and derived the equation to calculate the mean Sharpe ratio in dependence of the number of portfolio inputs, the sample size L used to estimate Sharpe ratios of each particular subset of inputs and the number of times the portfolio inputs were generated randomly. It was demonstrated that with the increase in portfolio complexity, and complexity of optimization procedure we can observe the over-fitting phenomena. Theoretically based conclusions were confirmed by experiments with artificial and real world 60,000-dimensional 12 years financial data.

Keywords: Complexity, financial portfolio, overfitting, sample size, variable selection

1. Introduction.

Data mining methods are slowly making their way into finance, where trading is mainly dominated by econometric and statistical models. The same is with financial portfolio construction. Markowitz mean variance portfolio optimization was proposed

Received: 11 December 2016 / Accepted: 30 December 2016

many years ago and many practitioners still use it in original form to create (learn from data) optimal portfolios [1, 2]. Mean variance portfolio optimization (MVO) is typically used to construct portfolios from various investments. For example, how much stocks bonds and real estate one needs to have in its portfolio to achieve best risk reward ratio? The quality of the portfolio is typically measured by the *Sharpe ratio* (Sh) [2]. This ratio is a mean of the profits divided by the variance (standard deviation). It constitutes how much you earned and with what risk. Very few investigate nonlinear methods offered by machine learning community.

The MVO works with any time series of profit and loses (PNL). So people use it with artificial investments such as generated by automated trading systems. Automated trading is known as algorithmic trading, systematic trading and other names. It is the process where human puts his investments knowledge into the computer program and allows the program to make buy and sell decisions automatically. It varies by types, trading frequency and strategies that are used. The trading firms can employ many potential trading systems. Each trading strategy (TS) can be run in simulated mode and out of this simulation is the series of PNLs. These series correspond to the success of the trading systems to generate profits. The question is what algorithms to trade together to maximize profitability and minimize the risk. Such time series can be used by the MVO engine to calculate the best portfolio best set of trading strategies. Numerous factors influence this process.

Complex portfolio design rules having too large number of inputs for relative short learning sequences often lead to overfitting. It is very easy to get good results in simulations with training data, but notoriously difficult on the unseen data. Main factors that are affecting the overfitting are: the training set size, a number of portfolio inputs, inexact estimation of means values and correlations of the returns [3]. In view of that, it is also very important to verify strategies in out of sample manner and select such methods that will work well on the unseen data.

In present-day tasks we face extremely large number (say, $N = 60{,}000$) of trading strategies and need to construct an investment portfolio for trading during short future time interval. Obviously, we cannot include all N systems into the portfolio. Therefore, we are obliged to choose much smaller subset of the best systems and use simpler portfolio design methods [3]. *The simplest portfolio* is an equal weighted rule where one weights all selected investments equally. This non-trainable portfolio is called, $1/N$, or Naïve portfolio. Often it outperforms more sophisticated methods [4, 5].

Machine learning has numerous methods that allow to deal with imprecise data and to perform feature selection or extraction. Many methods select the "best" subset of N_b ($N_b \ll N$) trading strategies (TS) are suggested in the literature [6, 7].

The simplest way to generate N_b - dimensional subset is to sort N trading strategies according to *sample estimates* of the Sharpe ratio, $\widehat{Sh}$. In this approach, one selects N_b of them having the highest $\widehat{Sh}$ values (method **A**). Sadly, this method ignores correlations. Sophisticated way of the best subset selection is forward selection Comgen procedure [8] that takes into account the correlations (method **B**). It makes series of locally optimal solutions and hopes that it will lead to a near global optimum solution. The benefit of this system is in its simplicity and granularity, as it virtually creates integer portfolio weights $(0, 1, 2, \dots)$ that can be traded straight away.

An important alternative is *a random selection* (method **C**) where one generates

m independent random subsets composed of N_b TSs. One estimates Sharpe ratios of m subsets, and selects the best subset having the highest $\widehat{Sh}$ value. An advantage of this method is *a possibility to analyze the accuracy of the best selection procedure theoretically*. Moreover, this way allows taking into account the correlations and frequently leads to selection of good subset.

Analysis of diverse portfolio design schemas suggested by multiple authors showed that majority of them does not hold out-of-sample scheme [9]. To obtain reliable results, in selection of the best input subset or the portfolio design strategy one needs to use independent validation data set. Due to finite size, the validation data is also imprecise. Hence, *selection of the TSs subset is inexact*. In such circumstances, use of more complex selection algorithm or an increase in the size of the TSs subset, N_b, often does not lead towards the desirable result. It is worth noting that typical MVO assumes that correlation, variance and profitability of the time series will remain constant. Notoriously it changes and changes a lot. Therefore, in the portfolio input and design scheme selection we face notable adaptation to validation set (secondary overfitting).

To our knowledge the *secondary overfitting effect* was never considered in the portfolio design literature. An objective of the present paper is analytic, numerical and experimental clarification of reasons of this important for the practitioner phenomenon and choosing for research directions allowing to overcome this difficulty.

2. Theoretical analysis of accuracy.

An objective of the present section is to obtain an analytical formula to calculate a mean value of true Sharpe ratio when random selection procedure $\boldsymbol{C}$ is used to learn (find) the "best subset" of TSs. To examine the accuracy, one needs to define a distribution density function of true Sharpe ratio values $f_t(Sh)$, and conditional density of estimates, $f_c\left(\widehat{Sh} \mid Sh\right)$. To simplify theory and numerical analysis, we assume true values, Sh, and *estimates*, $\widehat{Sh}$, can take only discrete values, $Sh_1, Sh_2, \ldots, Sh_A$, and $\widehat{Sh}_1, \widehat{Sh}_2, \ldots, \widehat{Sh}_B$. If numbers A and B are sufficiently large, this simplification is not restrictive. Let the elements of both vectors are ranked in an increasing way. In the discrete model, instead of probability densities $f_t\,(Sh)$, and $f_c\left(\widehat{Sh} \mid Sh\right)$, we deal with probabilities of discrete values

$$P_{true}\,(Sh = Sh_i) = P_{true\,i}, \quad (i = 1, 2, \ldots, A)\,, \tag{1}$$

$$P_{cond}\left(\widehat{Sh} = \widehat{Sh}^{\,j} \mid Sh = Sh_i\right) = P_i^{cj}, \quad (i = 1, 2, \ldots, A; j = 1, 2, \ldots, B)\,, \tag{2}$$

where P stands for the probability.

To investigate relations between the sample size and accuracy analytically, we need to choose models $P_{true}(Sh = Sh_i)$ and $P_{cond}\left(\widehat{Sh} = \widehat{Sh}^j \mid Sh = Sh_i\right)$. To calculate the mean Sharpe ratio, $E\,(Sh)$, we need to derive two expressions:

1. conditional probabilities $P_{cond}\left(\widehat{Sh} = \widehat{Sh}^j \mid Sh = Sh_i\right)$ and
2. probabilities of the maximal values of m estimates $\widehat{Sh}_1, \widehat{Sh}_2, \ldots, \widehat{Sh}_m$ (here subscript indicates a serial number the of N_b - dimensional subset of TSs'). Note, each estimate, $\widehat{Sh}_l$, can get any of B values defined in Eq. 2.

Without loosing generality, we normalize values of Sh and $\widehat{Sh}$ to have them varying in interval $[0, 1]$. According to the theory of probabilities, a joint probability of two dimensional vector (Sh, $\widehat{Sh}$)

$$P_{joint}\left(\widehat{Sh} = \widehat{Sh}^j,\, Sh = Sh_i\right) = P_i^{cj} \times P_{true\,i} = P_{ij}^{joint}. \tag{3}$$

Then *unconditional* probability

$$\begin{aligned} P_{ucond}(\widehat{Sh} = \widehat{Sh}^j) = \sum_{i=1}^{A} (P_i^{cj}\left(\widehat{Sh} = \widehat{Sh}^j,\, Sh = Sh_i\right) \times P_{true}\,(Sh = Sh_i) = \\ \sum_{i=1}^{A}\left(P_{ij}^{joint} \times P_{true\,i}\right) = P_{uc}^j. \end{aligned} \tag{4}$$

Subsequently, *conditional* probabilities can be expressed as

$$P_{cond}\left(Sh_l = Sh_i \mid \widehat{Sh} = \widehat{Sh}_l^j\right) = \frac{P_{joint}\left(\widehat{Sh}_l = \widehat{Sh}^j,\, Sh_l = Sh_i\right)}{P_{ucond}(\widehat{Sh}_l = \widehat{Sh}^j)} = P_{ci}^j, \tag{5}$$

where the subscript index l means "any of $1, 2, \ldots, m$".
According to definition of the maximal values their probabilities can be expressed as

$$\begin{aligned} P_{cond}\left(\widehat{Sh}_{maximal} = \widehat{Sh}^j\right) = & P\left(\widehat{Sh}_1 < \widehat{Sh}^{j+1},\, \widehat{Sh}_2 < \widehat{Sh}^{j+1}, \ldots, \widehat{Sh}_m < \widehat{Sh}^{j+1}\right) - \\ & P\left(\widehat{Sh}_1 < \widehat{Sh}^j,\, \widehat{Sh}_2 < \widehat{Sh}^j, \ldots, \widehat{Sh}_m < \widehat{Sh}^j\right). \end{aligned}$$

In the random search selection, the probabilities $\widehat{Sh}_1, \widehat{Sh}_2, \ldots, \widehat{Sh}_m$ are independent. Thus,

$$P\left(\widehat{Sh}_{maximal} = \widehat{Sh}^j\right) = P\left(\widehat{Sh}_l < \widehat{Sh}^{j+1}\right)^m - P\left(\widehat{Sh}_l < \widehat{Sh}^j\right)^m \tag{6}$$

where $P\left(\widehat{Sh}_l < \widehat{Sh}^j\right) = \sum_{l=1}^{j-1} P_{uc}^l$.

As a result

$$P\left(\widehat{Sh}_{maximal} = \widehat{Sh}^{j}\right) = \left(\sum_{l=1}^{j} P_{uc}^{l}\right)^{m} - \left(\sum_{l=1}^{j-1} P_{uc}^{l}\right)^{m} \tag{7}$$

Then the mean value of the expected Sharpe ratio after the random selection procedure, $E(Sh)$, can be calculated from Equations (5) and (7)

$$E(Sh) = \sum_{i=1}^{A} Sh_i \times \sum_{j=1}^{B} P_{cond}\left(Sh_l = Sh_i | \widehat{Sh}_l = \widehat{Sh}^{j}\right) \times P\left(\widehat{Sh}_{max} = \widehat{Sh}^{j}\right) = \sum_{i=1}^{A} Sh_i \times \sum_{j=1}^{B} P_{ci}^{j} \times \left(\left(\sum_{l=1}^{j} P_{uc}^{j}\right)^{m} - \left(\sum_{l=1}^{j-1} P_{uc}^{j}\right)^{m}\right) \tag{8}$$

3. Numerical analysis of the two-dimensional Beta distribution model

Eq. (8) does not allow seeing a relationship between the decreasing of the Sharpe ratio due inexact selection of the best subset of trading strategies in an explicit way. It can be done numerically. To see the relationship of $E(Sh)$ and *validation set size, L, the portfolio inputs selection algorithms complexity parameters*, m, N, N_b, and sets $P_{true\,i}, P_i^{cj} (i = 1, 2, \ldots, A; j = 1, 2, \ldots, B)$ one needs to define them. A simple way to fulfill this requirement is to assume values $P_{true\,i}, (i = 1, 2, \ldots, A)$ to be calculated from Beta density

$$P_{true\,i}, = \gamma Sh^{\alpha}(1 - Sh)^{\beta} \tag{9}$$

where α and β are shape parameters and coefficient γ is chosen from requirement $\sum_{i=1}^{A} Sh_i = 1$. By simple scaling of two extra parameters the Sharpe ratio value can be made to vary in an arbitrary interval. Then we would have a generalized Beta distribution.

We assume conditional probabilities $P_{cond}\left(\widehat{Sh} = \widehat{Sh}^{j} \mid Sh = Sh_i\right)$ are defined by Beta distribution with parameters γ_c, α_c and β_c. In numerical analysis we define values of α_c and β_c according to $mean = Sh_i$ and $variance = V_0/L$ of the Beta distribution density (9)

$$\alpha_c = Sh_i(Sh_i(1 - Sh_i)L/V_0 - 1), \beta_c = \alpha_c(Sh_i^{-1} - 1) \tag{10}$$

where parameter L symbolizes the validation set size used to obtain estimates $\widehat{Sh}^{l}$ and parameter V_0 symbolizes the variance when $L = 1$;

Below we will examine an example with $\alpha = 18.218$, $\beta = 160.968$, $V_0 = 0.006$, $A = B = 1,000$ and $L = 42$ (number 42 symbolizes two months validation days used to estimate Sharpe ratio in automated financial trading). These values were chosen

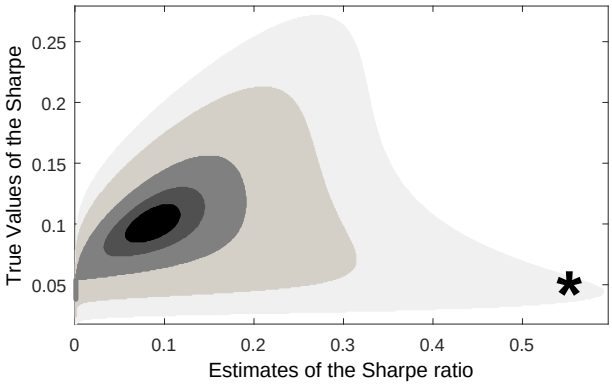

Figure 1. Schema of Two-D Beta distribution density.

while analyzing real world financial data with 50,356 trading strategies. In Figure 1 we present 2-D visualization of distribution of probability $P_{joint}(\widehat{Sh} = \widehat{Sh}^j, Sh = Sh_i)$ in variables $\widehat{Sh}$ (x axis) and Sh (y axis) space. We see the smallest Sharpe values (painted in black, here we have the highest P_{ij}^{joint} values) are much more correlated as the largest $\widehat{Sh}$, Sh values (painted in bright gray, here we have the smallest P_{ij}^{joint} values). Inspection of Figure 1 shows that an increase in the number of m random subsets TSs (entering the rights part of the gray area marked by "*") allows finding subsets characterized by high validation set based Sharpe values. The test estimates (Sh), however, are low in this area.

Calculations according to Eq. 8 confirm the conclusion made from visual analysis of Figure 1. Graphs for $L = 21$, 42 and 63 presented in Figure 2 indicate: 1) with lager value of validation set size, L, we obtain higher true Sharpe values, Sh, 2) the means of Sh are increasing with m at the very beginning, then saturate, and later start decreasing. Thus, training performed by random selection procedure confirms overtraining (peaking) effect known in the data mining and machine learning research [10]. The peaking effect appears earlier when validation set size, L, is small (inspect a curve marked by 21 in Figure 1). When L is extremely small, then $\widehat{Sh}$ and Sh become almost uncorrelated. Then peaking starts almost immediately. Contrary, for larger L values the peaking occurs later and is less expressed (see curve for $L = 64$ in Figure 2).

Note, *parameter m characterizes a complexity of the model selection procedure.* Thus, speaking in general, the theory and graphs in Figure 2 explain peaking relationship between accuracy and complexity of learning portfolio inputs selection algorithm.

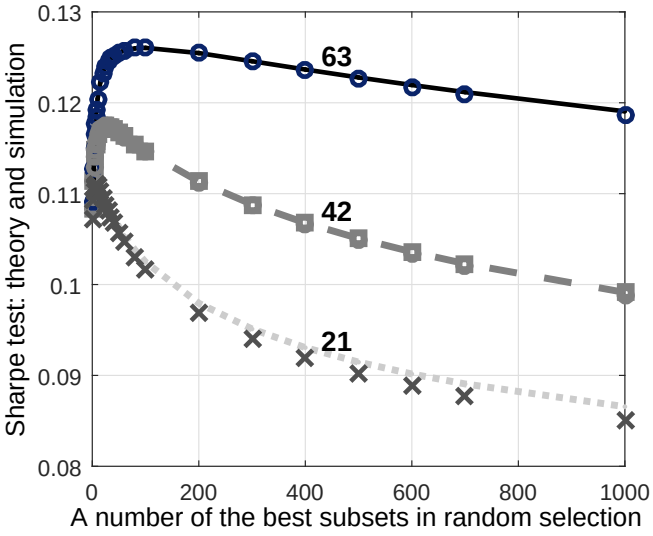

Figure 2. Mean of true Sharpe ratio after the best subset selection.

4. Experiments with 60,314-dimensional financial data

Our analysis is aimed to understand the influence of complexity on the $1/N$ portfolio design in real world financial trading when portfolio dimensionality is extremely high and the size of the data to be used for training (the best TSs subset selection) is relatively small. Therefore, for verification of conclusions presented above we performed experiments with the real world automated trading data.

Trading algorithms comes in many varieties. In our particular case, from prop trading firm we received series (60,314 – dimensions, years 2004 – 2016 data) of the PNLs generated by mean reversion type of strategies. Mean reversion strategies (MRS) are also known as contra trend strategies as they go against the trend. If market is moving upwards at some point MRS can decide that it moved too much and there will be a market correction. So algorithm will short sell and wait for correction. The same is for opposite direction. If the market falls down too much and/or too fast, then the MRS will buy with anticipation of some correction - at least short market movement upwards. If market moves upwards the trading strategy typically will sell and close the position with the profit. Sometimes, especially during market tumor and panic such strategies can generate sharp losses as market moves in one direction for extended period of time. Therefore, it is extremely important such strategies to trade in the portfolio. The risk in portfolio is divided among many strategies and quick loss in one of them makes only small loss in overall portfolio level.

Due to the presence of numerous economy and finance environments changes, in our investigations we used two months data for validation (estimation $\widehat{Sh}^l$ and selection of m best TSs). Later two months data were used for testing the trading

algorithm, i.e. estimation of Sh^l).

In real world trading with sudden environmental changes, we have *two sources of errors* that are affecting difference between evaluation of $\widehat{Sh}$ and true Sh. The first source is finite sizes of learning and validation segments of historical data. This problem was considered in previous section. The second source is the data variation. Investigation of contemporary financial time series showed that Sharpe ratios estimated for two earlier and later two months length data segments are very weakly correlated. Often correlations are even negative. In Figure 3 we present a "successful example" with absence of visible correlation (2+2 months of 2016 spring data). Like in the previous section we are interested in dependence of the mean Sh value on m. Due to random generation the subsets of TSs, single $\widehat{Sh}$ and Sh graphs are notably scattered.

To reduce fluctuations we generated $M = 5{,}0000{,}000$ subsets composed of N_b TSs for each of them. Then for each subset we calculated the estimates $\widehat{Sh}$ (two training months were used) and the true values, Sh (again two extra months were used). As a result, for further analysis we obtained $2 \times 5{,}000{,}000$ dimensional array of the Sharpe ratio values. For each particular m value chosen from *a priori* fixed vector $mm = [50, 100, 200, \ldots, 100{,}000]$ we used binomial coefficients to calculate true mean Sharpe ratio, $\widehat{Sh}(m)$, for all possible, $M!/m!/(M-m)!$, combinations. In Figure 4 we present graphs of the mean values calculated for three TSs subset sizes, $N_b = 20, 40$ and 90.

For each of the TSs subset size, N_b, we observe peaking effect. Curves in Figure 4 remind the curves, presented in Figure 2 calculated for the 2-D Beta model. The most obvious decrease in the performance of the TSs selected is observed for small sizes of the subsets. Small training strategies subsets result smaller portfolio accuracy and smaller correlations between the $\widehat{Sh}$ and, Sh values. Therefore, the peaking effect appears very early (curve for N_b=20 in Figure 4). With reserved increase in N_b, the portfolio performance increases (curve for N_b=40 in Figure 4). Notable increase in size of the TSs, m, increases the complexity of the optimization algorithm. In such a case, we obtain overtraining (overfitting) effect once more: for N_b=90 the true Sharpe ratio graph is notably below as that for N_b=40. This result confirms: too great increase in complexities of the TSs subset size, N_b, and in the algorithm used to select the best subset, m, causes a negative effect. Figure 4 illustrates that for successful employment of the random search procedure (N_b=40, $m \approx 400{,}000$) an average of the true (validation set estimates) the Sharpe ratio values obtained for these optimization procedures parameters, Sh_{best}=3.72, (see Figure 4). It is notably higher value as average of Sh_r ($r = 1, 2, \ldots, 5{,}000{,}000$). For practical application one needs to know optimal values of the optimization procedure parameters (N_b, m, L). Studies in this direction have already begun.

The peaking effect had been noticed in statistical, pattern recognition, data visualization, design of prediction rules, neural networks [10, 11, 12, 13]. We expect that this conclusion can be applicable also for other research and development disciplines where data mining and machine learning are used. Therefore, in dynamic Portfolio design, we need to pay attention both into development of faster optimization methods and ways to determine the optimal complexity of the optimization algorithms.

In the introduction we mentioned two heuristically based algorithms, the independent training strategy selection (**A**) and the feed forward selection algorithm Comgen

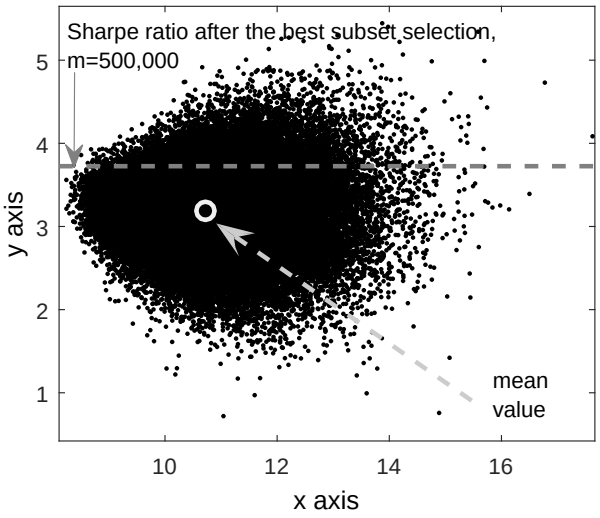

Figure 3. 2-Dscatter of Sharpe training and test Sharpe ratios calculated for M=100,000 subsets.

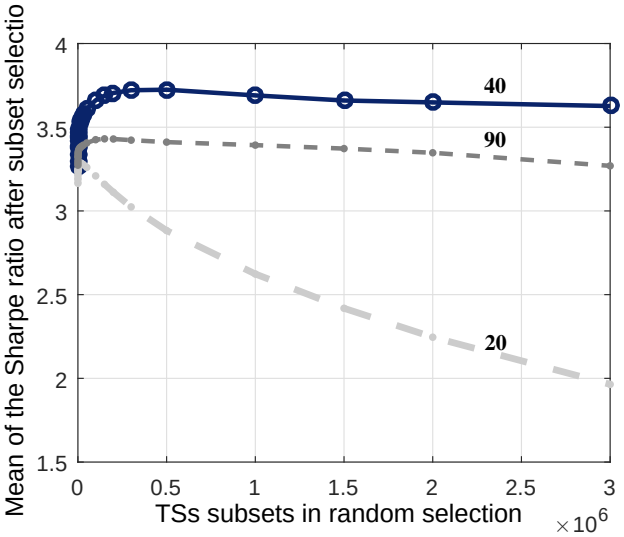

Figure 4. The Sharpe ratio after the random best subset selection.

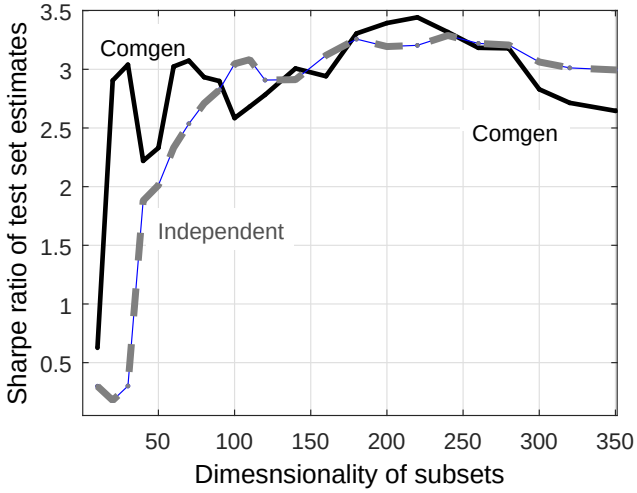

Figure 5. Sharpe ratio of the best TSs where trading systems were ranked individually or selected by the Comgen algorithm.

(**B**). In comparison with random search they work much faster. Their complexity can be characterized by the number of TSs selected, N_b. In Figure 5 we present variation of the test set Sharpe ratio in dependence of dimensionality of the TSs selected. The dependency curves are much more scattered as that in Figure 4, since only short, two month length data sequences were used to calculate $\widehat{Sh}$ and Sh.

Both graphs in Figure 5 exhibit the peaking behavior. For both algorithms the optimal complexity of the Portfolio inputs is found somewhere in the range $200 - 250$ trading strategies. The true Sharpe ratios detected are a bit lower as that obtained by relatively slow random search procedure. In the practical automated daily investment work the random search (algorithm **C**) can be easily applied since modern laptop computers calculate the family of curves similar to that in Figure 4 in 10 - 20 minutes.

5. Concluding remarks

This paper explores the secondary over-fitting effect that is very acute of $1/N$ portfolio design. It is an adaptation to validation data set used to select the best subset of inputs and/or the best algorithm to calculate the portfolio weights. Ignoring this up to now unexplored effect constitutes a big headache for portfolio managers as constructed portfolios do not behave in the way they were supposed to. In theoretical analysis we examined the random portfolio inputs optimization procedure and derived the equation to calculate the mean Sharpe ratio in dependence of (on) the number of portfolio inputs N_b, the validation sample size L used to estimate Sharpe ratios

of each particular subset of inputs and the number, m, of randomly generated N_b-dimensional portfolio inputs from their N-dimensional set. This equation was adapted for practical calculations of the mean Sharpe ratio when both, the probabilities of the true and estimate Sharpe ratios, are calculated from the 2-D Beta distribution model. It was demonstrated that with an increase of portfolio complexity, N_b, and complexity of optimization procedure, m, we can observe the over-fitting phenomenon.

Theoretically based conclusions were confirmed by experiments with high dimensional real world financial data and suggest several recommendations for future research and practical work. Due to the presence of numerous economy and finance environmental changes the 2-D scatters of Sharpe ration evaluated on training and validation data in diverse time periods often show zero or even negative correlations. Consequently, the practitioner should be careful: sometimes even a negligible optimization of the portfolio inputs subset can worsen the result. For that reason the practitioner should examine numerous subsequent in time 2-D Sharpe ratio scatters and be cautious with for the portfolio inputs optimization. Therefore a prudent analysis of changes in preceding historical data is very important. Preliminary experiments demonstrated that paying an attention to validation data size and knowledge about character of possible data changes could lead to novel useful ways of the portfolio management and determination of the portfolio size and parameters of optimization rules.

Acknowledgements

This work was supported by the Research Council of Lithuania (grant MIP-100/2015)

6. References

[1] Markowitz, H.M., *Foundations of portfolio theory.* The journal of finance, 1991, **46**(2), pp. 469–477.

[2] Reilly, F.K., Brown, K.C., *Investment analysis and portfolio management.* Cengage Learning, 2011.

[3] Raudys, S., *Portfolio of automated trading systems: Complexity and learning set size issues.* IEEE transactions on neural networks and learning systems, 2013, **24**(3), pp. 448–459.

[4] DeMiguel, V., Garlappi, L., Uppal, R., *Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?* Review of Financial Studies, 2009, **22**(5), pp. 1915–1953.

[5] Haley, M.R., *Shortfall minimization and the naive (1/n) portfolio: an out-of-sample comparison.* Applied Economics Letters, 2015, pp. 1–4.

[6] Guyon, I., Elisseeff, A., *An introduction to variable and feature selection.* Journal of machine learning research, 2003, **3**(Mar), pp. 1157–1182.

[7] John G H, Kohavi R, P.K., *Irrelevant features and the subset selection problem.* The journal of finance, 1994, pp. 121–129.

[8] Raudys, A., Pabarškaitė, Ž., Discrete portfolio optimisation for large scale systematic trading applications. In: *Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference on*, IEEE, 2012, pp. 1566–1570.

[9] Bailey, D.H., Borwein, J.M., de Prado, M.L., Zhu, Q.J., *Pseudomathematics and financial charlatanism: The effects of backtest over fitting on out-of-sample performance.* Notices of the AMS, 2014, **61**(5), pp. 458–471.

[10] Bradley, P.S., Fayyad, U.M., Mangasarian, O.L., *Mathematical programming for data mining: Formulations and challenges.* INFORMS Journal on Computing, 1999, **11**(3), pp. 217–238.

[11] Jackowski, K., Wozniak, M., *Algorithm of designing compound recognition system on the basis of combining classifiers with simultaneous splitting feature space into competence areas.* Pattern Analysis and Applications, 2009, **12**(4), pp. 415–425.

[12] Tetko, I.V., Livingstone, D.J., Luik, A.I., *Neural network studies. 1. comparison of overfitting and overtraining.* Journal of chemical information and computer sciences, 1995, **35**(5), pp. 826–833.

[13] Raudys, S., *Experts' boasting in trainable fusion rules.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, **25**(9), pp. 1178–1182.