Selecting Best Investment Opportunities from Stock Portfolios Optimized by a Multiobjective Evolutionary Algorithm

Krzysztof Michalak

Department of Information Technologies,
Institute of Business Informatics,
Wroclaw University of Economics, Wroclaw, Poland
krzysztof.michalak@ue.wroc.pl

Presentation Plan

- Introduction
- Construction and evaluation of portfolios
- Multiobjective optimization of portfolios
- Clustering of optimized portfolios
- Analysis of portfolio behaviour
- Portfolio selection strategies
- Experiments and results
- Conclusion & further work

Introduction

- Scenario
  - We want to invest a given sum of money
  - Here we consider investments (timeframe: days, months) as opposed to trading (timeframe: sometimes as short as (milli)seconds)
  - We are interested in achieving high return
  - Investments are not without risk
  - Note: higher returns often entail higher risk

Introduction

- Goal: maximize return, minimize risk
- How to do this: choose your assets wisely… but that's far from trivial
- Constructing portfolios of assets helps
  - we do not put all eggs in one basket
  - anticorrelations mitigate the risk (but lower the return)
- Approaches
  - maximize return for an acceptable level of risk
  - find Pareto-optimal trade-offs

Introduction

- Goal: maximize return, minimize risk
- How to do this: choose your assets wisely… but that's far from trivial
- Constructing portfolios of assets helps
  - we do not put all eggs in one basket
  - anticorrelations mitigate the risk (but
- Approaches
  - maximize return for an acceptable lev
  - find Pareto-optimal trade-offs

how to select a particular portfolio from the Pareto front

?

Construction and evaluation of portfolios

- A number N_s of stocks are available
- A portfolio = a vector $w \in R^{N_s}$
  - each coordinate w_i is a „weight" of i-th stock
  - we require that:

$$\sum_{i=1}^{N_s} w_i = 1$$

Construction and evaluation of portfolios

- The entire portfolio is an asset for which we can calculate a quotation at time *t* :

$$p(t) = \sum_{i=1}^{N_s} w_i q_i(t)$$

where:

$q_i(t)$ - quotation of i - th stock at time t.

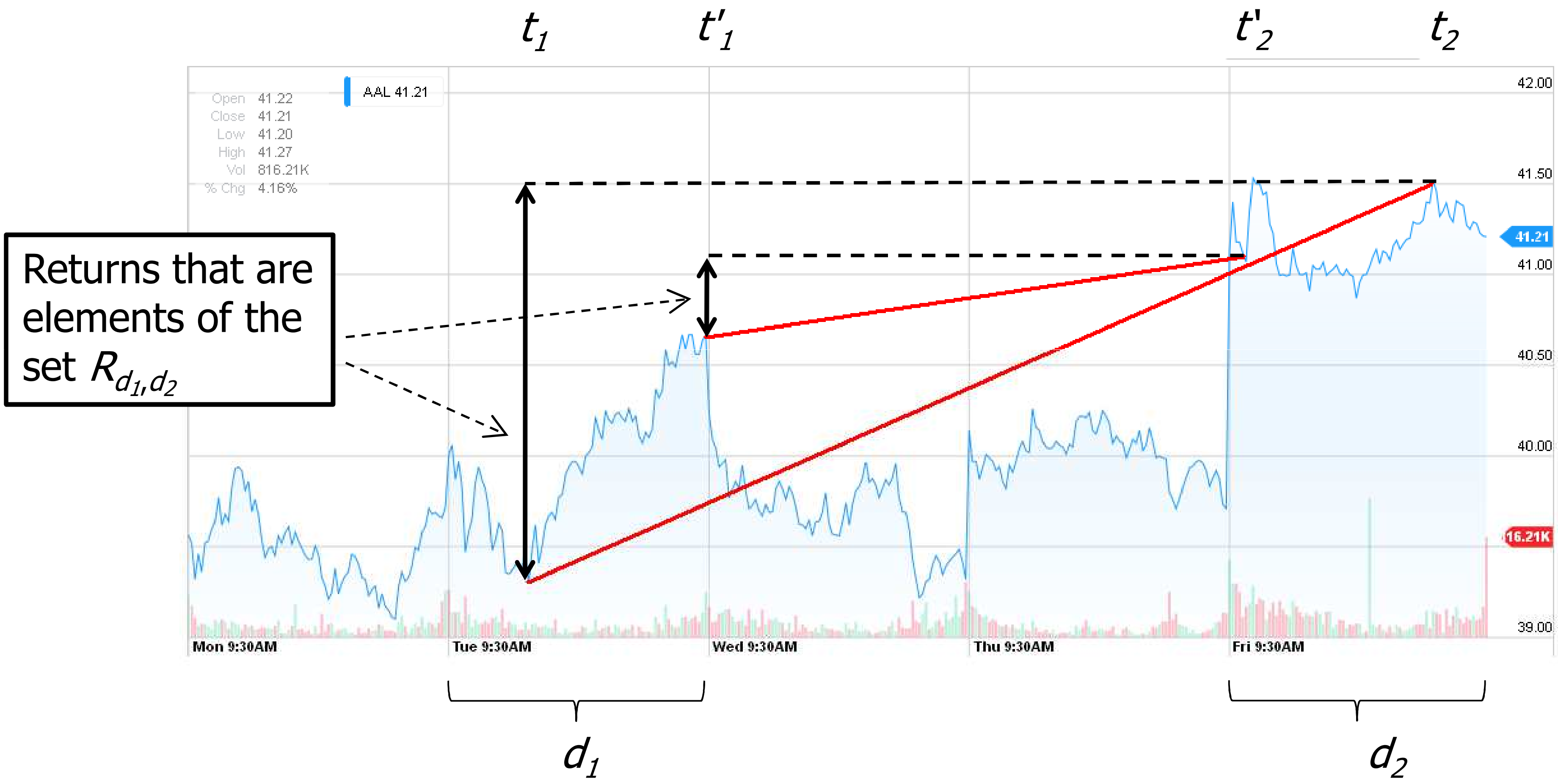
- Then the resolution of quotations of the portfolio would be the same as the resolution of stock quotations (e.g. daily or minute)

Construction and evaluation of portfolios

- In this paper we try to:
  - get some information from minute quotations
  - apply it for long-time investment
  - we want to know what happens when we buy on a day d_1 and sell on a day d_2
  - let R_{d_1,d_2} be a set of all possible returns that could be obtained if buying stocks in portfolio w on a day d_1 and selling on a day d_2 :

$$R_{d_1,d_2} = \left\{ \frac{\sum_{i=1}^{N_s} w_i m_i(t_2)}{\sum_{i=1}^{N_s} w_i m_i(t_1)} : t_1 \in d_1 \wedge t_2 \in d_2 \right\}$$

Construction and evaluation of portfolios



Based on a chart from: *http://finance.yahoo.com/echarts?s=AAL*

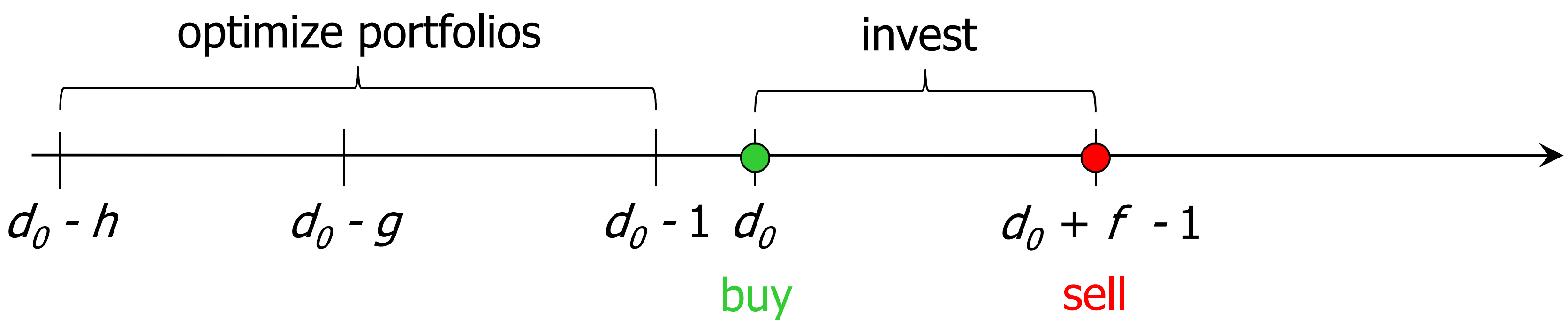
Construction and evaluation of portfolios

- Based on the elements of the set R_{d_1,d_2} we can calculate:
  - average return
  - variance
  - semi-variance - a variance calculated over these elements of the set R_{d_1,d_2} that are below the mean
  - Sharpe's ratio – measures, how well a given investment pays off for the risk taken

$$S = \frac{E\left[R_a - R_b\right]}{\sqrt{\mathrm{var}\left[R_a - R_b\right]}}$$

Optimization of portfolios

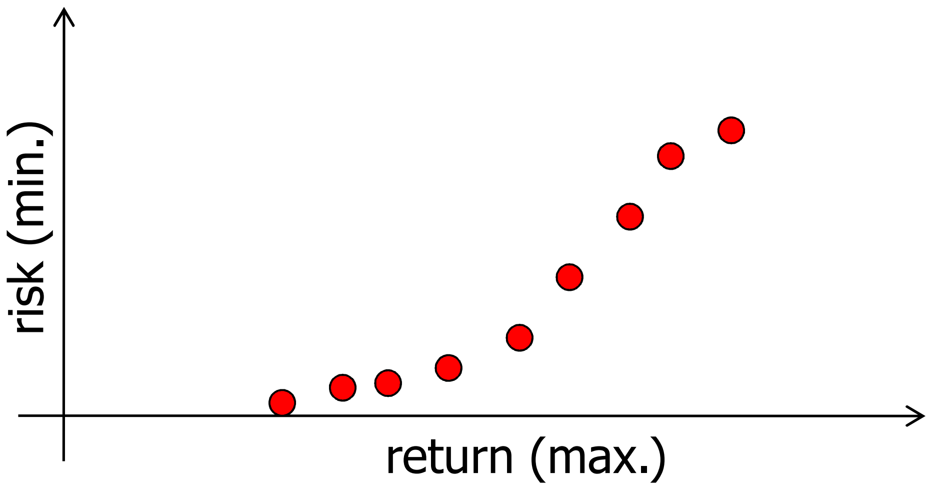
- Assume, that we want to invest for f days, for a period $[d_0, d_0 + f - 1]$
- We optimize portfolios on historical data on the interval $[d_0 - h, d_0 - 1]$, h days in the past



- We perform a multiobjective optimization w.r.t the objectives:
  - the average return
  - one of the risk measures

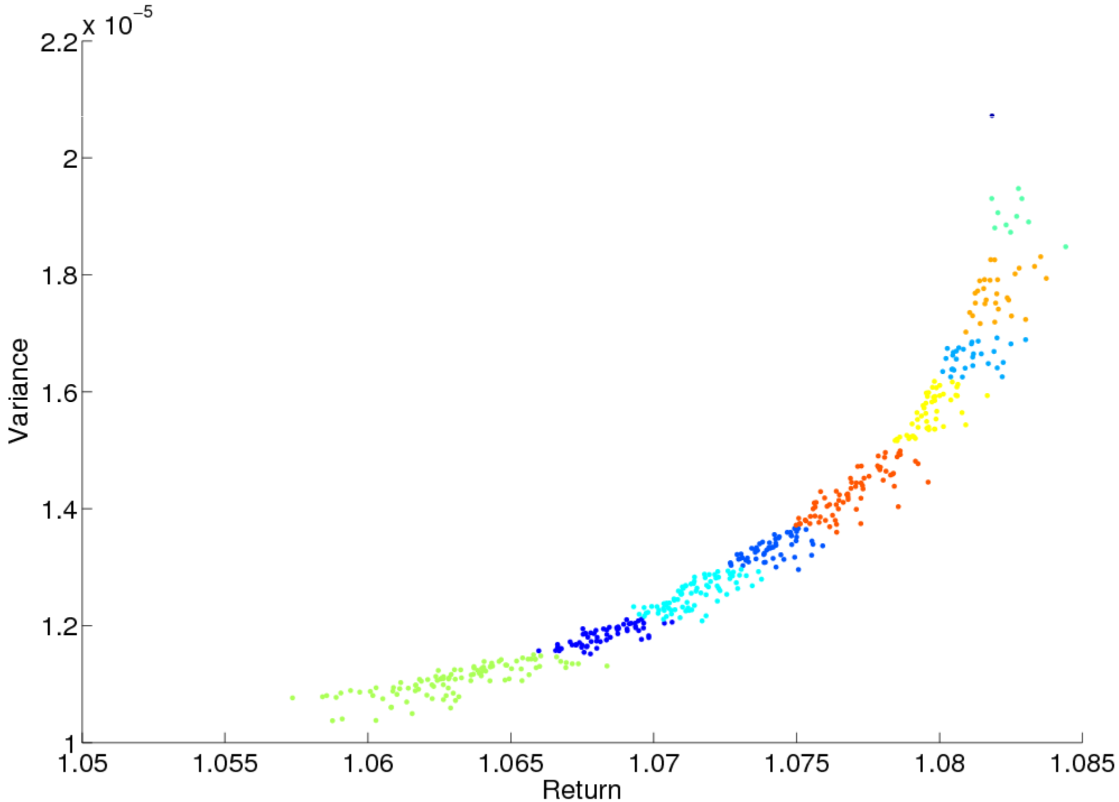
Optimization of portfolios

- NSGA-II was used in this paper
  - it uses non-dominated sorting and binary tournament selection
  - it works well with problems in which the objectives have disproportionate scales
  - Operators: SBX and polynomial mutation
  - Repair procedure to ensure that weights sum to 1
- We get a population of portfolios approximating the PF



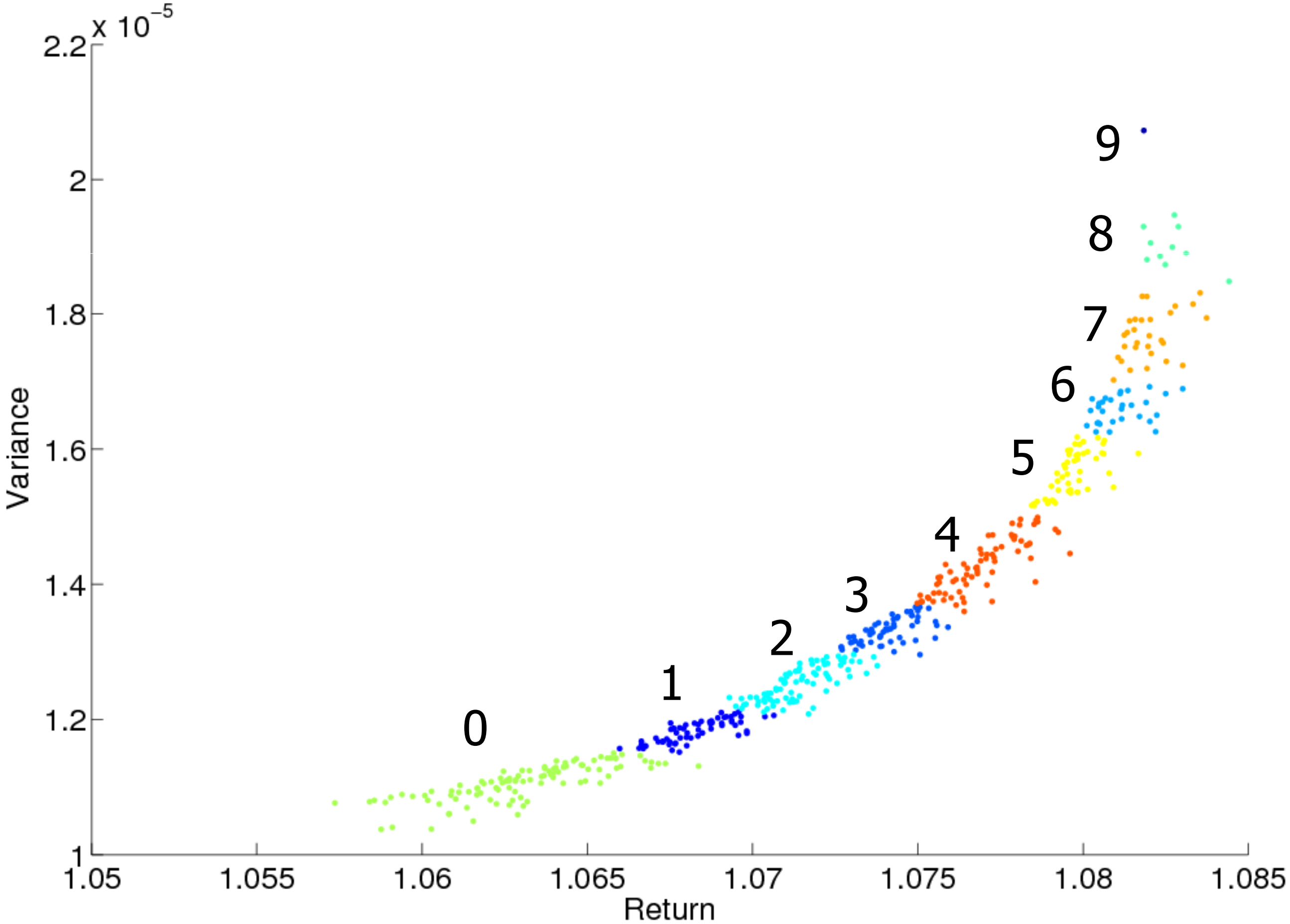
Clustering of portfolios

- Clustering of the solutions in the objective space
- $N_c = 10$ clusters
- Average Linkage Method (ALM)



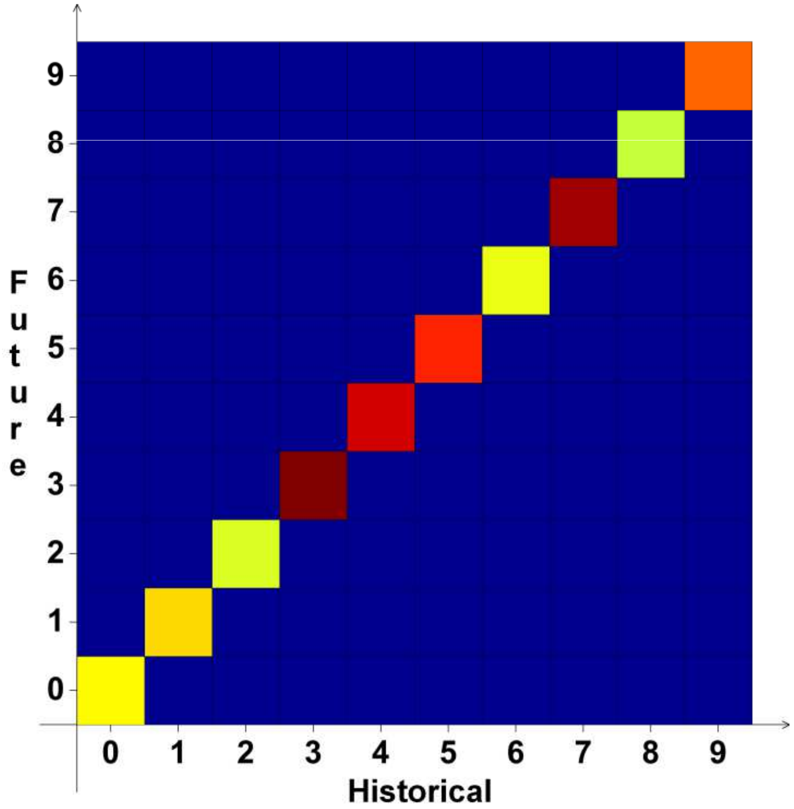
Analysis of portfolio behaviour

- We can order the clusters w.r.t. the increasing average return



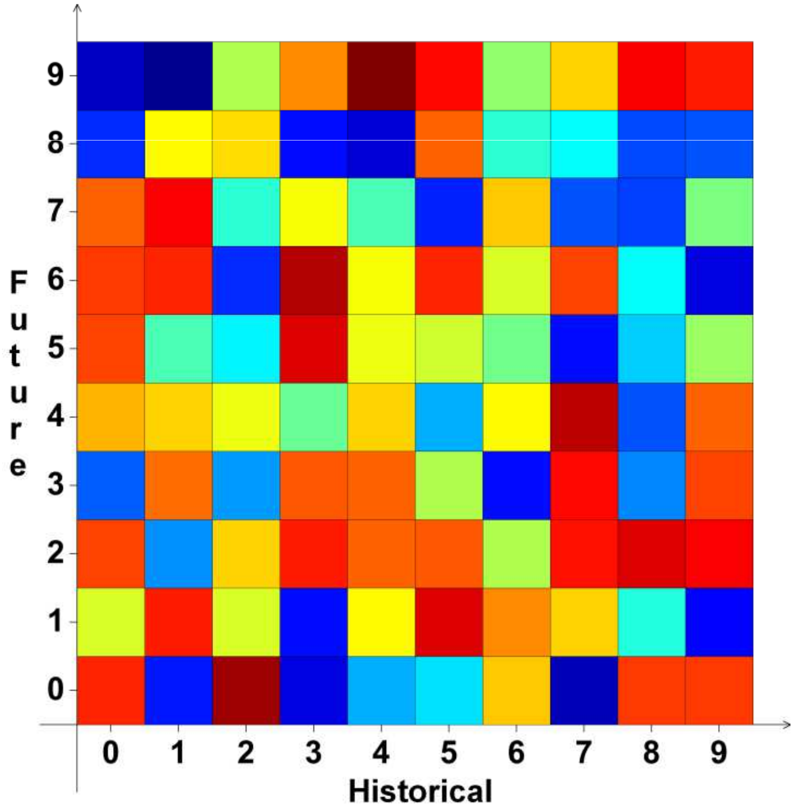
Analysis of portfolio behaviour

- Question: is the order of the clusters the same on future data?
- Can we expect a close correspondence?



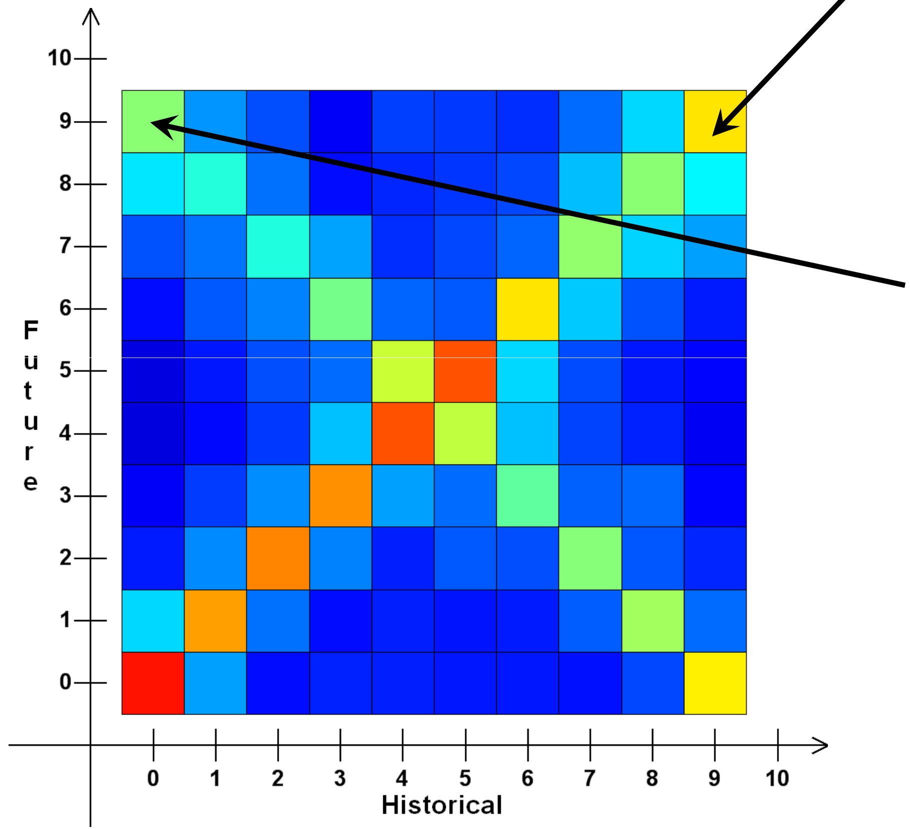
Analysis of portfolio behaviour

- Question: is the order of the clusters the same on future data?
- Or no correspondence at all?



Analysis of portfolio behaviour

- Actual observations



Portfolios with high return (but also high risk) on historical data, earn well in the future …

… but so do those with low return (but also low risk) on historical data

- From the graph we see, that selecting a portfolio properly is not trivial

Portfolio selection strategies

Several strategies of selecting the best porfolio have been proposed and tested in this paper

- C-RET - The centroid is taken from the cluster in which the past returns are highest on average
- C-RISK - The centroid is taken from the cluster in which the past risk measure is lowest on average
- IDX – Decision based on how a stock market index behaved in the last *g* days
  - increased $\Rightarrow$ the centroid is taken from the cluster with the highest average past returns
  - decreased $\Rightarrow$ the centroid is taken from the cluster with the lowest average risk measure
- S-RET - The portfolio which has shown the highest return on historical data interval is chosen.

Experiments

- Data set
  - Minute quotations of 200 stocks from NYSE
  - Time interval: from 2012.07.02 to 2014.10.17
  - 578 trading days (120 weeks)
  - 224,739 minute quotations were recorded for each stock
- Two different lengths of the investment period

  f = 30 and 60 trading days
- Six different lengths of the optimization period

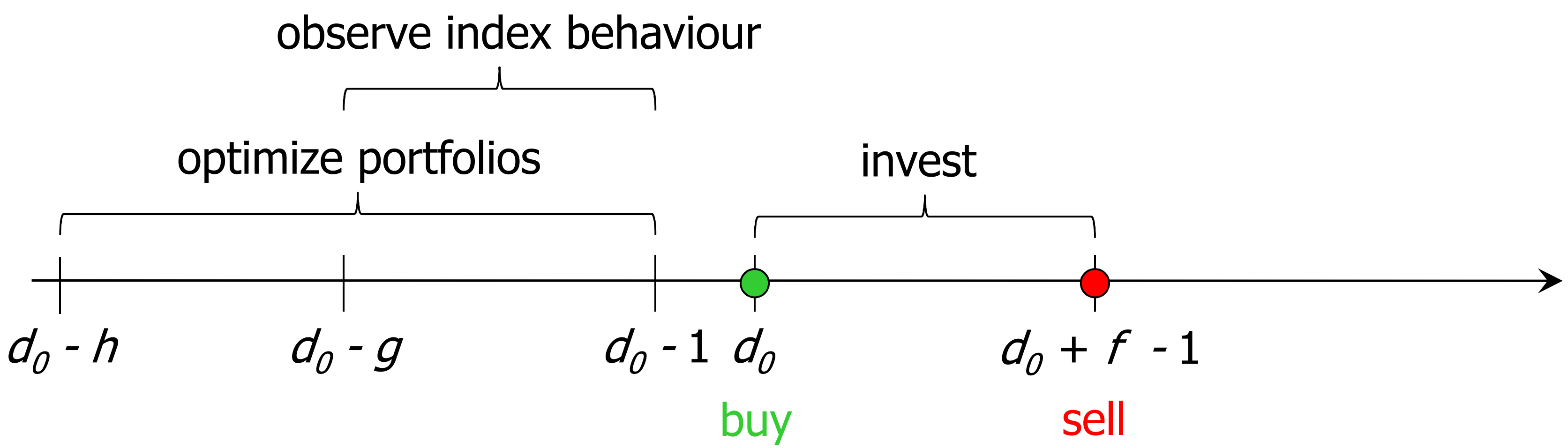
  h = 30, 60, 90, 120, 150 and 180 trading days

Experiments

- Evolutionary algorithm parameters
  - population size N_{pop} = 500
  - number of generations N_{gen} = 50
  - distribution index parameter η = 20 for both the crossover and the mutation

Experiments

- Simulated investments
  - start at day 181 in the data set (to accommodate $h = 180$)
  - optimize portfolios on historical data
  - select the best portfolio using the tested strategy
  - buy the assets, wait $f = 30$ or 60 days, and sell
  - reinvest immediately

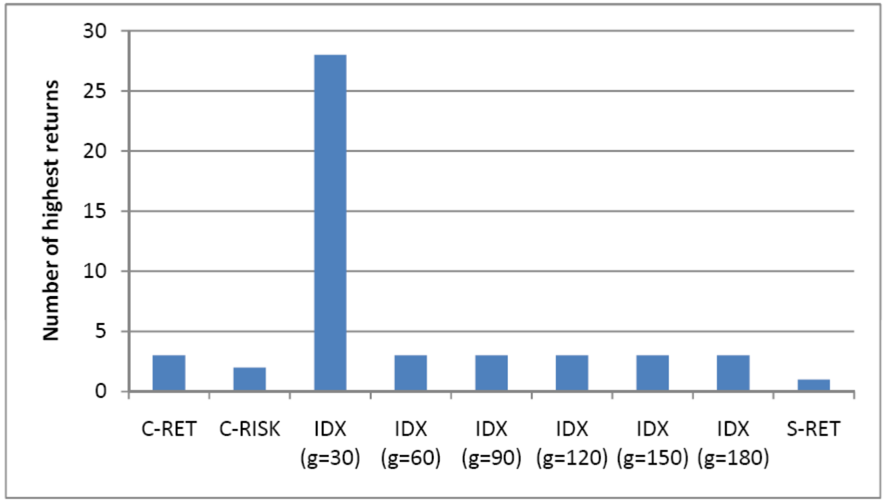


Results

- Detailed results are presented in the paper
- Summary of the results

The best returns obtained for the tested risk measures and for investment interval length of $f = 30$ and 60 trading days

Risk measure	Return	Strategy	h
$f = 30$			
Variance	1.1139	IDX ($g = 30$)	90
Semi-variance	1.1271	C-RISK	90
Sharpe ratio	1.1082	IDX ($g = 30$)	30
$f = 60$			
Variance	1.2150	IDX ($g = 30$)	120
Semi-variance	1.2241	IDX ($g = 30$)	60
Sharpe ratio	1.2106	IDX ($g = 30$)	60



The total number of times each strategy produced the highest return

Conclusion

- Past performance of porfolios is useful, but not in a straightforward way
- When the behaviour of portfolios located in various parts of the Pareto front is analyzed two distinct patterns can be found
  - high return, high risk on historical data ⇒ high return in the future
  - low return, low rist on historical data ⇒ high return in the future
- Based on these observations several investment strategies were proposed and tested

Conclusion

- The best-performing strategy is based on how a stock market index behaved in the last *g* days
  - increased $\Rightarrow$ the centroid is taken from the cluster with the highest average past returns
  - decreased $\Rightarrow$ the centroid is taken from the cluster with the lowest average risk measure

Further work

- Develop adaptive strategies
- Use risk also as a goal
  - other risk measures
  - on what period to measure
- Constrained optimization (e.g. cardinality constraints)
- Optimization of trading rules
  - lower cardinality of the PF

Thank you!

(questions?)