Final project report
Portfolio optimization

Juha Reinikainen

May 30, 2022

1 Problem

Portfolio optimization involves deciding how to use the available investment budget to maximize the total value of the investment and minimize its risk [9]. The investment budget is allocated to assets which can be for example stocks, gold, foreign exchange, real estate, bonds and cryptocurrencies [4]. For simplicity only stocks are considered in this project.

The problem is difficult because there is a large number of possible assets to include in the portfolio and even larger number of ways to divide the budget among them. Investing also has a lot of uncertainty as stock prices are affected by real world events which are hard to capture in the model [2].

2 Data

Data-driven approach to this problem involves predicting expected return and risk based on historical time-series data of stock prices over given time range, possibly years. Stock price can be sampled e.g. daily, weekly, monthly or quaterly. The stocks that are included in the data-set need to be selected.

The stock price data was collected from Yahoo Finance by calling its api with different stocks and merging these into one dataset. For example the weekly prices between 2021-05-09 20:55:46 and 2022-05-09 20:55:46 can be queried using following url.

`https://query1.finance.yahoo.com/v7/finance/download/ADS.DE?period1=1620582946&period2=1652118946&interval=1wk&events=history&includeAdjustedClose=true`

Where perdiod1 and period2 are the start and end times of the range as unix timestamps.

The stocks included in the problem where the stocks in eurostoxx 50 index.

3 Modelling

The problem can be modelled as a three objective optimization problem maximing expected return of the investment, minimizing its risk and maximizing social responsibility.

3.1 Variables

The decision vector consist of proportions of total budget allocated to n stocks $w = (w_1, w_2, ..., w_n)$ and binary variable for each stock which indicates whether the corresponding stock is included in the portfolio $y = (y_1, y_2, ..., y_n)$.

3.2 Constraints

It's assumed that the whole budget is used so the sum of weights should add up to one for the weight where the corresponding boolean flag is 1.

Boundary constraint requires that the weights of each stock is between w_{min} and w_{max}. Maximum limit makes it so that all budget is not allocated to too small number of stocks leading to diverse portfolio. Too small weights typically have little impact on the performance and weak liquidity and can be costly in respect to brokerage fess or monitoring costs [3].

Cardinality constraint requires that the number of stocks included in the portfolio is between some two numbers C_{min} and C_{max}.

3.3 Objectives

Popular way to model these objectives is Markowitz model which is also known as mean-variance model [6]. The risk measure from that model is used. For predicting the prices of individual stocks ARIMA model is used. ARIMA is a popular time series prediction model which also has been used in portfolio optimization [7].

Given time series data of n stocks for time period of 0...T with stock price p(t, i) of stock i at time t. prices of stock i are a series x_i.

$x_i = (p(0, i), p(1, i), ..., p(T, i))$

Prices can be computed into future for wanted number of steps using ARIMA. ARIMA depends of three parameters: p,d,q. p is the order of the auto-regressive model (AR), d is the degree of differencing (I) and q is the order of moving average (MA) model. Since there are 50 models to be trained (for each stock), these parameters are optimized using lbfgs optimizer to find good parameters from range 1-5 for these three parameters.

This gives us a new series. $x_i^{future} = (p(T + 1, i), p(T + 2, i), ..., p(T + steps, i))$

Return of an investment for stock i is calculated by.

$roi(i) = \frac{p(T+steps,i)-p(T,i)}{p(T,i)}$

Expected return for portfolio is calculated as weighted sum of all individual returns.

$er(w, y) = \Sigma_{i=1}^{n}(w_i * y_i * roi(i))$

Picking stocks where the price changes a lot is risky. This problem is increased by picking stocks where this happens similarly. This can be modelled using covariance of roi values [6]. n x n square covariance matrix Cov contrains covariances of each roi(i) values. Given weights and inclusion flags, the risk for portfolio is given by.

$risk(w, y) = (w^T * y)Cov(w * y)$

Third objective included in the problem is social responsibility which can be modelled using environmental, social and governance (ESG) score [1]. Taking social responsibility into consideration can lead to long-term advantages as risk of a company losing face over its actions is common. There are multiple ways to compute the ESG score and multiple organizations that do it. For this project the scores were acquired from Kaggle dataset `https://www.kaggle.com/datasets/debashish311601/esg-scores-and-ratings?resource=download`. The needed 50 'Overall ESG SCORE'-values for the stocks were then gathered.

The overall social responsibility value for a portfolio is calculated as weighted sum of stock esg scores and weight of each stock.

$esg(w, y) = \Sigma_{i=1}^{n}(w_i * y_i * esg_i)$

3.4 Problem

Putting all this together the whole problem is.

$$\begin{aligned} minimize\{risk(w, y), -er(w, y), -esg(w, y)\} \\ \Sigma_{i=1}^{n}(w_i * y_i) = 1 \\ w_{min} \leq w_i \leq w_{max}, i = 1...n \\ C_{min} \leq \Sigma_{i=1}^{n} y_i \leq C_{max} \\ y_i \in \{0, 1\}, i = 1...n \end{aligned} \tag{1}$$

4 Algorithm and settings

NSGA-II algorithm is used to solve the multiobjective constrained mixed-integer problem. The algorithm is run for 1000 generations with population size 100.

For the first n real variables simulated binary crossover and polynomial mutation are used. For the last n binary variables two-point binary crossover and bitflip mutation are used. Crossover probability is set to 1, mutation probability is set to $\frac{1}{50}$. Distribution index for real mutation and crossover are set to 3.

The constraint that the weight need to add up to one can be enforced using repair method [5]. The weights are first clamped to w_{min} w_{max} range. Then weight vector is element-wise multiplied by vector y to remove unselected stocks. Then each weight is divided by sum of all weights. It's easy to see that now weights add up to one. Special case where all weights are zero can be handled by assigning all weights with value $\frac{1}{n}$.

$$\begin{aligned}
&\frac{w_1}{w_1 + w_2 + ... + w_n} + \frac{w_2}{w_1 + w_2 + ... + w_n} + \frac{w_n}{w_1 + w_2 + ... + w_n} \\
&= \frac{w_1 + w_2 + ... + w_n}{w_1 + w_2 + ... + w_n} \\
&= 1
\end{aligned} \tag{2}$$

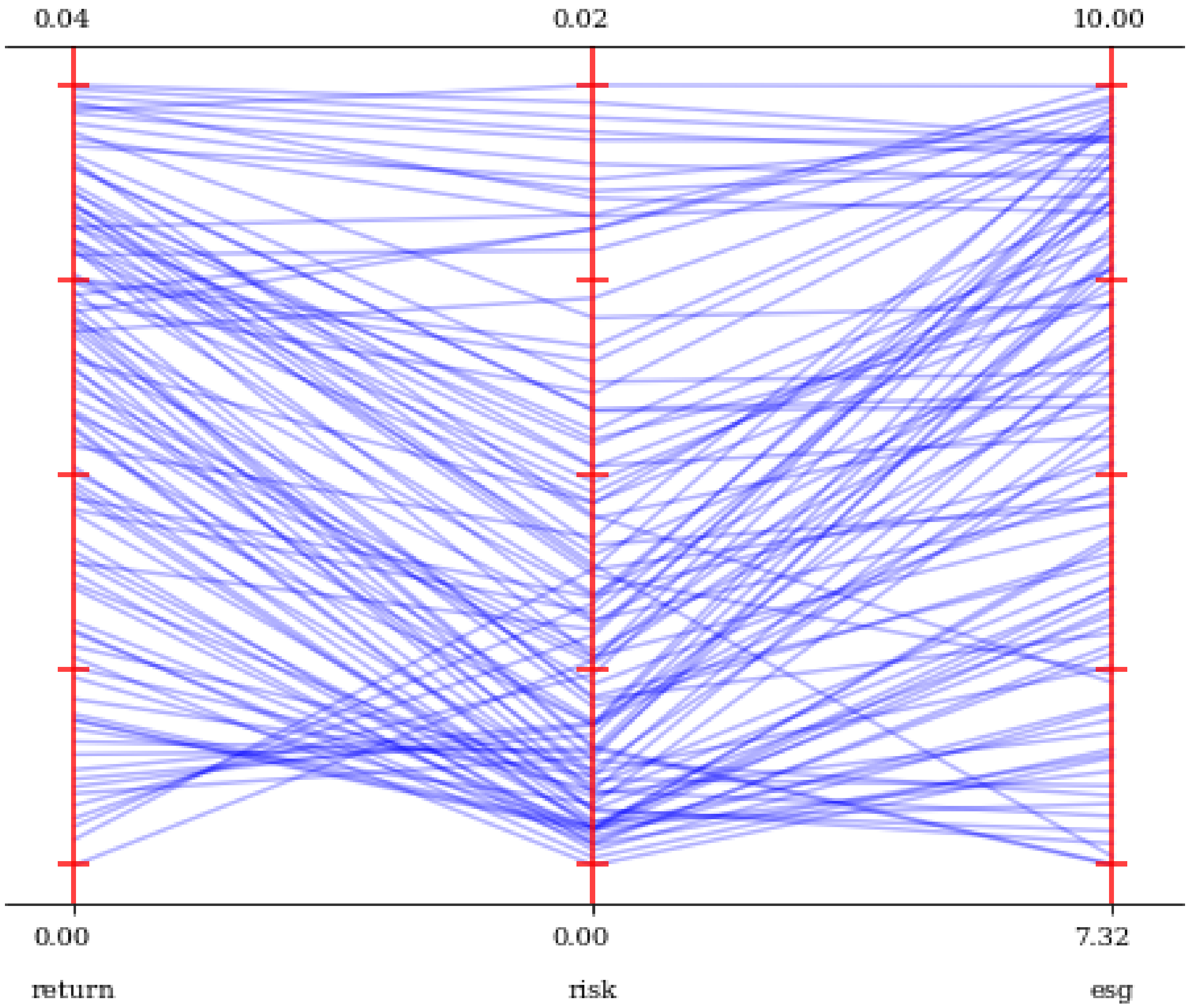
C_{min} was set to 4 and C_{max} to 20 to make sure that the budget is allocated to sufficient amount of stocks. w_{min} was set to 0.01 and w_{max} to 0.8.

The ARIMA models are trained on stock prices data from 1.4.2018 to 1.4.2022. Prediction is done 3 months into future (1.7.2022) with these models to get price prediction. The models are also cross validated using sliding window technique and mean symmetric mean absolute percentage error (smape) score is computed for the models.

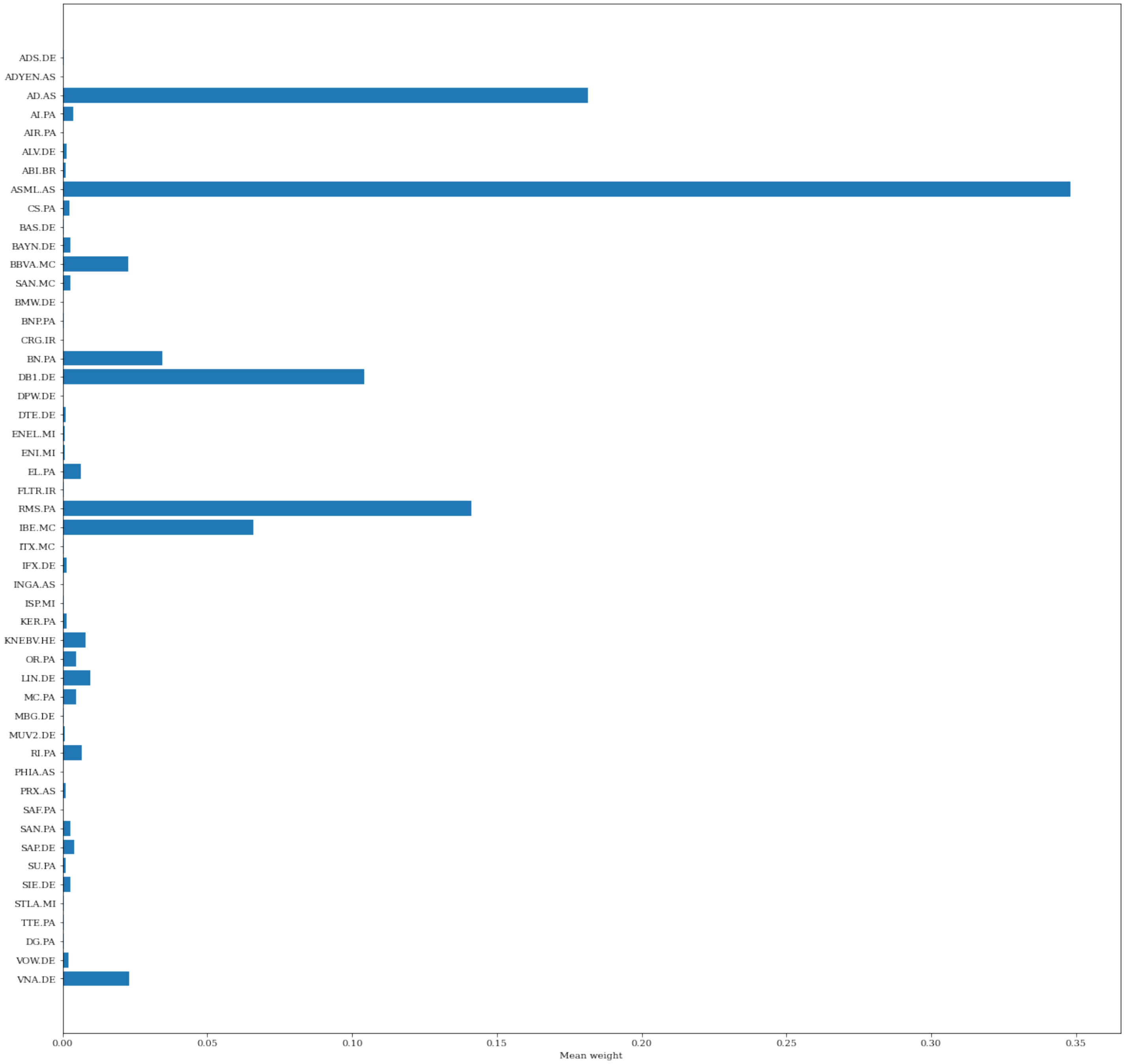
5 Results

The error of the model in the validation was 0.1. After 1000 generations and 100000 function evaluations all the 100 solutions were feasible and nondominated in respect to each other. The algorithm was able to find a diverse set of solutions in respect to all the objectives. The following parallel coordinate plot depicts the 100 solutions in objectives space.

stock	SMAPE	stock	SMAPE
ADS.DE	8.178385746708626	IBE.MC	6.152322262080069
ADYEN.AS	10.641090818052325	ITX.MC	7.219939644782287
AD.AS	5.408601054821343	IFX.DE	11.013903240141746
AI.PA	4.786506648996402	INGA.AS	12.135879754372953
AIR.PA	11.846596331941681	ISP.MI	9.915104274969671
ALV.DE	7.019777303513461	KER.PA	8.717096271195077
ABI.BR	9.809535329106653	KNEBV.HE	6.111059398376292
ASML.AS	8.526407784730827	OR.PA	5.782518002224039
CS.PA	8.955262184806415	LIN.DE	6.220693243454447
BAS.DE	7.996981181671698	MC.PA	7.2218918200092075
BAYN.DE	8.671662640055118	MBG.DE	11.535518384043854
BBVA.MC	11.018076694228448	MUV2.DE	7.50525191886549
SAN.MC	10.875880871596674	RI.PA	5.10132777229795
BMW.DE	8.921097861343082	PHIA.AS	7.702788221951842
BNP.PA	11.041334277715546	PRX.AS	6.7844902920183365
CRG.IR	7.355815124187261	SAF.PA	9.667711825509839
BN.PA	5.411205090121721	SAN.PA	4.833909652382033
DB1.DE	5.620575889394867	SAP.DE	7.825813071268087
DPW.DE	8.566878522693125	SU.PA	6.757823303514715
DTE.DE	5.363621193290268	SIE.DE	8.350806969589772
ENEL.MI	6.760362644618639	STLA.MI	10.773282376469194
ENI.MI	8.772337667774167	TTE.PA	8.226394503407779
EL.PA	6.815978917387625	DG.PA	6.682632458975507
FLTR.IR	10.613981495085303	VOW.DE	8.87749659412905
RMS.PA	6.811196472577539	VNA.DE	5.716103089466694



The horizontal bar chart contains the mean weight given to each stock from the 100 solutions.



The final solution was picked using Nautilus navigator.

References

[1] Li Chen et al. "Social responsibility portfolio optimization incorporating ESG criteria". In: *Journal of Management Science and Engineering* 6.1 (2021), pp. 75–85.

[2] Ningning Du, Yankui Liu, and Ying Liu. "A new data-driven distributionally robust portfolio optimization method based on wasserstein ambiguity set". In: *IEEE Access* 9 (2020), pp. 3174–3194.

[3] Okkes Ertenlice and Can B Kalayci. “A survey of swarm intelligence for portfolio optimization: Algorithms and applications”. In: *Swarm and evolutionary computation* 39 (2018), pp. 36–52.

[4] Muhammad Azfar Faizan. “Multiobjective portfolio optimization including sentiment analysis”. In: (2019).

[5] Massimiliano Kaucic, Mojtaba Moradi, and Mohmmad Mirzazadeh. “Portfolio optimization by improved NSGA-II and SPEA 2 based on different risk measures”. In: *Financial Innovation* 5.1 (2019), pp. 1–28.

[6] Petter N Kolm, Reha Tütüncü, and Frank J Fabozzi. “60 Years of portfolio optimization: Practical challenges and current trends”. In: *European Journal of Operational Research* 234.2 (2014), pp. 356–371.

[7] Raghavendra Kumar, Pardeep Kumar, and Yugal Kumar. “Multi-step time series analysis and forecasting strategy using ARIMA and evolutionary algorithms”. In: *International Journal of Information Technology* (2021), pp. 1–15.

[8] Panos Xidonas et al. “Multiobjective portfolio optimization: bridging mathematical theory with asset management practice”. In: *Annals of Operations Research* 267.1 (2018), pp. 585–606.

[9] TL van Zyl, Matthew Woolway, and Andrew Paskaramoorthy. “ParDen: Surrogate Assisted Hyper-Parameter Optimisation for Portfolio Selection”. In: *2021 8th International Conference on Soft Computing & Machine Intelligence (ISCMI)*. IEEE. 2021, pp. 101–107.