Assignment 1 - Minimum Variance Portfolios

Aliaksandr Panko, Nikolai Agafonov

November 29, 2018

Contents

1 Presentation

- Aliaksandr is ready to present
- Nikolai is ready to present

2 Objectives

1. Describe the data. Are there missing values? Decide on a strategy to deal with missing values and discuss possible consequences.

2. Compute three estimates of the covariance matrix:
    (a) sample covariance matrix
    (b) SAC estimate
    (c) PC estimate
3. A covariance matrix has to be symmetric and positive semi-definite. For further computation we also need nonsingularity (full rank). Check these properties
4. Compute the weights of the stocks in the minimum variance portfolio using both matrices in 2 cases: with short sales restriction and without
5. Describe the minimum variance portfolios
    (a) How many of the stocks have positive, zero and negative weights?
    (b) How similar/different are the MVP using different methods? What are the most important stocks and their weights?
    (c) Calculate the average monthly return and standard deviation of each portfolio and compare it with the index.
6. Is it a good choice to use of 5 years of monthly data for the calculation? What are the pro and cons for the use of shorter / longer periods, and higher/lower frequency?
7. Calculate the MVP each month, starting at the point in time when you have 5 years of historic data for the stocks that are in the index at your starting point. Each subsequent month, stick to 5 years of monthly data for calculation of covariances (so-called rolling windows).
8. Calculate the performance of each portfolio that you have formed over the subsequent month. This way you get a time series of returns that an investor could have obtained in real-time.
9. Compare the portfolio returns and standard deviations to the S&P 500 index and an equally weighted (1/N) portfolio. Plot diagrams.
10. Interpret and discuss.

3 Describing the Data

In this assignment we used provided data. The file contains monthly data for companies that have been a constituent of the S&P 500 index at some time from Jan 1970 to June 2018.

We used next tables:

1. cnam - to find company names
2. DT.members - to find relevant stocks
3. returns - to determine returns
4. spx - to get SP500 returns

In the first part of the exercise we need to use only a recent index composition (as of Dec 2017) and monthly return data over the last 5 years (60 months ending in Dec 2017). First step was creating unique keys (*id*) for shares by concatenating *gvkey* and *iid* fields (with dropping leading zeros in *iid*) and delete all useless data.

We found that according to $DT.members$ there were 500 stocks in the index. We use $DT.members$ result in our further analysis.

There are some missing values. We replace them by the mean of the existing values. Of course, this technique may lead to some deviations from the real data, but since we use the mean, there should not be any systemic biases. The important assumption of using this is that the data is missing completely at random. Another technique is to fully drop shares with at least one missing value, but since there are relatively many shares with such issues, it would lead to loss of information. We think that replacing of the missing values by mean leads to less harmful consequences than complete dropping of shares.

4 Three estimates of the covariance matrix

Sample covariance matrix calculation is fairly easy for SAC estimate and PC estimate we used the guides provided in the task description.

- For SAC estimate Σ_{SAC} $\delta_{SAC} = 0.5$ was used
- For PC estimate Σ_{PC} $n = 3$ was used

A covariance matrix has to be:

- symmetric

- positive semi-definite
- non-singular

Our findings are:

- Σ_S is symmetric, positive semi-definite, but not full-rank.
- Σ_{SAC} is symmetric, positive semi-definite and full-rank as required.
- Σ_{PC} is symmetric, positive semi-definite and full-rank as required.

```
For Original covariance matrix:
Symmetric: TRUE
Singular: TRUE
Positive semi-definite: FALSE

For SigmaPC covariance matrix:
Symmetric: TRUE
Singular: FALSE
Positive semi-definite: TRUE

For SigmaSAC covariance matrix:
Symmetric: TRUE
Singular: FALSE
Positive semi-definite: TRUE
```

4.1 Explanations

Σ_S is symmetric, positive semi-definite, but not full-rank. The explanation is as follows according to Dangl, Kashofer (2013): we have $N = 60$ time periods and $K = 500$ constituents. So $N < K$. In such a case, however, the sample covariance matrix is subject to large estimation errors and is singular by construction. The weak determination of the sample covariance matrix is evident when recognizing that one has to estimate $K(K+1)/2$ elements of it from only $N * K$ observations. Singularity follows immediately from the formula the sample covariance matrix which implies that its rank is bounded from above by min(K;N-1).

The other 2 matrices are proper by the construction (that is why we are using them)

5 The weights of the stocks using Σ_{SAC}

There are 4 weights vectors:

1. using PC without restrictions
2. using PC with restrictions
3. using SAC without restrictions
4. using SAC with restrictions

The following data frame summarize our findings:

```
For PC non-restricted:

Number of weights > 0:
 314
Number of weights < 0:
 183
Number of weights == 0:
 3
For PC restricted:

Number of weights > 0:
 51
Number of weights < 0:
 0
Number of weights == 0:
 449
For SAC non-restricted:

Number of weights > 0:
 235
Number of weights < 0:
 262
Number of weights == 0:
 3
For SAC restricted:

Number of weights > 0:
 32
Number of weights < 0:
 0
Number of weights == 0:
 468
```

The most important stocks are the following:

	w_PC.id	w_PC.conm	w_PC.weights
1	0317741	BRIGHTHOUSE FINANL INC	0.0806
2	0265901	FORTIVE CORP	0.0708
3	0321061	BAKER HUGHES A GE CO	0.0543
4	0309231	TECHNIPFMC PLC	0.0483
5	0047231	U S BANCORP	0.0259
6	0098501	SOUTHERN CO	0.0243
7	0055684	KRAFT HEINZ CO	0.0200
8	0067811	LOEWS CORP	0.0188
9	1121681	REPUBLIC SERVICES INC	0.0170
10	0082451	PNC FINANCIAL SVCS GROUP INC	0.0155

w_PC_c.id	w_PC_c.conm	w_PC_c.weights
0317741	BRIGHTHOUSE FINANL INC	0.1912
0265901	FORTIVE CORP	0.1705
0321061	BAKER HUGHES A GE CO	0.1263
0309231	TECHNIPFMC PLC	0.1009
0098501	SOUTHERN CO	0.0355
0055684	KRAFT HEINZ CO	0.0287
0318461	DARDEN RESTAURANTS INC	0.0254
1134901	CROWN CASTLE INTL CORP	0.0232
0281921	ARCONIC INC	0.0212
0075851	MOTOROLA SOLUTIONS INC	0.0204

w_SAC.id	w_SAC.conm	w_SAC.weights
0317741	BRIGHTHOUSE FINANL INC	0.1537
0265901	FORTIVE CORP	0.1393
0321061	BAKER HUGHES A GE CO	0.1119
0309231	TECHNIPFMC PLC	0.0945
1121681	REPUBLIC SERVICES INC	0.0378
0067811	LOEWS CORP	0.0347
0084791	PEPSICO INC	0.0341
0055684	KRAFT HEINZ CO	0.0300
0142821	AMPHENOL CORP	0.0286
0613881	EVEREST RE GROUP LTD	0.0286

w_SAC_c.id	w_SAC_c.conm	w_SAC_c.weights
0265901	FORTIVE CORP	0.2416
0317741	BRIGHTHOUSE FINANL INC	0.2292
0321061	BAKER HUGHES A GE CO	0.1366
0309231	TECHNIPFMC PLC	0.0722
0084791	PEPSICO INC	0.0427
1134901	CROWN CASTLE INTL CORP	0.0286
0055684	KRAFT HEINZ CO	0.0274
0613881	EVEREST RE GROUP LTD	0.0225
0111151	VARIAN MEDICAL SYSTEMS INC	0.0191
1121681	REPUBLIC SERVICES INC	0.0181

where:

- w_PC - PC without restrictions
- w_PC_c - PC with restrictions

- w_SAC - SAC without restrictions
- w_SAC_c - SAC with restrictions

Basically, the composition of restricted portfolio is close to the unrestricted one, just having larger weights for the main assets.

6 Time horizon discussion

We think that 5 years is reasonable. If one uses longer period, it may be better because it this case one better captures the increases and decreases of volatility that quite often characterize stock returns. Another benefit is that, for example, with more than 41 years of data one can get full-rank sample covariance matrix. If one uses shorter periods, it takes too short economic horizons, therefore could end up with too low or too high volatility used in the model. Using of lower frequency, e.g. year returns, implies losing of information, and it may be not a good idea, if the investment horizon is not very long. Using of higher frequency, like weekly or daily data, would provide more exact data than monthly one. But this would likely also result in larger perceived volatility. Use of intraday volatility is especially prone to this problem because of bid-ask bounce. Basically, use of monthly returns smooths the data and leads to more normalized estimates.

7 Portfolio Performance

The make the working process efficient we united all portfolio returns into one data frame:

	PC	PC_c	SAC	SAC_c	EQ	SP500
Jan 2013	1.22998434	0.574806205	0.005463936	0.811605219	6.56789429	5.17905858
Feb 2013	0.93808851	0.650782379	0.184523772	0.372851826	1.69681855	1.35685435
Mar 2013	1.32791166	1.190696114	0.145595270	0.775054501	4.34375276	3.75048147
Apr 2013	1.10749258	0.453143623	0.216394042	0.673496844	1.83278086	1.92646976
May 2013	0.22659732	-0.667559446	-0.031943055	-0.750625365	2.30815766	2.33802032
Jun 2013	2.85929801	-0.072890025	0.360941062	-0.088139275	-0.85956673	-1.34262044
Jul 2013	0.70477940	0.122305487	0.174899030	0.748916346	5.64058767	5.08838165
Aug 2013	-1.36042790	-0.800370301	-0.333860224	-0.913813467	-2.44691942	-2.89723410
Sep 2013	-0.70197796	0.007324246	0.176026793	0.569661650	4.22315399	3.13622407
Oct 2013	0.53146914	0.549207880	0.135915555	0.538156483	4.08588841	4.59629197
Nov 2013	2.12540266	0.885499256	0.327729903	0.293662262	2.39970756	3.04678344
Dec 2013	0.52647405	0.296916314	0.171067594	-0.211092058	2.84835834	2.52479381
Jan 2014	-0.25323658	-0.465499367	-0.186881815	-0.503246110	-1.89526578	-3.45742503
Feb 2014	0.71241181	0.800261826	0.073816378	0.912166998	5.34891867	4.57349480

Below we represent portfolios performance in terms of average returns and standard deviations:

Average returns:

	PC	PC_c	SAC	SAC_c	EQ	SP500
1	0.80208	0.1852034	0.1578506	0.122299	1.429425	1.265132

Standard Deviations:

	PC	PC_c	SAC	SAC_c	EQ	SP500
1	1.085746	1.061306	0.2122935	1.014411	2.845967	2.740868

The conclusion here is clear and expected:

1. Minimum variance portfolios have lower return (compared to naive and market)

2. Minimum variance portfolios have lower standard deviation (compared to naive and market)

3. SAC portfolio did the best from Min. variance portfolios

4. naive and index did approximately the same

8 Rolling Windows

Next step was using rolling window calculate average returns and standard deviations for all 6 portfolios:

1. w_PC - PC without restrictions

2. w_PC_c - PC with restrictions

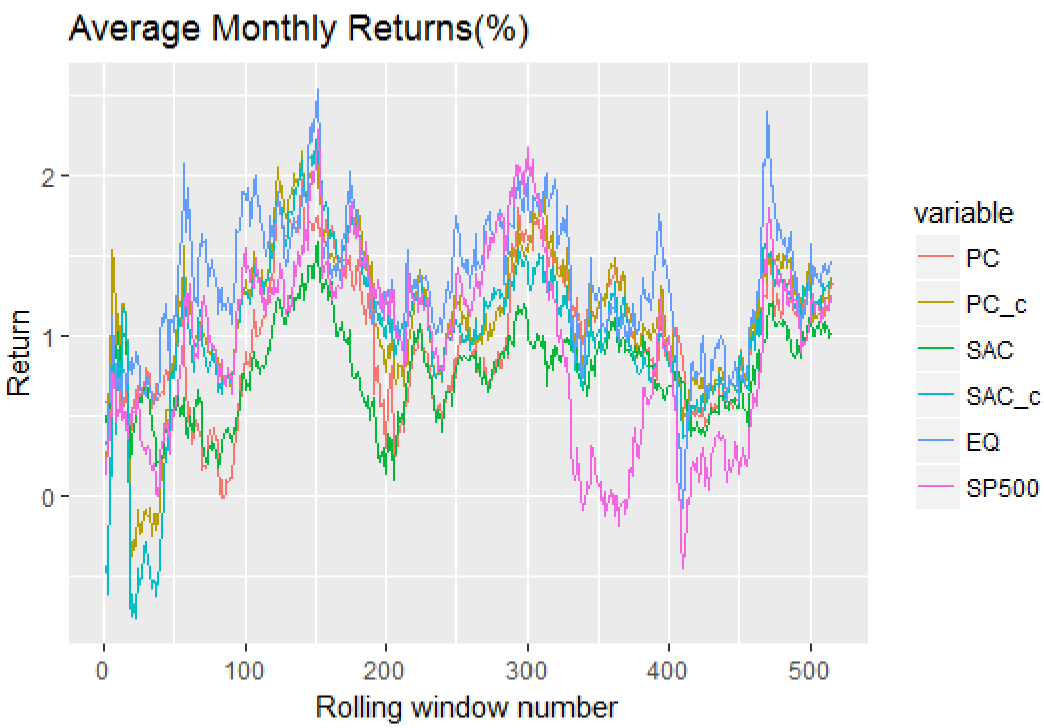
3. w_SAC - SAC without restrictions

4. w_SAC_c - SAC with restrictions

5. EQ - naive

6. SP500 - index

The results are represented in two charts below:



8.1 Comments

The results are expected:

- Naive portfolio shows the highest average returns (beating the market)
- In a bad times market portfolio shows the lowest average returns (naive is still beating the market)
- In terms of standard deviations, naive a bit more volatile and both market and naive are significantly more volatile that MVPs (as expected)
- The lowest standard deviation shows SAC non-constrained portfolio.