Investment Portfolio Optimization

Michał Podsiadło
WhyR?
28th September 2019
Warsaw

Plan Of Presentation

- **Intro**
  - ❖ Idea
  - ❖ Industry Specification
  - ❖ Assets Specification
  - ❖ Packages & Libraries
- **Data Set & ETL**
  - ❖ Data Set Specification
  - ❖ ETL Methodology
  - ❖ Key Variables
- **Price Predictions**
  - ❖ Regression Metrics
  - ❖ Prediction Models
  - ❖ Modelling Outcomes
- **Portfolio Optimization**
  - ❖ Key Formulas
  - ❖ Business Scenarios
  - ❖ Optimization Outcomes
- **Summary**
  - ❖ Short & Concise
  - ❖ Sources

INTRO

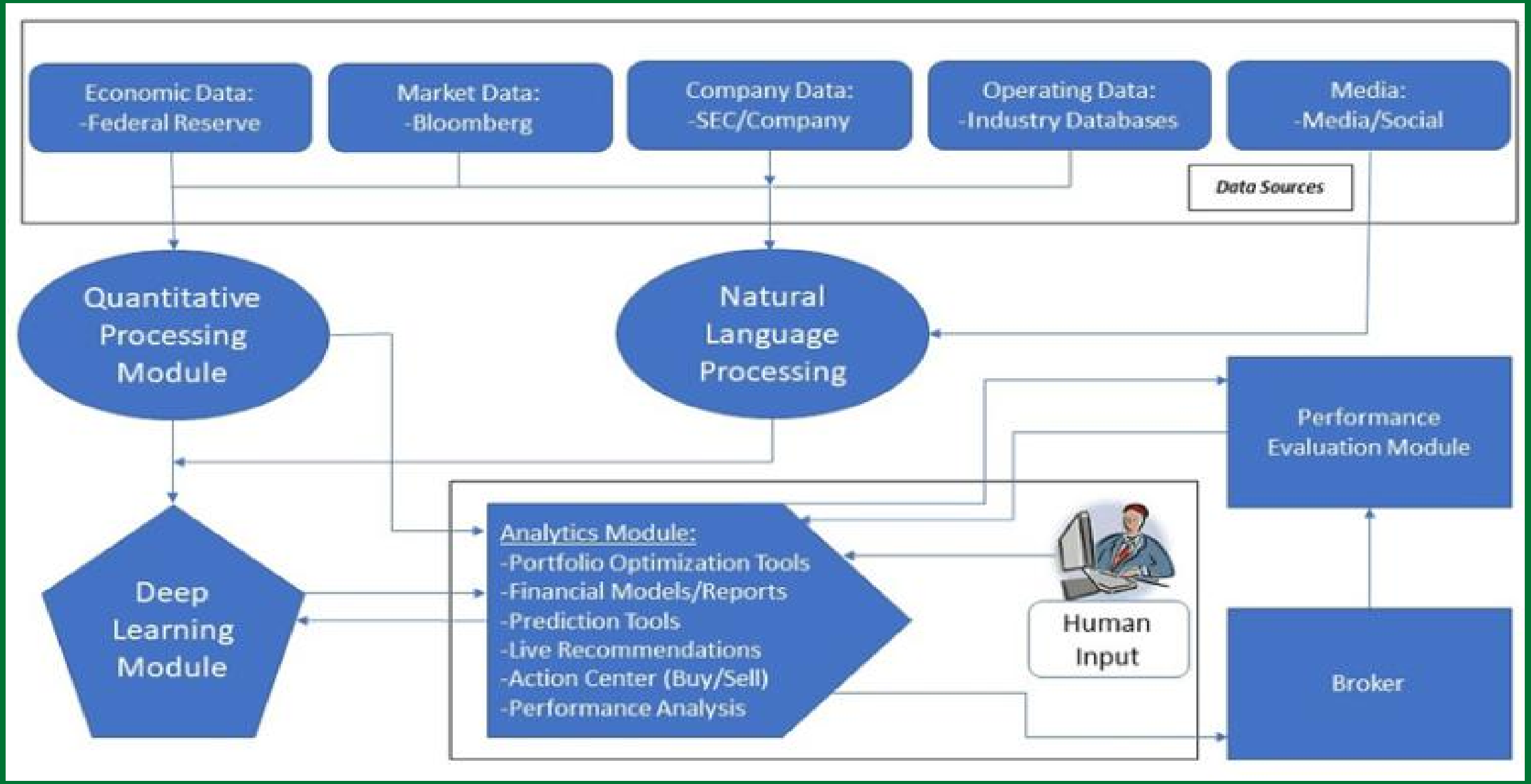
Idea

- The presentation's goal is just to explain the problem of constructing efficient portfolios and show how to impelement a solution for the best fitted portfolios which will give maximized returns with minimized risks.
- Based on the postgraduate thesis - Investment Portfolio Optimization based on ETF funds and traditional Mutual funds

Industry Specification (1/2)

- **Processing of collected data**
  - ETL
  - Various data sources
  - NLP

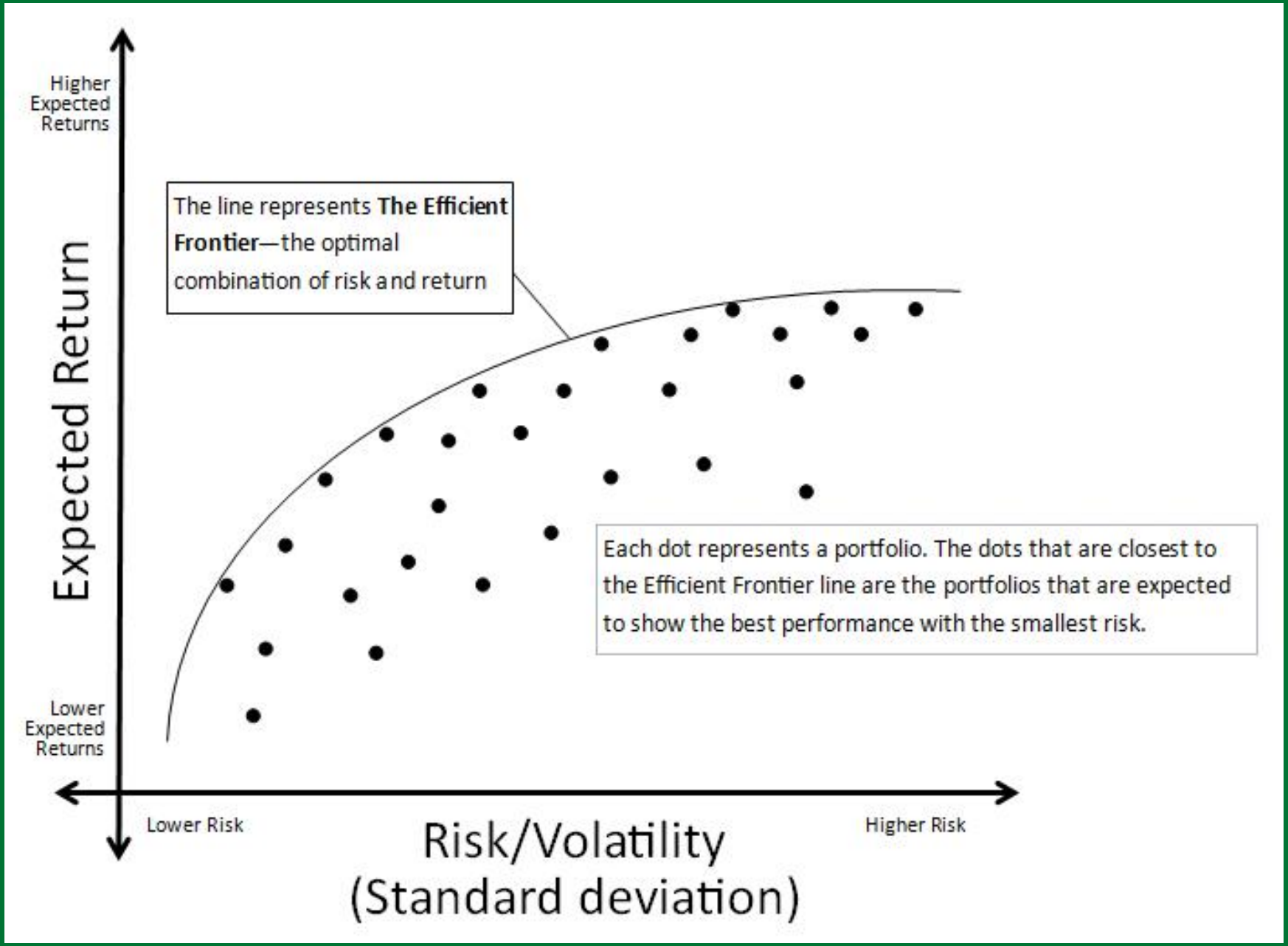
- **Automated portfolio construction**
  - Analytical tools
  - Portfolio optimization tools
  - Financial modelling
  - Dashboards / KPI / portfolio performance



SOURCE: https://blogs.cfainstitute.org/investor/2018/02/08/data-science-will-transform-the-investment-industry-are-you-prepared/

Industry Specification (2/2)

- **Modern theory of portfolio construction**
  - ❖ Efficient frontier
    - ➢ Maximized returns
    - ➢ Minimized risks
  - ❖ IPS (Investment Policy Statement)
    - ➢ individual investor needs
    - ➢ tailored assets & portfolio
    - ➢ any exceptions must be discussed



SOURCE: http://www.gettingyourich.com/blog/what-is-the-efficient-frontier-theory

Assets Specification (1/3)

Bonds

- Quite “safe”
- Form of loan to the companies, governments, etc.
- Coupon or non-coupon
- Ratings
  - ❖ Investment grade
  - ❖ Non-investment grade

Credit Rating Scales by Agency, Long-Term

Moody's	S&P	Fitch	
Aaa	AAA	AAA	Prime
Aa1	AA+	AA+	High grade
Aa2	AA	AA	
Aa3	AA-	AA-	
A1	A+	A+	Upper medium grade
A2	A	A	
A3	A-	A-	
Baa1	BBB+	BBB+	Lower medium grade
Baa2	BBB	BBB	
Baa3	BBB-	BBB-	
Ba1	BB+	BB+	Non-investment grade speculative
Ba2	BB	BB	
Ba3	BB-	BB-	
B1	B+	B+	Highly speculative
B2	B	B	
B3	B-	B-	
Caa1	CCC+	CCC	Substantial risk
Caa2	CCC		Extremely speculative
Caa3	CCC-		Default imminent with little prospect for recovery
Ca	CC	CC	
	C	C	
C	D	D	In default
/			
/			

"Junk"

WOLFSTREET.com

SOURCE: https://wolfstreet.com/credit-rating-scales-by-moodys-sp-and-fitch/

Assets Specification (2/3)

Equities

- Risky
- Part of companies' finances
- Possibility of dividends
- **Accounting principle**
  - ❖ Assets = Liabilities + Equities
- **Sectors**
  - ❖ Along **Global Industry Classification Standard**

Assets Specification (3/3)

Sectors by Global Industry Classification Standard

- Energy
- Materials
- Industrials
- Consumer Discretionary
- Consumer Staples
- Health Care
- Financials
- Information Technology
- Communication Services
- Utilities
- Real Estate

Packages & Libraries

- **Caret**
  - Well known
  - Multiple models
  - Predictions
- **PortfolioOptimizer**
  - Very simple, minimalistic
  - Contains portfolios as separate objects
  - You just only add up constraints, objectives

Data Set & ETL

Data Set Specification

- **Data sets from Kaggle**
  - Mutual Funds and ETFs
  - As at the 03/05/2019
  - Based on Yahoo Finance
  - Contain
    - General (basic) details
    - Sector weights (equities)
    - Rating weights (bonds)
    - Return rates for each year
    - Return rates for certain asset classes
    - Morningstar's portfolio classifications
    - Bonds maturities

- **ETFs - CSV file [;]**
  - 2.352 instances (rows)
  - 104 attributes (columns)

- **Mutual Funds - CSV file [;]**
  - 25.308 instances (rows)
  - 125 attributes (columns)

ETL Methodology

- Deleted “certain” ratios and measures
- Remained only factual data with weights and returns
  - ❖ Weights had to be adjusted to percents - e.g. 6.93 → 0.0693
  - ❖ Weights had to be adjusted in the context of the whole portfolio, not the certain asset class
  - ❖ Returns were adjusted in context of certain assets and asset classes - it will be needed in Portfolio Optimization

Key Variables (1/2)

General specific variables

- **Categorization**
  - investment
  - size
- **Asset classes**
  - portfolio_stocks
  - portfolio_bonds

- **Return rates**
  - fund_return_2018
  - fund_return_2017
  - fund_return_2016
  - fund_return_2015
  - fund_return_2014
  - fund_return_2013
  - fund_return_2012
  - fund_return_2011
  - fund_return_2010

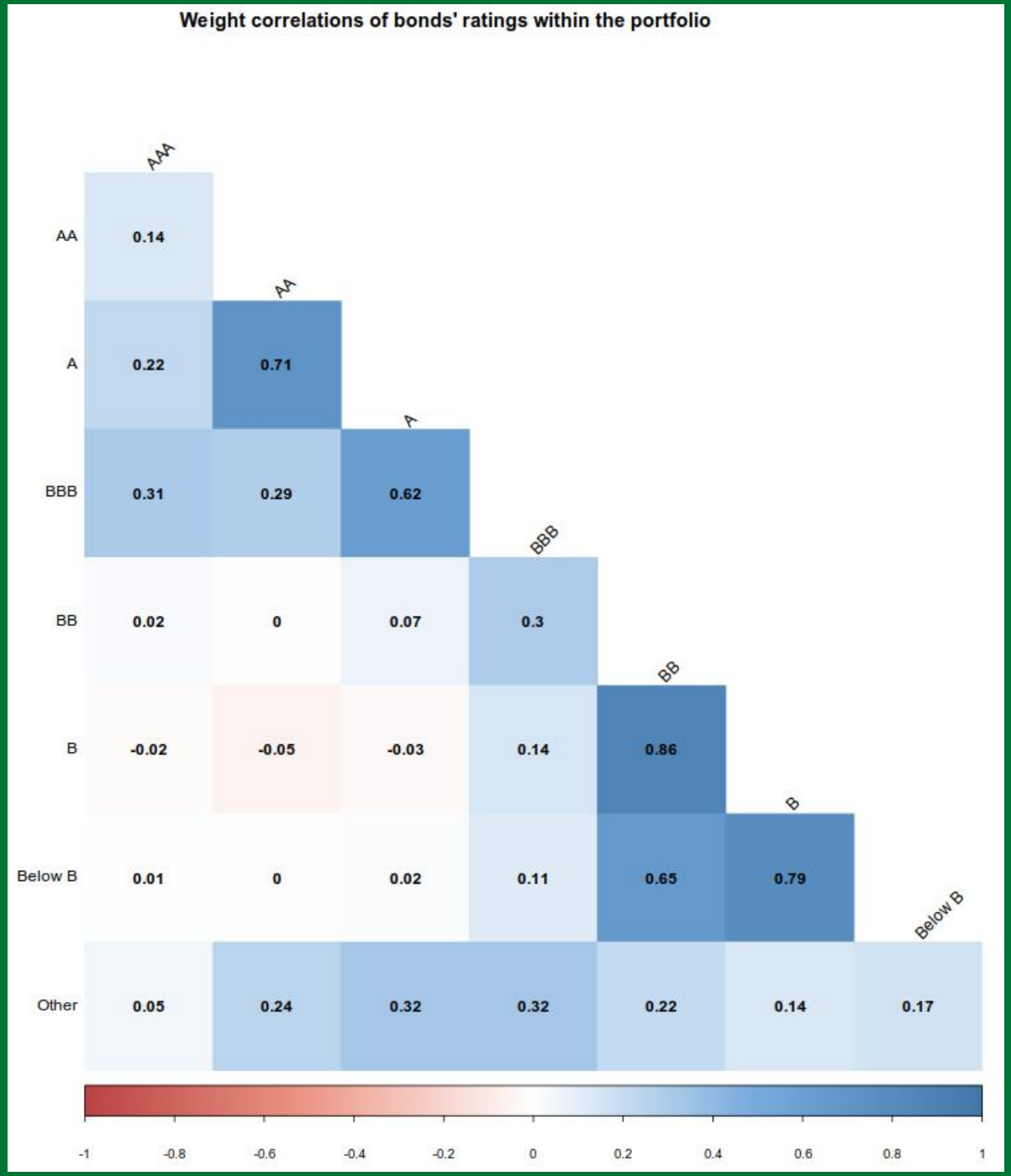
Key Variables (2/2)

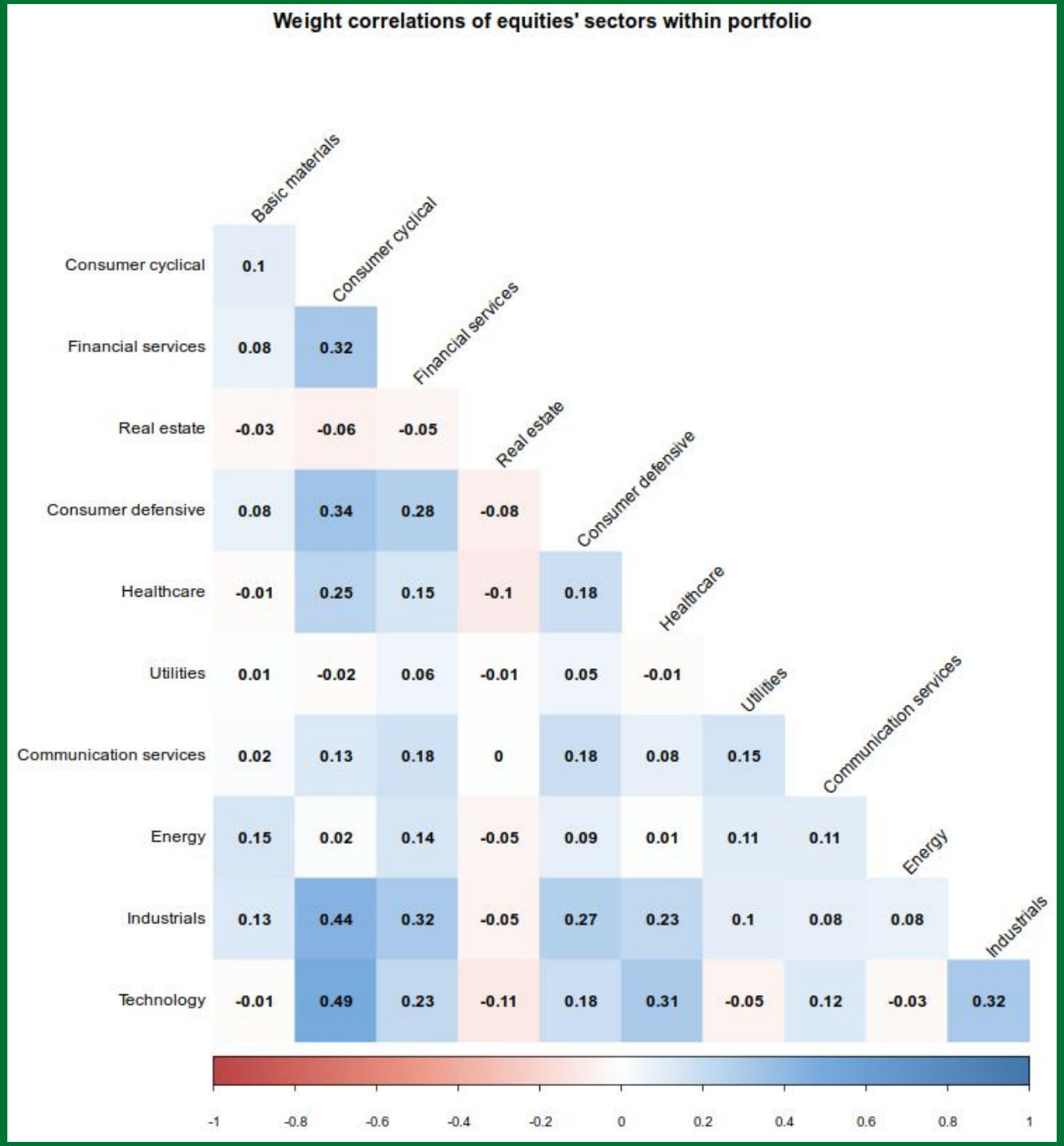
Sectors (Equities)

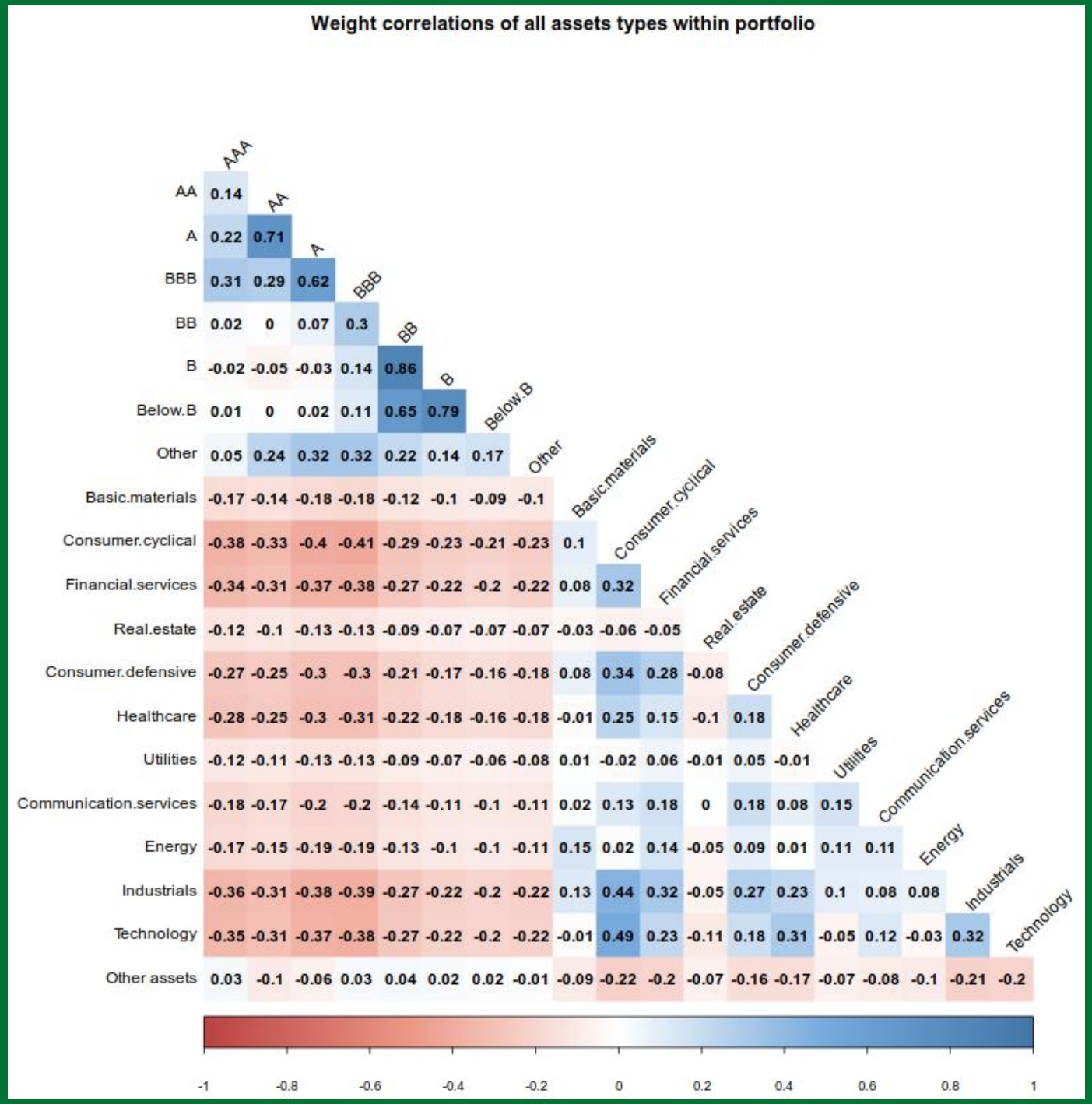
- basic_materials
- consumer_cyclical
- financial_services
- real_estate
- consumer_defensive
- healthcare
- utilities
- communication_services
- energy
- industrials
- technology

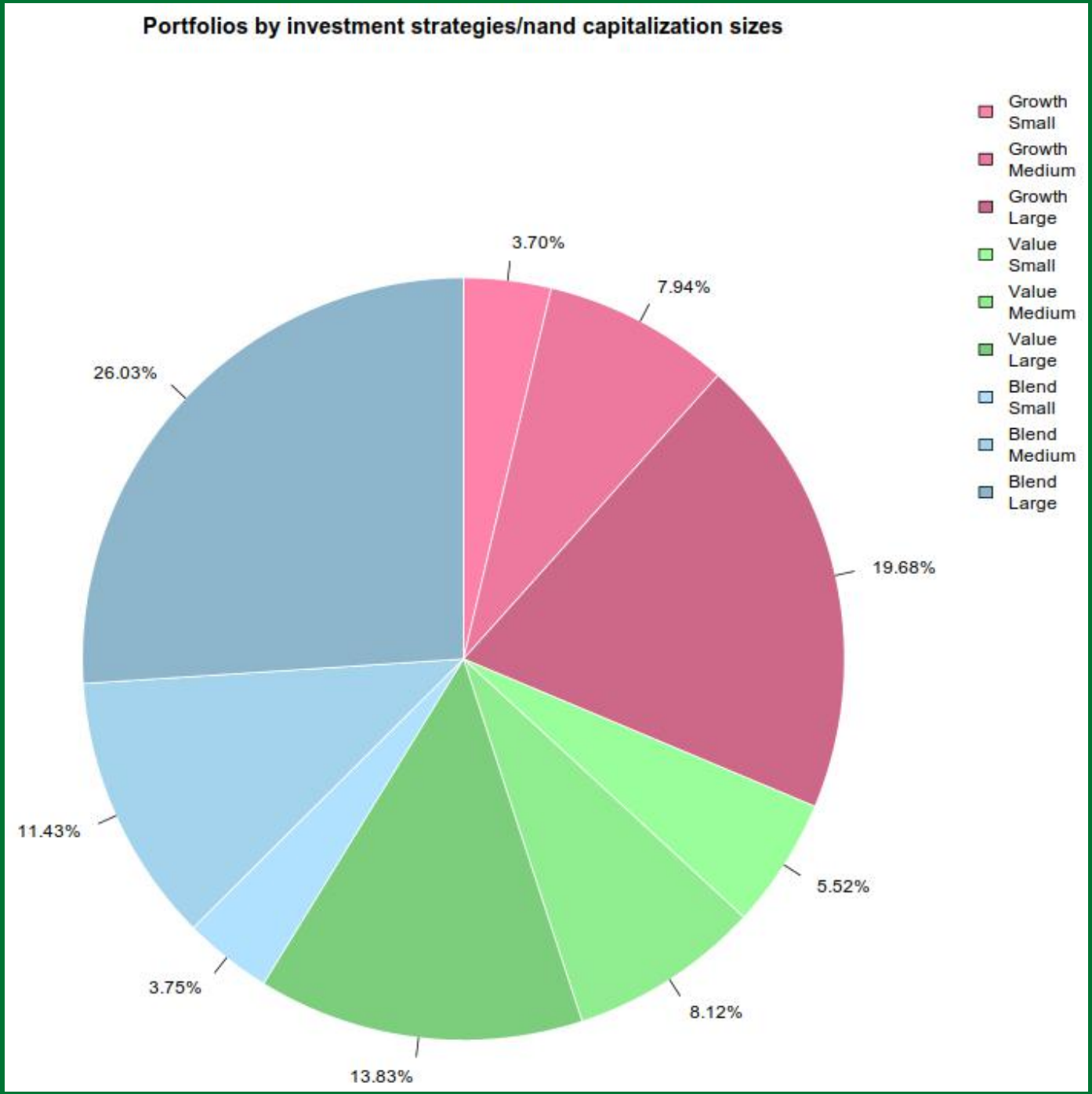
Ratings (Bonds)

- rating_aaa
- rating_aa
- rating_a
- rating_bbb
- rating_bb
- rating_b
- rating_below_b
- rating_others


Weight correlations of bonds' ratings within the portfolio
AAA
AA
A
BBB
BB
B
Below B
AA
A
BBB
BB
B
Below B
Other
0.14
0.22
0.71
0.31
0.29
0.62
0.02
0
0.07
0.3
-0.02
-0.05
-0.03
0.14
0.86
0.01
0
0.02
0.11
0.65
0.79
0.05
0.24
0.32
0.32
0.22
0.14
0.17
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1


Weight correlations of equities' sectors within portfolio
Basic materials
Consumer cyclical
Financial services
Real estate
Consumer defensive
Healthcare
Utilities
Communication services
Energy
Industrials
Consumer cyclical 0.1
Financial services 0.08 0.32
Real estate -0.03 -0.06 -0.05
Consumer defensive 0.08 0.34 0.28 -0.08
Healthcare -0.01 0.25 0.15 -0.1 0.18
Utilities 0.01 -0.02 0.06 -0.01 0.05 -0.01
Communication services 0.02 0.13 0.18 0 0.18 0.08 0.15
Energy 0.15 0.02 0.14 -0.05 0.09 0.01 0.11 0.11
Industrials 0.13 0.44 0.32 -0.05 0.27 0.23 0.1 0.08 0.08
Technology -0.01 0.49 0.23 -0.11 0.18 0.31 -0.05 0.12 -0.03 0.32
-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1


Weight correlations of all assets types within portfolio
AAA
AA AA 0.14
A 0.22 0.71 A
BBB 0.31 0.29 0.62 BBB
BB 0.02 0 0.07 0.3 BB
B -0.02 -0.05 -0.03 0.14 0.86 B
Below.B 0.01 0 0.02 0.11 0.65 0.79 Below.B
Other 0.05 0.24 0.32 0.32 0.22 0.14 0.17 Other
Basic.materials -0.17 -0.14 -0.18 -0.18 -0.12 -0.1 -0.09 -0.1 Basic.materials
Consumer.cyclical -0.38 -0.33 -0.4 -0.41 -0.29 -0.23 -0.21 -0.23 0.1 Consumer.cyclical
Financial.services -0.34 -0.31 -0.37 -0.38 -0.27 -0.22 -0.2 -0.22 0.08 0.32 Financial.services
Real.estate -0.12 -0.1 -0.13 -0.13 -0.09 -0.07 -0.07 -0.07 -0.03 -0.06 -0.05 Real.estate
Consumer.defensive -0.27 -0.25 -0.3 -0.3 -0.21 -0.17 -0.16 -0.18 0.08 0.34 0.28 -0.08 Consumer.defensive
Healthcare -0.28 -0.25 -0.3 -0.31 -0.22 -0.18 -0.16 -0.18 -0.01 0.25 0.15 -0.1 0.18 Healthcare
Utilities -0.12 -0.11 -0.13 -0.13 -0.09 -0.07 -0.06 -0.08 0.01 -0.02 0.06 -0.01 0.05 -0.01 Utilities
Communication.services -0.18 -0.17 -0.2 -0.2 -0.14 -0.11 -0.1 -0.11 0.02 0.13 0.18 0 0.18 0.08 0.15 Communication.services
Energy -0.17 -0.15 -0.19 -0.19 -0.13 -0.1 -0.1 -0.11 0.15 0.02 0.14 -0.05 0.09 0.01 0.11 0.11 Energy
Industrials -0.36 -0.31 -0.38 -0.39 -0.27 -0.22 -0.2 -0.22 0.13 0.44 0.32 -0.05 0.27 0.23 0.1 0.08 0.08 Industrials
Technology -0.35 -0.31 -0.37 -0.38 -0.27 -0.22 -0.2 -0.22 -0.01 0.49 0.23 -0.11 0.18 0.31 -0.05 0.12 -0.03 0.32 Technology
Other assets 0.03 -0.1 -0.06 0.03 0.04 0.02 0.02 -0.01 -0.09 -0.22 -0.2 -0.07 -0.16 -0.17 -0.07 -0.08 -0.1 -0.21 -0.2
-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1


Portfolios by investment strategies/nand capitalization sizes
Growth Small
Growth Medium
Growth Large
Value Small
Value Medium
Value Large
Blend Small
Blend Medium
Blend Large
3.70%
7.94%
19.68%
5.52%
8.12%
13.83%
3.75%
11.43%
26.03%

Price Predictions

Regression Metrics (1/3)

RMSE (Root Mean Squared Error)

- square root of Mean Squared Error
- error rate

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}$$

where:

$\hat{y}$ – predicted value of y
$\overline{y}$ – mean value of y

Regression Metrics (2/3)

Rsquared (Coefficient of determination)

- how properly predicted values fit to original ones
- interpreted as percentages

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \overline{y})^2}$$

where:

$\hat{y}$ – predicted value of y
$\overline{y}$ – mean value of y

Regression Metrics (3/3)

MAE (Mean Absolute Error)

- original and predicted values differences
- averaged absolute difference (along data set)

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \widehat{y}|^2$$

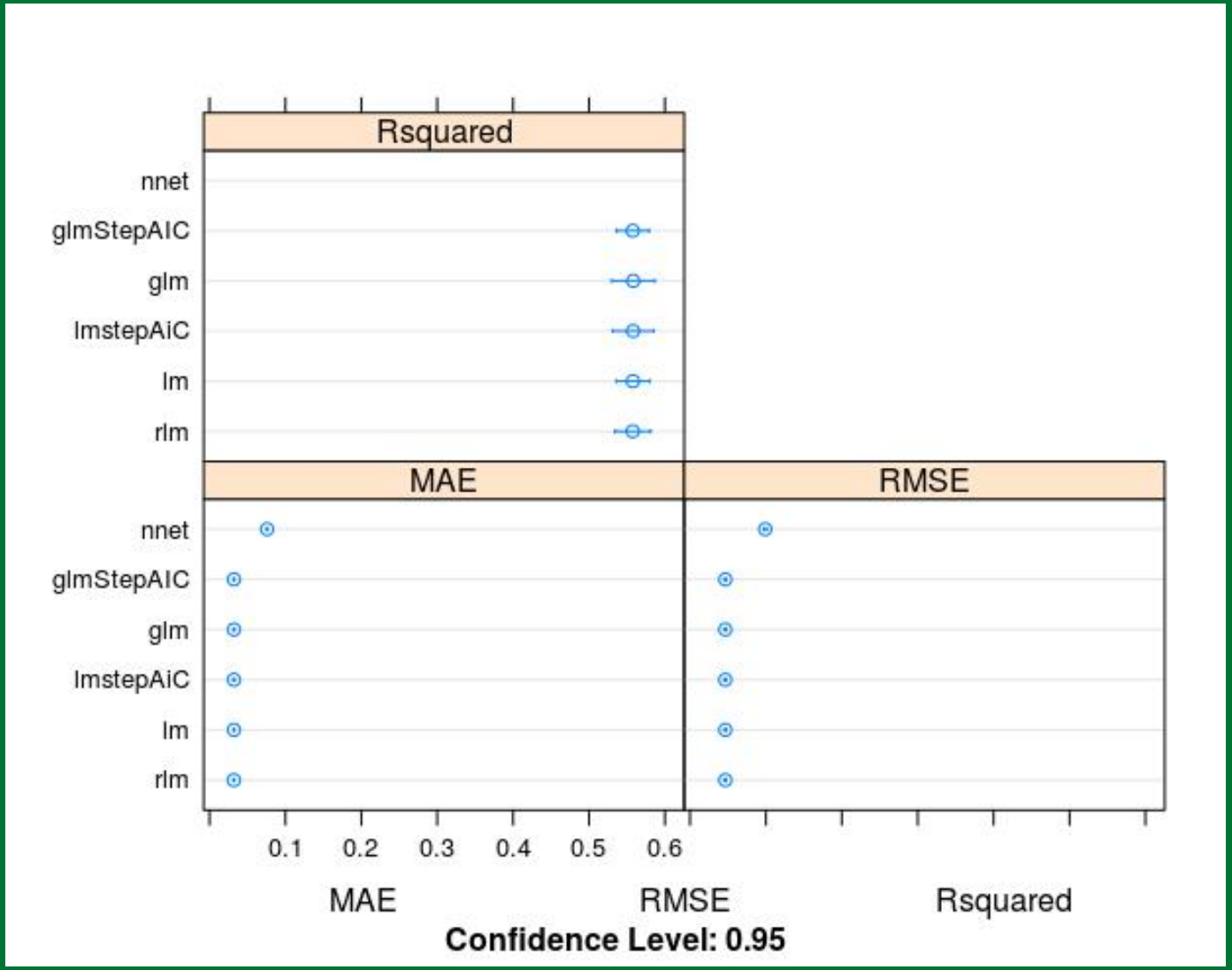
where:

$\widehat{y}$ – predicted value of y
$\overline{y}$ – mean value of y

Prediction Models

- **Time frames**
  - 2010 - 2018
  - Linearity is obvious - not only for prediction, but also for portfolio construction
- **135 Possible Linear Models to use**
  - Chosen only 1 model
  - RSquared used in decision making

Modelling Outcomes



Modelling Outcomes

Linear Regression

8867 samples
8 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 7982, 7980, 7979, 7981, 7980, 7981, ...
Resampling results:

RMSE	Rsquared	MAE
0.0464308	0.5579009	0.03218755

Portfolio Optimization

Key Formulas (1/2)

Expected Portfolio Return

- Return rates
- Assets weights
- Weighted average

$$E(R_p) = \sum_{i=1}^{N} w_i E(R_i) = w_1 E(R_1) + w_2 E(R_2) + \ldots + w_n E(R_n)$$

where:

$E(R_p)$ – expected return rate for the whole portfolio

$\sum_{i=1}^{N} w_i E(R_i)$ – sum of products for expected return rates and asset weights

w_n – weight of a certain asset

$E(R_n)$ – expected return rate of a given asset

Key Formulas (2/2)

Portfolio risk

- Correlations between assets
- Assets weights

$$\sigma_P = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2P_{1,2}}$$

or

$$\sigma_P = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2Cov_{1,2}}$$

where:

w_1^2 – weight of asset 1 in portfolio

σ_1^2 – individual variance for asset 1

$P_{1,2}$ – correlation between assets 1 and 2

$Cov_{1,2}$ – covariance between assets 1 and 2

Business Scenarios

- There can be many and lots of combinations or needs
  - ❖ depend on customer's needs and market flavours
  - ❖ by portfolio / capitalization sizes
    - ➢ small
    - ➢ medium
    - ➢ large
  - ❖ by investment strategy
    - ➢ value
    - ➢ growth
    - ➢ blend

Optimization Outcomes

	Strategy - GROWTH Size - SMALL	Strategy - GROWTH Size - MEDIUM	Strategy - GROWTH Size - LARGE	Strategy - VALUE Size - SMALL	Strategy - VALUE Size - MEDIUM	Strategy - VALUE Size - LARGE	Strategy - BLEND Size - SMALL	Strategy - BLEND Size - MEDIUM	Strategy - BLEND Size - LARGE
OTHER ASSETS	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Basic materials	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Consumer cyclical	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Financial services	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Real estate	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Consumer defensive	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Healthcare	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Utilities	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Communication services	0,0000	0,0000	0,0000	0,6500	0,0000	0,0000	0,0000	0,0000	0,0000
Energy	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Industrials	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Technology	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
AAA	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
AA	0,3499	0,6500	0,0000	0,3499	0,6500	0,6500	0,6500	0,6500	0,3499
A	0,0000	0,3501	0,0000	0,0000	0,0000	0,0000	0,0000	0,3501	0,0000
BBB	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
BB	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
B	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Below B	0,0000	0,0000	0,6500	0,0000	0,0000	0,0000	0,0000	0,0000	0,6500
Other	0,6500	0,0000	0,3499	0,0000	0,3499	0,3499	0,3499	0,0000	0,0000

Summary

Short & Concise

- **Diversification does matter!**
  - ❖ Both for **assets** and **asset classes**
  - ❖ We want **maximized returns** and **minimized risks**
- Only **proper selection** of **weights** and **asset classes** can **minimize** and **dilute risk** of invested capital
- **Capitalization size** and **investment strategy** do matter!

Sources (1/2)

- Andrecut, M. (2013). Portfolio Optimization in R. Unlimited Analytics Inc.
- Balcewicz, J. (2019). Cyberpolicy NASK. Retrieved from https://cyberpolicy.nask.pl/sztuczna-inteligencja-godna-zaufania-rekomendacje-grupy-ekspertow-komisji-europejskiej/
- Bennett, R. (2018). Introduction to PortfolioAnalytics.
- Everitt, B. S., & Hothorn, T. (2006). A Handbook of Statistical Analyses Using R. Chapman & Hall-CRC.
- Frątczak, E. (2012). W Zaawansowane metody analiz statystycznych (strony 45-47). Warszawa: Oficyna Wydawnicza SGH.
- Frątczak, E. (2015). In Statistics for Management and Economics (pp. 409-412). Warszawa: Oficyna Wydawnicza SGH.
- Frątczak, E. i Korczyński, A. (2014). Statystyka od podstaw z systemem SAS wersja 9.2 i 9.3. Warszawa: Oficyna Wydawnicza SGH.
- Garnry, P. (2017). Skarbiec. Retrieved from https://www.skarbiec.biz/inwestycje/gielda/zmienic-portfel-bardziej-defensywny.html
- Grzegorzewski, P., Gagolewski, M. i Bobecka-Wesołowska, K. (2014). Wnioskowanie statystyczne z wykorzystaniem srodowiska R. Warszawa: Biuro ds. Projektu „Program Rozwojowy Politechniki Warszawskiej”.
- Harrington, C. (2017). CFA Institute. Retrieved from https://www.cfainstitute.org/en/research/cfa-magazine/2017/speaking-data-science-with-an-investment-accent
- IShares. (2019). Retrieved from https://www.ishares.com/us/about-etfs/what-is-an-etf

Sources (2/2)

- KupAkcje.pl. (2007). Retrieved from http://kupakcje.pl/?p=159
- Leone, S. (2019). Kaggle. Retrieved from https://www.kaggle.com/stefanoleone992/mutual-funds-and-etfs
- Mankiw, G. N., & Taylor, M. P. (2016). Makroekonomia. Warszawa: Polskie Wydawnictwo Ekonomiczne.
- Panimalar, A., Shree, V., & Kathrine, V. (2017). Irjet. Retrieved from https://www.irjet.net/archives/V4/i9/IRJET-V4I957.pdf
- Przanowski, K. (2014). Credit Scoring w erze Big-Data. Warszawa: Oficyna Wydawnicza SGH.
- Przanowski, K. (2015). In Credit Scoring – studia przypadków procesów biznesowych (p. 36). Warszawa: Oficyna Wydawnicza SGH.
- Saidov, U. (2018). CFA Institute. Retrieved from https://blogs.cfainstitute.org/investor/2018/02/08/data-science-will-transform-the-investment-industry-are-you-prepared/
- Świerkot, M. (2011). Bankier. Retrieved from https://www.bankier.pl/wiadomosc/Ratingi-Wielkiej-Trojki-hustaja-gieldami-2393655.html
- Świerkot, M. (2012). Bankier. Retrieved from https://www.bankier.pl/wiadomosc/Przeglad-inwestycji-Warrena-Buffetta-2668741.html
- Wolfstreet (2019). Retrieved from https://wolfstreet.com/credit-rating-scales-by-moodys-sp-and-fitch/
- Würtz, D., Setz, T., Chalabi, Y., Lam, L., & Ellis, A. (2015). Portfolio Optimization with R-Rmetrics. Rmetrics Association & Finance Online Publishing.

THANK YOU FOR ATTENTION :)

Contact details:

name:	Michał Podsiadło
mobile:	506-962-069
mail:	Michal.Podsiadlo.88@gmail.com
insta:	von_sexcel
github:	github.com/Ravn88
kaggle:	kaggle.com/ravn88