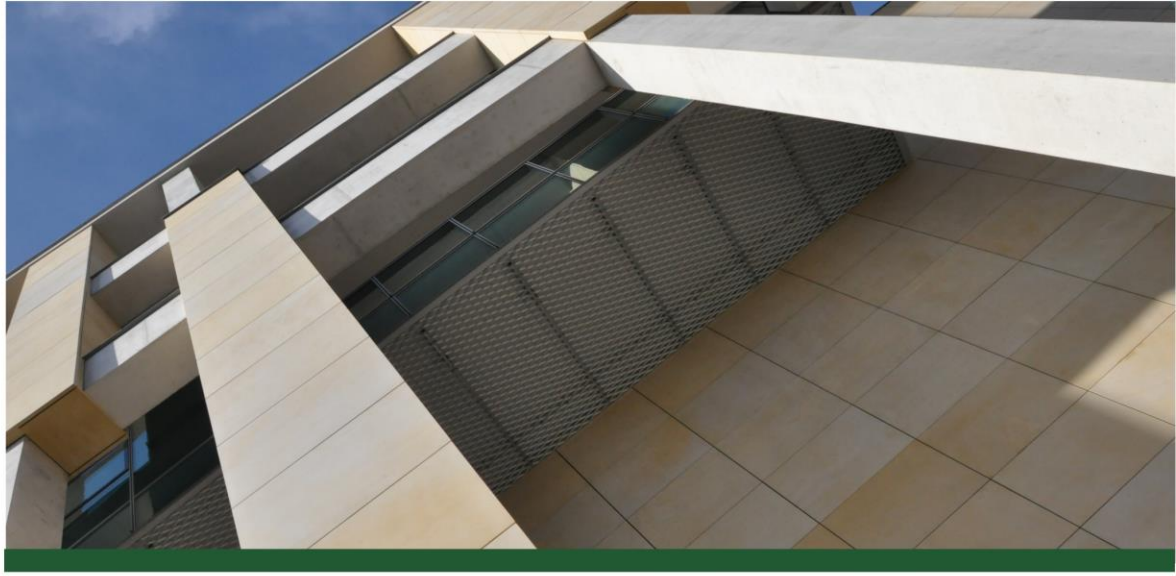


Classifying equity mutual funds according to various risk measures using neural networks

Katarzyna Perez & Małgorzata Szczyt (referee)

Investment and Financial Markets Department
Institute of Finance
Poznan University of Economics and Business

Motivation for research

- The level of risk determines the investment goals and investment style of mutual funds,

  but it is often changed by managers to earn higher returns – they do so called *style drift*
  - Brown et al. (1996), Chevalier i Allison (1997), Bams, Otten & Ramezanifar (2016), Cao, Iliev & Velthuis (2017), Sha (2020)

Motivation for research

- Rankings of ranking agencies, e.g. Morningstar, vary from eachother
  and there is no consensus whether investors rely on them or not:
  - Ben-David et al. (2018) – investors rely on rankings more than on a fund's exposure to risk factors
  - Barber, Huang& Odean (2016), Berk & Van Binsbergen (2016): only the best informed investors have tools to monitor fund managers
  - Huang et al. (2020) – investors are rational and react optimally on the ranking stars of Morningstar

Motivation for research

- Our previous study on Polish equity mutual funds confirms that various attributes explain the risks taken by these funds and that classification of funds according to risk is correct
  - Perez i Szczyt (2021): more applicable results in case of *standard deviation* than in case of *beta ratio*

Research question

- Which attributes/variables of mutual funds explain their classification according to risk measures?
  - Are these attributes *informative/valuable* to investors, so they can use them when they buy mutual funds?

- Does machine learning helps to answer this question?

Data

Morninstar Direct database:

- Open end investment funds registered in Europe
- *equity* funds
- *oldest share* funds
- inception before 01.01.2021 (24 917 funds)
- Sales region – European cross border (very popular recently)

6689 equity funds of European cross border sales region

- Monthly data from January 1990 to July 2023
- Data without survivalship bias

32 899 observations

Data

Quantitative variables:

- Rt – annual rate of return obtained on the basis of monthly data
- Age – operating time counted in years
- Size – fund's net assets (NAV)
- CF – cash flow

$$CF_{i,t} = \frac{NAV_{i,t} - NAV_{i,t-i}(1 + R_{i,t})}{NAV_{i,t-i}}$$

Additionally, the following was used as an input variable:

- Time (year to which a specific observation refers)

Data

	N	Average	Mediane	25% perc.	75% perc.	St. dev.
NAV (euro)	54623	131 777 898	28 558 862	7 528 274	98 464 946	42 299 045
CF	54623	0,74	0,88	-0,48	1,92	4,53
Rt	78067	0,528	0,809	-0,477	1,764	2,178
Age (Years)	78067	8,8	7	3	13	8

Data

Qualitative variables:

- Currency – the base currency of a fund
- Category – the category to which a fund belongs
- Domicile – location of fund registration
- Legal form of a fund

Data

Qualitative variables:

- Currency – the base currency of a fund (17 categories)

Currency	share in database
Euro	55%
US Dollar	29%
Pound Sterling	9%
Swiss Franc	3%
Japanese Yen	2%
Swedish Krona	1%
Norwegian Krone	1%
Other	1%

Data

Qualitative variables:

- Category – the category to which a fund belongs (38 categories)

Category	share in database
Global Equity Large Cap	19%
Europe Equity Large Cap	17%
Global Emerging Markets Equity	8%
Europe Equity Mid/Small Cap	6%
Long/Short Equity	4%
US Equity Large Cap Blend	4%
Asia ex-Japan Equity	4%
Japan Equity	4%
Global Equity Mid/Small Cap	3%
Other	27%
No info	4%

Data

Qualitative variables:

- Domicile (25 categories)

Domicile	share in database
Luxembourg	54%
Ireland	18%
France	6%
United Kingdom	5%
Germany	5%
Austria	4%
Belgium	3%
Liechtenstein	2%
Switzerland	1%
Netherlands	1%
Denmark	1%
Other	2%

Data

Qualitative variables:

- Legal form of a fund (7 categories)

Legal form	share in database
SICAV	48%
Open Ended Investment Company	32%
FCP	14%
Unit Trust	5%
Other	0,04%

Data

Output variables - classifiers:

1. Beta – coefficient determining the degree of relationship between the fund's return and the market's return (MSCI Europe)

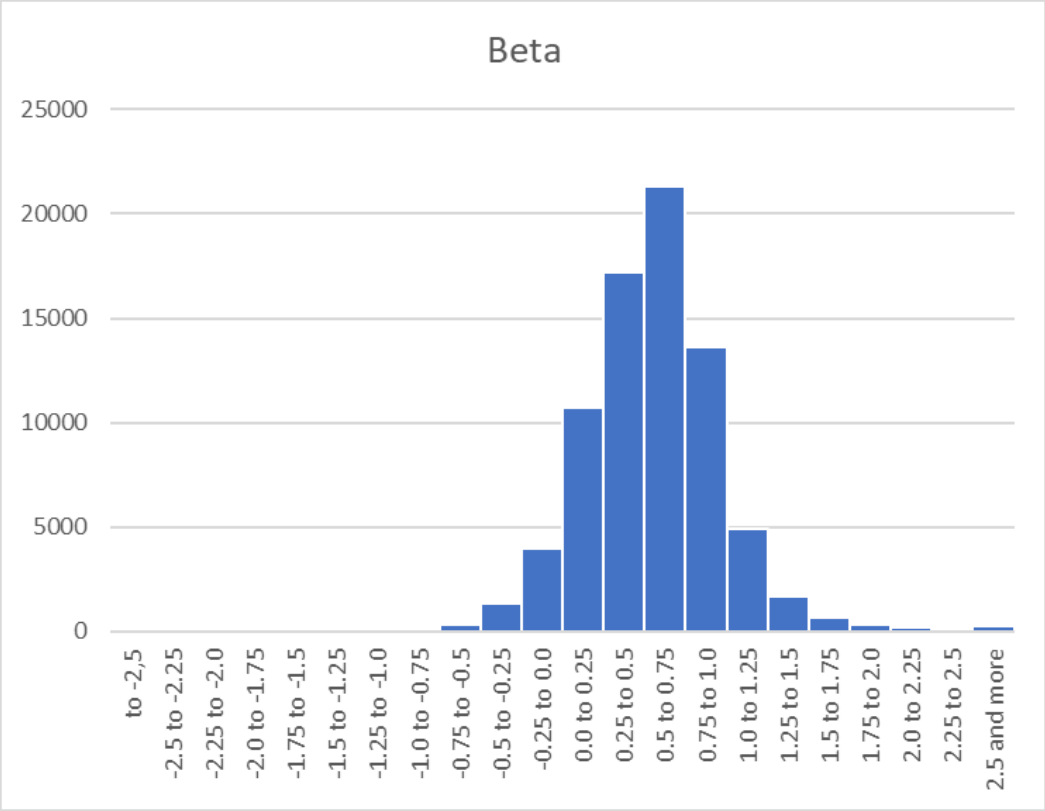
1: Beta < 0
2: Beta $\in$ <0 ; 0,5)
2: Beta $\in$ <0,5 ; 1)
4: Beta > 1

2. Sigma – standard deviation of monthly returns in particular years

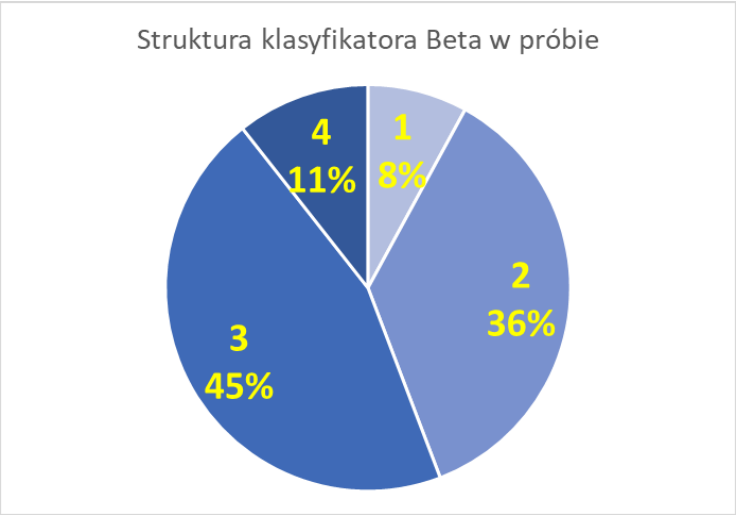
1: Sigma < 2
2: Sigma $\in$ <2 ; 4)
2: Sigma $\in$ <4 ; 6)
4: Sigma > 6

Data

Output variables - classifiers:



średnia	0,54
mediana	0,55
25% perc.	0,29
75% perc.	0,79



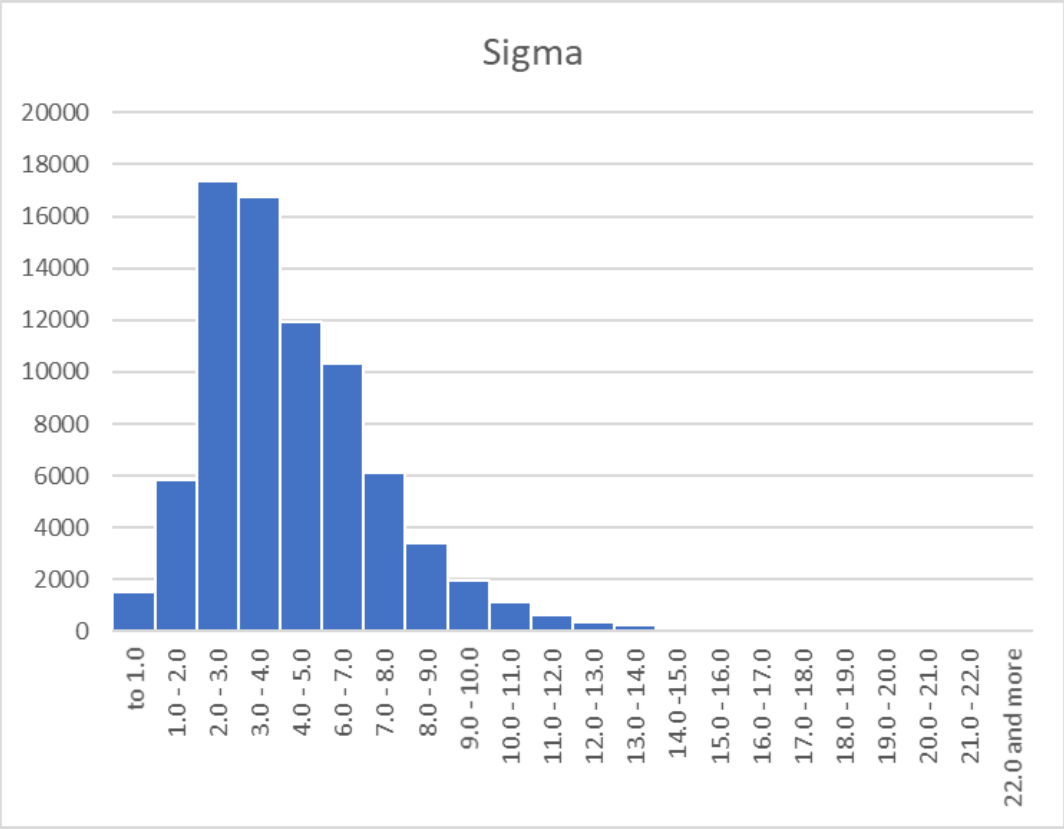
POZNAŃ UNIVERSITY OF ECONOMICS AND BUSINESS

SGH

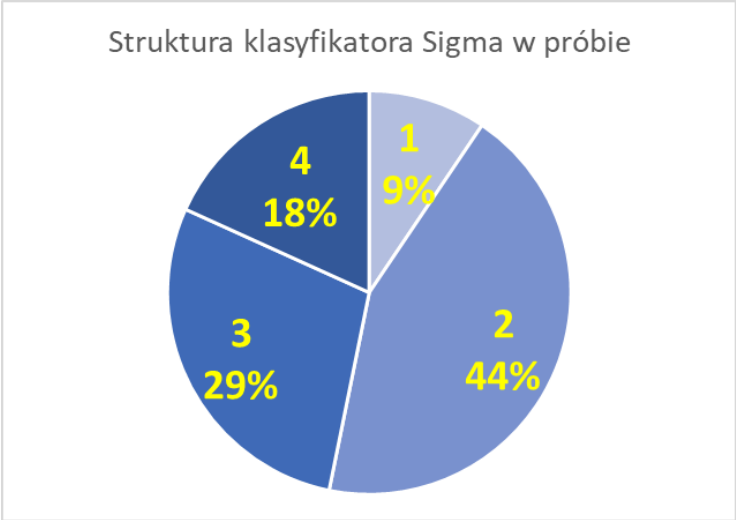
EERIA Eastern European Risk and Insurance Association

Data

Output variables - classifiers:



średnia	4,30
mediana	3,82
25% perc.	2,72
75% perc.	5,45



Poznań University of Economics and Business

SGH

Eastern European Risk and Insurance Association

Data

Output variables - classifiers:

3. SSD – semistandard deviation of monthly returns from the mean

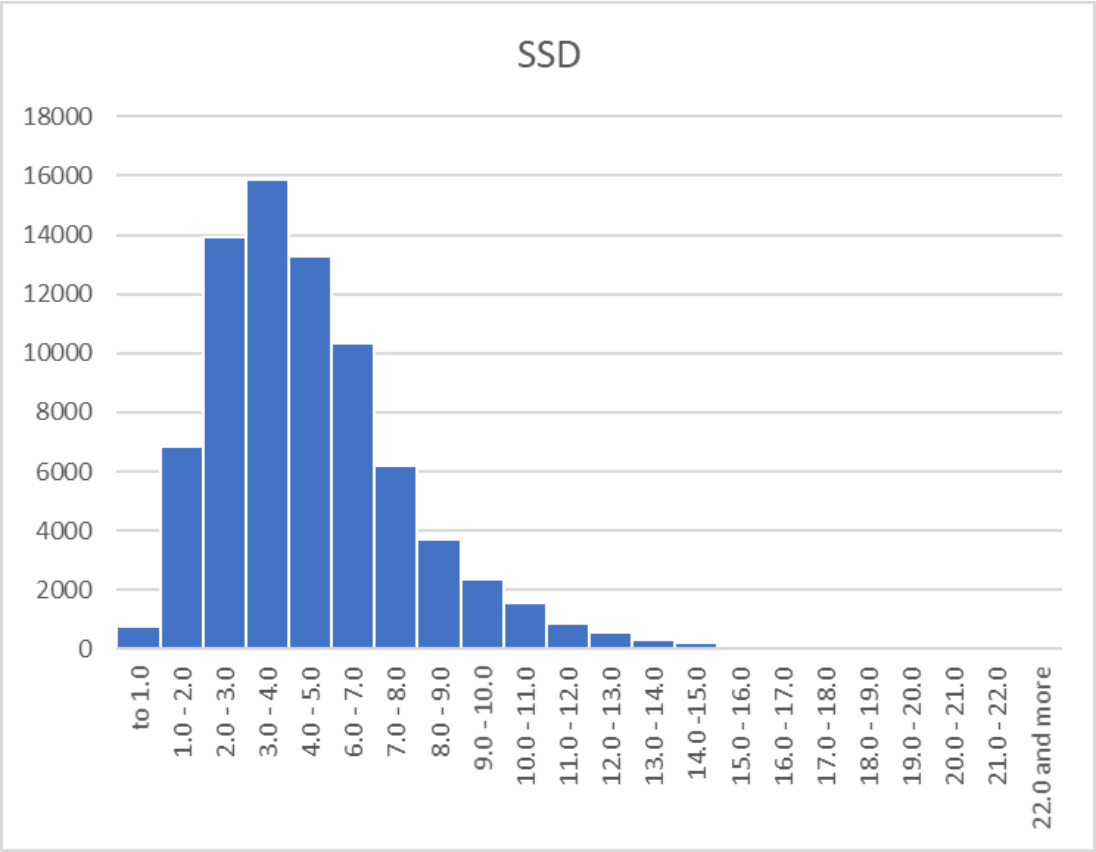
1: SSD < 2,5
2: SSD $\in$ <2,5 ; 4)
2: SSD $\in$ <4 ; 6,5)
4: SSD > 6,5

2. VaR – estimated using the parametric method, with a confidence level of 0.95

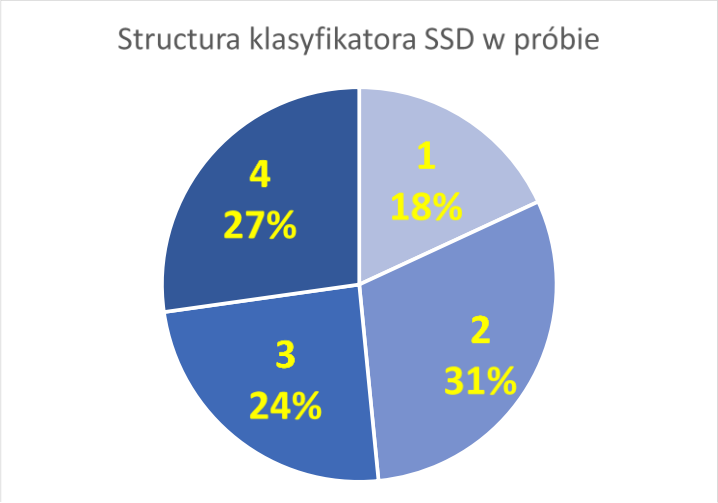
1: VaR < -9
2: VaR $\in$ <-9 ; -6)
2: VaR $\in$ <-6 ; -3)
4: VaR > -3

Data

Output variables - classifiers:



średnia	4,54
mediana	4,08
25% perc.	2,86
75% perc.	5,67



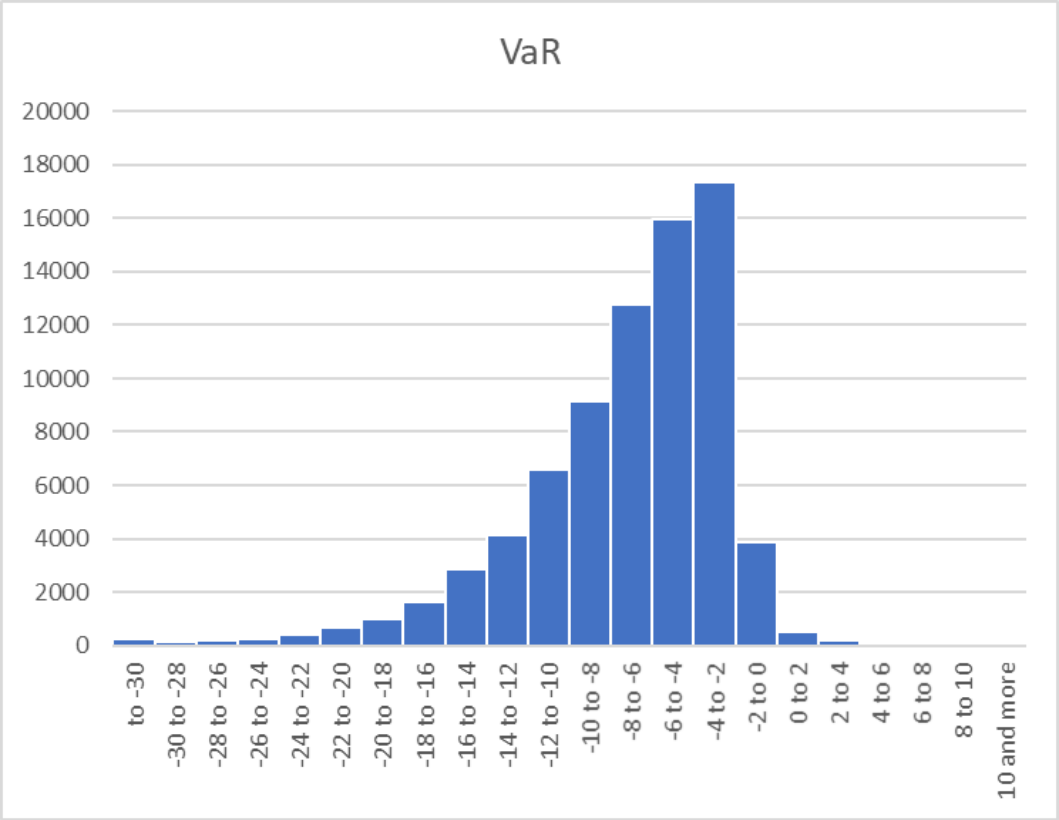
Poznań University of Economics and Business

SGH

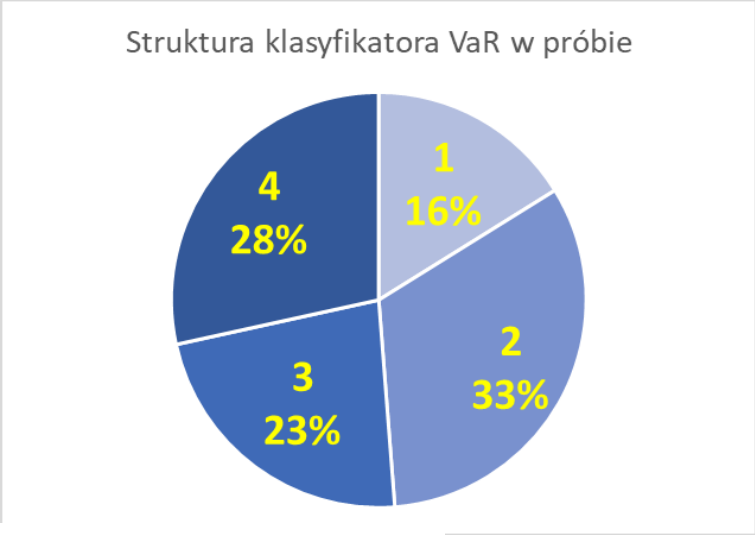
Eastern European Risk and Insurance Association

Data

Output variables - classifiers:



średnia	-7,26
mediana	-6,13
25% perc.	-9,64
75% perc.	-3,72



Poznań University of Economics and Business

SGH

Eastern European Risk and Insurance Association

Research procedure

- *Statistica.PL,* artificial neural networks module
- Automatic network search procedure was used
- It was allowed to build both MLP (multi-layer networks) and RBF (networks with one hidden layer).
- The number of hidden neurons was limited to the range of 4-13 for the MLP and 10-20 for the RBF
- Various activation functions have been adopted (linear, logistic, exponential, hyperbolic tangent)

Research procedure

- From all possible combinations of parameters, 30 networks were randomly selected, of which the 5 best ones were selected for the final classification (providing the best classification correctness).
- Classification was performed based on the network team
  - i.e. the fund was assigned to the class that was indicated by the largest number of networks

Results

<u>Beta classificator</u>

Overall classification accuracy – **68%**

Class 1 – 25% Class 2 - 66%

Class 3 – 85% Class 4 – 9%

- Global sensitivity analysis showed that in this classification, the attributes that differentiated the groups were primarily:

  currency (1,83), category (1,62), domicile (1,55), legal form (1,31) i time (1,24).

Results

<u>Sigma classificator</u>

Overall classification accuracy – **60%**

Class 1 – 9% Class 2 - 88%

Class 3 – 41% Class 4 – 32%

- Global sensitivity analysis showed that in this classification, the attributes that differentiated the groups were primarily:

  currency (1,38), legal form (1,31), category (1,30), domicile (1,18) and return (1,12).

Results

SSD classificator

Overall classification accuracy – **47%**

Class 1 – 41% Class 2 - 69%

Class 3 – 25% Class 4 – 46%

- Global sensitivity analysis showed that in this classification, the attributes that differentiated the groups were primarily:

  currency (1,26), category (1,17) and domicile (1,17).

Results

<u>VaR classificator</u>

Overall classification accuracy – **54%**

Class 1 – 29% Class 2 - 84%

Class 3 – 17% Class 4 – 63%

- Global sensitivity analysis showed that in this classification, the attributes that differentiated the groups were primarily:

  legal form (1,30), return (1,22), category (1,20), currency (1,18), domicile (1,14).

Results

Classificator	Accuracy	Main explaining variables/attributes
Beta	68%	Currency, category, domicile, legal form, time
Sigma	60%	Currency, legal form, category, domicile, return
SSD	47%	Currency, category, domicile
VaR	54%	Legal form, return, category, currency, domicile

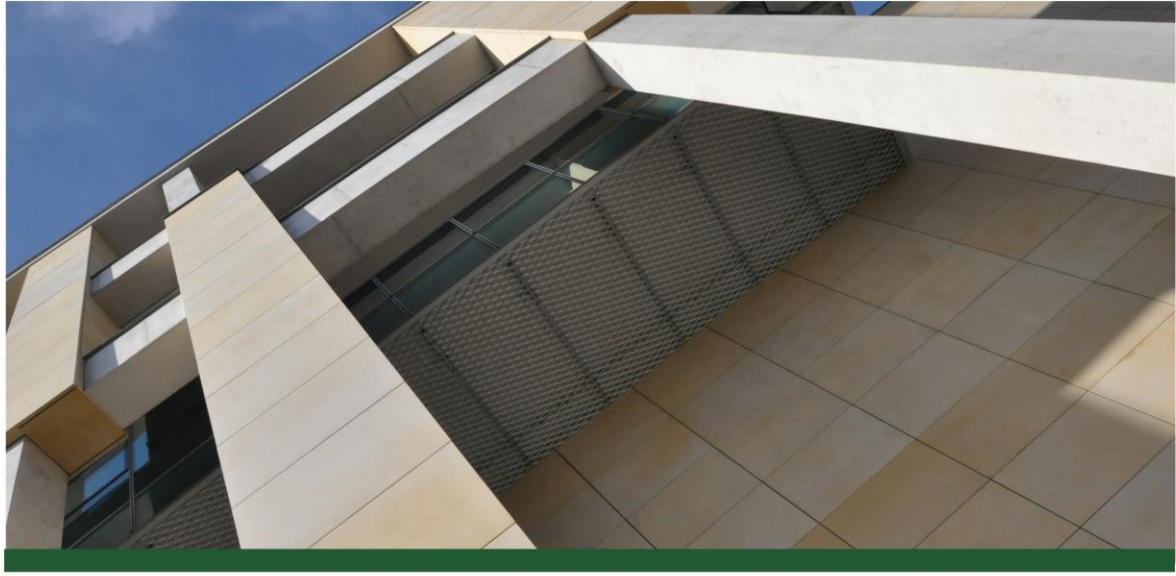
- The lowest value for classification showed Cash Flow (0,99).

Conclusions

- The methods used and the input variables used for classification **do not give very good results**, although these results are significantly better than random classification.
- The qualitative variables used, which are rather constant characteristics of funds over time, work much better in predicting the fund's risk - compared to quantitative variables.
- Classifying a fund into a category determining its risk, based on data that are not typical risk measures, may involve a noticeable probability of making an error.
- The classification based on the proposed methods is still better than random classification

Further research

- Switch from analysis based on annual data to monthly data.
- Expanding the sample to other stock markets and other types of mutual funds
- Searching for additional variables characterizing funds



POZNAŃ UNIVERSITY OF ECONOMICS AND BUSINESS

Thank you!

We appreciate comments and suggestions!

Classifying equity mutual funds according to various risk measures using neural networks

Katarzyna Perez & Małgorzata Szczyt

POZNAŃ UNIVERSITY OF ECONOMICS AND BUSINESS

SGH

EERIA Eastern European Risk and Insurance Association