

SCRAPING THE SYNTHETIC RISK AND REWARD INDICATOR FROM FUNDS' KEY INFORMATION DOCUMENTS

BY FABIAN BLASCH

University of Vienna

INSERT ABSTRACT

1. Introduction. According to BGBl. II Nr. 265/2011, effective 2011, every undertaking for collective investment in transferable securities (UCITS), has to publish a key information document (KID)(FMA, 2011). Said document is supposed to inform potential and current investors about various characteristics of the investments at hand. Besides a short description of the investments as well as performance measured against a benchmark, this document also contains a synthetic risk- and reward indicator (SRRI). As the name suggests, the purpose of this indicator is to measure the risk associated with an investment in the respective fund. How the SRRI is derived depends on the class of investment fund and will be described in greater detail in a separate subsection. Independent of the exact procedure of derivation the underlying interpretation of risk is measured via the volatility of returns.

The aim of this thesis is to firstly, give a short and precise introduction into the calculation of the SRRI. Then the data is briefly presented and the approach to measuring extraction performance is discussed. The following sections focus shifts to extracting the SRRI which is displayed as a graph, usually on the first page of every KID. The description of the extraction is based on pseudo-code which aims to aid the reader in understanding the approach taken to obtain the SRRI. In order to keep the structure as simple as possible the process is split into a parent function which calls a variety of helper functions.

MSC 2010 subject classifications: Primary Key1, Key2; secondary Key3

Keywords and phrases: Scraping, L^AT_EX 2_ε

2. SRRI. The SRRI aims to measure risk via the volatility of weekly returns from the last 5 years, should weekly returns be unobtainable, the calculation may be executed using monthly returns. Accordingly, by ordinance the SRRI may be obtained as follows

$$\sigma_f = \sqrt{\frac{m}{T-1} \sum_{t=1}^T (r_{f,t} - \bar{r}_f)^2},$$

where $r_{f,t}$ is the fund's return and \bar{r}_f represents the mean of returns over T periods. Then scaling via m , the return frequency within a year, yields the standard deviation of yearly returns. For illustrative purposes, the calculation using weekly returns would result in $m = 52$, as there are 52 months in a year and $T = 5 \times 52 = 260$ the number of months in 5 years. To obtain the SRRI which is measured on a scale from 1 to 7 the regulating authority provides a table.

TABLE 1
SRRI Classification

SRRI	σ_f
1	$0\% \leq \sigma_f < 0.5\%$
2	$0.5\% \leq \sigma_f < 2\%$
3	$2\% \leq \sigma_f < 5\%$
4	$5\% \leq \sigma_f < 10\%$
5	$10\% \leq \sigma_f < 15\%$
6	$15\% \leq \sigma_f < 25\%$
7	$25\% \geq \sigma_f$

3. Measuring Scraping Performance. The SRRI is measured on an increasing ordinal scale, with 7 being associated with the highest risk and simultaneously also with the highest return. This would in theory allow for a measurement of predictive performance utilizing the absolute or squared difference between prediction and actual value. However, in this case we want to measure whether the read-out was successful or not. Hence the performance evaluation will be based on predicting correctly. Formally this means, we will use a discrete metric, i.e

$$Accuracy = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{pred_i = act_i\}}.$$

Put differently, the accuracy measure is equivalent to the relative amount of correctly predicted cases.

4. Data. As previously mentioned the KIDs are available to download from the respective fund managing firm. The sample used for measuring the extracting performance contains 121 Key Information documents, obtained from the websites of the respective KAG (Kapitalanlagegesellschaft). The distribution of the SRRI as well as the amount of KIDs per KAG within the sample are displayed below.

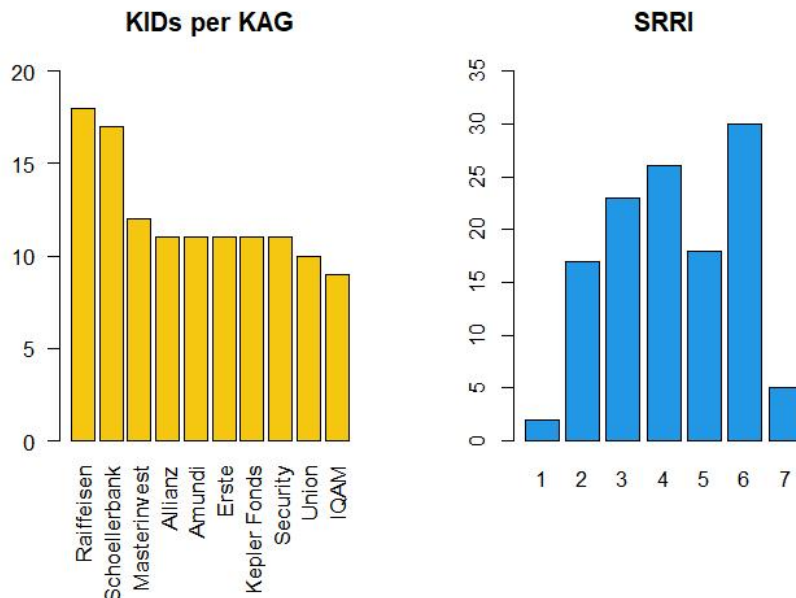


FIG 1. Number of KIDs per KAG and distribution of SRRI

The two depicted plots give a good impression of the sample at hand. We can immediately observe that two KAGs are overrepresented in the sample, namely Raiffeisen and Schoellerbank. Further, we can see that for the remaining fund managing firms around 10 KIDs are available in the sample. Regarding the distribution of the SRRI, the graph depicted on the right enables us to notice that the sample seems to contain very little funds that are associated with an SRRI of 1 or 7. The low number of funds classified into the first class may be due to the way the standard deviation of returns is converted into the SRRI. The intervals which are used to determine the SRRI increase in length quite significantly. The first interval for which $0\% \leq \sigma_f < 0.5\%$ spans a range of half a percent, while the interval for the classification into the 6th class $15\% \leq \sigma_f < 25\%$ spans 10%. The other end, namely the SRRI of 7, could be underrepresented because of a lack in demand for such high risk products. Of course this is just mere speculation but interesting nonetheless.

In regards to the extraction, to ease illustration of the coming chapters, an example document from Erste can be found [here](#). Additionally, an example of the SRRI graph as it can be found in an Erste KID is provided below.

← Potenziell niedrigere Rendite ← Niedrigeres Risiko				Potenziell höhere Rendite → Höheres Risiko →		
1	2	3	4	5	6	7

FIG 2. *Example of the SRRI graph from an Erste KID*

5. Naive Approach.

6. Agglomerative Clustering.

APPENDIX A: CODE

Appendix for code and additional illustrations

A.1. Functions.

REFERENCES

FMA. (2011). Ordinance of the financial market authority. *KID-V, BGBl II 265/2011*.