IDENTIFYING RISK SOURCES WITH PRINCIPAL COMPONENT ANALYSIS

Bogdan Romenskii – Tomáš Löster

Abstract

Risk is an important characteristic that plays a huge role in capital markets decision making. Thus, correctly determining and specifying the risk sources on a given market may shape investors' decisions. This thesis aims to identify and interpret the sources of financial risk on the German equity market. The goal is achieved using principal component analysis as a tool for building the portfolios that allocate their budgets to distinct uncorrelated risk sources. Throughout the research, correlation matrix was compared with the covariance matrix as an underlying foundation for PCA, effectively examining if standardising the returns is an important step. It was also revealed that the first five principal components have different interpretations, indicating the underlying equity characteristics that may not be obvious on a first sight, effectively revealing the biggest risk sources on the German equity market. The market component was extracted that accounted for over 30% of the overall variance, representing a big portion of systematic risk. The findings of the study also suggest that normalising the returns might be a beneficial approach.

Key words: Principal component analysis, Financial risk, Equity markets

JEL Code: C32, C38, G11

Introduction

In the dynamic landscape of global finance, capital markets stand as the bedrock upon which economics flourish, investments thrive, and innovation is fuelled. The intricacies of capital markets, comprising stocks, bonds, commodities, and derivatives, wield significant influence over the allocation of resources, the valuation of assets, and determination of risk and return profiles. As the focal point of myriads of economic activities, understanding the mechanisms, behaviours, and interdependencies within capital markets is not only crucial for investors and financial institutions but also holds profound implications for policymakers, regulators, and society at large. The general aim of modern portfolio theory is to have diversified assets portfolios that have maximised expected return for a given financial risk exposure. The theory assumes that an average investor should be risk-averse and give preferences to less risky portfolios.

(Sharpe, 1964) followed up on the said ideas by introducing the Capital Asset Pricing Model, where he constructed a market equilibrium theory of asset prices under the condition of risk. It was observed that the risk of a portfolio consists of two components: systematic (undiversifiable) and idiosyncratic risks. It was proven that diversification enables the investor to escape all but the systematic exposure. In practice, most financial portfolios consist of multitude of various assets and their yields pivots on many external and internal variables. Since the rise of computational power in the mid-late 20th century and the abundance of available economic and financial data, factor models were emerging. Following (Connor, 1995) classification, one may explore three types of financial factor models: macroeconomic, which consider such factors as inflation rate, unemployment etc.; fundamental, which build up their explanation of returns based on various attributes internal to assets: firm size, dividend yield, book-to-market ratio etc.; the last are the statistical models, which use numerous maximum-likelihood and principal-components-based approaches on time series samples of asset returns to identify the pervasive factors. (Partovi and Caputo, 2004) delved into the idea of reformulating the original efficient portfolio problem by reorganising the correlated assets set into an uncorrelated mixture of assets with the help of principal component analysis, thus coining the term “principal portfolio”. It was also highlighted that with the allowance of short trades, it is always possible to reorganise the entire asset set as an equivalent set of principal portfolios and that the elimination of correlations is emphasized as the key factor in achieving a major reduction of price volatility, simultaneously hedging and leveraging the principal portfolios. In this paper, the application of Principal Component Analysis (PCA) as a tool for identifying uncorrelated risk sources on German equity market will be explored further. The said analysis was done by many academics and practitioners alike, one can note the works of (Avellaneda and Lee, 2010), (Lohre et al., 2012) and (Avellaneda, Healy, et al., 2020) that partially conducted similar type of work for the U.S equity market. The paper explores following questions:

1. What is the difference between using correlation and covariance matrix as an underlying foundation for PCA in the context of financial markets?
2. What can be the possible interpretations of the first N principal components?

1 Principal Portfolios

Of particular interest for this work is a concept of principal portfolios, which was coined in the works of (Partovi and Caputo, 2004). The said idea bears multiple names, one of which is eigenportfolio, which was used by (Avellaneda and Lee, 2010). As was mentioned in the Introduction, the original concept was to simplify the structure of the efficient portfolio by introducing the uncorrelated portfolios via the principal component analysis, which would effectively function as a new asset set. The authors mention that one of the constraints that makes the whole idea possible is the allowance of short trades, i.e. investing in such a way that one will profit if the value of the asset fails. It is also worth noticing that the result principal portfolios are leveraged and hedged. Finally, an important property that the authors emphasise is that the variance of a typical principal portfolio is about the same as that of a single asset in the original set. (Avellaneda and Lee, 2010) propose two more ways how to choose the number of principal components to retain: (a) consider a fixed number of eigenvalues to extract the factors (assuming a number close to the number of industry sectors) or (b) take a variable number of components, in such a way that the truncation explains a given percentage of the total variance.

It is important to note that the principal component analysis is used for constructing the eigenportfolios as opposed to factor analysis because it has multiple advantages that benefit the whole concept according to (Tsay, 2010). Firstly, PCA produces orthogonal (uncorrelated) components, whereas factor analysis typically does not. The orthogonality is beneficial for principal portfolios, because it allows for each portfolio to represent a different risk source. Thus, if an investor decides to allocate their risk budgets to multiple eigenportfolios, they will not be exposed to one source of volatility multiple times. Secondly, PCA aims to maximise the variance captured by each principal component. This property is desirable because it ensures that the resulting portfolios capture as much volatility as possible. Lastly, PCA provides a simpler and more transparent framework for constructing eigenportfolios, since the resulting principal portfolios are liner combinations of the original variables and are uncorrelated, the result is easier to interpret and implement in practice. However, it should be emphasised that the factor analysis approach also has its application in financial time series as a tool for factor investing, which is out of scope for this paper.

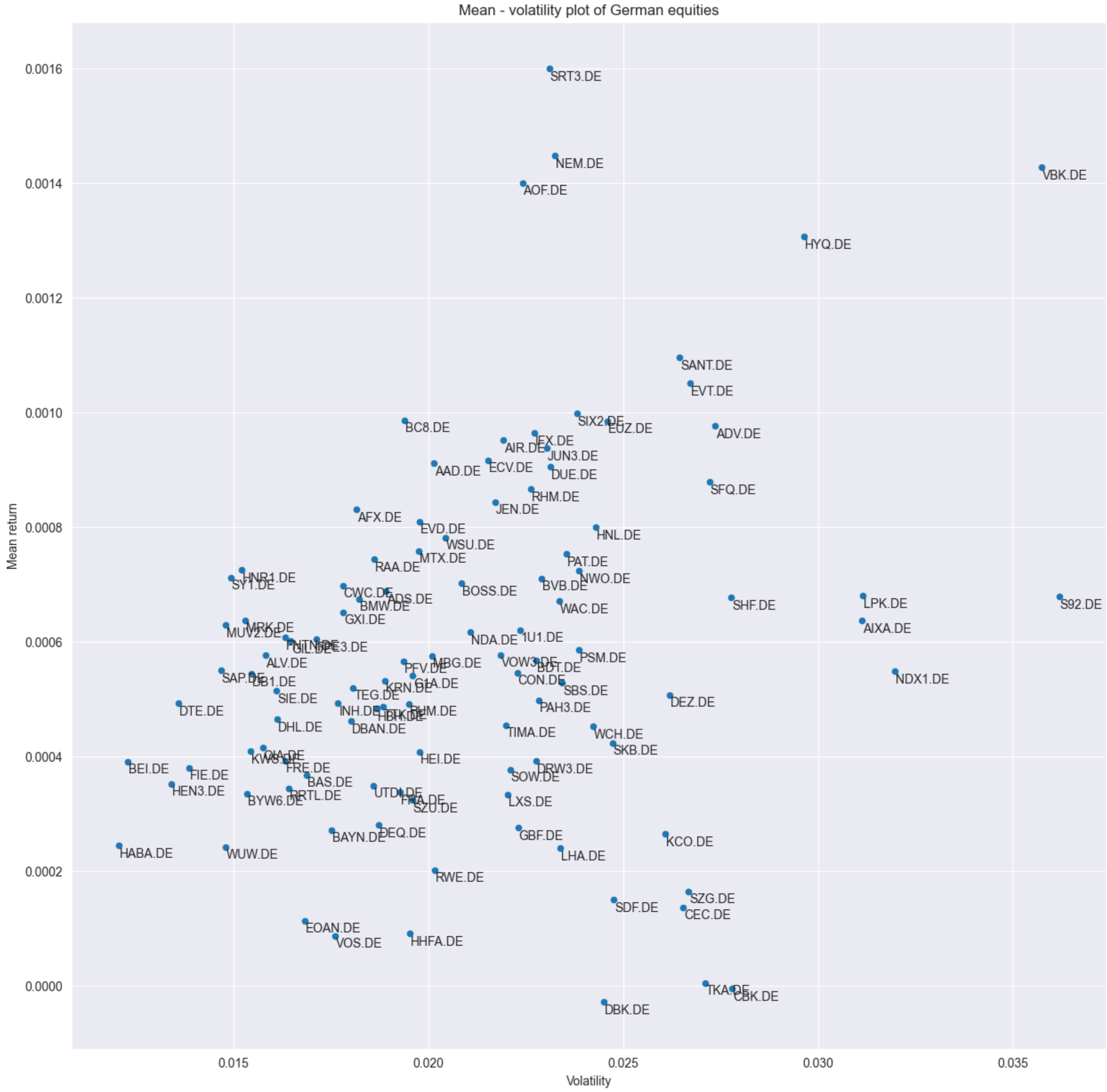
2 Fama-French Factors

For the sake of the interpretation of the principal components, as was stated in one of the research questions, one can make use of already existing factors that were created using multiple financial fundamentals. An example of which was introduced by (Fama and French, 1993), where the authors proposed common risk factors in the stocks' returns. The original 3 factors that were introduced are Market Risk Factor ($M\,kt - RF$), Size Factor (SMB - Small minus Big) and Value Factor (HML - High minus Low). The Market Risk Factor represents the excess return of the overall market portfolio over the risk-free rate, which was done by effectively incorporating market risk as a factor influencing equity returns. The SMB factor represents the difference in returns between the nine small stock portfolios and the nine big stock portfolios, where the size is measured by the equity's respective market capitalisation. The last of the original ones, the HML factors is the difference in average returns between the two value portfolios and two growth portfolios, where the respective metrics are being measured using the book-to-market ratio. High BE/ME ratio indicates a value stock and a low ratio - a growth stock. Later in (Fama and French, 2015), the authors have introduced two more factors, effectively composing a five-factor model. The added factors are the Profitability Factor (RMW - Robust minus Weak) and the Investment Factor (CMA - Conservative minus Aggressive). The first one is a difference between the average returns on the two robust operating profitability portfolios and the two weak operating portfolios. The last factor is the difference between two conservative investment portfolios and two aggressive ones. All the aforementioned factors were preconstructed for the European market by Kenneth French and provided on his website. The markets that were taken into consideration are the ones from the Western Europe and Northern Europe plus some Eurozone countries. Originally, (Fama and French, 1993) proposed the factors mentioned above based on empirical observations from historical stock returns. The idea of their research was to challenge the conventional wisdom of the CAPM, where it was suggested that only systematic (market) risk influences expected returns. Instead, it was argued that other factors might also play a significant role in explaining the variance in equity returns. In the scope of this paper, the resulting principal portfolios will be compared with different factors to find a possible interpretation of the resulting portfolios using the methodology proposed by (Lohre et al., 2012), where the authors used the Fama-French factors as explanatory variables.

3 Exploratory Data Analysis

The timeframe for the research was set up from the beginning of 2010 until the September 2023, effectively capturing a decade worth of data and market structure. Based on the conducted exploratory analysis the mean - volatility structure of the given asset universe was uncovered, stating that a big portion of the equities has mean return less than 0.008 and volatility below 0.025. The scatter graph on Figure 1 shows the distribution of stocks by average return on the y-axis and its volatility on the x-axis. Each dot represents an equity, the name is stated next to each dot. After examining this graph, one can divide the asset universe into four quadrants, starting from top left, counterclockwise: high return - high risk, high return - low risk, low return - low risk, low return - high risk. the direction of preference for a rational investor should be from the centre to the top right, since it maximises the return and minimises the risk. However, one can notice that this quadrant is emptier than the others in the given sample period. Most of the equities are grouped in the bottom left quadrant. Nonetheless, there are some noteworthy outliers present in the graphs, particularly VBK.DE, which is an example of high risk - high return stock, or SRT3.DE, which can be a better choice for a more rational investor. The returns' distributions of the assets were examined further to reveal that their kurtosis and skewness has peculiar values that indicate how the assets behave in the given timeframe. Lastly, the said distributions were tested against the normal distribution, where it was proven that none of the assets' returns indicate normality. In the sample cross-correlation analysis of the returns it was revealed that the average correlation coefficient between the assets' returns is 0.302. It was also uncovered that the companies from the same industry have the highest correlation among them, e.g. Volkswagen AG and Porsche AG. With this thought in mind, it was also decided to examine the cross-correlation between the companies coming from the same sector, which led to following insights: the most intercorrelated industries are insurance, banking and automobile manufacturing, where the average correlation coefficients rise as high as 0.891, 0.878, 0.823 respectively. The stationarity test revealed that the most important assumption for conducting the principal component analysis, having the underlying data stationary, was successfully fulfilled, thus ensuring that all following analysis is correctly specified.

Fig. 1: Mean - volatility plot of German equities



Source: Yahoo! Finance data, Author's calculation

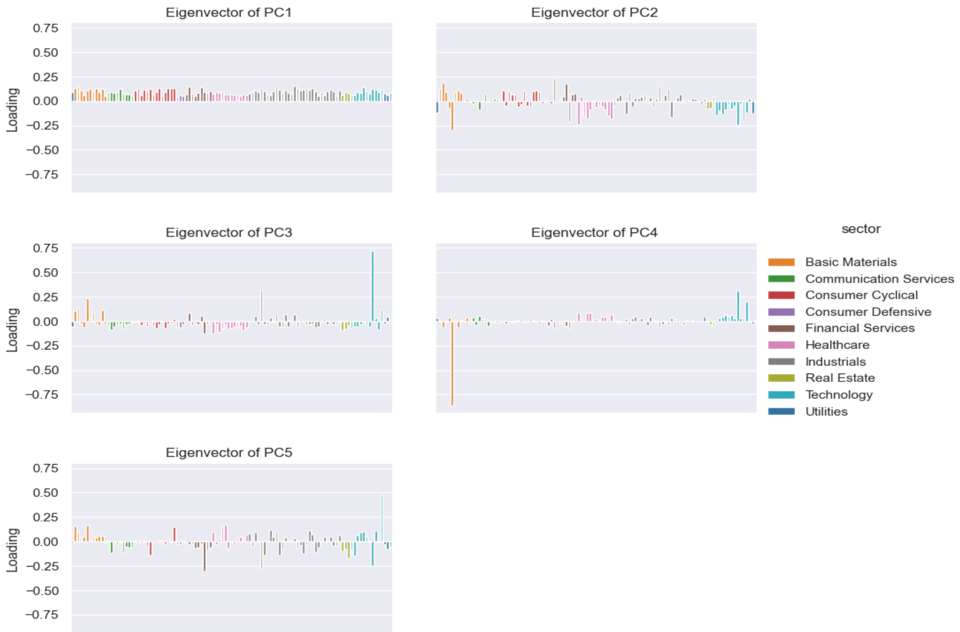
4 Principal Component Analysis

This chapter provides analyses of the eigenvectors based on correlation and covariance matrices and their comparison. It was discovered that the first 10 components explain roughly half of the variance of the original asset set for both approaches. The very first component carries around 1/3 of the original variance, with all subsequent components bringing less than 4% to overall explainability, this is not a usual behaviour for principal component analysis on panel data, where one would not expect such a sudden drop in explainability, albeit for time series data, this might make sense, since co-movement is tried to be explained. The obtained results are in line with the findings of other authors that have conducted their research on the same topic, most notably (Avellaneda and Lee, 2010), where the authors have explored the U.S. equity market and have obtained the first component that had explained 20% of overall variance with all succeeding components explaining less than 4% each. The first component has an interest interpretation (Tsay, 2010) gives it a “market” component name, since it represents the general movement of the equity market. Most notably, (Laloux et al., 2000) state that the first principal component can be a "market" component, since it has roughly equal elements on all the N stocks in its eigenvector. The figures 2 and 3 highlight that the values of the first principal component are roughly equal in size despite the approach, which further supports the aforementioned idea of a “market” component, thus implying that the first principal component displays a risk source coming from the overall market movement itself. In other components, both approaches leveraged the possibility to create both positive and negative values, thus indicating long and short weights respectively. However, the approaches diverge in the eigenvectors’ loadings allocation: the standardisation of the returns allowed for more equal weights distribution, whereas the lack of thereof enabled the high volatility assets to gain bigger weights, thus creating dominant equities in each portfolio. The second principal portfolio can be interpreted as "Healthcare and Technology" for long positions vs. "Basic Materials and Industrials" for short positions in case of correlation matrix. The third principal component favours Industrials and Technology for long positions, and Communication and Consumer Defensive for short. The fourth and fifth principal components cannot be said to clearly favour one sector over the other, thus it was proposed that the underlying variance may come from other sources, not only from the respective sectors.

Fig. 2: Eigenvectors of the first five principal components based on sample correlation matrix, coloured by their respective sectors

Source: Yahoo! Finance data, Author’s calculation

Fig. 3: Eigenvectors of the first five principal components based on sample covariance matrix, coloured by their respective sectors



Source: Yahoo! Finance data, Author’s calculation

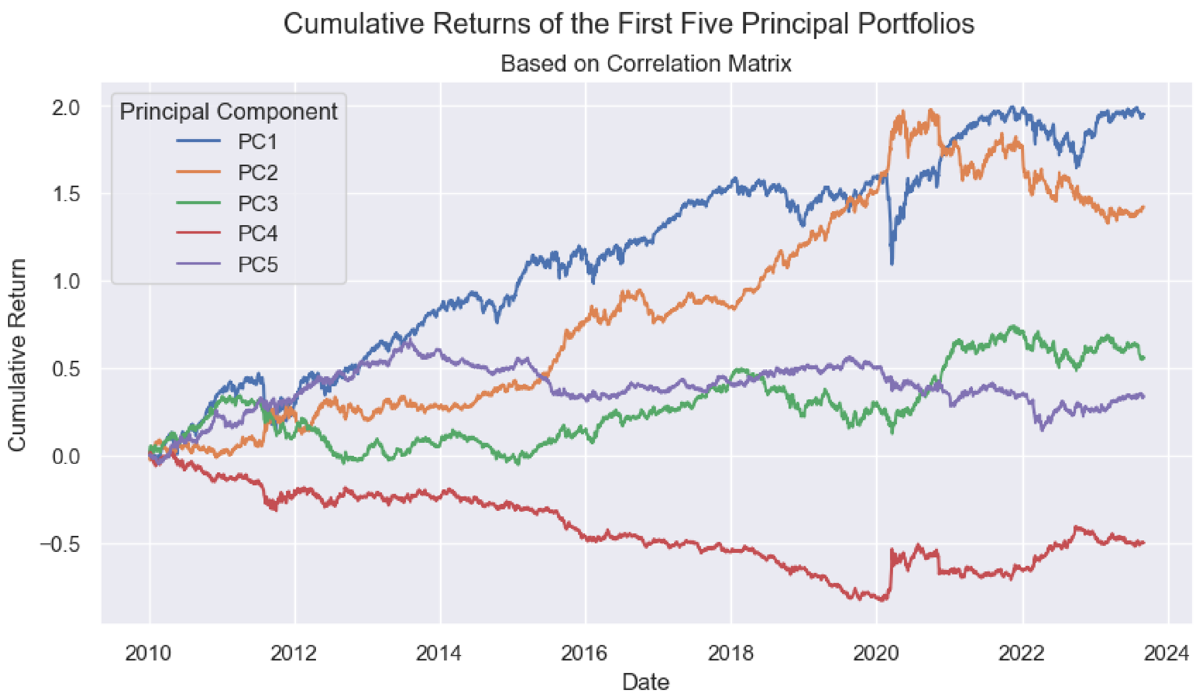
As was previously mentioned, the covariance matrix approach has some equities that dominate over the rest of the asset universe, which may obstruct in defining a clear picture in case of which sector is long, and which is short. Albeit the second principal component goes short on Healthcare and Technology. The third principal component goes long on one company from the Technology sector and one from Industrials sector, and then goes short on healthcare sector. The fourth principal component has a dominant short position on a company from Basic Materials, and then it goes long on Healthcare and Technology. The fifth principal component does not seem to have a clearly defined structure based on sectors, as was the case previously.

4.1 Assessing the Principal Portfolios

Figures 4 and 5 display the cumulative returns of the first five principal components based on correlation and covariance matrices' weights respectively. The comparison was done following the methodology of (Lohre et al., 2012), where the authors were using cumulative returns. Both first principal components outperform the German equity market benchmark for the given period. This further proves that despite the approach, the most dominant eigenvalue produces the eigenvector that will mimic the market, hence the name "market" component. In case of correlation matrix approach, from the first five principal portfolios only the fourth one is unprofitable, as its cumulative returns fall below 0 to the end date. All others do generate different magnitudes of profit. Quite interestingly, the second principal portfolio was underperforming from 2010 to 2020 compared to the first one, then the second component exceeded its counterpart briefly for less than a year, and then fell below again. This sudden spike in performance of the second principal component and decline of the first one might find an explanation in the worldwide COVID epidemic, since the second principal component has a significant number of its weights allocated to long positions on Healthcare and Technology sector, whilst the first one distributes its weights roughly equally, thus reflecting the overall situation of the market. The third and the fifth principal components are less performant than the first two, overall, they end up having around 0.5 cumulative return to date.

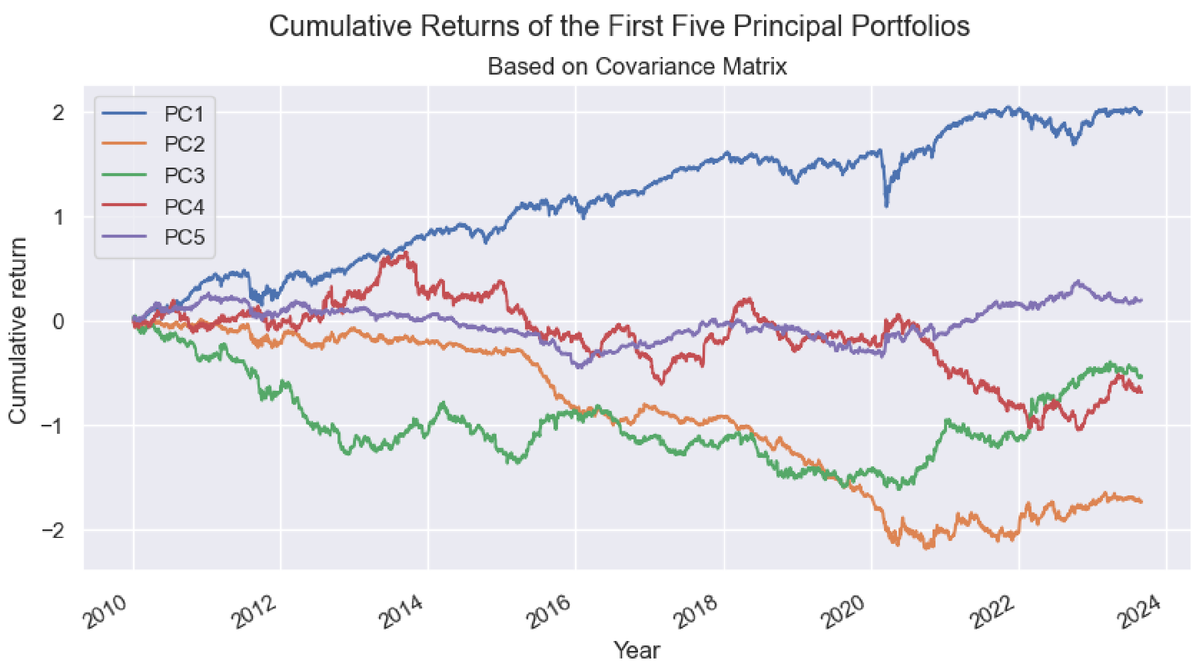
In case of covariance matrix approach, only the very first principal component is profitable, as it reaches a cumulative return of roughly 2.0 to date. The fifth principal component has positive returns as well, albeit they fall beneath 0.5. All other three components are negative in their returns, with the second principal component being the lowest in cumulative returns. The other two portfolios fall between $[-1, 0]$ range. With all that being said, one may conclude that the first five principal components based on covariance matrix are less profitable compared to their counterparts based on sample correlation matrix.

Fig. 4: Cumulative returns of the first five principal portfolios based on sample correlation matrix



Source: Yahoo! Finance data, Author's calculation

Fig. 5: Cumulative returns of the first five principal portfolios based on sample covariance matrix



Source: Yahoo! Finance data, Author's calculation

4.2 Interpreting the Principal Portfolios

There are multiple ways how one can interpret the result portfolios, one of which was explored on the figures 2 and 3, where it was established that some of the principal portfolios go long on some industries and go short on another ones, effectively maintaining a pattern of being industry components. However, as it was previously mentioned, this methodology may have some drawbacks, since it considers only one dimension of exploring the components, which is by their industry sector affiliation. Another approach is to use linear regression with known factors as was explored in the Chapter 2.

Fig. 6: Linear regression analysis results for the principal portfolios for the period from the beginning of 2010 till September 2023.

	PC_1	PC_2	PC_3	PC_4	PC_5
Coefficients					
Intercept	0.000	0.000	0.000	0.000	0.000
Mkt-RF	0.857	−0.157	0.351	−0.027	0.042
SMB	−0.052	0.294	0.836	−0.562	−0.303
HML	0.191	−1.292	0.197	−0.049	0.119
RMW	0.165	−0.216	−0.136	−0.381	0.498
CMA	−0.210	−0.080	−0.506	−0.208	−0.283
p-values					
Intercept	0.002	0.000	0.508	0.457	0.944
Mkt-RF	0.000	0.000	0.000	0.027	0.000
SMB	0.073	0.000	0.000	0.000	0.000
HML	0.000	0.000	0.000	0.141	0.000
RMW	0.001	0.000	0.011	0.000	0.000
CMA	0.000	0.118	0.000	0.000	0.000
Adjusted R^2	0.742	0.602	0.292	0.146	0.116

Source: Author's own calculation

Figure 6 displays the results of the linear regression analysis against predefined factors. Here it was confirmed that the first component is indeed a "market" component, since the respective linear regression coefficient has a value of 0.857, all other coefficients do not exceed a magnitude of 0.2 in absolute value. The second principal component had a negative loading on HML factor with a value of −1.292, effectively meaning that this principal component may be labelled as Low minus High. The third principal component had the highest coefficient on the

SMB factor with a value of 0.836, thus getting an interpretation of the Small minus Big component. The fourth and the fifth principal component had their biggest absolute coefficients on SMB and RMW coefficient respectively. With the fourth principal component getting a Big minus Small label, and the fifth Robust minus Weak label. Albeit the overall variance of these components was explained only partly - roughly 10% in both cases, as was indicated by their respective R^2. The obtained results are in line with those of (Lohre et al., 2012), where the authors achieved the same outcomes, they proposed that most of the principal portfolios' time series variation cannot be accounted for by the common factors only, since either the higher order principal components are not meaningful or the factor structure of 1.14 is incomplete and may be lacking some important factors.

Conclusion

After comparing both approaches, one can be ready to answer the first research question: indeed, when speaking of financial time series, it is important to standardise the returns beforehand, to produce the portfolios that have their weights allocated evenly. This positively contributes to the diversification ratio as well. Moreover, the correlation matrix approach produced portfolios that performed better during the backtesting.

All in all, the first principal component can be interpreted as a market component, despite the chosen approach. For given timeframe and asset universe, the second principal component can be effectively interpreted as a "Healthcare and Technology" vs. "Basic Materials and Industrials" and a Low minus High component, thus identifying and leveraging the said risk sources. The third component can be labelled as "Industrials and Technology" vs. "Communication and Consumer Defensive" and Small minus Big component. The fourth principal component could not be linked with some sectors, albeit it can still get a Big minus Small label. The fifth principal component did not have any affiliations with sectors structure as well, however it can still be interpreted as Robust minus Weak component. Effectively, these are the top five dominant risk sources for the German equity market for a given timeframe.

The research underscores the importance of standardising the asset returns before conducting a principal component analysis on financial time series data by making a comparison between both approaches. This study also contributes insights to revealing the underlying risk sources that were present on the German equity market for the period of the last 10 years. Furthermore, it is important to explore the ramifications of the given methodology for the portfolio construction field in the future, since principal component analysis can be proven to be a valuable tool for constructing the portfolios and identifying the risk sources.

References

Avellaneda, M., Healy, B., Papanicolaou, A., Papanicolaou, G., & Xu, T. (2020). Principal eigenportfolios for U.S. equities. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3738769

Avellaneda, M., & Lee, J.-H. (2010). Statistical arbitrage in the US Equities Market. *Quantitative Finance*, *10*(7), 761–782. https://doi.org/10.1080/14697680903124632

Connor, G. (1995). The three types of factor models: A comparison of their explanatory power. *Financial Analysts Journal*, *51*(3), 42–46. https://doi.org/10.2469/faj.v51.n3.1904

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk*. *The Journal of Finance*, *19*(3), 425–442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and Bonds. *Journal of Financial Economics*, *33*(1), 3–56. https://doi.org/10.1016/0304-405x(93)90023-5

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1–22. https://doi.org/10.1016/j.jfineco.2014.10.010

Laloux, L., Cizeau, P., Potters, M., & Bouchaud, J.-P. (2000). Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance*, *03*(03), 391–397. https://doi.org/10.1142/s0219024900000255

Lohre, H., Neugebauer, Dr. U., & Zimmer, C. (2012). Diversified Risk Parity Strategies for Equity Portfolio Selection. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2049280

Partovi, M. Hossein and Caputo, Michael, (2004), Principal Portfolios: Recasting the Efficient Frontier, *Economics Bulletin*, 7, issue 3, p. 1-10,
https://EconPapers.repec.org/RePEc:ebl:ecbull:eb-04g00003.

Tsay, R. S. (2010). *Analysis of Financial Time Series*. Wiley.

Contact

Bogdan Romenskii
Prague University of Economics and Business,
W. Churchill Sq. 1938/4,
130 67 Prague 3, Czech Republic
romb02@vse.cz

Tomáš Löster
Prague University of Economics and Business,
W. Churchill Sq. 1938/4,
130 67 Prague 3, Czech Republic
tomas.loster@vse.cz